College of Polytechnics Jihlava

40th International Conference

Mathematical Methods in Economics 2022

Proceedings

College of Polytechnics Jihlava

7 – 9 September 2020

Jihlava, Czech Republic

40th International Conference Mathematical Methods in Economics 2022

Proceedings

Publisher: College of Polytechnics Jihlava, Tolstého 1556/16, Jihlava, Czech Republic

Editor: Hana Vojáčková

This publication is not a subject of language check.

Papers are sorted by authors names in alphabetical order.

All papers passed a blind review process.

Technical production: College of Polytechnics Jihlava, Tolstého 1556/16, Jihlava, Czech Republic

First edition

2022

ISBN: 978-80-88064-62-6 (online ; pdf)

Programme Committee

doc. RNDr. Ing. Miloš Kopa, Ph.D.

President of the Czech Society for Operations Research

Charles University, Faculty of Mathematics and Physics

prof. Dr. Ing. Miroslav Plevný

Vice-president of the Czech Society for Operations Research

University of West Bohemia, Faculty of Economics

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

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

Ing. Martina Kuncová, Ph.D.

College of Polytechnics Jihlava

prof. RNDr. Jan Pelikán, CSc.

Prague University of Economics and Business, Faculty of Informatics and Statistics

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

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

Ing. Martina Kuncová, Ph.D. (chair)

Ing. Simona Činčalová, Ph.D.

Ing. Roman Fiala, Ph.D.

Ing. Andrea Kubišová, Ph.D.

Ing. Lenka Lízalová, Ph.D.

Ing. Marie Slabá, Ph.D.

Mgr. Hana Vojáčková, Ph.D.

Bc. Martin Bulák

Markéta Stejskalová

Conference website: https://mme2022.vspj.cz/

List of Referees:

Vladislav Bína
Adam Borovička
Ivan Brezina
Helena Brožová
Michal Černý
Andrea Čížků
Martin Dlouhý
Jan Fábry
Petr Fiala
Tomáš Formánek
Lukáš Frýd
Andrea Furková
Jana Hančlová
Milan Hladík
Robert Hlavatý
Vladimír Holý
Michaela Chocholatá
Lucie Chytilová
Josef Jablonský
Petr Jiříček
Jana Klicnarová
František Koblasa
Miloš Kopa
Igor Krejčí
Andrea Kubišová
Jaromír Kukal
Martina Kuncová
Martin Lukáčik
Adriana Lukáčiková
Markéta Matulová
Denisa Mocková
Ivana Olivková
Václava Pánková
Juraj Pekár
Jan Pelikán
Miroslav Plevný
Miroslav Rada
Jaroslav Ramík
Marián Reiff
Alena Rybičková
Pavla Říhová
Jana Sekničková
Ondřej Sokol
Radek Stolín
Milan Svoboda
Karol Szomolányi
Dušan Teichmann
Petra Tomanová
Quang Van Tran
Jan Zouhar
Miroslav Žižka

Contents

Indexes of Desirable Properties of a Pairwise Comparison Matrix with Fuzzy Elements

David Bartl[1], Jaroslav Ramík[2]

Abstract. In the Analytic Hierarchy Process (AHP), pairwise comparisons are used to quantify the relative importance of the elements, i.e. the criteria and/or alternatives. Fuzzy elements are appropriate whenever the decision maker is uncertain about the value of his/her evaluation of the relative importance of the elements in question. In this paper, we deal with the general case when the elements of the pairwise comparison matrix are fuzzy subsets of an Abelian linearly ordered group (alo-group). We then propose some desirable properties – consistency, intensity, and coherence – of the fuzzy pairwise comparison matrix and we also propose indexes to measure these desirable properties. Based on these indexes, a new solution algorithm to find the priority vector satisfying these desirable properties can be formulated.

Keywords: multi-criteria optimization, Analytic Hierarchy Process (AHP), pairwise comparison matrix, fuzzy elements, consistency, intensity, coherence, priority vector, alo-group

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

1 Introduction

The main subproblem of the Analytic Hierarchy Process (AHP) is to calculate the priority vectors, i.e. the weights assigned to the elements of the hierarchy (criteria, subcriteria, and alternatives or variants), by using the information provided in the form of a pairwise comparison matrix. Given a set of elements and corresponding pairwise comparison matrix, whose entries evaluate the relative importance of the elements with respect to a given criterion, the purpose is to calculate the priority vector characterizing the ranking of the elements. There are various methods for calculating the vector of weights, e.g. Saaty's Eigenvector Method, the Geometric Mean Method, and others, see [3, 4].

Fuzzy sets as the elements of the pairwise comparison matrix can be applied whenever the decision maker is not sure about the preference degree of his/her evaluations of the pairs in question. Such an approach is also well known in the Fuzzy Analytic Hierarchy Process (FAHP) originated by Thomas Saaty in [6]. Recent development of the problem can be found in [2]. Comparing to [4], here, we propose newly reformulated desirable properties – consistency, intensity, and coherence – of the priority vector and we also propose indexes to measure these desirable properties of the given fuzzy pairwise comparison matrix. Then, a new algorithm for deriving the priority vector satisfying the desirable properties can be devised.

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].

A *fuzzy subset* S of a nonempty set X (or a *fuzzy set* on X) is a family $\{S_\alpha\}_{\alpha\in[0;1]}$ of subsets of X such that $S_0 = X$, and $S_\beta \subset S_\alpha$ whenever $0 \le \alpha \le \beta \le 1$, and also $S_\beta = \bigcap_{0\le\alpha<\beta} S_\alpha$ whenever $0 < \beta \le 1$. The *membership function of* S is the function μ_S from X into the unit interval $[0;1]$ defined by $\mu_S(x) = \sup\{\, \alpha \mid x \in S_\alpha \,\}$. Given an $\alpha \in]0;1]$, the set $[S]_\alpha = \{\, x \in X \mid \mu_S(x) \ge \alpha \,\}$ is called the *α-cut of the fuzzy set S*. In order to unify various approaches and to prepare a more flexible presentation, we apply alo-groups, see [1]. Recall that an Abelian group is a set, G, together with an operation $\odot$ and corresponding "group axioms" that combine any two elements

[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

$a, b \in G$ to form another element in G denoted by $a \odot b$. The well known examples of alo-groups can be found in [4, 3], or, [1].

3 Fuzzy pairwise comparison matrices, reciprocity and consistency

Let $\mathcal{G} = (G, \odot, \leq)$ be an alo-group over an open interval G of $\mathbf{R}$, see [1]. Let $\tilde{A} = \{\tilde{a}_{ij}\}$ be an $n \times n$ matrix where each element is a bounded fuzzy interval of the alo-group $\mathcal{G}$, let $\alpha \in [0; 1]$, and let $[\tilde{a}_{ij}]_\alpha = [a^L_{ij}(\alpha); a^R_{ij}(\alpha)]$ be the α-cut of $\tilde{a}_{ij}$. The matrix $\tilde{A} = \{\tilde{a}_{ij}\}$ is said to be *α-$\odot$-reciprocal* if the following two conditions hold for each $i, j \in \{1, \ldots, n\}$:

$$a^L_{ii}(\alpha) = a^R_{ii}(\alpha) = e, \qquad \text{and} \qquad a^L_{ij}(\alpha) \odot a^R_{ji}(\alpha) = e. \tag{1}$$

If $\tilde{A} = \{\tilde{a}_{ij}\}$ is α-$\odot$-reciprocal for all $\alpha \in [0; 1]$, then it is called *$\odot$-reciprocal*. If $\tilde{A} = \{\tilde{a}_{ij}\}$ is $\odot$-reciprocal, then $\tilde{A} = \{\tilde{a}_{ij}\}$ is called a *fuzzy pairwise comparison matrix*, *fuzzy PC matrix*, *FPC matrix*, or, shortly, *FPCM*.

Now, we turn to the concept of consistency of FPC matrices.

Definition 1. Let $\alpha \in [0; 1]$. A FPC matrix $\tilde{A} = \{\tilde{a}_{ij}\}$ is said to be *α-$\odot$-consistent* if the following condition holds: There exists a crisp matrix $A' = \{a'_{ij}\}$ with $a'_{ij} \in [\tilde{a}_{ij}]_\alpha$ such that $A' = \{a'_{ij}\}$ is consistent, i.e. for each $i, j, k \in \{1, \ldots, n\}$ it holds

$$a'_{ik} = a'_{ij} \odot a'_{jk}. \tag{2}$$

The FPC matrix $\tilde{A} = \{\tilde{a}_{ij}\}$ is said to be *$\odot$-consistent* if $\tilde{A}$ is α-$\odot$-consistent for all $\alpha \in [0; 1]$. If for some $\alpha \in [0; 1]$ the FPC matrix $\tilde{A} = \{\tilde{a}_{ij}\}$ is not α-$\odot$-consistent, then $\tilde{A}$ is called *α-$\odot$-inconsistent*. If for all $\alpha \in [0; 1]$ the FPC matrix $\tilde{A} = \{\tilde{a}_{ij}\}$ is α-$\odot$-inconsistent, then $\tilde{A}$ is called *$\odot$-inconsistent*.

The next proposition gives an equivalent condition for a FPC matrix to be α-$\odot$-consistent, see, e.g. [4]. The proof of the following proposition is easy, or can be found in [4].

Proposition 1. *Let $\alpha \in [0; 1]$, let $\tilde{A} = \{\tilde{a}_{ij}\}$ be a FPC matrix, and let $[\tilde{a}_{ij}]_\alpha = [a^L_{ij}(\alpha); a^R_{ij}(\alpha)]$ be an α-cut of $\tilde{a}_{ij}$. Then $\tilde{A} = \{\tilde{a}_{ij}\}$ is α-$\odot$-consistent iff there exists a vector $w = (w_1, \ldots, w_n)$ with $w_i \in G$, for $i \in \{1, \ldots, n\}$, such that for each $i, j \in \{1, \ldots, n\}$, it holds:*

$$a^L_{ij}(\alpha) \leq w_i \div w_j \leq a^R_{ij}(\alpha). \tag{3}$$

Notice that another, i.e. stronger, concept of α-$\odot$-consistency has been defined in [4, 5].

4 Desirable properties of the priority vector

In this section, we shall investigate PC matrices with fuzzy elements, and extend some concepts and properties of PC matrices on alo-group with crisp elements.

From now on we shall assume that $\mathcal{G} = (G, \odot, \leq)$ is a continuous alo-group in $\mathbf{R}$. Then G is an open interval of $\mathbf{R}$, see [1]. We denote the set of the first n positive integers by $\mathcal{N}$, i.e. we put $\mathcal{N} = \{1, 2, \ldots, n\}$. Let $\tilde{A} = \{\tilde{a}_{ij}\}$ be an $n \times n$ fuzzy pairwise comparison matrix (FPCM), which is $\odot$-reciprocal.

The result of the pairwise comparisons method based on the FPC matrix $\tilde{A} = \{\tilde{a}_{ij}\}$ is a rating of the set $C = \{c_1, c_2, \ldots, c_n\}$ 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 C (or the *ranking* of C) is a function $w\colon C \to G$ that assigns to every element from $C = \{c_1, c_2, \ldots, c_n\}$ a value from the linearly ordered set G of the alo-group $\mathcal{G} = (G, \odot, \leq)$.

Here, $w(c)$ represents the ranking value for $c \in C$. The function w is usually written in the form of a vector of *weights*, i.e. $w = (w(c_1), w(c_2), \ldots, w(c_n))$, or, simply $w = (w_1, w_2, \ldots, w_n)$, and it is called the *priority vector*. Also, we say that the priority vector w is associated with the FPC matrix $\tilde{A}$, or that the priority vector w is generated by a priority generating method based on the FPC matrix $\tilde{A}$.

The priority vector $w = (w(c_1), w(c_2), \ldots, w(c_n))$ is *$\odot$-normalized*, if $\bigodot_{i=1}^n w(c_i) = e$.

We start with definitions of three various concepts of priority vectors: consistent, intensity, and coherent ones.

Definition 2. Let $\tilde{A} = \{\tilde{a}_{ij}\}$ be a FPC matrix on the alo-group $\mathcal{G} = (G, \odot, \leq)$, let $w = (w_1, w_2, \ldots, w_n)$, with $w_j \in G$, be a priority vector, let $\alpha \in [0; 1]$, and let $[\tilde{a}_{ij}]_\alpha = [a^L_{ij}(\alpha); a^R_{ij}(\alpha)]$ be the α-cut of $\tilde{a}_{ij}$.

(i) We say that the vector w is an *α-consistent vector* (α-CsV) of the FPC matrix $\tilde{A}$ if the following condition holds:

$$a_{ij}^L(\alpha) \leq w_i \div w_j \qquad \text{for all} \quad i,j \in \mathcal{N}. \tag{4}$$

Moreover, the vector w is a *consistent vector* (CsV) of the FPC matrix $\tilde{A}$ if condition (4) holds for all $\alpha \in [0;1]$. If there exists an α-consistent vector or consistent vector of the FPC matrix $\tilde{A}$, then $\tilde{A}$ is called an α-consistent FPC matrix or consistent FPC matrix, respectively.

(ii) We say that the vector w is an *α-intensity vector* (α-InV) of the FPC matrix $\tilde{A}$ if the following condition holds:

$$a_{ij}^L(\alpha) > a_{kl}^R(\alpha) \quad \text{implies} \quad w_i \div w_j > w_k \div w_l \qquad \text{for all} \quad i,j,k,l \in \mathcal{N}. \tag{5}$$

Moreover, the vector w is an *intensity vector* (InV) of the FPC matrix $\tilde{A}$ if condition (5) holds for all $\alpha \in [0;1]$. If there exists an α-intensity vector or intensity vector of the FPC matrix $\tilde{A}$, then $\tilde{A}$ is called an α-intensity FPC matrix or intensity FPC matrix, respectively.

iii) We say that the vector w is an *α-coherent vector* (α-CoV) of the FPC matrix $\tilde{A}$ if the following condition holds:

$$a_{ij}^L(\alpha) > e \quad \text{implies} \quad w_i > w_j \qquad \text{for all} \quad i,j \in \mathcal{N}. \tag{6}$$

Moreover, the vector w is a *coherent vector* (CoV) of the FPC matrix $\tilde{A}$ if condition (6) holds for all $\alpha \in [0;1]$. If there exists an α-coherent vector or coherent vector of the FPC matrix $\tilde{A}$, then $\tilde{A}$ is called an α-coherent FPC matrix or coherent FPC matrix, respectively.

Remark 1. In definition (4), it looks like the concept of α-consistent FPC matrix depends only on the left sides $a_{ij}^L(\alpha)$ of the α-cuts of the elements of the matrix $\tilde{A}$, and not on the right sides. Notice that by the reciprocity property of the elements it is easy to see that w is an α-CsV of $\tilde{A} = \{\tilde{a}_{ij}\}$ if and only if

$$a_{ij}^L(\alpha) \leq w_i \div w_j \leq a_{ij}^R(\alpha) \qquad \text{for all} \quad i,j \in \mathcal{N}. \tag{7}$$

We obtain the following modification.

Proposition 2. *Let $\tilde{A} = \{\tilde{a}_{ij}\}$ be a FPCM on $\mathcal{G} = (G, \odot, \leq)$, let $\alpha \in [0;1]$, and let $[\tilde{a}_{ij}]_\alpha = [a_{ij}^L(\alpha); a_{ij}^R(\alpha)]$. A priority vector $w = (w_1, w_2, \ldots, w_n)$, with $w_j \in G$, satisfies*

$$a_{ij}^L(\alpha) \div a_{kl}^R(\alpha) \leq (w_i \div w_j) \div (w_k \div w_l) \qquad \text{for all} \quad i,j,k,l \in \mathcal{N} \tag{8}$$

if and only if w is an α-consistent vector of the FPC matrix $\tilde{A}$.

Remark 2. Notice that if $w = (w_1, w_2, \ldots, w_n)$ satisfies (8), then w is also an α-intensity vector of the FPC matrix $\tilde{A}$. To see that, let $a_{ij}^L(\alpha) > a_{kl}^R(\alpha)$, which is equivalent to $a_{ij}^L(\alpha) \div a_{kl}^R(\alpha) > e$. In other words, if w is an α-consistent vector, then w is an α-intensity vector of the FPC matrix $\tilde{A}$. Notice also that if $w = (w_1, w_2, \ldots, w_n)$ is an α-intensity vector, then it is also an α-coherent vector of the FPC matrix $\tilde{A}$. Indeed, condition (6) follows from (5) easily by considering $k = l$.

5 Measuring desirable properties of FPC matrices

The condition of consistency of FPC matrices is the strongest condition of the three: consistency condition, intensity condition, and coherence one. In practice, FPC matrices are often inconsistent, even more, the intensity condition is not satisfied and/or they are incoherent. Hence, it is useful to know "how much" these desirable conditions are satisfied. This is why we measure the inconsistency, non-intensity, or incoherence of a given FPC matrix $\tilde{A} = \{\tilde{a}_{ij}\}$ by special indexes, see, e.g., [6, 4]. Let $\mathcal{G} = (G, \odot, \leq)$ be a continuous alo-group in the set of the real numbers. Then $G =]g^-; g^+[$ is an open interval in $\mathbf{R}$.

For $a \in \mathcal{G} = (G, \odot, \leq)$, define

$$a^+ = \max\{a, e\}, \qquad a^- = \max\{a^{-1}, e\}, \tag{9}$$

the *absolute value* $|a|$ of a as

$$|a| = a^+ \odot a^- = \max\{a^{-1}, a\}. \tag{10}$$

Moreover, for a $w = (w_1, \ldots, w_n) \in G^n$, define the *norm* $\|w\|$ of w as

$$\|w\| = \max\{\, |w_i| \mid i \in \mathcal{N} \,\}. \tag{11}$$

Finally, for an $r \in G$, set the *ball* $B(r)$ as

$$B(r) = \{\, w \in G^n \mid \|w\| \leq r \,\}. \tag{12}$$

Definition 3. Let $\tilde{A} = \{\tilde{a}_{ij}\}$ be a FPC matrix on an alo-group $\mathcal{G} = (G, \odot, \leq)$. For each pair $i, j \in \mathcal{N}$, for a priority vector $w = (w_1, w_2, \ldots, w_n)$, with $w_j \in G$, and for $r \in G$ and $\alpha \in [0; 1]$, define:

(i) Let the *local α-inconsistency grade of an element of FPC matrix $\tilde{A}$ and vector w* be defined as

$$\varepsilon_{ij}^{Cs}(\tilde{A}, w, \alpha) = a_{ij}^{L}(\alpha) \div (w_i \div w_j), \tag{13}$$

and define the *global α-inconsistency grade of FPC matrix $\tilde{A}$ and vector w* as

$$E^{Cs}(\tilde{A}, w, \alpha) = \max\{\, \varepsilon_{ij}^{Cs}(\tilde{A}, w, \alpha) \odot \varepsilon_{kl}^{Cs}(\tilde{A}, w, \alpha) \mid i, j, k, l \in \mathcal{N} \,\}. \tag{14}$$

Now, we define the *α-inconsistency index* of $\tilde{A}$ as

$$I^{Cs}(\tilde{A}, \alpha) = \inf\{\, E^{Cs}(\tilde{A}, w, \alpha) \mid w \in B(r) \,\}. \tag{15}$$

The smaller the α-inconsistency index $I^{Cs}(\tilde{A}, \alpha)$ is, the higher the α-consistency of the matrix $\tilde{A}$ is, i.e. the less α-inconsistent the matrix $\tilde{A}$ is.

(ii) Let the *local α-non-intensity grade of two elements of FPC matrix $\tilde{A}$ and vector w* be defined as

$$\varepsilon_{ijkl}^{In}(\tilde{A}, w, \alpha) = \begin{cases} (w_k \div w_l) \div (w_i \div w_j) & \text{if } a_{ij}^{L}(\alpha) > a_{kl}^{R}(\alpha), \\ g^{-} & \text{otherwise,} \end{cases} \tag{16}$$

and define the *global α-non-intensity grade of FPC matrix $\tilde{A}$ and vector w* as

$$E^{In}(\tilde{A}, w, \alpha) = \max\{\, \varepsilon_{ijkl}^{In}(\tilde{A}, w, \alpha) \mid i, j, k, l \in \mathcal{N} \,\}. \tag{17}$$

Now, we define the *α-non-intensity index* of $\tilde{A}$ as

$$I^{In}(\tilde{A}, \alpha) = \inf\{\, E^{In}(\tilde{A}, w, \alpha) \mid w \in B(r) \,\}. \tag{18}$$

The smaller the α-non-intensity index $I^{In}(\tilde{A}, \alpha)$ is, the higher the α-intensity of the matrix $\tilde{A}$ is, i.e. the less α-non-intensive the matrix $\tilde{A}$ is.

(iii) Let the *local α-incoherence grade of an element of FPC matrix $\tilde{A}$ and vector w* be defined as

$$\varepsilon_{ij}^{Co}(\tilde{A}, w, \alpha) = \begin{cases} w_j \div w_i & \text{if } a_{ij}^{L}(\alpha) > e, \\ g^{-} & \text{otherwise,} \end{cases} \tag{19}$$

and define the *global α-incoherence grade of FPC matrix $\tilde{A}$ and vector w* as

$$E^{Co}(\tilde{A}, w, \alpha) = \max\{\, \varepsilon_{ij}^{Co}(\tilde{A}, w, \alpha) \mid i, j \in \mathcal{N} \,\}. \tag{20}$$

Now, we define the *α-incoherence index* of $\tilde{A}$ as

$$I^{Co}(\tilde{A}, \alpha) = \inf\{\, E^{Co}(\tilde{A}, w, \alpha) \mid w \in B(r) \,\}. \tag{21}$$

The smaller the α-incoherence index $I^{Co}(\tilde{A}, \alpha)$ is, the higher the α-coherence of the matrix $\tilde{A}$ is, i.e. the less α-incoherent the matrix $\tilde{A}$ is.

Remark 3. As every FPCM is reciprocal, we can obtain a slightly different formulas for the above indexes for a FPC matrix $\tilde{A} = \{\tilde{a}_{ij}\}$, replacing $i, j \in \mathcal{N}$ by $1 \leq i < j \leq n$.

Proposition 3. *Let $\tilde{A} = \{\tilde{a}_{ij}\}$ be a FPC matrix, let $r \in G$, and let $\alpha \in [0; 1]$. Then it holds:*

A vector $w = (w_1, \ldots, w_n) \in B(r)$ is α-consistent if and only if

$$I^{Cs}(\tilde{A}, \alpha) \leq E^{Cs}(\tilde{A}, w, \alpha) \leq e. \tag{22}$$

A vector $w = (w_1, \ldots, w_n) \in B(r)$ is an α-intensity vector if and only if

$$I^{In}(\tilde{A}, \alpha) \leq E^{In}(\tilde{A}, w, \alpha) < e. \tag{23}$$

A vector $w = (w_1, \ldots, w_n) \in B(r)$ is α-coherent if and only if

$$I^{Co}(\tilde{A}, \alpha) \leq E^{Co}(\tilde{A}, w, \alpha) < e. \tag{24}$$

Proposition 4. *Let $\tilde{A} = \{\tilde{a}_{ij}\}$ be a FPC matrix, let $w = (w_1, \ldots, w_n)$, with $w_j \in G$, and let $\alpha \in [0; 1]$. Then*

$$E^{Co}(\tilde{A}, w, \alpha) \leq E^{In}(\tilde{A}, w, \alpha) < E^{Cs}(\tilde{A}, w, \alpha). \tag{25}$$

For any vector $w = (w_1, \ldots, w_n)$, with $w_j \in G$, and for any element $r \in G$, consider the vector $w' = (w'_1, \ldots, w'_n)$ with $w'_j = w_j \odot r \div \|w\|$, where the norm $\|w\|$ of w is defined by (11), and observe that $w' \in B(r)$, where the ball is defined by (12). Consider yet a FPC matrix $\tilde{A}$ and $\alpha \in [0; 1]$. By Definition 3, it is easy to see that the global α-inconsistency / α-non-intensity / α-incoherence grade of $\tilde{A}$ and w is equal to the global α-inconsistency / α-non-intensity / α-incoherence grade of $\tilde{A}$ and w', respectively. Hence, as a clear consequence of Proposition 4, particularly (25), we obtain the following result.

Proposition 5. *Let $\tilde{A} = \{\tilde{a}_{ij}\}$ be a FPC matrix, let $r \in G$, let $\alpha \in [0; 1]$, and let $w = (w_1, \ldots, w_n) \in B(r)$. Then*

$$I^{Co}(\tilde{A}, \alpha) \leq I^{In}(\tilde{A}, \alpha) < I^{Cs}(\tilde{A}, \alpha). \tag{26}$$

Moreover, if $\alpha, \alpha' \in [0; 1]$ are such that $\alpha \leq \alpha'$, then

$$I^{Co}(\tilde{A}, \alpha) \leq I^{Co}(\tilde{A}, \alpha'), \qquad I^{In}(\tilde{A}, \alpha) \leq I^{In}(\tilde{A}, \alpha'), \qquad I^{Cs}(\tilde{A}, \alpha) \leq I^{Cs}(\tilde{A}, \alpha'). \tag{27}$$

The proofs of Propositions 3, 4 and 5 are not difficult, they are left to the reader.

Example 1. Consider the multiplicative alo-group $\mathcal{R}_+$. Let $\tilde{A} = \{\tilde{a}_{ij}\}$ be a 4×4 FPC matrix given by the α-cut representation, see the definition of fuzzy sets in Section 2, for $\alpha \in [0; 1]$ as follows:

$$\tilde{A} = \begin{bmatrix} [\,1\,;\,1\,] & [\,1 + \frac{1}{2}\alpha\,;\,2 - \frac{1}{2}\alpha\,] & [\,1 + \alpha\,;\,4 - 2\alpha\,] & [\,1 + \alpha\,;\,3 - \alpha\,] \\ [\,\frac{2}{4-\alpha}\,;\,\frac{2}{2+\alpha}\,] & [\,1\,;\,1\,] & [\,1 + \alpha\,;\,4 - 2\alpha\,] & [\,2 + 6\alpha\,;\,9 - \alpha\,] \\ [\,\frac{1}{4-2\alpha}\,;\,\frac{1}{1+\alpha}\,] & [\,\frac{1}{4-2\alpha}\,;\,\frac{1}{1+\alpha}\,] & [\,1\,;\,1\,] & [\,1 + \alpha\,;\,3 - \alpha\,] \\ [\,\frac{1}{3-\alpha}\,;\,\frac{1}{1+\alpha}\,] & [\,\frac{1}{9-\alpha}\,;\,\frac{1}{2+6\alpha}\,] & [\,\frac{1}{3-\alpha}\,;\,\frac{1}{1+\alpha}\,] & [\,1\,;\,1\,] \end{bmatrix}.$$

The α-cuts, for $\alpha = 1$, of the elements of the FPC matrix $\tilde{A}$ consist of the crisp elements of the PC matrix A. Notice that the elements located above the main diagonal of the FPC matrix $\tilde{A}$ are triangular elements with the piece-wise linear membership functions, whereas the elements located under the diagonal are reciprocal, hence the corresponding membership functions are non-linear.

Consider the priority vector $w^* = (1.515, 1.344, 0.975, 0.504)$. Evidently, $\tilde{A}$ is α-·-consistent FPCM for $\alpha = 0$. Moreover, it can be verified that $\tilde{A}$ is α-·-consistent FPCM, if and only if $\alpha \in [0; 0.117]$, otherwise, $\tilde{A}$ is not α-·-consistent. It can be easily verified that $\tilde{A}$ is α-·-coherent for all $\alpha \in [0; 1]$, i.e. by Definition 3, $\tilde{A}$ is ·-coherent.

6 Deriving priority vectors of FPC matrices with the desirable properties

In this section, we present an Algorithm to generate a *crisp* priority vector, therefore no defuzzification is necessary for the final ranking of the elements, i.e. criteria or alternatives $c_1, c_2, \ldots, c_n \in C$. The Algorithm consists of the following six steps:

STEP 1. Choose radius $r \in G$, for example $r := \max\{\|a_{ij}^R(0)\| \mid i,j \in \mathcal{N}\}$, and set $\alpha := 0$. Find an optimal solution w^α to Problem 1:

$$E^{Cs}(\tilde{A}, w, \alpha) \longrightarrow \min \qquad \text{subject to} \quad w \in B(r). \tag{28}$$

If the minimal value of the objective function $E^{Cs}(\tilde{A}, w^\alpha, \alpha) > e$, then there is no 0-consistent vector; go to Step 3. Otherwise, if the minimal value of the objective function $E^{Cs}(\tilde{A}, w^\alpha, \alpha) \le e$, then there exists a 0-consistent vector. Look for an α-consistent vector with the maximal $\alpha \in [0;1]$, i.e. proceed with Step 2.

STEP 2. Find an optimal solution α^*, w^{α^*} to Problem 2:

$$\alpha \longrightarrow \max \qquad \text{subject to} \quad E^{Cs}(\tilde{A}, w, \alpha) \le e, \quad w \in B(r), \quad \alpha \in [0;1]. \tag{29}$$

The optimal solution $w^{\alpha^*} \in B(r)$ is an α^*-consistent priority vector such that $\alpha^* \in [0;1]$ is maximal. At the same time, it is an α^*-intensity vector, such that $E^{In}(\tilde{A}, w^{\alpha^*}, \alpha^*) < e$. Look for an α-intensity vector with the maximal $\alpha \in [0;1]$, i.e. go to Step 4.

STEP 3. Set $\alpha := 0$. Find an optimal solution w^α to Problem 3:

$$E^{In}(\tilde{A}, w, \alpha) \longrightarrow \min \qquad \text{subject to} \quad w \in B(r). \tag{30}$$

If the minimal value of the objective function $E^{In}(\tilde{A}, w^\alpha, \alpha) \ge e$, then there is no 0-intensity vector; go to Step 5. Otherwise, if the minimal value of the objective function $E^{In}(\tilde{A}, w^\alpha, \alpha) < e$, then there exists a 0-intensity vector. Look for an α-intensity vector with the maximal $\alpha \in [0;1]$, i.e. proceed with Step 4.

STEP 4. Find an optimal solution $\alpha^{**}, w^{\alpha^{**}}$ to Problem 4:

$$\alpha \longrightarrow \max \qquad \text{subject to} \quad E^{In}(\tilde{A}, w, \alpha) < e, \quad w \in B(r), \quad \alpha \in [0;1]. \tag{31}$$

The optimal solution $w^{\alpha^{**}} \in B(r)$ is an α^{**}-intensity priority vector such that $\alpha^{**} \in [0;1]$ is maximal. At the same time, it is an α^{**}-coherent vector, such that $E^{Co}(\tilde{A}, w^{\alpha^{**}}, \alpha^{**}) < e$. Look for an α-coherent vector with the maximal $\alpha \in [0;1]$, i.e. go to Step 6.

STEP 5. Set $\alpha := 0$. Find an optimal solution w^α to Problem 5:

$$E^{Co}(\tilde{A}, w, \alpha) \longrightarrow \min \qquad \text{subject to} \quad w \in B(r). \tag{32}$$

If the minimal value of the objective function $E^{Co}(\tilde{A}, w^\alpha, \alpha) \ge e$, then there is no 0-coherent vector; change some elements of the FPCM $\tilde{A}$ and go to Step 1. Otherwise, if the minimal value of the objective function $E^{Co}(\tilde{A}, w^\alpha, \alpha) < e$, then there exists a 0-coherent vector. Look for an α-coherent vector with the maximal $\alpha \in [0;1]$, i.e. proceed with Step 6.

STEP 6. Find an optimal solution $\alpha^{***}, w^{\alpha^{***}}$ to Problem 6:

$$\alpha \longrightarrow \max \qquad \text{subject to} \quad E^{Co}(\tilde{A}, w, \alpha) < e, \quad w \in B(r), \quad \alpha \in [0;1]. \tag{33}$$

The optimal solution $w^{\alpha^{***}} \in B(r)$ is an α^{***}-coherent priority vector such that $\alpha^{***} \in [0;1]$ is maximal.

END.

The proposed Algorithm provides three priority vectors with the desirable properties, i.e. the α^*-consistent vector w^{α^*}, the α^{**}-intensity vector $w^{\alpha^{**}}$, and the α^{***}-coherent vector $w^{\alpha^{***}}$, such that their membership grades α^*, α^{**}, and α^{***} are maximal and non-decreasing $(0 \le \alpha^* \le \alpha^{**} \le \alpha^{***} \le 1)$.

7 Conclusion

In this paper we propose the method for deriving the priority vector consisting of six steps. With respect to natural logical requirements, we reformulated "desirable properties" of FPC matrix when compared with [4]. Then, an algorithm was proposed to obtain the priority vectors with the newly formulated properties. Such an approach is more natural from the DM perspectives and enables us also to extend various MCDM approaches known from the literature.

Acknowledgements

This research has been supported by GAČR project No. 21-03085S.

References

[1] Cavallo, B. & D'Apuzzo, L. (2009). A general unified framework for pairwise comparison matrices in multicriteria methods. *International Journal of Intelligent Systems*, 24, 377–398.

[2] Liu, Y., Eckert, C. M. & Earl, C. (2020). A review of fuzzy AHP methods for decision-making with subjective judgements. *Expert Systems with Applications*, 161, 113738.

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

[4] Ramík, J. (2020). *Pairwise comparisons method*: *Theory and Applications in Decision Making*. Switzerland, Cham—Heidelberg—New York—Dordrecht—London: Springer Internat. Publ. 253 pp.

[5] Ramík, J. (2020). Desirable Properties of Weighting Vector in Pairwise Comparisons Matrix With Fuzzy Elements. In S. Kapounek & H. Vránová (Eds.), *38th International Conference on Mathematical Methods in Economics* (pp. 481–487). Brno: Mendel University Brno, Czech Republic.

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

Factors Affecting the Level of Housing Expenditures in V4 Countries

Jitka Bartošová[1], Vladislav Bína[2]

Abstract. Currently, the substantial inflation growth of prices leads to the continuous increase in the expenditures of households, of which a significant proportion is caused by the increase in housing costs. This paper aims to identify relevant factors affecting the housing costs in so-called Visegrád Group or V4 countries (Czech Republic, Slovakia, Poland, and Hungary) and to quantify the intensity of their influence. To perform this task, regression models were developed providing a possibility to explore the similarities and differences in the effect of particular factors on the housing costs in this group of post-communist countries. The analysis uses the newest available data from the broad sample survey of incomes and living conditions of households, the EU-SILC 2019 survey. The model parameters were estimated using the R software.

Keywords: EU SILC, logistic regression model, housing costs

JEL Classification: C35, I31, R21
AMS Classification: 62J12

1 Housing affordability

Housing affordability is a broad notion incorporating the financial affordability of purchasing the property and some measure of satisfaction with its quality, but also the cost burden of household connected with the use of the property. Housing affordability can thus be assessed either in an objective way (employing a financial burden) or in a subjective manner (measuring individual satisfaction). The notion of housing affordability according to price is a market concept connected with the solvency of households. The affordability of housing is in this concept given by the disposable finance of the household.

Housing costs are among the most important expenditures in the financial balance of households. The high burden given by the housing cost can lead to severe economic difficulties and significantly decrease the well-being of households in the long term. Households are pushed to decrease other important constituents of their budget like expenditures on health, education, food, clothing, etc. Thus, housing costs are one of the main determinants of poverty and the socio-economic suffering of households [1]. The affordability of housing according to its price is now topical. The high inflation (among other factors caused by a significant increase in energy prices) makes for numerous households the question of housing affordability crucial.

1.1 Measurement of housing affordability

Measurement of housing affordability is always based on a choice of some limit or threshold distinguishing the burden loaded on the household budgets into bearable and unbearable. On the international level, the housing affordability measurement is based on two main concepts: ratio and residual [8].

Indicators of housing affordability are based on a relation between the level of disposable household incomes and total costs of housing including water consumption and energies. The ratio concept is based on the proportion of incomes and the costs related to housing and this provides a basis for the expression of measure of housing cost burden. Expenditures exceeding a certain threshold are considered a disproportional burden. According to the Eurostat definition, the household is considered disproportionally burdened by the housing costs if such costs (not considering the housing benefits) exceed 40 percent of the household's disposable income. On contrary, the residual approach is based on the difference between household incomes and housing costs.

The ratio concept uses an objective measure of the housing cost burden, namely the HCB (**Household Cost Burden**) indicator given by the ratio of the total monthly cost of housing $HH070$ (**Total housing cost (including electricity, water, gas, and heating)**) multiplied by 12 and decreased by $HY070G$ (**Gross housing allowances**) and yearly disposable income $HY020$ (**Total disposable household income**) again decreased by $HY070G$ (**Gross housing allowances**). Therefore:

[1] Prague University of Economics and Business, Faculty of Management, Jarošovská 1117/II, 37701 J. Hradec, jitka.bartosova@vse.cz.
[2] Prague University of Economics and Business, Faculty of Management, Jarošovská 1117/II, 37701 J. Hradec, vladislav.bina@vse.cz.

$$HCB = \frac{HH070 \cdot 12 - HY070G}{HY020 - HY070G} \cdot 100\ \% \qquad (1)$$

All variables employed in Formula (1) are contained among the indicators assessed every year in a large sample survey EU-SILC (European Union Statistics on Income and Living Conditions). Therefore, it is possible to use this data for the calculation of household cost burden. This relative concept of housing cost measurement is a base for the definition of the indicator **Housing cost overburden rate** ($HH_{\text{OVERBURDEN}}$) (Mulliner, 2012) which calculates the percentage of inhabitants living in a household where the total housing costs (without gross housing allowances) exceed 40% of the total disposable incomes (again without gross housing allowances):

$$HH_{\text{OVERBURDEN}} = \frac{\sum_{\forall\, i\ in\ relevant\ breakdown\ with} RB050_i}{\sum_{\forall\, i\ in\ the\ same\ breakdown} RB050_i} \cdot 100\% \qquad (2)$$

The indicator $HH_{\text{OVERBURDEN}}$ thus providing a possibility to compare the share of individuals burdened by housing costs above a certain threshold throughout the EU countries.

2 Analysis of individuals burdened by housing costs in V4 countries

2.1 Database and variable choice

The database consists of the cross-sectional part of a survey EU-SILC 2019. This large sample survey is organized every year throughout the countries of the European Union according to a unified methodology and contains a wide scale of relevant indicators allowing the manifold choice of possible factors in analyses of their impact on the measure of housing cost burden in households. Moreover, according to the extent and design of this sample survey, it is possible to generalize the results of analyses based on this database.

The choice of variables comes from an assumption that the housing cost burden of households is influenced not only by financial factors like incomes and expenditures of households, amount of rent, etc. but also by a wide scale of other factors – characteristics of households and their members, regional, demographical and temporal factors, etc. For example, Li in [6] concludes in his paper that the problem of housing affordability is interconnected with multi-faceted economic, social, political, and demographic aspects. The previous research (among other conclusions) resulted in a finding that one of the important factors related to the higher housing cost burden of individuals and households is the life phase, ie. the **age** of household members [7]. The reason is that at higher age a transition occurs from the economically active phase of life to the economically inactive period. This is usually related to a decrease in disposable incomes. As a consequence, the households of seniors are forced to decrease total expenditures and thus also limit their housing costs [3, 7]. Several other factors significantly influencing the probability of high housing cost burden include the **Household type** [2] and **Tenure status** [5]. A reference category in the model is usually the first variant, in case of region it is a capital (or central region), in other cases it was in a way default or the most frequent possibility. In case of age, the "active" category was chosen as the reference similarly to status of economic activity. Let us remark, that the categories can be easily rearranged.

Choice of variables for the dependence model:

<u>**Explained variable (response):**</u>

- **$HH_{\text{OVERBURDEN}}$: Housing cost overburden rate**:
  – 0 – $HCB \leq 40\%$ (<u>reference category</u>)
  – 1 – $HCB > 40\%$

<u>**Explanatory variables (factors):**</u>

- **$HY020$: Total disposable household income (in EUR thousands)**
- **$DB040$: Region NUTS (CZ – level 2; HU and PL – level 1):**
  – CZ01 – Prague (Praha, <u>reference category</u>)
  – CZ02 – Central Bohemia (Střední Čechy)
  – CZ03 – Southwest (Jihozápad)
  – CZ04 – Northwest (Severozápad)
  – CZ05 – Northeast (Severovýchod)
  – CZ06 – Southeast (Jihovýchod)
  – CZ07 – Central Moravia (Střední Morava)
  – CZ08 – Moravian-Silesian (Moravskoslezsko)

- PL4 – North-west macro-region (Makroregion północno-zachodni)
- PL5 – South-west macro-region (Makroregion południowo-zachodni)
- PL6 – Northern macro-region (Makroregion północny)
- PL7 – Central macro-region (Makroregion centralny)
- PL8 – Eastern macro-region (Makroregion wschodni)
- PL9 – Mazovian voivodeship macro-region (Makroregion województwo mazowieckie)
- HU1 – Central Hungary (Közép-Magyarország, reference category)
- HU2 – Transdanubia (Dunántúl)
- HU3 – Great Plain and North (Alföld és Észak)

- **$DB100$: Degree of urbanization:**
  - 1 – densely-populated area (reference category)
  - 2 – intermediate area
  - 3 – thinly-populated area
- **$HH010$: Dwelling type (DW_T):**
  - 1 – detached house (reference category)
  - 2 – semi-detached or terraced house
  - 3 – apartment or flat in a building with less than 10 dwellings
  - 4 – apartment or flat in a building with 10 or more dwellings
- **$HX060$: Household type:**
  - 5 – single person (reference category)
  - 6 – two adults younger than 65 years
  - 7 – two adults, at least one aged 65 years or over
  - 8 – households without dependent children
  - 9 – single person with dependent children
  - 10 – two adults with one dependent child
  - 11 – two adults with two dependent children
  - 12 – two adults with three dependent children
  - 13 – other households without dependent children
- **$HX070$: Tenure state (TENURE_STAT):**
  - 1 – owner or free (reference category)
  - 2 – full or reduced rents
- **$PX050$: Status of economic activity:**
  - 2, 3, 4 – economically active (reference category)
  - 5 – unemployed
  - 6 – retired
  - 7, 8 – economically inactive
- **$RX020$: Age:**
  - 26 – 55 – productive age adults (reference category)
  - 0 – 25 – young
  - 56 – 65 – pre-seniors
  - 66 – – seniors
- **$PE040$: Highest education level:**
  - 1 – none or primary (reference category)
  - 2 – secondary or post-secondary
  - 3 – tertiary

2.2 Goals and methodology of research

This paper aims to analyze the influence of individual and household characteristics on the high housing cost burden (HCB > 40%). The primary aim is to define relevant factors influencing the housing cost burden of households in particular countries of the V4 group, i.e. the Czech Republic, Slovakia, Poland, and Hungary and quantify the intensity of their influence on the individual overburden. The complementary aim is to compare the results in V4 countries and to reveal national differences in the effect of particular factors.

For this purpose, logistic regression models for all V4 countries were constructed. The analysis employs the newest data available for all four countries from the European sample survey of incomes and living conditions of households, the EU-SILC 2019 survey. The model parameters were estimated using the R statistical software.

Since the dichotomous variable HCB is a standard measure of individual overburden by housing costs, it serves as an explained (response) variable in a logistic regression model as a method for parameter estimation. In our case, we chose as a response the variable HCB in a binomial form which informs, whether the ratio of housing costs to incomes is higher than 40 percent (HCB value 1) or not (HCB value 0). If we denote the probability of an observed event, i.e. $P(HCB = 1) = \pi$, then the complementary event has a probability $1 - \pi$ and their ratio represents the odds that the share of housing costs does not exceed the given threshold. The relation between conditional probability $P(HCB = 1|X_1, X_2, \ldots, X_k) = \pi$ and predictors are not linear and the estimation of parameters is based on linearizing transformation – namely using the natural logarithm of odds, the so-called *logit*. For k explanatory variables (quantitative, or binary) $X_1, X_2, \ldots, X_k$ the linearized model is

$$logit = \ln(odds) = ln\frac{\pi}{1-\pi} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k. \tag{3}$$

Since we model quantities derived from probabilities (not individual values of response), the model does not contain an error term. For the interpretation of the influence of selected factors on the existence or non-existence of overburden by excessive housing cost the linearized variant (3) is difficult to use. The interpretation is more straightforward after a transformation of the model to the shape for the estimation of *odds* described by the formula

$$odds = \frac{\pi}{1-\pi} = \exp(\beta_0) \cdot \exp(\beta_1)^{X_1} \cdot \exp(\beta_2)^{X_2} \cdot \ldots \cdot \exp(\beta_k)^{X_k}. \tag{4}$$

The interpretation uses *odds ratio* $OR = \frac{odds_1}{odds_2}$ where $odds_1$ and $odds_2$ are probabilities that the given event takes place for the first or the second of the compared objects. For categorical predictor, one category is chosen as a reference, and parameters of the model (4) represent average coefficients of growth/decrease of odds relative to the reference category. For the quantitative predictor, the reference value decreased by one unit and the parameters of the model (4) represent average coefficients of growth/decrease for an increase of predictor by one unit.

Parameters of generalized linear models, thus also in the case of the logistic regression function, are estimated using the maximum likelihood principle (maximization of likelihood function) which also considerably determines the methods used for assessing the quality and significance of the models. Significance tests of predictors in the model are usually performed using the Wald statistics which allows for testing categorical predictors as a whole. Qualitative predictors are tested as multidimensional variables with degrees of freedom given by the number of categories. Another suitable test statistic is a likelihood ratio based on the values of the likelihood function.

The quality of generalized models (particularly in the model of logistic regression) is in general given by the value of the logarithm of the likelihood function. Several different evaluation criteria are based on this measure like deviance, pseudo-R-squared, and Hosmer-Lemeshow statistics. The assessment of the quality of the generalized linear model during the process of changing the number of predictors is facilitated by information criteria (Akaike – AIC, or Schwarz – BIC). For more detail, see, e.g., [4]. The quality of the model can be assessed also using AUC statistics (Area under the Receiver operating character curve) based on the sensitivity and specificity of an estimate and used to evaluate the overall ability of the model to predict the explained variable.

2.3 Results achieved

The analysis of factors influencing significantly the overburden of individuals by housing cost is performed using the logistic regression models. The models were constructed for particular countries of the V4 group, namely the Czech Republic, Slovakia, Poland, and Hungary. The calculation of the maximum likelihood estimates was performed using the *glm* procedure within the R software which employs an iterative procedure of Fisher scoring. Before the estimation outliers were removed from the data file resulting in values of HCB logically restricted to the interval ⟨0,100⟩. The construction results are summarized in Tables 1 and 2. The quality of estimated regression models is assessed using deviance and pseudo R2 coefficient. For results see Table 1.

State	Null deviance	Degrees of freedom	Residual deviance	Degrees of freedom	Pseudo R^2
CZ	7734.7	15890	4509.6	15862	0.4170
SK	4806.3	12598	3309.2	12577	0.3115
PL	15867	32329	9179.0	32299	0.4215
HU	3647.6	12780	2464.3	12757	0.3244

Table 1 The assessment of the quality of logistic regression models

Logistic regression	CZ 2019		SK 2019		PL 2019		HU 2019	
Characteristic	**exp(β)**	**p-value**	**exp(β)**	**p-value**	**exp(β)**	**p-value**	**exp(β)**	**p-value**
Intercept	21.12	<0.001	5.42	<0.001	11.82	<0.001	0.87	0.761
***HY*020**	0.68	<0.001	0.77	<0.001	0.52	<0.001	0.61	<0.001
***DB*040**								
CZ02	0.45	<0.001						
CZ03	0.23	<0.001						
CZ04	0.41	<0.001						
CZ05	0.25	<0.001						
CZ06	0.51	<0.001						
CZ07	0.36	<0.001						
CZ08	0.26	<0.001						
PL2					—	—		
PL4					0.93	0.491		
PL5					0.90	0.324		
PL6					0.66	<0.001		
PL7					0.45	<0.001		
PL8					0.32	<0.001		
PL9					0.54	<0.001		
HU1							—	—
HU2							0.53	<0.001
HU3							0.42	<0.001
***DB*100**								
2	0.69	<0.001	0.45	<0.001	0.75	<0.001	0.74	0.046
3	0.35	<0.001	0.21	<0.001	0.47	<0.001	0.50	<0.001
***HH*010**								
2	1.04	0.797	0.17	<0.001	1.21	0.107	0.76	0.293
3	0.79	0.080	0.69	0.048	0.55	<0.001	0.86	0.584
4	0.70	0.003	0.50	<0.001	0.78	0.002	0.69	0.027
***HX*060**								
6	1.49	0.018	1.02	0.895	1.77	<0.001	1.58	0.009
7	0.68	0.776	0.21	<0.001	1.32	0.002	0.91	0.657
8	1.14	0.776	0.30	<0.001	2.25	<0.001	1.82	0.030
9	1.73	0.002	0.96	0.860	1.68	<0.001	0.96	0.894
10	2.08	<0.001	1.01	0.984	2.56	<0.001	1.30	0.329
11	2.53	<0.001	1.52	0.044	2.77	<0.001	3.03	<0.001
12	5.75	<0.001	1.54	0.114	5.05	<0.001	1.22	0.670
13	3.25	0.017	0.78	0.273	2.97	<0.001	1.23	0.607
***HX*070**								
2	6.36	<0.001	2.01	<0.001	9.03	<0.001	10.59	<0.001
***RX*020**								
0-25	0.83	0.275	0.76	0.105	1.48	<0.001	0.94	0.786
56-65	0.79	0.146	0.39	<0.001	0.90	0.268	0.83	0.215
65-	0.51	0.001	0.35	<0.001	0.77	0.027	0.46	<0.001
***PE*040**								
sec. & posts.					1.84	<0.001	2.25	0.047
terc.					3.00	<0.001	4.62	<0.001
***PX*050**								
5	0.50	0.007	2.27	<0.001	0.59	<0.001		
6	1.19	0.380	1.21	0.438	0.87	0.149		
7, 8	0.96	0.802	1.36	0.039	0.72	<0.001		

Table 2 Transformed values of parameter estimates in logistic regression models and their significance

Values summarized in Table 2 present the most important characteristics of the four logistic regression models in a transformed variant $exp(\beta)$ corresponding to the form of model usable for the estimation of odds (Formula 4). The assessment of the statistical significance of predictors incorporated in the model the Wald statistics was used. Table 2 contains corresponding p-values for model parameters.

3 Conclusion

The analysis performed in the paper documented that in all V4 group countries exist several factors influencing significantly the probability that an individual is overburdened by the housing costs (the housing cost of a household exceeds 40 percent of its income). Among the set chosen at the beginning, the only influence of gender appeared to be insignificant in all countries. Then in the Czech Republic and Slovakia also the Highest education level (*PE040*) appeared insignificant and in Hungary, it was the Status of economic activity (*PX050*). All remaining factors showed to be significant in all V4 group countries, but let us remark that quite naturally not all categories of qualitative variables are significantly different from the reference level. Among the significant factors appearing in models for all countries ranks particularly **Total disposable household income** (*HY020*). With the growing disposable income, the odds of passing the threshold of 40% proportion of housing costs on household incomes decreases. Namely, with the 1000 EUR increase, these odds decrease in the Czech Republic (CZ) at about 32% on average, in Slovakia (SK) at about 23%, in Poland (PL) at about 48%, and in Hungary (HU) about 39%. The characteristic of **Region NUTS 1 or 2** (*DB040*) leads also to a statistically significant decrease in the endangerment of individuals by a high proportion of housing costs in comparison with the reference category of Prague (CZ), South macro-region (PL), and Central Hungary (HU). Data from Slovakia do not contain information about the NUTS region. A decrease in comparison with the reference category occurs in all V4 countries also in the case of variable **Degree of urbanization** (*DB100*). As could be expected, the increase in comparison with the reference category (owner or free) occurs in the second category (full or reduced rents) in the variable **Tenure state** (*HX070*). An increase against the reference category (none or primary) occurs in Poland and Hungary in the case of higher education levels in the variable **Highest education level** (*PE040*). In the case of the Czech Republic and Slovakia is this factor insignificant. The remaining variables show different behavior in different countries possibly reflecting national differences.

Acknowledgements

The second author was supported by grant project 19-00015S of the National Science Foundation of the Czech Republic (GAČR). The EU-SILC data were provided under project RPP 289/2018-EU-SILC.

References

[1] Deidda, M. (2013) Economic hardship, housing cost burden and tenure status: evidence from EU-SILC. *Working Papers* 2013/23. Cagliari: CUEC 2013.

[2] Elmelech, Y. (2004) Housing inequality in New York City: Racial and ethnic disparities in homeownership and shelter-cost burden. *Housing, Theory, and Society*, 2004, 21(4), pp. 163–175. DOI: 10.1111/j.1533-8525.2001.tb00028.x.

[3] Godarová, J., Trbola, R., Jahoda, R. & Rákoczyová, M. (2017) *Dostupnost bydlení a stávající bytová situace seniorů*. Praha: Výzkumný ústav práce a sociálních věcí, v. v. i., 2017.

[4] Hosmer, D. V. & Lemeshow, S. (2000) *Applied Logistic Regression*. New York: John Wiley & Sons, 2000.

[5] Labudová, V. & Sipková, L. (2019) Housing Affordability in Slovakia: what Factors Affect it? *Statistika*, 99 (4). 399 –416.

[6] Li, J. (2014) Recent Trends on Housing Affordability Research: Where are we up to? [online]. Urban Research Group – CityU on Cities *Working Paper series*, WP No. 5/2014. [cit. 15.10.2019].

[7] McConnell, E. D. (2013) Who has housing affordability problems? Disparities in Housing Cost burden by Race, Nativity and Legal Status in Los Angeles. *Race and social problems*, 2013, 5(3), pp. 173–190. DOI: 10.1007/s12552-013-9086-x.

[8] Mulliner, E. K. (2012) *A model for the complex assessment of sustainable housing affordability: Doctoral thesis* [online]. Liverpool: John Moores University, 2012.

Application of L-Moments into Labour Market Data Analysis: Czech Sectors with the Lowest Earnings

Diana Bílková[1]

Abstract. The aim of this paper is to demonstrate the application of the L-moment method to labour market data. In the past, this method of point estimation of parameters was used mainly in the field of climatology, meteorology, or hydrology, for example in connection with the study of extreme precipitation. This paper deals with the use of the L-moment method in order to estimate the parameters of three- parameter lognormal curves used to model the distribution of monthly earnings (wages and salaries together) in the two sectors of the Czech economy with the lowest level of earnings. These are Accommodation and Food Service Activities and Administrative and Support Service Activities sectors. The level of earnings in the Administrative and Support Service Activities sector is slightly higher than the level of earnings in the Accommodation and Food Service Activities sector. This procedure succeeded in capturing the development of the distribution of earnings in these two sectors of the Czech economy during the years 2009–2020.

Keywords: three-parameter lognormal curve, L-moments of probability distribution, sample L-moments, sectors with the lowest earnings, earning distribution model

JEL Classification: E24, C51
AMS Classification: 62P20, 91B39

1 Introduction

The method of L-moments of point estimation of parameters ([5], [6]) has been used in the past, for example, in connection with the study of extreme precipitation [8] or monthly precipitation [4]. Studies [10], [7], [3], [9] and [2] examine the issue of L-moments in connection with flood frequency and rainfall frequency analyses. In connection with the distribution of income, for example, the L-moments method was used by a team of authors [1].

This paper deals with the application of the L-moment method to economic data from the labour market. The L-moment method was used for point estimation of parameters of three-parameter lognormal curves representing models of distribution of monthly earnings (wages and salaries together) of employees in two sectors of the Czech economy with the lowest earnings in the period 2009 –2020. There are Accommodation and Food Service Activities and Administrative and Support Service Activities sectors.

Year	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**
Minimum wage amount	8,000	8,000	8,000	8,000	8,208	8,500
Year	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**
Minimum wage amount	9,200	9,900	11,000	12,200	13,350	14,600

Source: www.mpsv.cz

Table 1 Development of the minimum wage amounts in 2009–2020[2]

The importance of the lognormal distribution as a model for sample distributions cannot be questioned. This model has found application in various fields, from astronomy, through technology, economics to sociology. The characteristic features of the process described by the lognormal model are gradual action of interdependent factors, the tendency to develop in geometric sequence and the growth of random variability into systematic variability, i. e. differentiation.

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

[2] In 2013, the minimum wage was CZK 8,000 in the period from the 1st January to the 31st July and CZK 8,500 in the period from the 1st August to the 31st December. The amount of the minimum wage in Table 1 represents the average minimum wage in 2013.

The beginning of these curves was estimated using the minimum wage in the corresponding year. The development of the minimum wage amounts in the period under review is shown in Table 1. However, the question of the suitability of a given curve for the earning distribution model is not a completely common mathematical- statistical problem, in which we test the null hypothesis “H_0: Random sample comes from the assumed theoretical distribution” against the alternative hypothesis “H_1: non H_0”, since in the case of the distribution of earnings, typical samples are huge (in this case tens to hundreds of thousands of respondents), therefore the goodness-of- fit test always leads to the rejection of the null hypothesis about the assumed distribution. This is due not only to the fact that at such large sample sizes, the power of the test is so strong that the goodness-of-fit test reveals all the slightest deviations of the actual distribution of earnings and model, but also from the principle of the test. However, we are not practically interested in small deviations, so only the approximate conformity of the model with reality suffices. When evaluating the suitability of the model, it is necessary to proceed to a large extent subjectively and rely on logical analysis and experience. For this reason, goodness-of-fit tests are not performed in this study.

The main aim of this research is to present the application of the L-moment method to economic data, specifically to data from the labour market. Another objective is to compare the development of the earnings distributions of two Czech economy sectors with the lowest level of earnings.

The data come from the official website of the Czech Statistical Office. These were data in the form of interval frequency distribution with unequally wide earning intervals and with extreme open intervals. The data were processed using a Microsoft Excel spreadsheet and the SPSS statistical programming environment.

2 L-Moments

L-moments represent an alternative system describing the shape of the probability distribution. They are an analogy of conventional moments, but they can be estimated on the basis of a linear combination of order statistics, i. e. L-statistics.

Let X be a random variable having a continuous distribution with a distribution function $F(x)$ and a quantile function $x(F)$. Let $X_1, X_2,\dots, X_n$ be a random sample of size n from this distribution. Then $X_{1:n} \le X_{2:n} \le \dots \le X_{n:n}$ are the order statistics of random sample of size n, which comes from the distribution of a random variable X.

2.1 L-Moments of Probability Distribution

L-moment of the r-th order of the random variable X is defined

$$\lambda_r = \frac{1}{r}\cdot \sum_{j=0}^{r-1} (-1)^j \cdot \binom{r-1}{j} \cdot E(X_{r-j:r}), \quad r = 1, 2, \dots, \tag{1}$$

the expected value of the order statistic has the form

$$E(X_{r:n}) = \frac{n!}{(r-1)!\cdot(n-r)!}\cdot \int_0^1 x\cdot[F(x)]^{r-1}\cdot[1-F(x)]^{n-r} \mathrm{d}F(x). \tag{2}$$

The first four L-moments of the probability distribution are now defined

$$\lambda_1 = E(X_{1:1}) = \int_0^1 x(F)\,\mathrm{d}F, \tag{3}$$

$$\lambda_2 = \frac{1}{2}E(X_{2:2} - X_{1:2}) = \int_0^1 x(F)\cdot(2F-1)\,\mathrm{d}F, \tag{4}$$

$$\lambda_3 = \frac{1}{3}E(X_{3:3} - 2X_{2:3} + X_{1:3}) = \int_0^1 x(F)\cdot(6F^2 - 6F + 1)\,\mathrm{d}F, \tag{5}$$

$$\lambda_4 = \frac{1}{4}E(X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}) = \int_0^1 x(F)\cdot(20F^3 - 30F^2 + 12F - 1)\,\mathrm{d}F. \tag{6}$$

The probability distribution can be specified by its L-moments even if some of its conventional moments do not exist, but the opposite is not true. It is often appropriate to standardize higher L-moments λ_r, $r \ge 3$ to be independent of the unit of measure of random variable X. The ratio of L-moments of random variable X is defined

$$\tau_r = \frac{\lambda_r}{\lambda_2}, \quad r = 3, 4, \ldots. \tag{7}$$

Using equations (3)–(5) and equation (7), we obtain equations for the case of a three-parameter lognormal distribution[4]

$$\lambda_1 = \theta + \exp\left(\mu + \frac{\sigma^2}{2}\right), \tag{8}$$

$$\lambda_2 = \exp\left(\mu + \frac{\sigma^2}{2}\right) \cdot \operatorname{erf}\left(\frac{\sigma}{2}\right),^{3} \tag{9}$$

$$\tau_3 = \frac{6}{\sqrt{\pi} \cdot \operatorname{erf}\left(\frac{\sigma}{2}\right)} \cdot \int_0^{\sigma/2} \operatorname{erf}\left(\frac{x}{\sqrt{3}}\right) \cdot \exp(-x^2)\,\mathrm{d}x. \tag{10}$$

2.2 Sample L-Moments

Let $x_1, x_2, \ldots, x_n$ be a random sample of size n and $x_{1:n} \leq x_{2:n} \leq \ldots \leq x_{n:n}$ be an ordered sample. Then the r-th sample L-moment can be defined as

$$l_r = \binom{n}{r}^{-1} \sum_{1 \leq i_1 < i_2 < \ldots < i_r \leq n} \cdots \sum \frac{1}{r} \cdot \sum_{j=0}^{r-1} (-1)^j \cdot \binom{r-1}{j} \cdot x_{i_{r-j}:n}, \quad r = 1, 2, \ldots, n. \tag{11}$$

Hence the first four sample L-moments have a form

$$l_1 = \frac{1}{n} \cdot \sum_i x_i, \tag{12}$$

$$l_2 = \frac{1}{2} \cdot \binom{n}{2}^{-1} \cdot \sum_{i>j}\sum (x_{i:n} - x_{j:n}), \tag{13}$$

$$l_3 = \frac{1}{3} \cdot \binom{n}{3}^{-1} \cdot \sum_{i>j>k}\sum\sum (x_{i:n} - 2x_{j:n} + x_{k:n}), \tag{14}$$

$$l_4 = \frac{1}{4} \cdot \binom{n}{4}^{-1} \cdot \sum_{i>j>k>l}\sum\sum\sum (x_{i:n} - 3x_{j:n} + 3x_{k:n} - x_{l:n}). \tag{15}$$

The natural estimate of the L-moment ratio (7) is the sample L-moment ratio

$$\tau_r = \frac{\lambda_r}{\lambda_2}, \quad r = 3, 4, \ldots. \tag{16}$$

2.3 Parameter estimation

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

$$\begin{aligned} f(x; \mu, \sigma^2, \theta) &= \frac{1}{\sigma \cdot (x-\theta) \cdot \sqrt{2\pi}} \cdot \exp\left[-\frac{[\ln(x-\theta) - \mu]^2}{2\sigma^2}\right], && x > \theta, \\ &= 0, && \text{else.} \end{aligned} \tag{17}$$

Let $\Phi^{-1}(\cdot)$ be distribution function of the standardized normal distribution. Parameter estimates obtained by the L-moment method for the case of a three-parameter lognormal distribution are obtained using the following equations[3] ("L" means L-moment estimation)

[3] Expression erf(z) is the so-called error function

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \cdot \int_0^z \exp(-t^2)\,\mathrm{d}t.$$

$$z = \sqrt{\frac{8}{3}} \cdot \Phi^{-1}\left(\frac{1+t_3}{2}\right),^{4} \tag{18}$$

$$\sigma^{L} \approx 0{,}999\,281z - 0{,}006\,118z^{3} + 0{,}000\,127\,z^{5}, \tag{19}$$

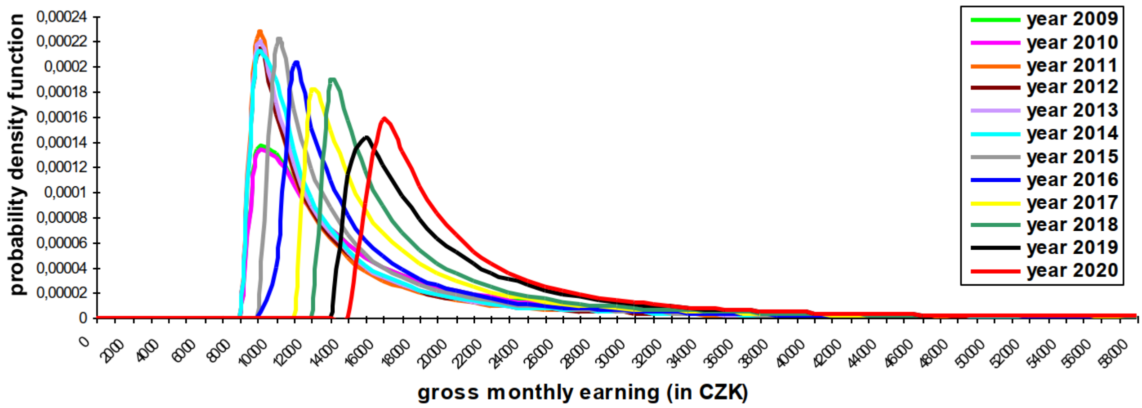
$$\mu^{L} = \ln\left[\frac{l_2}{\operatorname{erf}\left(\frac{\sigma^{L}}{2}\right)}\right] - \frac{\sigma^{2^{L}}}{2}, \tag{20}$$

$$\theta^{L} = l_1 - \exp\left(\mu^{L} + \frac{\sigma^{2^{L}}}{2}\right). \tag{21}$$

See [5].

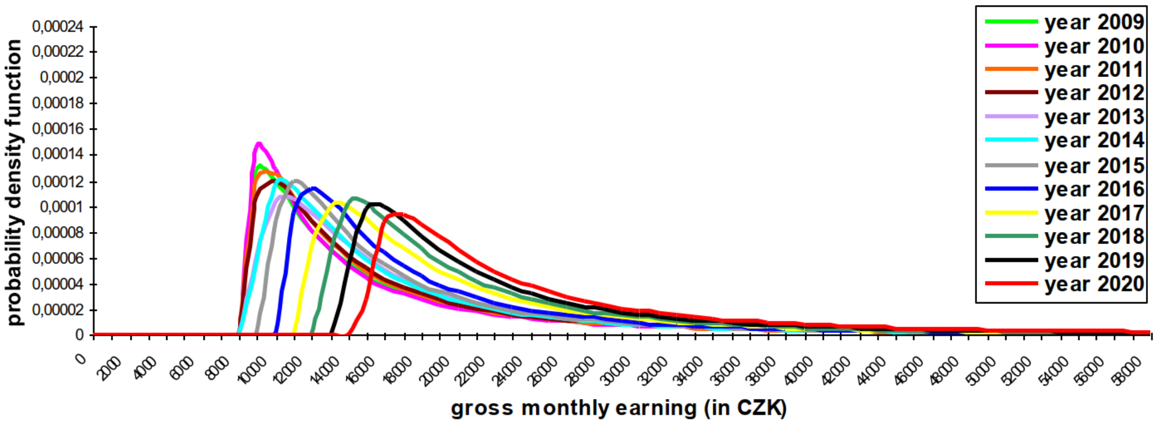
3 Results and Conclusion

Figures 1 –2 offer a comparison of the development of the model distribution of earnings (for wages of employees in the private sphere and salaries of employees in the public sphere together) of the two sectors in which the Czech Republic has the lowest earnings, in the period 2009 –2020. There are Accommodation and Food Service Activities and Administrative and Support Service Activities sectors.



Source: Own construction

Figure 1 Development of model earning distribution of Accommodation and Food Service Activities sector

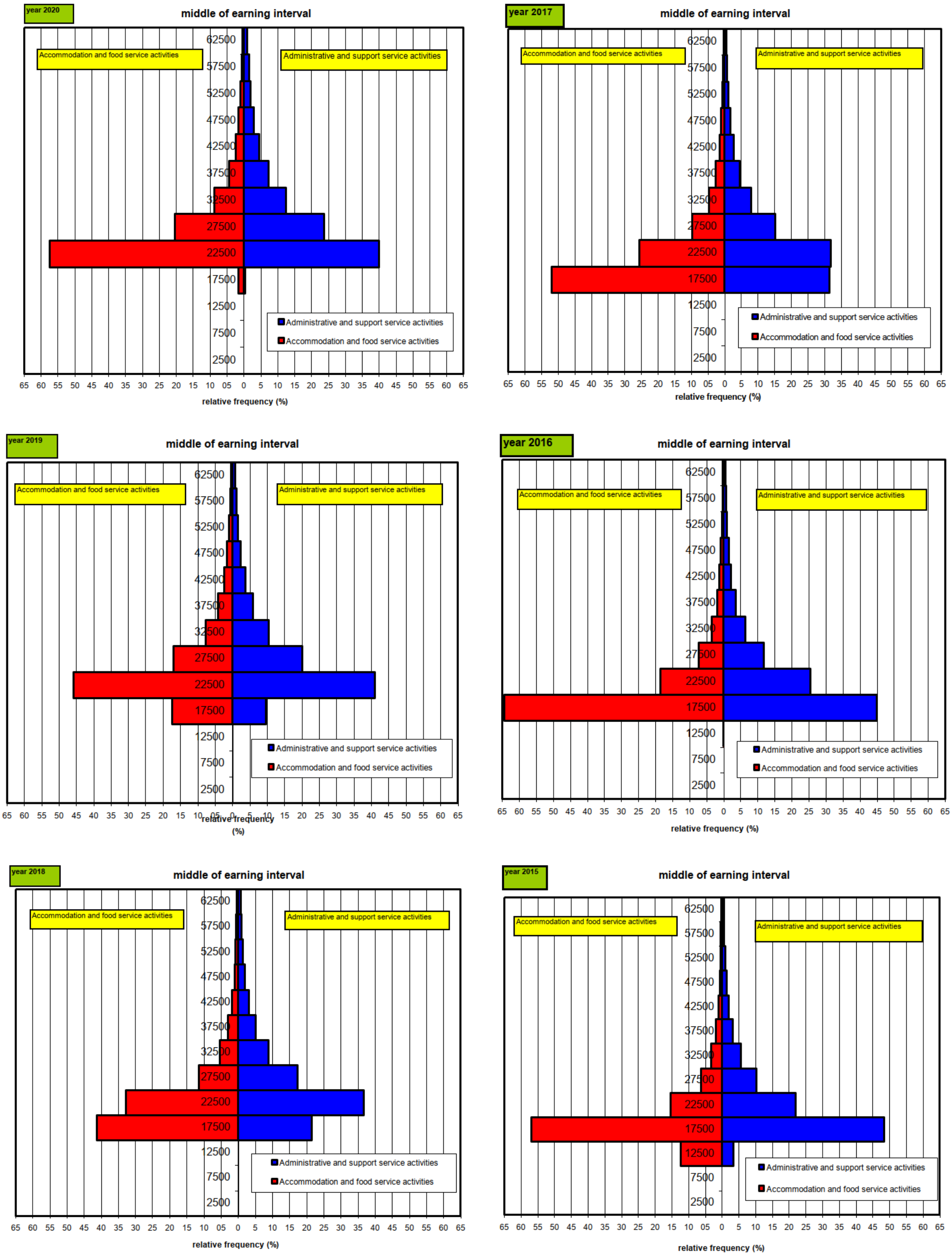


Source: Own construction

Figure 2 Development of model earning distribution of the Administrative and Support Service Activities sector

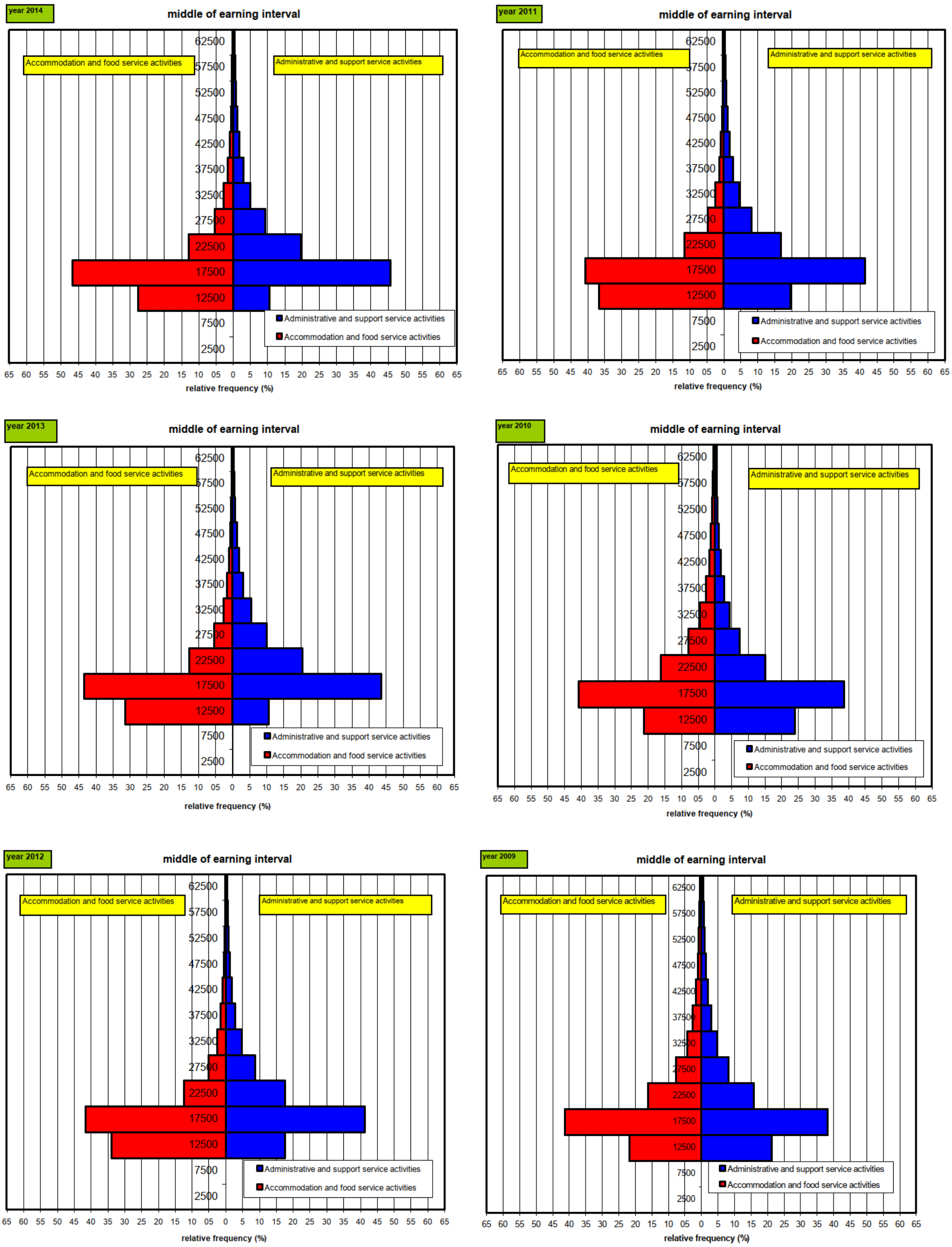
[4] Expression $\Phi^{-1}(\cdot)$ is the quantile function of standard normal distribution.

Figures 3–4 allow a comparison of development of model distribution of earnings between the two sectors over the years 2009–2020. Relative frequencies were calculated based on model distribution of earnings.



Source: Own construction

Figure 3 Comparison of the distribution of relative frequencies (%) of Accommodation and Food Service Activities and Administrative and Support Service Activities sectors during 2015–2020



Source: Own construction

Figure 4 Comparison of the distribution of relative frequencies (%) of Accommodation and Food Service Activities and Administrative and Support Service Activities sectors during 2009–2014

From Figures 1–2, this is clear that the model distributions of earnings in the Accommodation and Food Services Activities sector are characterized by a lower level and variability compared to the model distributions of earnings in the Administrative and Support Service Activities sector. It is also clear that the model distributions of earnings in the Accommodation and Food Service Activities sector are more skewed, and they have more kurtosis than the model distributions of earnings in the Administrative and Support Service Activities sector. In Accommodation and Food Service Activities sector, the frequencies of lower earning intervals are higher, then in Administrative and Support Service Activities sector and vice versa, see Figures 3‑4.

Acknowledgements

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

References

[1] Alam, M., Jan, S. U. & Zeb, A. (2019). Gender Wise Distribution of Income Using L-Moments Method. *Communication, Society and Media*, 2(1), 29–40.
[2] Alahmadi, F. (2017). Regional Rainfall Frequency Analysis by L-Moments Approach for Madina Region, Saudi Arabia. *International Journal of Engineering Research and Development*, 13(7), 39–48.
[3] Eslamian, S. S. & Feizi, H. (2007). Maximum Monthly Rainfall Analysis Using L-Moments for an Arid Region in Isfahan Province, Iran. *Journal of Applied Meteorology and Climatology*, 46(4), 494–503.
[4] Guttman, N. B. (1994). On the Sensitivity of Sample L Moments to Sample Size. *Journal of Climate*, 7(6), 1 026–1 029.
[5] Hosking, J. R. M. (1990). L-moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society (Series B)*, 52(1), 105–124.
[6] Hosking, J. R. M. (2006). On the Characterization of Distributions by Their L-Moments. *Journal of Statistical Planning and Inference*, 136(1), 193–198.
[7] Kumar, R. & Chatterjee, C. (2005). Regional Flood Frequency Analysis Using L-moments for North Brahmaputra Region of India. *Journal of Hydrologic Engineering*, 10(1), 1–7.
[8] Kyselý, J. & Picek, J. (2007). Regional Growth Curves and Improved Design value Estimates of Extreme Precipitation Events in the Czech Republic. *Climate research*, 33(3), 243–255.
[9] Noto, L. V. & La Loggia, G. (2009). Use of L-Moments Approach for Regional Flood Frequency Analysis in Sicily, Italy. *Water Resources Management*, 23(9), 2 207–2 229.
[10] Pearson, C. P. (1991). New Zealand Regional Flood Frequency Analysis Using L-Moments. *Journal of Hydrology (New Zealand)*, 30(2), 53–64.

DSS Capabilities Evaluation Using ANP with Methods for Consistency Improvement

Petra Bláhová[1], Helena Brožová[2]

Abstract. This paper focuses on decision-makers' preferences for Decision Support Systems. A survey questionnaire and ANP method was used for Decision support systems (DSS) capabilities evaluation. Knowledge of DSS capabilities preferences is essential for organizations using, implementing or planning to implement DSS. ANP method is based on the pairwise comparison (Saaty's method). Unfortunately, human judgement is often inconsistent and finding methods to improve inconsistency while keeping the unprecise comparisons representing the preferences of human decisions are essential. Traditionally, the Saaty's method identifies inconsistency and suggests the deduction unless revision can be made with a decision-maker which leads to changing his/her original opinion. Here, three methods for improving consistency were tested. The minimum deviation method, which highly preserves the original decision-maker's preferences while achieving the required consistency was selected for inconsistent pairwise comparisons modification. Easy application of this method shall improve ANP usage in large scale surveys.

Keywords: ANP, DSS, survey, inconsistency improvement, Saaty's matrix

JEL Classification: C69
AMS Classification: 90B50, 90B90

1 Introduction

Integrated automation and informatization affect the organisations' decision-making process [1] and quality management decisions, effective information and analytical support are required. Decision Support Systems (DSS) are intended to improve the quality of decisions by, for example, processing and analysing data and documents, using quantitative models to identify problems, expediting problem-solving and increasing organizational control [13]. DSS implementation reflects the specific needs of an organization and often represents significant investment. Specifically, in large business corporations, managers are the DSS users, and therefore evaluation of their preferences of the DSS capabilities is crucial. Willingness to use DSS reflects various aspects as well as capabilities being aligned with managers' preferences. There are multiple researches focused on how DSS capability effects decision-makers, such as task motivation to use DSS [5], business simulation game where correct and defective DSS are analyzed in terms of trust in automation, usefulness and intention to use [4].

DSS capabilities have been selected for further evaluation based on recent research. *Explanation* capability is covered by several authors, such as how different explanation treatments cause a revision of initial decision [10], and how explanation length affects confidence level [8]. *Cooperation* capability covered, for example: how shared knowledge creation effect strategies [6], the knowledge creation capability of a company [20], and cooperative planning [6]. *Experience* capability described how past experience improves business processes by [7], relevant prior experience information increases confidence by [18], an experiment where the existence of previous experience with similar decision was preferred more than DSS [2]. DSS Impacts have been selected similarly. *Certainty* of a decision-maker covered by [2], where prior experience with a given decision problem resulted in the highest certainty of a decision-maker, [9] higher satisfaction with a decision in relation to automation support system. *Quality* of a decision included as a fundamental DSS related impact. According to [14], DSS are intended to improve accuracy, quality and overall effectiveness of a specific decision or a set of related decisions. *Efficiency* of a decision-making process covered by [9], where automated decision support systems and task performance, [12] examines DSS impact on the organizational intelligence and structure and decision-making speed. While countless research papers focus in-depth on a specific capability and its impact, there is a lack of evaluation regarding managers preferences for DSS capabilities.

Evaluation of DSS capabilities and related impacts 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

[1] CZU Prague, Department of Systems Engineering, blahovap@pef.czu.cz.
[2] CZU Prague, Department of Systems Engineering, brozova@pef.czu.cz.

level of automation as well as the ability to cover a complex problem. Regarding complex relations among DSS capabilities and impacts problem the ANP method is used. The weakness and strength of ANP are pairwise comparisons because the decision-makers are often inconsistent. Data which inconsistency is above a given threshold need to be either excluded, or revisited with the decision makers as per [15], near consistent pairwise comparison matrices (PCM) are essential. Reviewing the original PCM with a decision-maker is often impossible, impractical, or even contra productive since the original judgment is lost. At the same time, the change may not reflect the actual source of the inconsistency. Nevertheless, it is widely accepted that improved inconsistency increases a validity of the results and therefore finding methods to analyze and improve inconsistency for the pairwise comparison matrices have been studied extensively since Saaty introduced AHP in 1970s. While some of the existing methods are complicated and challenging to use when revising the inconsistent comparison matrix, others make it difficult to preserve most of the original comparison information as a new matrix has to be constructed.

The main aim of the paper is to evaluate managers preferences of DSS capabilities and impacts. To evaluate managers' preferences in complexity, ANP model structure was introduced and a questionnaire for including pairwise comparisons was used. Furthermore, several methods for improving consistency were tested, and the minimum deviation method was selected for inconsistent pairwise comparison modification. This method preserves the original decision-maker's preferences while achieving acceptable consistency and is suitable for survey data processing.

The structure of the paper is as follows: the second chapter describes the used materials and methods, including ANP model structure, PCM inconsistency and modification methods for consistency improvement. The third chapter includes results for methods testing and ANP results using the selected method, and the fourth chapter presents a discussion followed by conclusions in chapter five.

2 Methods

2.1 ANP model structure used for DSS evaluation

For this article, three DSS capabilities were chosen based on their potential influence on a decision-maker, the decision itself and the decision process. Also three DSS impacts were defined for this article. Definition of the DSS capabilities and impacts are provided in **Table 1.**

Criteria	**Definition**
DSS Capabilities	
1. Explanations	System can provide supplemental explanations alongside recommended alternatives.
2. Cooperation	System enables cooperation with others, sharing knowledge, transparent visibility.
3. Experience	System recommendations are using prior similar decisions, imitating experience.
DSS Impacts	
1. Certainty	Usage of the system increases decision-maker certainty with his/her decision.
2. Quality	Using the system increases decision quality (adequate, exact).
3. Efficiency	System improves the decision-making process in terms of time and cost.

Table 1: Decision criteria for ANP – Capabilities and Impacts

The first step is to construct a model to be evaluated. The ANP model consists of two clusters (Capabilities and Impacts) with elements connected by their dependence on one another. This structure and impacts with respect to the capabilities aims to evaluate DSS capabilities with respect to the impacts. The ANP network structure and the relationship between the DSS capabilities and DSS impacts are shown in **Figure 1**. There are inner-dependencies within each cluster and outer-dependencies between the elements of each cluster outside of the second cluster. This model consists of 2 clusters with 3 elements, together needing 8 PCMs. 2 PCMs for evaluating elements within each cluster, 3 PCMs for evaluating all elements from one cluster with respect to the elements from the second cluster and 3 PCMs for reversed evaluation.

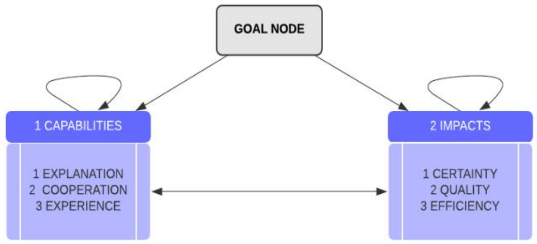


Figure 1 – Model structure

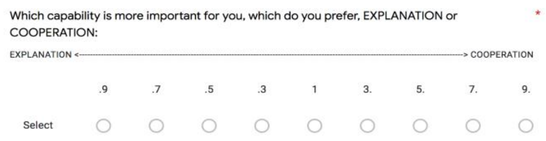


Figure 2 – One question from the survey

Data was obtained from a survey questionnaire that included 24 pairwise comparisons questions in form as shown in **Figure 2**. Respondents in this case study are managers from various areas and students of final year Master's level who focus on DSS as part of their study.

Responses data were exported into MS Excel via Google forms. For ANP purposes, Saaty's matrices were calculated automatically using sheet functions for each questionnaire and the consistency index was computed using Goal seeking for each pairwise comparison. Saaty's matrixes and weights are calculated for each respondent individually, and aggregated weights are calculated as average. The final weights are calculated for each respondent, and aggregated final weights are calculated from aggregated partial weights.

2.2 Inconsistency and modification methods tested

Pairwise comparison is the process of comparing pairs of items to judge which elements of each pair is preferred or has a greater amount of some quantitative property. One broadly used method is Saaty's pairwise comparison method [16]. Inconsistency index is calculated for all individual PCM for all respondents.

Saaty's matrix of pairwise comparison of 3 elements can be given

$$A = \begin{pmatrix} 1 & a & b \\ 1/a & 1 & c \\ 1/b & 1/c & 1 \end{pmatrix} \tag{1}$$

According to Saaty's approach, the inconsistency is measured by the consistency ratio CR given as

$$CR = \frac{CI}{RI_n} \tag{2}$$

where RI_n represents random inconsistency index. A given PCM is considered inconsistent for $CR > 0.1$, according to Saaty. Values of RI_n differs for particular n and was proposed by Saaty [16] based on simulating random matrix. The CI stands for consistency index and for matrix A is calculated as

$$CI = \frac{\lambda_{\max} - 3}{3 - 1} \tag{3}$$

and λ_{max} is the maximum eigenvalue of A.

Four methods are applied on the inconsistent Saaty's matrices and their results (received weights) are compared with original preferences including the inconsistent matrixes

Method 1 – expert adjustment (ADJ)

Revisiting each individual questionnaire is done as follows. The most inconsistent element is identified and modified. This survey questionnaire followed the same structure of the question, out of three criteria, first pairwise comparison compared 1st to 2nd elements (a), 1st to 3rd elements (b) and last is 2nd to 3rd elements (c). Therefore, naturally the last judgement (c) had to be modified. CR needs to be recalculated after each modification. This process is repeated until $CR < 0,1$.

Method 2 – Kockodaj'and Dusak (KD)

In the consistent cases one of the following equations must hold:

$$a = {}^{b}/_{c} \quad \text{or} \quad b = ac \quad \text{or} \quad c = {}^{b}/_{a} \tag{4}$$

It is always possible to produce three consistent matrices by computing the third value from the combination of the remaining two elements [3].

Method 3 – minimum deviation (minD)

The last method of Saaty's matrix modification is based on optimization model [11]. For the Saaty's matrix (1) the model is formulated as follows:

$$\min(D) \quad \text{subject to} \quad \begin{array}{l} \left|\frac{\bar{a}-a}{a}\right| \le D, \left|\frac{\bar{b}-b}{b}\right| \le D, \left|\frac{\bar{c}-c}{c}\right| \le D \\ \det(\bar{A}-\lambda I)=0 \\ \frac{\lambda-3}{(3-1).RI_3} \le 0.1 \\ \frac{1}{9} \le \bar{a}, \bar{b}, \bar{c} \le 9 \end{array} \tag{5}$$

where D is minimum deviation, a and $\bar{a}$, b and $\bar{b}$, c and $\bar{c}$, are the corresponding original and modified Saaty's values, $\bar{A}$ is modified Saaty's matrix A, λ is maximum eigenvalue, and RI_3 represents random index.

3 Results

This pilot study includes 14 respondents. Defined ANP model data consists of 8 PCM per respondent. Out of 14 respondents, 2 provided consistent judgments and 12 included some inconsistent judgments. Out of total 112 PCM, 21 were inconsistent. On average each respondent provided 2 inconsistent PCM while there was 1 respondent that provided 7 inconsistent PCM.

For consistency improving methods testing, firstly individual inconsistent PCM weights were modified using all the methods. Original and modified weights are shown in F**igure 3.** MinD method is preserving the original decision-maker's judgment while achieving required consistency

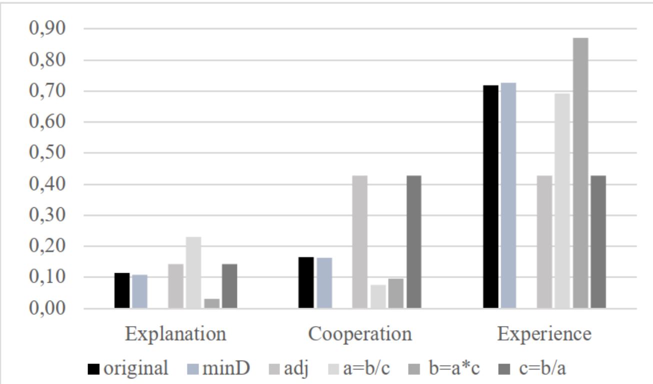


Figure 3 – Individual PCM weights

PCM – original inconsistent

	Expl	Coop	Exp	weights
Explanation	1	1/3	1/3	0,115
Cooperation	3	1	1/9	0,166
Experience	3	9	1	0,719
				CI = 0,28

PCM – minD modification method

	Expl	Coop	Exp	weights
Explanattion	1	3/7	1/4	0,110
Cooperation	2 1/3	1	1/7	0,163
Experience	4 1/4	7	1	0,727
				CI = 0,1

Figure 4 – PCM original vs minD method

Highest and lowest inconsistent respondents were selected to confirm the method validity on two extremes and their preferences of capabilities and impacts are shown on **Figure 5** and **Figure 6** respectively.

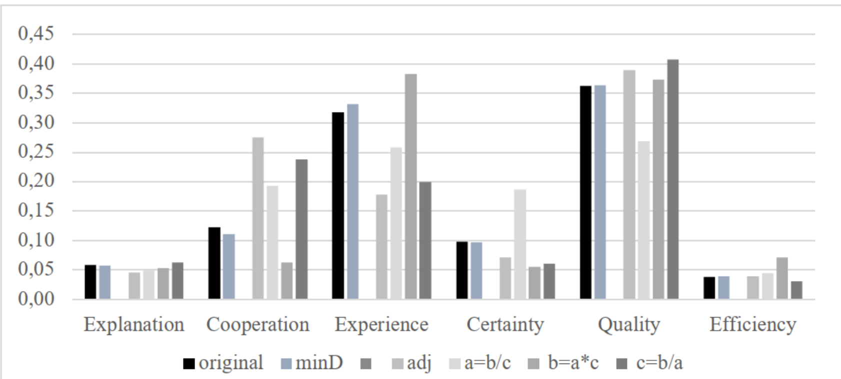


Figure 5 – Weights of highest inconsistency respondent

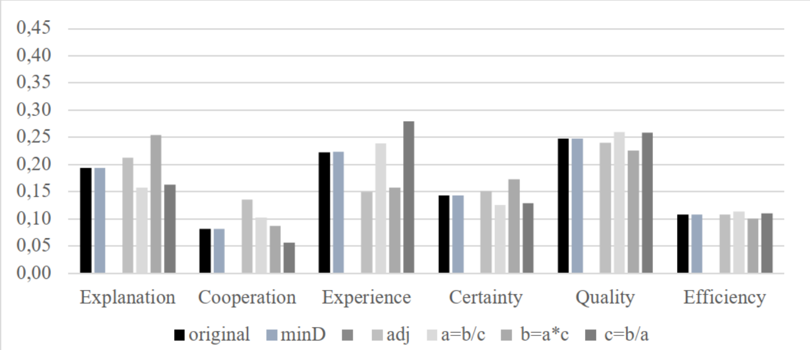


Figure 6 – Weights of lowest inconsistent respondent

To calculate DSS final weights, minimum deviation method was applied to improve consistency of Saaty's matrix. The final weights are shown **Figure 7**. Experience is the most preferred capability, followed by Cooperation and Explanation. Quality is the most preferred impact, followed by Certainty and Efficiency. Final weights are received while original decision-makers judgment is preserved.

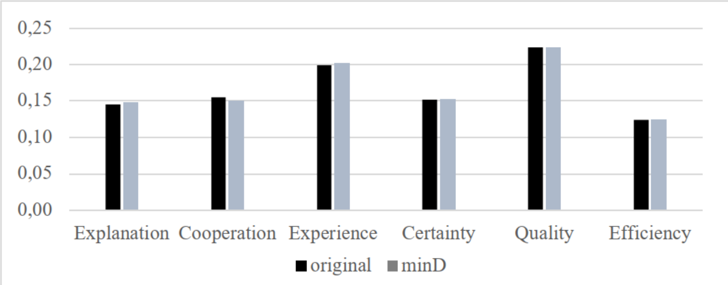


Figure 7 – DSS capabilities and impacts final weights

4 Discussion

Using ANP are usually focused on implementation on individual or smaller scale survey, usually utilizing an expert team information [13], [19]. In the case of using ANP for evaluating larger scale survey data, the data processing may be difficult, especially in larger models. The presented ANP evaluation of DSS seems to reflect such limitations by including only three capabilities and three impacts. This scale is sufficient to obtain valuable information from managers as well as challenges respondents to think and concentrate by adding a layer of complexity. The perceived limitation to use ANP on a survey data leads to DSS being evaluation only by experts such as for ERP system [13]. The model structure consisted of two clusters, each with three elements. The pairwise comparisons of three elements obtained via survey questionnaire yet can lead to inconsistency due to various combinations of pairwise questions. Out of 14 respondents, 12 included inconsistent PCM. While this model was not large in the scale, the inconsistent matrixes of this scale suggest a complex problem. This complexity can be further analyzed by focusing on inconsistent matrixes. Not only are survey designers interested in the level of inconsistency present in their surveys, they are also interested in the source of inconsistency. Are respondents making inconsistent choices because some attributes are ill-defined, or that a pairwise comparison between those attributes simply do not make sense? Regarding consistency threshold the 0,05 shall be used for 3x3 matrix, according to [17]. Apparently, the size of the matrix does not always relate to the problem complexity. Using 0,05 threshold in this ANP model, we would achieve majority of PCM inconsistent.

5 Conclusion

This paper evaluates the DSS preferences in terms of capabilities and impacts. Capabilities of a system were set as explanation, cooperation and experience and has implications of DSS were selected as the impact on quality of decision, certainty of a decision-maker and efficiency of a decision-making process. Inconsistent judgments were modified by using minimum deviation method achieving required consistency while preserving original decision-makers' judgements. Next research will focus on completing this DSS evaluation study by reaching out to dozens of managers. Evaluation of inconsistent judgements for the purposes of providing valuable information will be examined.

References

[1] Alieinykov, I. 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. doi:10.15587/1729-4061.2019.184394.

[2] Blahova, P. & Škoda, M. (2021). Decision Support Systems and Their Influence on Decision Choices and Results. In *Proceedings of the 14th International Conference on Strategic Management and its Support by Information Systems*. Ostrava.

[3] Bozóki, S. & Rapcsák, T. (2008). On Saaty's and Koczkodaj's inconsistencies of pairwise comparison matrices. *Journal of Global Optimization* 42(2), 157–175. doi: 10.1007/s10898-007-9236-z.

[4] 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 & Information Technology*, 38(12), 1225–1242. doi: 10.1080/0144929X.2019.1581258.

[5] 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 & Management* 54(7), 934–947, doi: 10.1016/j.im.2017.01.006.

[6] 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. doi: 10.1016/j.ijproman.2010.01.008.

[7] Ghattas, J., Soffer, P. & Peleg, M. (2014). Improving business process decision making based on past experience. *Decision Support Systems*, 59(1), 93–107. doi: 10.1016/j.dss.2013.10.009.

[8] Gönül, M. S., Önkal D. & Lawrence M. (2006). The effects of structural characteristics of explanations on use of a DSS, *Decis. Support Syst*., vol. 42, no. 3, pp. 1481–1493, 2006, doi: 10.1016/j.dss.2005.12.003.

[9] 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. doi: 10.1007/s10869-020-09711-6.

[10] Burkart, N., Robert, S., & Huber, M. F. (2020). Are you sure? Prediction revision in automated decisionmaking. *Expert Systems* 38(93), 1–19. doi: 10.1111/exsy.12577.

[11] Hlavatý, R. & Brožová, H. (2022). Fixing Inconsistent Saaty's Matrix using Optimisation Model., In: *Proceedings of the 40th Conference on Mathematical Methods in Economics.* VŠP Jihlava, 2022, working paper.

[12] Peixoto, L. de C., Golgher, A. B. & Cyrino, Á. B., Using information systems to strategic decision: an analysis of the values added under executive's perspective, *Brazilian J. Inf. Sci*., vol. 11, no. 2, pp. 54–71, 2017, doi: 10.36311/1981-1640.2017.v11n2.06.p54.

[13] Perçin, S. (2008). Using the ANP approach in selecting and benchmarking ERP systems. *Benchmarking 15*(5), 630–649. doi: 10.1108/14635770810903196.

[14] Power D. J., (2005). *Decision Support Systems: Frequenty Asked Questions*. iUniverse Publishing, Inc.

[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*, Creative Decisions Foundation.

[16] Saaty, T. L. (1980). *The Analytic Hierarchy Process*. New York, Mc Graw Hill Inc.

[17] Saaty, T.L. (1994). *Fundamentals of decision making*. RSW Publications.

[18] van der Waa, J., Schoonderwoerd, T., van Diggelen, J. & Neerincx, M. (2020). Interpretable confidence measures for decision support systems. *International Journal of Human-Computer Studies* 144, 11. doi: 10.1016/j.ijhcs.2020.102493.

[19] 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. doi: 10.1007/s10700-020-09323-y.

[20] 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. doi: 10.1016/j.ijinfomgt.2009.08.005.

Impact of the Incoming Pandemic on Investment Decision Discussed Through a Weighted Moving Mean-absolute Negative Deviation Model

Adam Borovička[1]

Abstract. The incoming COVID-19 pandemic affected various spheres of human life. Investment decision making was no exception. Significant uncertainty triggered a wave of sell-offs on the capital market. Thus, the main aim of the article is to discuss the effects of a pandemic on investment decisions through a very broadened investment strategy. Consider a stock portfolio made before the pandemic for medium or long-term investment horizon through the weighted moving mean-absolute negative deviation model. This developed linear model can account for the dynamics of instability of the prices on the capital market. Another advantage is the possibility to express preference on the importance of considered criteria – return and risk. Linear relations are application-friendly. Under the condition of a significant decline in the value of an investment in the first few weeks (February/March 2020) of a significant spread of a new disease, the investment strategy may change abruptly. A change in the attitude to the risk and turnaround in the development on the capital market with a significantly uncertain outlook encourages a change in the investment portfolio for which the flexibly adaptable model mentioned above will serve perfectly. Changing personal preferences is expressed through the weights. The dynamics of development is then included in the designed moving form of returns. The real effects of the portfolio re-optimization caused by the COVID-19 pandemic are demonstrated on the investment portfolio of stocks traded on the RM-System Czech Stock Exchange. The existing and new portfolios are compared in terms of both composition and characteristics, hence their performance over real time. Based on the results, the meaningfulness of the portfolio review decision is discussed.

Keywords: investment decision, moving, pandemic, portfolio, stock

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

1 Introduction

The COVID-19 pandemic has affected almost every area of human life. The capital market was not spared. In February/March 2020, due to very unclear future development caused by a relatively unknown devastating disease, investors began to divest or reallocate their investments. The change in investment strategies has led to significant sell-offs in capital markets around the world [4]. Even after a month of falling stock prices, a considerable uncertainty remained on the capital market [3]. What impact did the rapid change or adaption of investment behavior with the emerging pandemic actually have on the composition of investment portfolio? And over time, was the ‘panic’ decision right? The answer to these key questions can be beneficial for investment practice.

In order to empirically analyze the effect, or impact of the spreading disease COVID-19, one of the most common investment strategies is used. The study thus tries to be as close as possible to the investment reality. Specify a risk- averse investor who invests in the medium or longer-time period. This can be a typical investor in the middle productive age who (e.g.) is actively trying to secure financially for retirement age. In 2019, he decided to invest his free funds in stocks traded on the Czech capital market. After several weeks of very unstable capital market behavior caused by the outbreak of the pandemic, he decided to change his investment strategy. The result is a change in the portfolio composition in the spirit of new facts and adjustment to his risk perception. Today, after more than two years (in 2022), this change can be evaluated, albeit in a shorter time horizon.

To meet the goal of performing the above analysis, an adequate methodological concept is proposed. At the heart of this concept is, of course, the approach to making an investment portfolio. Investment experiences for a (longer-

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

term) investor indicates, in comparison with other characteristics (fees, liquidity, etc.), the fundamental influence of the expected return and risk on his decision making. The use of notorious mean-variance model [6] reflecting the investor's 'return-risk' profile is therefore offered. However, the application of this model is not without limits. In addition to the assumption of the return normality, it penalizes positive deviations from the average return. However, after thinking more deeply, the list of shortcomings can be expanded. Thus, the assumption of risk, or return stability, throughout the observed historical period, from which the monitored characteristics are determined, can be problematic. As the sudden pandemic ultimately demonstrates, the degree of uncertainty may evolve over time.

How to eliminate these shortcomings? The risk measure based on the variance of returns is replaced by absolute negative deviation concept which is inspired by Konno and Yamazaki's [5] absolute deviation. Then the model mean-absolute negative deviation model can be designed. This model, unlike the mean-variance form, is linear, which potentially simplifies the search for the global constrained extreme of the particular function. On the other side, it is necessary to realize that this formulation eliminates covariance relations, which can really narrow the applicability. And what about the aforementioned unstable uncertainty over time? The path leads through the dynamization of both characteristics. Inspired by Borovička [1], mean and absolute negative deviation is proposed in the form of moving average. Then the observed historical period is divided into a few shorter parts shifting by one-time subperiod. Local characteristics can be calculated in these time-overlapping periods. Finally, return, or risk of the portfolio are calculated as moving average of partials means, or absolute negative deviations over all subperiods. The proposed approach is less data complexity and user friendliness, e.g. compared to the concept using fuzzy numbers (see more in [2]). Its application friendliness is supported by the integration of both (standardized) characteristics into the objective function through weighted sum approach. Such a model allows the expression of preference about the importance (and its change) of both characteristics. In addition, the minimum required level of return, or the maximum acceptable risk of the investment, unlike the original Markowitz's model, does not have to be stated which can be especially helpful mainly for a less experienced investor.

What is the proposed methodological design for the empirical analysis? At the beginning of the investment process, a stock portfolio is made by the developed 'return-risk' model. With the advent of very unstable conditions, the investor begins to consider revising the investment because he fears more significant loss. Based on past and current data, or actual investor preferences, the portfolio is re-optimized using the proposed model. The impact of changes in investor behavior, or investment strategy, may be reflected in the composition of the portfolio. The reasonableness of the decision can be reflected in the future by comparing the performance of the original and revised portfolio.

To fulfill the main empirical goal, the introduced methodological approach is applied on the RM-System Czech Stock Exchange. The change in investment behavior due to fears of unclear, or unfavorable development caused by an unknown insidious disease is reflected in a change in the composition of the investment portfolio. It turns out that the effect of 'something unknown', potentially very dangerous, is significant. But what does this decision look like after two years? The performance of the re-optimized portfolio is lower than the original one. Thus, in the shorter(!) term, the decision seems to be reckless.

The structure of the article is as follows. After Introduction, the methodological concept for the empirical analysis is designed (Section 2). The impact of the pandemic on investment behavior (on the Czech capital market) is demonstrated in Section 3. Finally, the empirical and theoretical aspects of the analysis are summarized. Inspirational directions for potential research are outlined

2 Analysis methodology based on the weighted moving mean-absolute negative deviation model

To make the empirical analysis of the impact of the pandemic uncertainty to the investment behavior, the appropriate methodological procedure is developed. This approach is based on a designed weighted moving mean-absolute negative deviation model that is able to detect the changing dynamics of the capital market and related changes in the investor's preferences.

2.1 Investment strategy and its revision

Investment strategy is a basic pillar of the investment process. The strategy is shaped by the purpose, goal or time horizon of the investment. Another important aspect is the attitude to risk, from which the expectation of the investment performance is derived. Last but not least, is also depends on the attitude to investment management. In dramatic times of large fluctuations on the capital market, even a slightly active investor tends to restructure the investment portfolio. In extreme situations, such as the outbreak of a financial crisis or a pandemic, the investor may deviate significantly from the original intentions. The investment policy is then adapted to the current situation. The investment can then be significantly revised.

2.2 Weighted moving mean-absolute negative deviation model

As methodologically introduced above, the weighted moving mean-absolute negative deviation model is developed for the portfolio selection. Its formulation is as follows

$$\begin{aligned} &\max \; w\frac{r_{max}-\mathbf{r}^{\mathbf{T}}\mathbf{x}}{r_{max}-r_{min}}+(1-w)\frac{\mathbf{p}^{\mathbf{T}}\mathbf{x}-p_{min}}{p_{\max}-p_{\min}} \\ &\mathbf{x}\in X. \end{aligned} \tag{1}$$

In model (1), the most important of the portfolio characteristics, return $\mathbf{p}^{\mathbf{T}}\mathbf{x}$ and risk $\mathbf{r}^{\mathbf{T}}\mathbf{x}$, are included. The vector $\mathbf{x}=(x_1,x_2,...,x_M)^T$ contains the shares of M assets in the investment portfolio. The set of the feasible solutions includes all the necessary conditions expressing the conditions of a specific investment situation. The integral part is a portfolio and non-negativity conditions, formalized then as $\mathbf{e}^{\mathbf{T}}\mathbf{x}=1$ and $\mathbf{x}\geq\mathbf{0}$, where $\mathbf{e}^{\mathbf{T}}=(1,1,...,1)$ is a vector of M ones. Other conditions may, for example, limit the number of (selected) assets in the portfolio. The investor's attitude to the risk, or expected return, can be expressed by the weights w, or $1-w$.

For general comparison, the values of both functions are standardized, using the best and worst possible value. The main advantage of this not commonly used technique is the acceptance of negative function values. The best values of return $p_{max}=\mathbf{p}^{\mathbf{T}}\mathbf{x}_{\mathbf{p}}^*$ and risk $r_{min}=\mathbf{r}^{\mathbf{T}}\mathbf{x}_{\mathbf{r}}^*$ are, of course, determined through the following

$$\begin{aligned} &\mathbf{x}_{\mathbf{p}}^*=\arg\max\mathbf{p}^{\mathbf{T}}\mathbf{x} \qquad &&\mathbf{x}_{\mathbf{r}}^*=\arg\min\mathbf{r}^{\mathbf{T}}\mathbf{x} \\ &\mathbf{x}\in X, &&\mathbf{x}\in X. \end{aligned} \tag{2}$$

The worst possible values of the characteristics are determined in the context of simultaneous optimization. Thus, the following $p_{min}=\mathbf{p}^{\mathbf{T}}\mathbf{x}_{\mathbf{r}}^*$ and $r_{max}=\mathbf{r}^{\mathbf{T}}\mathbf{x}_{\mathbf{p}}^*$ holds.

The vector $\mathbf{p}^{\mathbf{T}}=(p_1,p_2,...,p_M)$, or $\mathbf{r}^{\mathbf{T}}=(r_1,r_2,...,r_M)$ reflects the (expected) return, or risk of M assets. Let T defines the number of equally long time periods with K observations of returns. Then (expected) return, or risk of the i-th asset in the t-th period is calculated as a mean, or absolute negative deviation

$$p_{it}=\frac{\mathbf{e}^{\mathbf{T}}\mathbf{p}_{\mathbf{it}}}{K}, \text{ or } r_{it}=\frac{\mathbf{e}_{\mathbf{it}}^{\mathbf{T}}\Delta\mathbf{r}_{\mathbf{it}}}{K_{it}^-} \qquad i=1,2,...,M;\, t=1,2,...,T, \tag{3}$$

where $\mathbf{p}_{\mathbf{it}}=(p_{it1},p_{it2},...,p_{itK})^T$, or $\Delta\mathbf{p}_{\mathbf{it}}=(p_{it}-p_{it1},p_{it}-p_{it2},...,p_{it}-p_{itK_{it}^-})^T$ is vector including K observations of return, *or* K_{it}^- positive differences between mean and partial return, for the i-th asset in the t-th time period. Adequately, the vector $\mathbf{e}^{\mathbf{T}}=(1,1,...,1)$, or $\mathbf{e}_{\mathbf{it}}^{\mathbf{T}}=(1,1,...,1)$ contains K, or K_{it}^- ones. The difference between the beginning, or end, of two consecutive periods is constant during the entire historical period. Neighboring periods overlap between the beginning of one period and the end of the previous one. And now, the return, or risk, of the i-th asset as the moving mean, or moving absolute negative deviation, can be designed as follows

$$p_i=\frac{\mathbf{e}^{\mathbf{T}}\mathbf{p}_{\mathbf{i}}}{T}, \text{ or } r_i=\frac{\mathbf{e}^{\mathbf{T}}\mathbf{r}_{\mathbf{i}}}{T} \qquad i=1,2,...,M, \tag{4}$$

where $\mathbf{p}_{\mathbf{i}}=(p_{i1},p_{i2},...,p_{iT})^T$, or $\mathbf{r}_{\mathbf{i}}=(r_{i1},r_{i2},...,r_{iT})^T$ is vector containing T returns, or risk of the i-th asset covering the entire 'moving' time period.

The model (1), compared to the original Markowitz's model, reflects the dynamically changing development on the capital market by integrating moving averages. Especially in the case of insignificant correlations between returns, the linear function of risk is an excellent alternative which is greatly simplifies finding the global extreme of the aggregate objective function of the model. In addition, the level of implemented measure of risk, absolute negative deviation, does not penalize positive deviations of returns from their average. Finally, the weighted form of the aggregated objective function allows the simultaneous expression of the investor's preferences about the expected characteristics of the investment according to his attitude to risk, or the 'return-risk' profile. The elimination of the thresholds for characteristics supports the applicability of the model in investment practice.

2.3 Evaluation of changes in investment behavior

The outcome of the investment decision making process is an investment portfolio made via the model (1). The investment is managed according to the investment plan (strategy). A sudden extreme situation (financial crisis, pandemic, terrorist attack) can trigger a wave of sales in the capital market. Rapid changes in the markets can also be caused by positive events. A non-passive investor can also react to the situation. He therefore decides to re-optimize the portfolio.

For this purpose, the model (1) with the updated data and preferences can be formulated and solved to revise investment portfolio under the prevailing unstable conditions. Data from the period of the current life of the investment are used in the spirit of the investment horizon. The observed historical period is then moved by this time into the present. The number and length of partial 'moving' periods remain the same. The preferences, especially the attitude to risk, can be significantly adjusted due to unexpected facts, which is reflected in the model (1) by a change in weights. Model (1) with updated inputs provides a new investment portfolio that may be more or less (not at all) different from the existing one.

After a sufficient time period, is possible to evaluate the effect of the above-mentioned change in the behavior of the investor who, on the basis of sudden events, reassessed his investment plans. By comparing the performance of the revised and original investment (portfolios), the benefits, or negatives of changing the investment strategy can be declared.

3 Portfolio (re)making under the pandemic uncertainty

In this section, the empirical analysis of the impact of the emerging pandemic on investment decision making is performed through the methodological procedure proposed above. The original portfolio is re-optimized under conditions of sudden market volatility. With a time lag, this move is soberly evaluated.

3.1 Investment strategy and its revision

To make the practical demonstration of the impact of sudden extreme situation as illustrative as possible, we will focus on a common investment strategy. This can be (not only) an investor in the earlier part of productive age, but already consciously thinking about the future. It is therefore a medium- or long-term investment with the aim of appreciating the available funds for use in the distant future (e.g. in retirement). Thus, the liquidity of the investment is not the key issue. The investor is not a speculator, the portfolio management is rather passive with irregular evaluation, hence revision. Such an investor is usually not very experienced in the capital market. The investment tends to be smaller, but can be made on a regular basis.

Let's move to the beginning of 2019, when the investor decides to invest his free cash. As it happens, a decision making on the capital market is significantly influenced by the past. Here, too, we will go back in time, specifically five years, in order to be representative of developments in the medium or long term. According to my personal investment experience or the results of various surveys and polls, the return and risk of the investment are absolutely crucial for the investor. These characteristics are monitored for stock titles traded in the RM-System. The RM-System Czech Stock Exchange is suitable for 'smaller' investors. Trading is possible in small standardized units of stocks. However, the supply of stocks is wide. The investors can use the user-friendly online trading system (see more [7]). Return and risk may not be the only criteria taken into account. In addition to (e.g.) liquidity, cost, currency, and others, the number of assets (stocks) in the portfolio may also be relevant. Especially for a less experienced investor, a 'less complex' portfolio is preferable. The minimum and maximum share in the portfolio is then set. In our case, the minimum, or maximum share is 15%, or 45%. Thus, there may be three to six stocks in the portfolio. Other characteristics, even with respect to the investment location, or market, are not relevant.

Thirteen stocks (listed in Table 1) with sufficient trading history have been selected for potential investment. As mentioned above, the return, and hence the risk, is calculated from the historical prices of the five-year period from 2014 to 2018, downloaded from [8] The return is calculated as a moving mean of the monthly returns over three-year overlapping periods (2014–2016, 2015–2017, 2016–2018) to reflect the unstable dynamics of the capital market. Risk of the stock can then be calculated as the moving absolute negative deviation of monthly returns, formalized as (3) and (4).

	01/2014–12/2018		04/2015–03/2020	
Stock	**Return** [%]	**Risk** [%]	**Return** [%]	**Risk** [%]
ČEZ	0.02	5.12	0.14	4.20
Deutsche Telekom	0.21	3.25	-0.50	3.37
Erste Group Bank	0.86	6.05	0.49	6.28
Exxon Mobil	-0.08	3.45	-0.76	4.03
Intel	1.11	4.57	1.23	5.37
Komerční banka	-0.04	4.22	-0.70	4.01
McDonald's	1.32	3.10	0.98	3.76
Microsoft	1.92	3.95	2.17	3.80
Nokia	-0.74	6.06	-0.84	5.40
O2 C.R.	0.88	5.92	0.57	5.16
Philip Morris	0.80	3.22	0.64	3.40
VIG	-0.93	5.45	-0.40	4.84
Volkswagen	-0.15	6.19	-0.26	4.40

Table 1 Return and risk of the stocks. Source: self-calculation

With the outbreak of the COVID-19 disease pandemic, capital markets have been shaken. Although the investor pursues longer-term goals, the fear of not meeting them caused by the panic-induced wave of sell-offs, overwhelms the investor. After several weeks of extreme uncertainty, or behavior of the capital markets, he decides to revise the portfolio at the end of March 2020. For this purpose, the data is updated with the duration of the investment. Then the historical five-year period is moved closer to the present in the form of 04/2015–03/2020. Three overlapping three-year periods, 04/2015–03/2018, 04/2016–03/2019 and 04/2017–03/2020, are again used to calculate the 'moving' investment characteristics. The (expected) return and risk in both periods is shown in Table 1.

3.2 Portfolio selection and its re-optimization

Considering the longer investment time horizon, the purpose of the investment or positive sentiment on the capital markets (before 2019), the investor perceives the risk rather neutrally. The fear of losing the investment is not strong, the desire for return prevails. Then the weight of the return slightly exceeds the weight of the risk of the investment. To make a portfolio, the following mathematical model along the lines of (1) is formulated

$$
\begin{aligned}
&\max \ 0.4\frac{3.70-\mathbf{r}^{\mathrm{T}}\mathbf{x}}{3.70-3.17}+0.6\frac{\mathbf{p}^{\mathrm{T}}\mathbf{x}-0.95}{1.56-0.95} \\
&\mathbf{e}^{\mathrm{T}}\mathbf{x}=1 \\
&0.15\mathbf{y}\le\mathbf{x}\le 0.45\mathbf{y} \\
&\mathbf{x}\ge 0 \\
&\mathbf{y}\in\{0,1\},
\end{aligned}
\tag{5}
$$

where $\mathbf{x}=(x_1,x_2,...,x_{13})^T$ is a vector of variables $x_i, i=1,2,...,13,$ expressing the share of the i-th stock sorted by the first columns of Table 1. Vectors $\mathbf{r}^{\mathrm{T}}=(r_1,r_2,...,r_{13})$ a $\mathbf{p}^{\mathrm{T}}=(r_1,r_2,...,r_{13})$ contain the risk $r_i, i=1,2,...,13,$ and return $p_i, i=1,2,...,13,$ of the i-th stock. The vector of thirteen ones is denoted as $\mathbf{e}^{\mathrm{T}}$. The binary variable $y_i, i=1,2,...,13,$ from the vector $\mathbf{y}=(x_1,x_2,...,x_{13})^T$ indicates (non-) presence of the i-th stock in the portfolio. It serves as an auxiliary element to express the conditions for participation in the portfolio in terms of minimum and maximum shares. The extreme values of both characteristics are determined through the optima (2) of one-objective models.

The optimal solution of the mixed binary problem (5) is found by brunch and bounds method implemented in the LINGO optimization software. The portfolio consists of the following components in their respective proportions: 40% *McDonald's*, 45% *Microsoft* and 15% *Philip Morris*. Its (expected) return is 1.51% and risk 3.50%. Looking at the data in Table 1, the composition of the portfolio is not surprising. Microsoft stock has the highest return with a solid level of risk. McDonald's stock characteristics are also great – lowest risk, second highest return. The maximum share in the portfolio is not reached because of the lower limit, hence the excellent performance of Microsoft stock. The portfolio is rounded out by the stock of tobacco company Philip Morris, thanks in particular to the second-lowest risk and solid return. The low level of risk reflects the combination of the lowest number of instances of negative return deviations from the mean and the smallest maximum deviation across all stocks. Some stocks at first glance could not be in the portfolio under the specified conditions, e.g. Nokia or VIG.

As mentioned above, the portfolio is being reoptimized at the end of March 2020. Model (5) is revised through actual (new) levels of return and risk data, and also changes in the weights to reflect the increase in risk aversion in the turbulent times of an accelerating pandemic. The weights are set in reverse, 0.4 for return and 0.6 for risk. Investor is reacting a bit timidly for now, not wanting to make any rush judgement in this new, hard-to predict event. The revised portfolio is as follows: 15% *McDonald's*, 45% *Microsoft* and 40% *Philip Morris*. Its (expected) return is 1.38% and risk 3.63%. The deterioration in the values of both characteristics in turbulent times is not surprising. The portfolio base remains the same. The group of three stocks mentioned above still outperforms the others. However, the shares have partially changed. Although the return is now a slightly less important criterion, Microsoft still holds the maximum possible share in the portfolio. Its return is significantly higher than that of other stocks. Moreover, there has not been such an increase in risk relative to other titles. McDonald's stock took a hit by unleashing the pandemic. It offers only the third best return and risk. Philip Morris, however, holds the second lowest risk with a very solid return, which, with stronger risk aversion, leads to its more significant participation in the portfolio at the expanse of McDonald's stock. The stock of Deutsche Telekom position should not go unnoticed. This stock provides the lowest risk. The chaos on the capital market caused by the unfolding pandemic did not rock this stock that much. While its return turned negative, the deviations of returns from the mean is not as dramatic as that of other stocks. A poor return, even if not as volatile, causes non-participation in the portfolio. To push a stock into the portfolio, the investor would have to exhibit a stronger risk aversion. Its weight would have to be at least 0.73. On the other hand, a stock with negative performance outlook is not desirable.

3.3 Portfolio performance evaluation

What is the current value of the investment? Has the portfolio revision, under the weight of the unknown, paid off in terms of performance? Although the investments have not yet had such a long life with respect to the chosen investment strategy, or investment horizon, some comparison of the performance or value of the investments (portfolios) can be made, at least indicatively. The current value of the investment is based on prices as of February 28, 2022. The original stock portfolio, purchased at the beginning of 2019, increased its value by 93.57%! Under the given conditions, this investment is relatively close to the best achievable performance (109.12%). The worst decision could even lead to a decrease in the value of the investment (-1.33%) over the whole period 2019–2022. The re-optimization forced by the pandemic panic caused a worse investment performance at 84.02%. For the record, combining the worst, or best initial investment decision with the worst, or best 'pandemic' revision would yield a loss of -41.16%, or a gain of 166.67% from 2019 to 2022.

4 Conclusion

The paper focused on a post-crisis empirical analysis of the impact of the COVID-19 disease pandemic on investment decisions. For this purpose, the procedure based on the weighted moving mean-absolute negative deviation model was designed. Stock portfolios made on the Czech capital market before and after pandemic outbreak are analyzed.

As recent history shows the outbreak of something unknown, potentially very dangerous, can have a profound effect on people's behavior. This has been no different in the capital market. A rethinking the investment strategy affects the composition of the investment portfolio, leading to lower performance than the original portfolio. The decision to change the investment strategy seems ill-advised. In defense of the reasonably spooked investor, he understandably 'defended' his investment in the spirit of the general market sentiment. Turns out if he had gone against the tide, invested in more 'volatile' stocks (e.g. Erste Group Bank), portfolio performance would have improved significantly. However, this 'speculative' approach is out of line with the beliefs of the typical investor defined by us. Although the performance is slightly reduced by the revision, under the weight of very negative circumstances the investor may not regret his decision.

Further research could enrich the analysis with the idea of adjusting not only the weights but also other preferences (e.g., minimum and maximum level for share). Considering the transaction costs of portfolio changes, especially with a different investment instrument base, could be beneficial. Finally, it would be interesting to consider different weights for particular subperiods in the computed moving averages to reflect the degree of participation in the expected development.

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] Borovička, A. (2021). Portfolio selection via a dynamic moving mean-variance model. In *Proceedings of the 39th International Conference on Mathematical Methods in Economics 2021* (pp. 42–47). Praha: Czech University of Life Sciences Prague.

[2] Borovička, A. (2021). Stock portfolio selection under unstable uncertainty via fuzzy mean-semivariance model. *Central European Journal of Operations Research*, 30, 595–616.

[3] Echo24. (2022). *Pád akcií se podařilo dočasně zastavit. Strach z dalšího vývoje ale zůstává*. [online] Available at: https://echo24.cz/a/SEfEE/pad-akcii-se-podarilo-docasne-zastavit-strach-z-dalsiho-vyvoje-ale-zustava, [Accessed 1 March 2022]

[4] iRozhlas. (2022). *Černý čtvrtek: Koronavirus srazil burzy po celém světě*. [online] Available at: https://www.irozhlas.cz/ekonomika/cerny-ctvrtek-koronavirus-srazil-burzy-po-celem-svete_2003130630_pek, [Accessed 1 March 2022]

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

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

[7] RM-System. (2022). *Basic information*. [online] Available at: https://www.rmsystem.cz/spolecnost/zakladni-informace, [Accessed 30 April 2022]

[8] RM-System. (2022). *On-line prices*. [online] Available at: https://www.rmsystem.cz/online-prices/stocks/easyclick, [Accessed 4 April 2022]

Optimal Scheduling of Vehicle Loading/Unloading Operations in Depots

Martin Branda[1]

Abstract. We deal with the problem of optimal scheduling of cargo loading or unloading for a fleet of vehicles in depots. We show that it can lead to fixed interval scheduling where starting and finishing times of jobs are prescribed and the goal is to assign them to a set of machines. However, in real applications the finishing times can be subject to uncertainty where the random delay can be caused by unpredictable complications. In our case this corresponds to problems during cargo loading/unloading or even delay on arrival. We propose a two-stage stochastic programming formulation and its robust coloring reformulation leading to a large mixed-integer programming problem. In the numerical study we solve several instances of the problem.

Keywords: vehicle loading, fixed interval scheduling, random delay, stochastic optimization, robust coloring

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Loading/unloading operation of cargo to vehicles in depots can be seen as important part of the whole supply chain, see [8, 10, 14] for recent contributions in this area. In this paper, we will focus on the special case when each loading/unloading operation has its prescribed starting and finishing time and must be assigned to available gates in a depot. This problem can be seen a special case of the fixed interval scheduling problem where the jobs must be processed during given fixed intervals on available machines. These problems have been elaborated by several papers, see, e.g., [9, 11, 12, 13, 15, 16, 20], which were focused on purely deterministic case where all parameters of the problem are given and fixed for the whole planning period. However, in our case, the processing intervals can be subject to uncertainty caused by various reasons. First, there can be some complications during the loading/unloading operations. Secondly, the vehicles need not to be available in the depot but can arrive from a previous trip with a delay. These two complications as well as any others can be incorporated into the mathematical programming model as random delays. This bring us to the area of fixed interval scheduling under uncertainty which was elaborated by Branda et al. [7] for the first time. The authors proposed two formulations taking into account random delays with a known probability distribution. Two-stage stochastic programming formulation with the expected number of overlaps objective was introduced and it was shown to be equivalent to a robust coloring problem [18] with a special choice of the penalties. The resulting problem can be then solved using standard mixed-integer techniques, cf. [19]. Moreover, they also delt with reliability type objective where the goal is to get the schedule with the highest probability of remaining feasible during the whole planning period. This led to a generalized robust coloring problem which was solved by a special tabu search heuristic algorithm. A network flow reformulation of the problem with probabilistic objective, which enables to solve significantly larger instances to optimality, was proposed by [4]. In [2], the tactical fixed interval scheduling problem where the goal is to find the minimal number of machines to process all jobs was solved by a special iterative algorithm. Robustness of operational fixed interval scheduling problems was elaborated by [3] where cardinality constrained uncertainty set was considered. This set enables to change a limited number of marginal distributions of random delays and the goal is to find a schedule which is the most robust with respect to these changes. New decomposition algorithm was based on a network flow reformulation with conditional value at risk measure in the objective. Most of the above works which delt with fixed interval scheduling problems under uncertainty considered homogenous machines. This means that any job can be assigned to any available machine. However, the loading/unloading operations, the gates as well as the vehicles may come from different classes which means that we are facing problem with heterogeneous machines and job classes. Below, we will propose an extension of the two-stage problem and introduce a reformulation based on robust coloring which can be solved by a mixed-integer solver.

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

The paper is organized as follows. In Section 2, we propose the problem formulation as an two-stage integer stochastic program. The robust coloring problem is generalized in Section 3 and it is shown that the problem is equivalent to the two-stage formulation from the previous section under special choice of the penalties. In Section 4, we propose a numerical study on several simulated instances and discuss the efficiency of the proposed approach. Section 5 concludes the paper.

2 Problem formulation: two-stage stochastic integer programming

We consider jobs which correspond to the loading/unloading operations and machines which correspond to places where the cargo can be loaded or unloaded, e.g., gates in a depot or a logistic center. This task can be seen as one of the sub-problems in a complex supply chain. One of the main difficulties lies in dealing with the uncertainty because it can increase the solution complexity significantly.

Figure 1 shows two schedules where 6 jobs are assigned to two machines. If the machines are identical then both schedules are feasible. In the first one, the random delay of job 1 can cause problem in processing not only job 2, but if the delay is long, also job 3 can be compromised. If we consider heterogeneous machines and, for example, job 2 cannot be processed by machine 1, then the second schedule becomes infeasible. Therefore we must use proper mathematical programming formulations.

We will denote the set of jobs by J and set of machines by C. Let sj denote the prescribed starting times of the loading/unloading operations and let fj0 be the prescribed finishing times which are subject to uncertainty represented by random delays Dj (ξ) leading to random finishing times.

$$f_j(\xi) = f_j^0 + D_j(\xi),$$

which are defined on probability space $(\Xi, \mathcal{F}, P)$. Paper [7] proposed two-stage integer stochastic programming formulation with simple recourse, see [17] for a general introduction, where identical machines were considered. In this paper, we generalize the problem for machine classes where only a subset of jobs can be processed by particular machine. We denote by $\mathcal{J}_c \subset \mathcal{J}$ the set of jobs which can be processed by machine c and by $C_j \subset C$ the set of machines which can process job j. To simplify the notation we will further denote the set of starting times of jobs by $\mathcal{S} = \{s_1, \ldots, s_J\}$. Note that it is necessary to check if the machine processes at most one jobs at these starting times only. The problem with the expected number of overlaps objective can be formulated as follows:

$$\begin{aligned}
\min_{x,\tilde{y}}\ & \mathbb{E}_P \left[\sum_{c\in C} \sum_{j\in\mathcal{J}} \tilde{y}_{jc}(\xi) \right] \\
\text{s.t.}\ & \sum_{j\in\mathcal{J}_c:\ s_j\le t<f_j^0} x_{jc} \le 1,\ t\in\mathcal{S},\ c\in C, \\
& \sum_{c\in C_j} x_{jc} = 1,\ j\in\mathcal{J}, \\
& \sum_{k:\ f_j^0\le s_k<f_j(\xi)} x_{kc} \le \tilde{y}_{jc}(\xi) + |\mathcal{J}|(1-x_{jc}),\ c\in C_j,\ j\in\mathcal{J},\ \xi\in\Xi, \\
& x_{jc}\in\{0,1\},\ c\in C_j,\ j\in\mathcal{J}, \\
& \tilde{y}_{jc}(\xi)\in\mathbb{N},\ c\in C_j,\ j\in\mathcal{J},\ \xi\in\Xi.
\end{aligned} \tag{1}$$

The first constraint ensures that at most one job is assigned to the machine at each starting time $t \in \mathcal{S}$ taking into account the prescribed finishing times of jobs f_j^0. The second constraint prescribes that each job is processed by exactly one machine. The third constraint expresses the random number of overlaps $\tilde{y}_{jc}(\xi)$ caused by random delay $D_j(\xi)$ of job j. The expectation of all these overlaps is then penalized in the objective function. Finally, we consider restrictions on binary and integer variables. Note that the problem does not contain any first stage costs and the objective corresponds to a simple recourse in the two-stage stochastic programming settings, cf. [17].

3 Tractable reformulation: generalized robust coloring

In previous section, we introduced the two-stage stochastic programming formulation with an expectation in the objective which cannot be directly solved. In general, we would need to employ an approximation technique which enables us to formulate and solve the problem by available optimization tools, e.g., sample average approximation, cf. [17], where the random distribution is replaced by a finite sample. However, in [7], it was shown that the two-stage problem can be reformulated as a robust coloring problem with a special choice of penalties which use the underlying distribution of random delays without any approximation. We will generalize the problem to cover the heterogeneous machines and job classes.

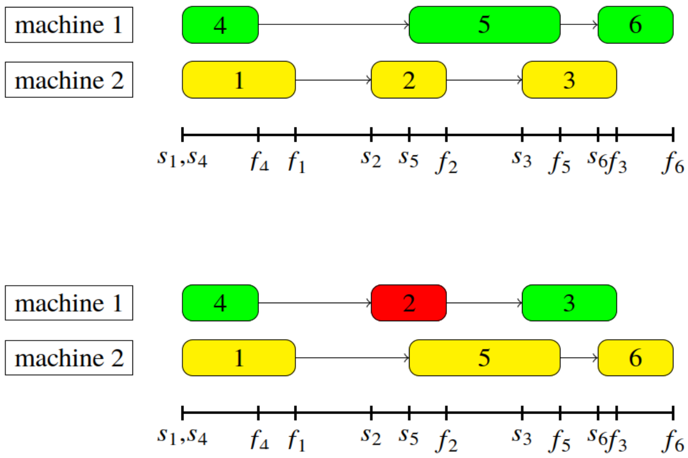


Figure 1 FIS schedules: two possible assignments of 6 jobs to 2 machines. If the machines are identical, both schedules are feasible. However, if job 2 cannot be processed by the first machine, then the second schedule is not feasible.

Robust coloring problem [18] aims at assigning colors $c \in C$ to vertices $j \in \mathcal{J}$ where two vertices cannot be colored by the same color if they are connected by a hard edge in E, whereas two vertices can share the same color for a penalty q if they are connected by a soft edge in $\overline{E}$. The goal is to minimize the sum of penalties under preserving the hard edges. This problem is related to the fixed interval scheduling if we consider the colors as machines, the vertices as the jobs and hard edges connect jobs which cannot be processed by the same machine because of interval overlaps, i.e.

$$E = \{(j, j') : [s_j, f_j) \cap [s_{j'}, f_{j'}) \neq \emptyset\}.$$

Note that Branda et al. [7] considered identical machines only, however, we would like to take into account heterogeneous machines and job classes. We can realized that machines correspond to colors and vertices to jobs, so in our case the vertices cannot be colored by all colors. So we can use the notation introduced in the previous section, in particular $\mathcal{J}_c \subset \mathcal{J}$ denotes the jobs which can be colored by color c, and vice versa $C_j \subset C$ denotes the colors which can be used to color vertex j. Moreover, the set of hard edges E can be extended by edges which correspond to jobs with disjunctive processing intervals, but which cannot be processed by the same machine. These edges are then removed from the set of soft edges $\overline{E}$. This leads us to the following generalized robust coloring formulation

$$\begin{aligned}
\min_{x,y} \quad & \sum_{\{j,j'\} \in \overline{E}} q_{jj'} y_{jj'} \\
\text{s.t.} \quad & \sum_{c \in C_j} x_{jc} = 1, \ j \in \mathcal{J}, \\
& x_{jc} + x_{j'c} \leq 1, \ \{j, j'\} \in E, c \in C, \\
& x_{jc} + x_{j'c} \leq 1 + y_{jj'}, \ \{j, j'\} \in \overline{E}, c \in C, \\
& x_{jc} \in \{0, 1\}, \ c \in C_j, j \in \mathcal{J}, \\
& y_{jj'} \in \{0, 1\}, \ \{j, j'\} \in \overline{E}.
\end{aligned} \tag{2}$$

The objective is to minimize the penalty for same colored vertices subject to that each vertex is colored by exactly one color, hard edges are preserved whereas soft edges can be violated for a penalty which is assigned using the binary additional binary variables $y_{jj'}$. Note that variables y have different meaning than variables $\tilde{y}$ which were used in the previous section in two-stage problem (1). We can show that if the penalties are chosen in a proper way than problem (2) is equivalent to the two-stage program (1). These penalties are set to

$$q_{jj'} = P(D_j(\xi) > s_{j'} - f_j^0).$$

Instance	No. of machines	No. of jobs	Time (s)	Opt. value	Rel. gap
1	10	40	20	0.86	0.0097
2	10	40	19	17.21	0.0088
3	10	40	18	2.64	0.0083
4	10	40	28	14.51	0.0097
5	10	40	74	0.25	0.0099
6	10	40	**3600**	0.51	0.0336
7	10	40	281	0.26	0.01
8	10	40	22	0.12	0.0099
9	10	40	37	28.76	0.0099
10	10	40	13	0.14	0.0092

Table 1 Results of solving 10 test instances (instance sizes, computational time in seconds, optimal value and relative optimality gap)

We can modify the original proof which was proposed in [7] where it was shown that the expected number of overlaps caused by the delay $D_j(\xi)$ of job j is equal to

$$\mathbb{E}_P\left[y_{jc}(\xi)\right] = \sum_{n=1}^{N_j} P(D_j(\xi) > s_{k_n} - f_j^0),$$

where N_j is the number of jobs which are assigned to the same machine after job j and s_{k_n} denote the starting times of these jobs. This means that all penalties q_{jk_n} must be summed in the objective. These correspond directly to the penalty which is assigned to vertices with the same color in the generalized robust coloring problem. Moreover, we can realize that the assignment of jobs as well as coloring of vertices are restricted by the job classes and eligible machines.

4 Numerical study

In the numerical study, we will use simulated instances to verify performance of the proposed robust coloring reformulation. We use the exponential distributions with parameter $\lambda_1 = 0.2$ for the simulation of job lengths and $\lambda_2 = 0.05$ for the lengths of breaks between the jobs. We simulated two sets of 10 instances with 40 and 50 jobs. These jobs are then assigned to 10 available machines under the following limitations: originally the jobs are simulated to particular machines, whereas in the optimization problem, they are allowed to be assigned to the neighboring machines only. This naturally creates the job classes and heterogeneous machines. The following formula is employed for the cumulative distribution function of the random delay

$$P(D_j \leq x) = F_j(x) = p_j + (1 - p_j)(1 - e^{-\lambda_j x}),\ x \geq 0. \tag{3}$$

where parameter values $\lambda_j = 0.2$ and $p_j = 0.9$ are used for all jobs $j \in \mathcal{J}$. Remind that no approximation of the probabilistic distribution is necessary and the formula (3) is used directly to compute the penalties q.

To solve the simulated instances and resulting mixed-integer linear programming problems, we used GAMS modelling system and the solver Cplex installed on PC with Windows 10 operational system, 32 GB RAM memory and Intel(R) Xeon(R) CPU E3-1220 v5, 3.00 GHz. The relative optimality gap was set to 0.01 and maximal solution time is restricted to one hour.

Tables 1, 2 show the summary of the numerical experiment. We can observe from Table 1 that almost all instances with 40 jobs (except to one) were solved to optimality in reasonable amount of time. On the other hand, only two out of ten instances with 50 jobs reached the relative optimality gap of 1 % in the one hour time limit. This means that – at least under our choice of the parameters – we arrived at a "breakpoint" (between 40 and 50 jobs) from which this kind of generalized robust coloring problem is difficult to solve to optimality using available PC.

Instance	No. of machines	No. of jobs	Time (s)	Opt. value	Rel. gap
1	10	50	20	12.45	0.0057
2	10	50	**3600**	95.38	0.5456
3	10	50	**3600**	43.31	0.0326
4	10	50	2259	310.23	0.01
5	10	50	**3600**	21.65	0.0336
6	10	50	**3600**	17.84	0.1437
7	10	50	**3600**	8.80	0.0617
8	10	50	**3600**	7.78	0.2725
9	10	50	**3600**	75.58	0.0135
10	10	50	**3600**	3.05	0.0471

Table 2 Results of solving 10 larger test instances (instance sizes, computational time in seconds, optimal value and relative optimality gap)

5 Conclusion

In this paper, we have dealt with the optimal scheduling of loading/unloading operation in depots where the problem can be seen as a special case of general fixed interval scheduling problems. In our case, this problem involves job classes and heterogeneous machines which corresponds to the case that not all gates can serve all vehicles. We have proposed a two-stage integer stochastic programming problem formulation. Moreover, a generalized robust coloring problem has been introduced and it has been shown to be equivalent to our problem under suitable choice of the penalties. In the numerical study, we have shown that the middle-sized instances can be solved to optimality by mixed-integer programming solvers in reasonable time. However, if the size of the problem increases, a special designed algorithm will be necessary. For example, a strong decomposition algorithm can help us to solve larger problem instances to optimality, cf. [1]. In future research, we will focus also on alternative criteria which can take into account the uncertainty inspired by papers [5, 6].

Acknowledgements

The work was supported by the Grant Agency of the Czech Republic under the grant project 22-11867S. This support is gratefully acknowledged.

References

[1] Adam, L., Branda, M., Heitsch, H. & Henrion, R. (2020). Solving joint chance constrained problems using regularization and Benders' decomposition. *Annals of Operations Research*, 292, 683–709.

[2] Branda, M.(2017).Astochastic–integerprogrammingapproachtotacticalfixedintervalschedulingproblems. In P. Pražák (Ed.), *Proceedings of the 35th International Conference on Mathematical Methods in Economics 2017* (pp. 78–83), Hradec Králové.

[3] Branda, M. (2018). Distributionally robust fixed interval scheduling on parallel identical machines under uncertain finishing times. *Computers & Operations Research*, 98, 231–239.

[4] Branda, M. & Hájek, Š. (2017). Flow-based formulations for operational fixed interval scheduling problems with random delays. *Computational Management Science*, 14(1), 161–177.

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

[6] Branda, M. & Kopa, M. (2016). DEA models equivalent to general N-th order stochastic dominance efficiency tests. *Operations Research Letters*, 44(2), 285–289.

[7] Branda, M., Novotný, J. & Olstad, A. (2016). Fixed interval scheduling under uncertainty – a tabu search algorithm for an extended robust coloring formulation. *Computers & Industrial Engineering*, 93, 45–54.

[8] Calzavara, G., Iori, M., Locatelli, M., Moreira, M.T.C & Silveira, T. (2021). Mathematical models and heuristic algorithms for pallet building problems with practical constraints. *Annals of Operations Research*, online first, DOI: 10.1007/s10479-021-04349-w.

[9] Eliiyi, D.T. (2013). Integrating tactical and operational decisions in fixed job scheduling. *Engineering Optimization*, 45(12), 1449–1467.

[10] Gajda, M., Trivella, A, Mansini, R. & Pisinger, D. (2022). An optimization approach for a complex real-life container loading problem. *Omega*, 107, 102559.
[11] Kolen, A.W., Lenstra, J.K., Papadimitriou, C.H. & Spieksma, F.C. (2007). Interval scheduling: a survey. *Naval Research Logistics*, 54(5), 530–543.
[12] Kovalyov, M.Y., Ng, C. & Cheng, T.E. (2007). Fixed interval scheduling: models, applications, computational complexity and algorithms. *European Journal of Operational Research*, 178(2), 331–342.
[13] Kroon, L., Salomon, M. & Van Wassenhove, L. (1997). Exact and approximation algorithms for the tactical fixed interval scheduling problem. *Operations Research*, 45, 624–638.
[14] Nascimento, O.X., Queiroz, T.A. & Junqueira, L. (2021). Practical constraints in the container loading problem: comprehensive formulations and exact algorithm. *Computers & Operations Research*, 128, 105186.
[15] Ng, C.T., Cheng, T.C.E., Bandalouski, A.M., Kovalyov, M.Y. & Lam, S.S. (2014). A graph-theoretic approach to interval scheduling. *Journal of the Operational Research Society*, 65(10), 1571–1579.
[16] Smet, P., Wauters, T., Mihaylov, M. & Berghe, G.V. (2014). The shift minimisation personnel task scheduling problem: a new hybrid approach and computational insights. *Omega*, 46, 64–73.
[17] Shapiro, A., Dentcheva, D. & Ruszczyński, A. (2009). *Lectures on stochastic programming: Modeling and theory*. SIAM, Philadelphia.
[18] Yánez, J. & Ramírez, J. (2003). The robust coloring problem. *European Journal of Operational Research*, 148(3), 546–558.
[19] Wolsey, L.A. (1998). *Integer programming. Wiley-Interscience series in discrete mathematics and optimization*, J. Wiley & sons, New York (N.Y.), Chichester, Weinheim.
[20] Zhou, S., Zhang, X., Chen, B. & van de Velde, S. (2014). Tactical fixed job scheduling with spread-time constraints. *Computers & Operations Research*, 47, 53–60.

Mathematical Model for Allocation of Aircraft on Airport´s Apron

Jakub Cíleček[1], David Mičulka[2], Dušan Teichmann[3]

Abstract. The presented article deals with the allocation of handling spots on the international regional airport´s apron. In seasonal months the ground handling capacity seems to decrease and one of the possible solutions how to avoid this problem is to allocate the handling spot appropriately. With optimal handling spots, we can decrease the time of embarking and disembarking of passengers which leads to a decrease in time needed for the process of technical handling as well as increasing the airport´s handling capacity. For calculation of the solution to this problem, we created a linear mathematical model of the allocation of aircraft on the airport´s apron presented in this article. Optimizing criterium in the mentioned model was the time of embarking and disembarking of passengers. The value of optimizing criterium was minimalized. A calculation experiment was conducted in the conditions of the international regional airport Ostrava.

Keywords: Apron management, Aircraft Allocation, Mathematical Model

JEL Classification: C69, C44
AMS Classification: 90C10

1 Introduction

One of the elementary problems which appear in airports is the allocation of aircraft on the airport´s apron. The solution to this problem is individual for each airport and it depends on some factors -technical equipment of airport, organization of apron, and type of handled aircraft. The hub airports with a high level of technical amenities are the solution to this problem partially influenced by the contract between airport and airline. It means that if the airline is capable of paying a higher fee then they can have a better stand at the apron. In international regional airports with lower levels of amenities, it is appropriate to allocate a stand at the apron based on the time aspect of onboarding and disembarking times.

2 Current State Analysis

The process of allocation of the stand is done before the actual arrival of the aircraft at the airport, normally this process is done at the beginning of the day or work shift. If there is only one aircraft in any time slot, then the problem of allocation of the stand is not needed and the process of allocation of the stand does not need to be conducted. If there is more than one aircraft planned to be at the apron in any time slot then there is a need for the decision-making tool in any form. In [3], we can see the tool needs to abide by all the safety measures mentioned in manuals. At the international regional airports, the allocation of the handling spots is done by the manager of airport traffic who usually allocates the stands based on internal rules or his experience. The optimal allocation of stands directly influences the length of ground handling and on handling capacity of the airport. At international hub airports which has more stands and different level of amenities, the decision-making process is complex and the airport´s responsible worker is not able to allocate the stand-in same manner, which is the reason that decision-making tools exist.

In [4], we can see that the authors are dealing with gate allocation problems for various flights in a manner to lower customer dissatisfaction. This dissatisfaction is characterized by the merging of two factors – the walking time of passengers and delays in flights. For passengers and transiting passengers mathematical model was created and it considers the arrival and departure times of aircraft. There are implemented conditions that assure operational safety, namely the restriction of allocating one stand to two aircraft in the same time slot. The solution was found using the Benders decomposition method at San Francisco airport in one day. The results show the percentual

[1] VSB-TUO/Institute of Transport, 17. listopadu 15/2172, 708 33 Ostrava-Poruba, Jakub.cilecek@vsb.cz.
[2] VSB-TUO/Institute of Transport, 17. listopadu 15/2172, 708 33 Ostrava-Poruba, David.miculka@vsb.cz
[3] VSB-TUO/Institute of Transport, 17. listopadu 15/2172, 708 33 Ostrava-Poruba, Dusan.teichmann@vsb.cz

contribution of walking times of transit, non-transit passengers, and delays of arriving and departing aircraft on the value of the objective function.

In [5], we can see that the authors deal with dynamic allocation of stand-in situations when allocation cannot be used because of the turbulent operational situation. The authors suggested a multicriterial dynamic model for the allocation of stands. Their model provides an effective solution in the decision-making process on short time notice and the model can react to deals of aircraft, emergencies, and non-standard situations. The model uses fuzzy logic as an approach that allows us to make decisions even under uncertainty.

In [1], we can see that the authors deal with the allocation of flights to the gate. This problem is solved on daily basis and the operational condition of airports affects its complexity. Authors with their research aimed at the possibility of a solution to the said problem to increase the availability of gates for passengers and handling capacity. Authors analyzed theoretical and practical approaches for the solution of the problem of gate allocation with mathematical modeling, algorithms, heuristic, and metaheuristic algorithms.

In [2], we can see that the authors deal with an innovative approach to the allocation of a handling spot. These handling spots are divided into two categories, remote handling spots and handling spots that are close to the terminal and have higher levels of amenities for example boarding bridges. Authors are solving this problem for a few possible operational situations with the use of mathematical models. The solution is minimalization of walking distance of passengers, minimalization of tows of aircraft, maximization of outbound passengers who board or disembark the aircraft via the bridge, and maximization of potential profit. The research was conducted at the international airport in Sao Paulo, where the allocation of handling spots was tested for six days. The result was an increase in overall effectivity with the application of mathematical models.

The novelty is the creation of a decision tool for the international regional airports where handling spots are allocated based on internal procedures and responsible worker´s experience. With optimal handling spot allocation, we can achieve shorter walking times for passengers from gate to aircraft and vice versa. With mentioned decision tool we can help the responsible worker to allocate the handling spots and we can lower his workload as well. Even though some variation of decision-making tool is implemented in international hub airports, at the international regional airports this decision-making is burdened with the human factor, which can lead to not optimal allocation of handling spots. This can lead to prolongation of the handling process, blocking of another handling spot or necessary delay of aircraft can be created.

3 Proposed Mathematical Model

3.1 Formulation of Optimization Task

The set of aircrafts I is given and their handling has to be completed in chosen time frame. Next set J consists of handling spots at the airport's apron. For each handled aircraft, we defined modes of passengers based on the airport of arrival and departure, mainly if they are in the Schengen area or not. At the terminals, we have 4 gates which are for departure to the airport within the Schengen area, departure outside of the Schengen area, arrival from the Schengen area and arrival outside the Schengen area. The matrix of time consumptions of transfers $\boldsymbol{D}$ is given. The element d_{ij} the represents the time complexity of passenger movements from flights operated by the aircraft $i \in I$ in case it will be allocated to the handling spot $j \in J$. The task is to design and allocate the handled aircraft to the handling spot on the apron so that the total time required by the passenger flows between the handling spot and the terminal is minimal.

For decision modelling purposes, a group of decision variables x_{ij} will be introduced. The variable x_{ij} will model the assignment of flight $i \in I$ to handling spot $j \in J$. If $x_{ij} = 1$ is valid after the optimization calculation, then the aircraft operating the departing flight or the arriving flight or the pair of arriving and departing flights $i \in I$ will be assigned to the operating position $j \in J$. If $x_{ij} = 0$ is valid after the optimization calculation, then the aircraft operating the departing flight or the arriving flight or the pair of arriving and departing flights $i \in I$ will not be assigned to the handling spot $j \in J$.

The constraints in the model must ensure that:

- each aircraft will be assigned exactly one handling spot,
- passenger flows do not intersect at boarding and alighting of individual aircraft,
- aircrafts that will be checked in at the same time will not be assigned adjoining handling spot,
- for time-colliding flights, the model will not assign the same service point to multiple flights,

- passengers will flow from/to the gate characterizing the destination of departure or arrival (schengen, non-schengen).

Furthermore, a $\boldsymbol{A}$-incidence matrix will be introduced into the model. The elements of the matrix $\boldsymbol{A}$ indicate whether the flight $i \in I$ conflicts with the flight $k \in I$. If the flights $k \in I$ and $i \in I$ are time colliding, then $a_{ki} = 1$, otherwise $a_{ki} = 0$.

Furthermore, for the proposed model, the values of the elements of the matrix $\boldsymbol{D}$ represent the transfer times that the passengers of the arriving flight / departing flight / pair of arriving and departing flights must $i \in I$ when the aircraft serving said flight / pair of flights is assigned to the service position $j \in J$.

3.2 Mathematical Model

The mathematical model has the form:

$$\min f(x) = \sum_{i \in I} \sum_{j \in J} d_{ij} x_{ij} \tag{1}$$

Under the constraints:

$$\sum_{j \in J} x_{i,j} = 1 \qquad \text{for } i \in I \tag{2}$$

$$x_{k,j} + x_{i,j} \leq 1 \qquad \text{for } k \in I, i \in I, j \in J, a_{ki} = 1 \tag{3}$$

$$x_{k,1} + x_{i,2} \leq 1 \qquad \text{for } k \in I, i \in I, a_{ki} = 1 \tag{4}$$

$$x_{k,j} + x_{i,j-1} + x_{i,j+1} \leq 1 \qquad \text{for } k \in I, i \in I \text{ a } j \in J \setminus \{1, n\}, a_{ki} = 1 \tag{5}$$

$$x_{k,n-1} + x_{i,n} \leq 1 \qquad \text{for } k \in I, i \in I, a_{ki} = 1 \tag{6}$$

$$x_{ij} \in \{0; 1\} \qquad \text{for } i \in I, j \in J \tag{7}$$

Function (1) represents the optimization criterion, the total distance that passengers of all flights must cover. The restraint group (2) shall ensure that each incoming flight / departure flight / pair of flights is assigned to only one handling spot. The group of restrictive conditions (3) will ensure that in the case of time-collision flights, these flights will not be filled by the same handling spot. The groups of restrictive conditions (4) - (6) will ensure that in time collision flights there will be no situation where the aircraft will be operated in parallel at adjoining handling spots (due to operating restrictions - the need for so-called "self-maneuvering"). The group of constraint conditions (7) defines the domains of variables used in the model.

4 Computational Experiment

The computational experiment with the mathematical model was designed for operating conditions at the international regional airport Ostrava-Mošnov. The list of flights in Table 1 is taken from the real flight schedule valid for the summer season of 2022. The arrival and departure times in Table 1 represent the arrival times of aircraft on the apron and departures of aircraft from the apron

	Scheduled time of arrival	Destination	Scheduled time of departure	Destination
Flight 1	01:25	Mallorca (SA)	04:45	Rhodos (SD)
Flight 2	10:45	Lefkada (SA)	11:45	Lefkada (SD)
Flight 3	11:15	Rhodos (SA)	12:20	Heraklion (SD)
Flight 4	11:50	Marsa Alam (NSA)	12:40	Marsa Alam (NSD)
Flight 5	16:45	Burgas (NSA)	17:35	Burgas (NSD)
Flight 6	18:20	Heraklion (SA)	19:10	Korfu (SD)
Flight 7	18:45	Bodrum (NSA)	19:30	Bodrum (NSD)

Table 1 Flight schedule from summer season 2022 at Ostrava Airport

For the needs of the model, a flight is always understood as a pair of destinations (the destination from which the flight arrives and the destination to which the flight departs).

The model also includes values from the time-consuming matrix $\boldsymbol{D}$ determined empirically, see Table 2. In Table 2, the time-consuming values of passenger movements are given in minutes. For example the value at the position of element d_{11} represents the sum of time requirements of passenger movements on the apron from handling spot 1 to the gate designated for passenger entry for check-in after the arrival of the flight from the destination in the

Schengen Area and the gate designated for passenger entry to the apron before departure of the flight to the destination in Schengen area.

Flight \ Handling spot	1	2	3	4	5	6
1	27	23	25	29	31	35
2	27	23	25	29	31	35
3	27	23	25	29	31	35
4	25	21	23	25	31	37
5	25	21	23	25	31	37
6	27	23	25	29	31	35
7	25	21	23	25	31	37

Table 2 Values of time-consuming movements in minutes

It is clear from the flight schedule for the given day that flights 1 and 5 will occur separately on the airport apron. The collision flights will be flight 2 and flight 3, flight 3 and flight 4, and flight 6 and flight 7. An example of a collision situation is shown in Figure 1.

After completion of the optimization calculation, flights were assigned to the given service points according to Table 3 and the cumulative value of passenger crossings was set at 165 minutes, see Table 3.

Flight	1	2	3	4	5	6	7
Allocated handling spot	2	2	4	2	2	2	4

Table 3 Results of the experiment

The solutions for individual collision flights are shown in Figures 1 - 3. Add pictures for the remaining collision situations and explain the types of arrows.

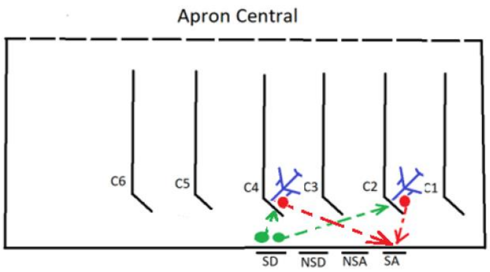


Figure 1 Deployment scheme of aircraft operating the flights 2 and 3 on the apron

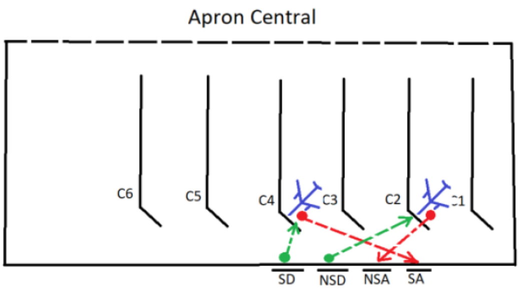


Figure 2 Deployment scheme of aircraft operating flights 3 and 4 on the apron

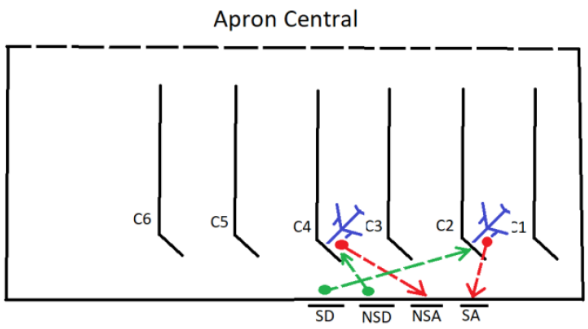


Figure 3 Deployment scheme of aircraft operating flight 6 and 7 on the apron

Figure 1 shows the location of the aircraft for the first conflict situation (aircraft operating flights 2 and 3). Passengers from flight 2 will, after arrival, walk on the apron in the time interval from 10:45 to 10:53 and switch to departure in the time interval from 11:30 to 11:45. Passengers from flight 3 will, after arrival, walk on the apron in the time interval from 11:15 to 11:32 and switch to departure in the time interval from 12:08 to 12:20. The flow of departing passengers of the second flight intersects with the arriving passengers of the third flight, therefore one of the flights will have to be transported by airport bus to ensure operational safety (passenger crossing in front of a stopping or departing aircraft, boarding of passengers on the wrong plane, etc.)

Figure 2 shows the location of the aircraft for the first conflict situation (aircraft operating flights 3 and 4). Passengers from flight 3 will, after arrival, walk on the apron in the time interval from 11:15 to 11:32 and switch to departure in the time interval from 12:08 to 12:20. Passengers from flight 4 will, after arrival, walk on the apron in the time interval from 11:50 to 11:59 and switch to departure in the time interval from 12:28 to 12:40. There will not be intersecting flows of passengers between flights on apron.

Figure 3 shows the location of the aircraft for the first conflict situation (aircraft operating flights 6 and 7). Passengers from flight 6 will, after arrival, walk on the apron in the time interval from 18:20 to 18:28 and switch to departure in the time interval from 18:55 to 19:10. Passengers from flight 7 will, after arrival, walk on the apron in the time interval from 18:45 to 19:00 and switch to departure in the time interval from 19:20 to 19:30. The flow of departing passengers of the sixth flight intersects with the arriving passengers of the seventh flight, therefore one of the flights will have to be transported by airport bus to ensure operational safety (passenger crossing in front of a stopping or departing aircraft, boarding of passengers on the wrong plane, etc.)

5 Conclusion

The operation at the Ostrava Regional Regional Airport is mainly seasonal, and in order to increase the handling capacity and efficiency of the manager's work, a mathematical model was designed to serve as an aid in the process of deciding on the allocation of handling spot for aircraft. The input quantities are information on the number of service points at the selected airport, the matrix of transition times from the airport gate to the service point to which the aircraft operating the relevant flight is assigned. The result of the mathematical model designed to solve this problem and described in the article is to obtain information on the allocation of handling spot to aircraft that will be served at the airport on the selected day of the summer season according to the applicable flight schedule so that the cumulative transition distance traveled by passengers is minimized.

Acknowledgements

This paper was funded from project SP2022/62 Development and research in transport and logistics.

References

[1] Bouras, A., Ghaleb, M.A., Suryahatmaja, U.S. & Salem, A.M., 2014. The airport gate assignment problem: a survey. *The scientific world journal*.
[2] Dijk, B., Santos, B.F. & Pita, J.P., 2019. The recoverable robust stand allocation problem: a GRU airport case study. *Or Spectrum*, 41(3), pp.615-639.
[3] IATA *Ground Handling Manual*. (2 018)

[4] Rashidi Komijan, A. & Nasrollahpourniazi, F., 2020. A new mathematical model for gate assignment problem considering transit passengers and safety constraints: Benders decomposition approach. *Journal of Industrial Engineering International*, 16(4), pp.82-92.
[5] Skorupski, J. & Żarów, P., 2021. Dynamic management of aircraft stand allocation. *Journal of Air Transport Management*, 90, p.101964.

Potential Output during Great Recession and Covid-19 Crisis

Andrea Čížků[1]

Abstract. Potential output is an important economic concept representing the level of economic activity sustainable in the long run. Potential output is traditionally considered to be driven by supply-side factors such as labor supply, capital investments and new technologies. An associated concept of the output gap is usually seen as short-run deviations of actual output from its potential due to demand-side fac- tors. The great recession of 2008 slightly altered this traditional view and suggested that a demand-side factors might also play an important role on potential output. This paper proposes a modelling framework in which short-run demand-driven fluctuations of the output gap might have an important effect on the potential output and its long-run growth. Unobserved components methodology is applied and original nonlinear conditionally Gaussian state space model is formulated and econometrically estimated for the Czech Republic by maximum likelihood methodology for the time period 1996Q1-2021Q4 including great recession of 2008 and recent covid-19 crisis.

Keywords: output gap, potential output, economic crisis, unobserved components methodology, state space form, maximum likelihood.

JEL Classification: C51
AMS Classification: 90C15

1 Introduction

The global economic crisis of 2008 has initiated suggestions that demand-side factors might have permanent impact on output (Ball [3], Blanchard [5]). Andersson et al. [2] argue on the basis of hysteresis hypothesis according to which demand-driven shortfalls can perpetuate themselves by lowering level of potential output or its growth rate. Demand-driven recessions might erode skills of unemployed workers and firms may cut their innovation budgets which in turn lowers (growth of) potential output.

Empirical verification of the hysteresis hypothesis is complicated by the fact that potential output is an unobservable variable which cannot be measured directly and can be only estimated. Simple statistical procedures such as the Hodrick-Prescott (HP) filter proposed by Hodrick, Prescott [12] or Beveridge-Nelson decomposition (Beveridge, Nelson [4]) have many disadvantages which are summarized by Harvey, Jaeger [10]. Unobserved components (UC) methodology pioneered by Watson [14] and advocated by Harvey [9] is considered to be superior to simple statistical procedures and is commonly applied in empirical literature (e.g. Andersson et al. [2], Borio et al. [7], Kuttner [13]).

This paper applies unobserved components approach to model potential output and the gap. Specifically, the original nonlinear conditionally Gaussian state space model is formulated and estimated by the method of maxi- mum likelihood. Unlike other empirical literature of this kind, the formulated model explicitly describes the effect of output gap on the (growth of) potential output. The strength of this effect will also be quantified by performing econometric estimation of the model parameters and statistically tested by associated methods of statistical inference.

Structure of the paper is as follows. Firstly, the model is described in chapter 2. The subsequent chapter 3 shows how the model can be written in a state space form which enables applications of Kalman filter algorithm and maximum likelihood estimation of the model parameters. Chapter 4 describes the data and their sources. The results of econometric estimation of the parameters and unobserved state variables are presented and discussed in chapter 5. Final chapter 6 summarizes main findings and conclusions.

[1] Prague University of Economics and Business, Department of Econometrics, sqr. W. Churchille 4, Prague 3, cizeko@vse.cz.

2 Model

Unobserved component methodology decomposes (log of) output y_t as follows:

$$y_t = \bar{y}_t + \hat{y}_t\,, \tag{1}$$

where $\bar{y}_t$ is potential output and $\hat{y}_t$ represents the output gap.

It is usually assumed (Harvey [9], Watson [14]) that trend component follows a random walk with drift and a cyclical component is commonly modelled as AR(2) process. Nonetheless, the model presented in this paper will differ from this popular specification. Trend of the potential output is modelled as dependent on the output gap, which describes the hysteresis effect of the short-run fluctuations on the long run growth of the potential output:

$$\bar{y}_t = \bar{y}_{t-1} + \mu + \alpha_{t-1} \cdot \hat{y}_{t-1}\,, \tag{2}$$

where the expression $\mu + \alpha_{t-1} \cdot \hat{y}_{t-1}$ models potential output growth rate, parameter μ represents its constant part and the second term $\alpha_{t-1} \cdot \hat{y}_{t-1}$ describes asymmetrical influence of the output gap on potential output growth. Asymmetry is modeled by time-varying parameter

$$\alpha_{t-1} = \begin{cases} 0, & \hat{y}_{t-1} \geq 0, \\ \alpha, & \hat{y}_{t-1} < 0, \end{cases} \tag{3}$$

which is equal to zero if $\hat{y}_{t-1} \geq 0$ while $\alpha_{t-1} = \alpha > 0$ when $\hat{y}_{t-1} < 0$. This specification introduces nonlinearity into the model. Economic reason for this nonlinear specification is that there are upper bounds on the growth of potential output which cannot be exceeded by short-run demand driven fluctuations. On the other hand, it is assumed with accordance to the hysteresis hypothesis that short-run recessions do have a negative impact on the growth of potential output.

The output gap is modelled according to a stationary AR(1) process. Nonetheless, this process is modified by an inertial variable v_t representing an impact of external shocks caused by the great recession of 2008 and recent covid-19 crisis:

$$\hat{y}_t = \lambda \cdot \hat{y}_{t-1} - v_t + u_t\,, \tag{5}$$

where λ describes inertia of the output gap, $u_t \sim N\left(0, \sigma^2\right)$ is i.i.d. random error and the variable v_t is modeled by a stationary AR(1) process:

$$v_t = \omega \cdot v_{t-1} + \varepsilon_t\,, \tag{5}$$

where ω represents inertia of the variable v_t and ε_t can attain only two values – it is equal to $\varepsilon > 0$ at the beginning of the great recession of 2008 or recent covid-19 crisis and zero otherwise:

$$\varepsilon_t = \begin{cases} \varepsilon, & t = 2008Q2,\ 2020Q1, \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

Model parameters μ, α, λ, ε and σ are econometrically estimated by the method of maximum likelihood and unobserved state variables $\bar{y}_t$, $\hat{y}_t$ are estimated by Kalman filter. Application of this methodology necessitates to put the model in state space form which is described in the next subchapter.

3 State space form

Formulated nonlinear model with time-varying parameter α_{t-1} can be expressed as a conditionally Gaussian state space model with the following transition and measurement equation:

$$\mathbf{x_t} = \mathbf{A_t}(\mathbf{Z_{t-1}}) \cdot \mathbf{x_{t-1}} + \mathbf{R_t}(\mathbf{Z_{t-1}}) \cdot \mathbf{u_t}, \tag{7}$$

$$\mathbf{z_t} = \mathbf{D_t}(\mathbf{Z_{t-1}}) \cdot \mathbf{x_t} + \boldsymbol{\upsilon}_\mathbf{t}, \tag{8}$$

where $\mathbf{x_t}$ is a vector of unobserved state variables, $\mathbf{z_t}$ is a vector of observed indicator variables, $\mathbf{u_t} \sim N(0;\Sigma_{uu})$, $\boldsymbol{\upsilon}_\mathbf{t} \sim N(0;\Sigma_{\upsilon\upsilon})$ are i.i.d. vectors of mutually uncorrelated random errors with Σ_{uu} and $\Sigma_{\upsilon\upsilon}$ being their covariance matrices and $\mathbf{A_t}(\mathbf{Z_{t-1}})$, $\mathbf{R_t}(\mathbf{Z_{t-1}})$, $\mathbf{D_t}(\mathbf{Z_{t-1}})$ are system matrices with parameters. The symbol $\mathbf{Z_{t-1}}$ in $\mathbf{A_t}(\mathbf{Z_{t-1}})$ indicates that the matrix $\mathbf{A_t}$ depends on the values of the observed variables till the time $t-1$, i.e. on the vector $\mathbf{Z_{t-1}} \equiv (\mathbf{z_1}',\ldots,\mathbf{z_{t-1}}')'$.

The state vector is defined as $\mathbf{x_t} = (\bar{y}_t \quad \hat{y}_t \quad v_t \quad 1_t)'$ and the system matrices take the form:

$$\mathbf{A_t}(\mathbf{Z_{t-1}}) = \begin{pmatrix} \mathbf{e_1} + \mu \cdot \mathbf{e_4} + \alpha_{t-1} \cdot \mathbf{e_2} \\ \lambda \cdot \mathbf{e_2} - (\omega \cdot \mathbf{e_3} + \varepsilon_t \cdot \mathbf{e_4}) \\ \omega \cdot \mathbf{e_3} + \varepsilon_t \cdot \mathbf{e_4} \\ \mathbf{e_4} \end{pmatrix} \qquad \mathbf{R_t}(\mathbf{Z_{t-1}}) = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \qquad \mathbf{D_t}(\mathbf{Z_{t-1}}) = (\mathbf{e_1} + \mathbf{e_2})$$

where $\mathbf{e_k}$ is a $(1 \text{ x } 4)$ vector, which has a number 1 at the k-th position and zeros elsewhere.

Random error in the transition equation is $\mathbf{u_t} = u_t$ and in the measurement equation $\boldsymbol{\upsilon}_\mathbf{t} = 0$.

Because the output gap $\hat{y}_t$ is unobservable, the time-varying parameter α_{t-1} had to be slightly redefined as:

$$\alpha_{t-1} = \begin{cases} 0, & \hat{y}_{t-1|t-1} \geq 0 \\ \alpha, & \hat{y}_{t-1|t-1} < 0 \end{cases}$$

where $\hat{y}_{t-1|t-1}$ represents the Kalman filter estimate of the unobserved variable $\hat{y}_{t-1}$.

Once written in the state space form, maximum likelihood method was used to estimate the parameters. Firstly, square root version of the Kalman filter algorithm was applied to construct likelihood function which was maximized by standard numerical optimization techniques in Matlab. Standard deviations of the estimated parameters were estimated using the Fisher matrix. For details on the square root version of the Kalman filter and the construction of the likelihood function see Anderson, Moore [1] and Harvey [9]. Secondly, Kalman filter was applied once again after the econometric estimation of the parameters to estimate unobserved state variables (potential output and the gap).

4 Data

Quarterly real GDP (in millions of CZK) in the Czech Republic for the time period 1996Q1-2021Q4 were used and obtained from the Czech Statistical Office database [8]. Seasonal adjustment was performed by the Census X-13 method. The model variable y_t was calculated as logarithm of seasonally adjusted quarterly real GDP.

5 Results and discussion

Results from econometric estimation are summarized in the following table 1. Standard errors are indicated in parentheses below estimated coefficients and symbols ** and *** denote that the given parameter is statistically significant at 5% and 1% level of significance.

$\hat{\mu}$	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\omega}$	$\hat{\varepsilon}$	$\hat{\sigma}$
0.021 (0.001)***	0.380 (0.140)***	0.300 (0.134)**	0.976 (0.012)***	0.034 (0.010)***	0.022 (0.002)***

Table 1 Estimation results of the UC model (1)–(6) for the Czech Republic (1996Q1-2021Q4)

Firstly, all the estimated coefficients are statistically significant even at 1% level of significance except the parameter λ which is statistically significant at 5% level of significance. This results clearly shows the relevancy of the formulated model. Especially important is the statistical significance of the parameter α representing the hysteresis effect of demand driven recessions on the growth of potential output.

Long-run growth of the potential output in times of economic booms ($\hat{y}_{t-1} \geq 0$) was estimated to be equal to $\hat{\mu} = 0.021$. Potential output thus growths by 2.1% each quarter when output is above potential. In times of economic recessions, growth of potential output is estimated by the expression $\hat{\mu} + \hat{\alpha}_{t-1} \cdot \hat{y}_{t-1|t-1} = 0.021 + 0.380 \cdot \hat{y}_{t-1|t-1}$. The estimated coefficient $\hat{\alpha} = 0.380$ says that a decrease of output below its potential by 1% will cause a decrease in the growth of potential output by 0.38 percentage points. The growth of potential output together with the estimated output gap is illustrated at the following figure 1.

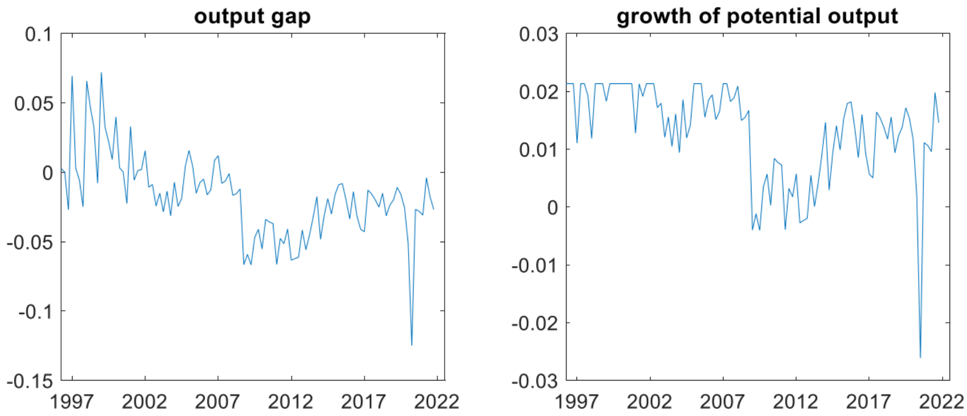


Figure 1 Estimates of output gap and growth of potential output

Output gap was above or only slightly below its potential before 2008. Therefore, growth of potential out- put was oscillating between 0.01 and 0.02 during this period. Output fell below its potential by approximately 5% in 2008 as a consequence of global recession. Output gap of about -5% was rather persistent and stayed at this level for 5 years until 2013 when it began to slowly return to its equilibrium value. The growth of potential output fell approximately to zero during 2008-2013 as a consequence of the fall in the gap and began to slowly rise after 2013 as the gap was getting better. Covid-19 crisis in 2020 caused huge decrease of the output gap. At first glance, it seems that the gap recovered quickly this time. Nonetheless, a closer look at the dynamics of the variable v_t representing the impact of great recession of 2008 and covid-19 crisis of 2020 will reveal that a rather long period of slow growth of potential output can be expected. The dynamics of v_t shows that the economy had not yet fully recovered from the global crisis of 2008 when the covid-19 crisis hit the economy in 2020 which is illustrated at the figure 2.

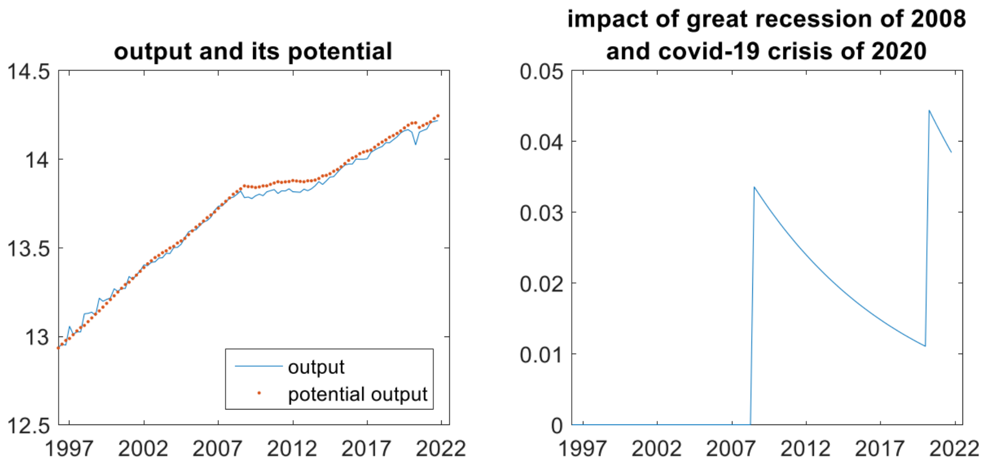


Figure 2 Log of output y_t together with its potential $\bar{y}_t$ and the impact of the crises from 2008 and 2020 (v_t)

The variable ν_t attained a value of about 0.01 in 2019. Then the covid-19 crisis hit the economy in 2020Q1 and ν_t increased by $\hat{\varepsilon} = 0.034$ and attained a value of 0.044. Taking into account that this variable is highly persistent (as $\hat{\omega} = 0.976$), it will take approximately 7-8 years to reduce the value of ν_t by half, i.e. from the value of 0.044 to 0.022. Moreover, this calculation assumes that no other crisis will hit the economy in the future. This is not a realistic assumption due to the war in Ukraine which started in February 2022.[2] The value of $\nu_t = 0.044$ decreased the value of the output gap $\hat{y}_t$ by 4.4 percentage points in 2020 and we can optimistically expect that this crisis variable ν_t will reduce the value of the variable $\hat{y}_t$ by 2.2 percentage points in 2028. An interesting finding is that the persistency of the output gap $\hat{\lambda} = 0.300$ is quite low compared to the estimated persistency of the crisis variable $\hat{\omega} = 0.976$.

6 Conclusion

The paper proposed a nonlinear time-varying parameter model based on unobserved components methodology in order to describe hysteresis effect of short-run demand-driven fluctuations on long-run growth of potential output. The formulated model was econometrically estimated for the Czech Republic using the data 1996Q1- 2021Q4 spanning both the global economic recession of 2008 and the recent covid-19 crisis. Nonlinearity of the model stems from the fact that recessions are assumed to decrease growth of potential output, but output above its potential do not lead to increased growth of the potential output due to intrinsic limits of the supply side of the economy.

All the econometrically estimated model parameters were statistically significant at 1% level of significance, except one coefficient which was statistically significant at 5% level of significance. These results clearly show the relevance of the formulated model and specifically the relevance of the negative impact that economic recessions have on the growth of the potential output. Kalman filter estimates of the output gap showed that output was below its potential by approximately 5% during 2008-2013 and the growth of the potential output was by roughly 2 percentage points lower compared to the pre-crisis period. When the gap began to return to its equilibrium zero value, the growth of the potential began to rise.

Covid-19 crisis caused a huge decrease of the output gap in 2020. At first glance, the recovery from this second crisis seems to be fast according to the output gap estimates. Nonetheless, a closer investigation of other model variables revealed that the economy had not fully recovered from the global recession of 2008 in the time when it was hit by another covid-19 crisis in 2020. This cumulative effect of crises caused output to be below its potential by approximately 4 percentage points in 2021 when compared to a hypothetical situation of no crises. Moreover, variable representing the effect of crises was estimated to be extremely persistent. Calculations based on results from econometric estimation suggest that it will take approximately 7-8 years to reduce the value of this crisis variable by half. Not to mention the fact that these calculations implicitly assumes that no other crisis will hit the economy. This is obviously not a realistic assumption due to the war in Ukraine which started in February 2022. Economic consequences of this war cannot be yet analyzed since no data are available for 2022 at the time of writing this paper and it is therefore left for future research.

The formulated model could be extended in some ways. Natural extension would be to model more observable variables like inflation which would represent additional information that could be useful when estimating the model. Nonetheless, this usual extension of the basic unobserved components model would be rather problematic for the time periods which includes the global recession of 2008 as well as current covid-19 crisis as usual economic relations like Phillips curve turned out to be rathe unstable during these crisis periods (Bobeica et al. [6]). For this reason, such an extension is left for future research.

Acknowledgements

Financial support of IP400040 is gratefully acknowledged.

References

[1] Anderson, B. D. O. & Moore, J. B. (1979). *Optimal filtering.* Englewood Cliffs, N.J: Prentice-Hall.

[2] No data were available after 2021 at the time of writing this paper.

[2] Andersson, M., Szörfi, B., Tóth, M. & Zorell, N. (2018). Potential Output in the Post-Crisis period. *Economic Bulletin Articles*, European Central Bank, 7.

[3] Ball, L. (2014). Long-term damage from the Great Recession in OECD countries. *European Journal of Economics and Economic Policies*: Intervention, 11(2),149-160.

[4] Beveridge, S. & Nelson, C. (1981). A New Approach to the Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the Business Cycle. *Journal of Monetary Economics*, 7, 151–74.

[5] Blanchard, O. (2017). Should We Reject the Natural Rate Hypothesis? *Working Papers*, No 17-14, Peterson Institute for International Economics, November 2017.

[6] Bobeica, E., Hartwig, B. & Nickel, C. (2022). The Euro Area Phillips Curve: Damaged But Not Dead, *VoxEU*. [online] [Accessed 24 April 2022] < https://voxeu.org/article/euro-area-phillips-curve-damaged-not-dead>

[7] Borio, C., Disyatat, P. & Juselius, M. (2013). Rethinking potential output: Embedding information about the financial cycle. Bank for International Settlements, *Working Paper* No. 404.

[8] Czech Statistical Office Database (2022), [online] [Accessed 10 March 2022] <https://www.czso.cz/csu/czso/hdp_cr>

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

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

[11] Harvey, A. & Trimbur, T. (2008). Trend Estimation and the Hodrick-Prescott Filter. *Journal of the Japan Statistical Society*, 38(1), 41-49.

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

[13] Kuttner, K. N. (1994). Estimating potential output as a latent variable. *Journal of Business and Economic Statistics*, 12(3), 361-68.

[14] Watson, M. W (1986). Univariate detrending methods with stochastic trends. *Journal of Monetary Economics*, 18, 49-75.

Non-Homogeneity in Data Envelopment Analysis and the Reference Set Restrictions

Martin Dlouhý[1]

Abstract. The limited comparability of production units is a practical managerial problem when conducting efficiency assessments and benchmarking. In this study, the problem of non-homogeneity is studied in the context of the data envelopment analysis, which is a well-known non-parametric method of efficiency evaluation. The original models of data envelopment analysis assume homogeneity of the production units under evaluation and homogeneity of the external environment. However, such assumptions do not often correspond to reality and may be too restrictive. The paper summarises twelve recommendations on how to deal with non-homogeneous production units and with non-homogeneous external environment. Above all, we investigate how the comparability of production units can be improved by introducing restrictions of the reference set into the model. However, no individual recommendation provides a universal solution, and it is necessary to consider the specific situation and the goal of the efficiency evaluation.

Keywords: data envelopment analysis, comparability, non-homogeneity, external environment, reference set

JEL Classification: C61, D24
AMS Classification: 91B38

1 Introduction

The methods of efficiency evaluation are used by organizations to identify sources of inefficiency and increase their competitiveness. It can be an internal evaluation, which involves a comparison of company departments, regional branches, clinics in a teaching hospital, and university departments. Alternatively, it can be an external evaluation, in which the company is compared with other competing companies in the market. In general, we will talk about production units, by which we mean specific units with similar inputs and outputs that are associated with similar production activity.

In any efficiency evaluation, it is assumed that an analyst makes a comparison of similar production units (firms, university departments, schools, bank branches, stores, farms, hospitals). The set of production units must be homogeneous; otherwise, the validity of the comparison may be questioned. However, the assumption of homogeneity is not true in many practical situations, and some heterogeneity is usually present amongst the production units under evaluation. One possibility is that a researcher ignores the problem and assumes for his or her convenience that the evaluated units are more or less homogeneous and thus fully comparable. However, such a comparison may be misleading. Another possibility is that a researcher accepts the limited comparability of production units due to many different internal and external factors.

In this paper, the problem of non-homogeneity in efficiency evaluation is studied in the theoretical framework of the data envelopment analysis (DEA). DEA is a well-known non-parametric method of efficiency evaluation with hundreds of applications from various fields, such as hospital efficiency [12], health policy analysis (health systems, equity [4]), efficiency in education (schools, programmes, university departments), the efficiency of bank branches [1], transport (e.g. efficiency of airports), agriculture (e.g., farms), retail (e.g., the efficiency of the food store chains [16]), efficiency of teams in the Major League Baseball [14], the efficiency of public libraries [8], etc. The traditional DEA models assume the homogeneity of production units and the homogeneity of the external environment in which the production units operate. However, such assumptions do not often correspond to reality and may be too restrictive. Not surprisingly, a great research effort is concentrated on the development of DEA models and procedures that can relax the assumption of the perfect homogeneity (e.g., [3], [9], [11]). The objective of this paper is to investigate how the comparability of production units can be improved in the case of existing non-homogeneity if an analyst decides to use the DEA method.

[1] Prague University of Economics and Business, Faculty of Informatics and Statistics, Department of Econometrics, 4 Winston Churchill Square, 13067 Prague 3, dlouhy@vse.cz.

2 Data Envelopment Analysis

The original DEA model was developed by Charnes, Cooper and Rhodes (the CCR model) in 1978 [2] in order to evaluate the relative technical efficiency of production units. Unlike econometric methods (for example, the stochastic frontier analysis [10]), the data envelopment analysis uses mathematical programming to construct the production frontier as the piecewise linear envelopment of the observed data. The method assumes that the production units use a clearly defined set of inputs to produce a defined set of outputs and that the weights (prices) of inputs and outputs are not known. Without any 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 defined as the best possible ratio of the total weighted output to the total weighted input or vice versa.

In the DEA model, each production unit is allowed to optimise its input and output weights to maximise the technical efficiency score. Technically efficient production units lie on the production frontier; inefficient units lie below the production frontier. We distinguish the output-oriented DEA model, which maximises quantities of outputs produced by the fixed levels of inputs, and the input-oriented model, which minimises quantities of inputs required to produce the fixed levels of output. According to the character of the returns to scale, there are two original DEA models known as the CCR model, in which constant returns to scale are assumed, and the BCC model, in which variable returns to scale are assumed. Many other DEA models have been developed since the formulation of the first DEA model in 1978 [2]. A lot of examples of the DEA models can be found, for example, in Dlouhý, Jablonský, and Zýková [5].

Both basic DEA models (the CCR and BCC models) have two equivalent mathematical formulations: the multiplier form and the envelopment form. For the purpose of this paper, we consider the envelopment formulation more appropriate. Let us suppose that we have a set of n production units that use m types of inputs to produce r types of outputs. The dual (envelopment) formulation of the input-oriented variable-returns-to-scale DEA model (1) for production unit q is below:

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

where θ_q denotes 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 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.

3 Dealing with Non-Homogeneity

Dyson et al. [6] systematically studied the problems with applications of DEA, including the problem of non-homogeneity in the DEA models. In their paper, they stated three assumptions of homogeneity:

(1) the production units perform similar activities and produce comparable outputs;

(2) the same set of resources (inputs) is available to all production units;

(3) the production units operate in similar external environments.

The first two assumptions of homogeneity are related to production units themselves, while the third property is related to a non-homogeneous external environment under which otherwise homogeneous production units operate. By the external environment, we understand any external factors that affect the efficiency of a production unit but are not usually considered as typical inputs in the DEA models and are not under the control of the management. Examples of such external factors are government regulation, socio-economic conditions, ownership (public/private), soil quality, geographic location, etc. We will use these two types of non-homogeneity in presenting the recommendations on how to deal with non-homogeneity in the DEA models.

3.1 Non-Homogeneity of Production Units

We will refer to the recommendations from this section as (PU.x). Dyson et al. [6] propose four recommendations on how to circumvent the non-homogeneity in the set of the production units (PU.1) – (PU.4).

(PU.1) Choose your set of production units carefully. In some cases, it can be possible to completely avoid the problem of non-homogeneity in the set of production units. For example, different faculties from the same university can be compared internally, but they are not homogeneous in terms of inputs and outputs. Therefore, the correct procedure would be an external comparison with equally focused faculties from other universities.

(PU.2) Divide your set of production units into smaller groups (clusters) that will be more homogeneous. For example, divide the set of hospitals into the group of teaching hospitals and the group of non-teaching hospitals. However, in many practical cases, the groups (clusters) are not explicitly given and have to be determined by an analyst, for example with help of cluster analysis.

(PU.3) Question the validity of the efficiency evaluation. In spite of non-homogeneity, it is possible to perform an efficiency analysis by DEA; however, an analyst should be aware of the limited validity of the results. Hence in the next step, the evaluation should be subjected to a critical analysis.

(PU.4) Another source of non-homogeneity can be a false assumption about the character of the economies of scale. Traditional DEA models include the constant returns-to-scale model and the variable returns-to-scale model. For example, the wrong choice of a model with the variable returns-to-scale leads to an overestimation of efficiency for the smallest and largest production units. Dyson et al. [6] recommend testing the data on the type of scale effects.

(PU.5) Cook et al. [3] studied non-homogeneous production units that differ in the output mix, i.e. with production units that can choose not to produce certain outputs. The developed model evaluates efficiency by dividing the production unit into a set of subunits, and then efficiency scores are calculated within each individual subgroup of subunits with a similar output mix. The overall efficiency of a production unit is a convex combination (weighted average) of its subunits. Cook et al. [3] believe that their approach performs better than simply breaking the set of production units into multiple homogeneous clusters.

(PU.6) Golany and Thore [7] describe the concept of dynamic clustering that establishes a different categorisation of the set of production units for each evaluated unit q. The set of all production units is divided into the subset P_q (permitted units) and the subset Nq (not-permitted units) for each evaluated production unit. Hence, each production unit constructs its own production frontier. The boundaries of the cluster for an evaluated unit can be defined in absolute terms, so the production unit belongs to the cluster only if its distances from the evaluated unit are smaller than the predefined distance in all input and output dimensions. The alternative possibility is to define the boundaries of the cluster in relative terms that are determined by the proportions taken from input-output values of the evaluated unit q. Mathematically, the following constraint on the reference set (2) is added to the basic DEA model (1):

$$\sum_{j \in N_q} \lambda_j = 0. \tag{2}$$

(PU.7) Boďa, Dlouhý, and Zimková [1] investigated the question of comparability in situations in which production units are organised in an ordered hierarchy with functions shared at different levels. In such a case, the production units from different levels of the hierarchy may have identical sets of inputs and outputs; however, they do not form an ideally homogeneous group, and their comparability in a benchmarking context is limited.

The proposed approach is demonstrated in the case study of a Slovak commercial bank with three hierarchical branch categories. Efficiency measurement for bank branches in a hierarchically ordered structure is implemented through a flexible comparability constraint. Under this comparability constraint, production units from the same category (level of hierarchy) as a production unit under evaluation contribute the most, and the contribution of production units from other categories (levels of hierarchy) is limited and decreases with their distance in the hierarchy.

Let us assume a set of production units organised into an ordered hierarchy. There are G categories $C_1, \ldots, C_G$. A series of conditions on variables λ_j is introduced for production units in a category C_Γ that may take the form:

$$\sum_{j \in C_\Gamma} \lambda_j \geq \lambda_{min}^{C_\Gamma} \sum_{j=1}^{n} \lambda_i ; \quad \sum_{j \in C_g} \lambda_j \leq \lambda_{max}^{C_g} \sum_{j=1}^{n} \lambda_j \quad \text{for} \quad g \in \{1, \ldots, G\}, \tag{3}$$

where $\lambda_{min}^{C_\Gamma}$, $\lambda_{max}^{C_g}$ are percentage levels regulating the presence of production units in the reference set by categories. These inequalities stipulate that the production units from the same category participate in the formation of benchmarks by no less than $\lambda_{min}^{C_\Gamma}$ (the minimum share condition) and the shares of other categories (including category C_Γ) are $\lambda_{max}^{C_g}$ (the maximum share condition). The comparability constraint (3) are quite general and include recommendations (PU.2), (EE.1), (EE.2), and (EE.3).

3.2 Non-Homogeneous External Environment

We will refer to the recommendations from this section as (EE.x).

(EE.1) Golany and Thore [7] studied the possibility of the restricted reference set in cases in which an analyst has to take into account institutional circumstances, externalities in production, equity considerations or other extraneous information. An analyst can treat all these factors as the non-homogeneous environment. The so-called categorisation constraints control the participation of production units in the reference set for other production units simply by separating the whole sample into clusters (groups).

In the dichotomous approach, described by recommendations (1.2) and (2.3), we will either include or not include a production unit in the subset of permitted units P_q. However, a more flexible approach is proposed in which production units from the set of non-permitted units N_q are partially included in the reference set by adding a constraint such as:

$$\sum_{j \in N_q} \lambda_j \Big/ \sum_{j=1}^{n} \lambda_j \leq \lambda_{max}^{N_q}, \tag{4}$$

where $\lambda_{max}^{N_q}$ is a constant that determines the maximal proportion of production units from N_q in the reference set. If the maximal proportion constant $\lambda_{max}^{N_q}$ is set to zero, then production units outside the category cannot be included in the reference set, and an evaluation of production units according to individual categories is carried out.

In their review paper, Mendelová and Kráľ [11] describe four possible approaches to deal with the non- homogeneous external environment. The application of these approaches depends on the character of environmental factors.

(EE.2) In the case of environmental factors that can be ordered, the production unit is compared with those production units that are from the same group or the groups with a less favourable environment. This approach prohibits situations in which a production unit is unfairly compared with production units with a more favourable environment.

(EE.3) In the case of the categorical character of the environmental factor (e.g., public or private ownership, EU or non-EU), the production unit is compared only with production units from the same category. However, in some cases, the categories are not explicitly given and have to be specified by a researcher.

For example, Holý [8] analysed the technical efficiency of 4660 Czech municipal libraries. The inputs were the total expenditures, the number of employees, and the number of books. The outputs were the total number of registered users, the number of book loans, the total number of visitors of events, and the collection additions. In the first stage of analysis, the DEA was applied to the whole dataset. In the second stage, in order to remove the effect of the operating environment, the municipal libraries were evaluated separately: (a) for 11 categories based on the decision tree analysis that determined the municipality population and the travel distance to town as criteria; (b) for 11 categories selected by an expert.

(EE.4) If the environmental factor is continuous, it can be directly included in the DEA model as the so-called non-controllable input or output.

(EE.5) If the environmental factors have a categorical or continuous character, the two-stage method can be applied. In the first stage, the traditional DEA model with the whole set of production units included is solved. In the second stage, the technical efficiency scores are regressed upon the environmental factors. The advantages of the two-stage DEA method are that an analyst can test if the environmental factor has a statistically significant influence on efficiency, the method can be used for more environmental factors, and an analyst does not need to make any prior assumptions on the direction of the influence of the environmental factor. Paradoxically, the result of the two-stage method can be that no environmental factor is statistically significant.

4 Conclusion

The limited comparability of production units is a practical managerial problem when conducting efficiency assessments and benchmarking. In this paper, we focused on the theoretical framework of data envelopment analysis, which is a well-known non-parametric method of efficiency evaluation. We summarise the list of seven recommendations on how to deal with non-homogeneous production units and five recommendations on how to deal with the non-homogeneous external environment. However, no individual recommendation presented here provides a universal solution. An analyst must consider the specific market situation and the goal of the efficiency evaluation.

Six out of the twelve recommendations presented in this paper ((PU.2), (PU.6), (PU.7), (EE.1), (EE.2), (EE.3)) assume that the original DEA models are extended by some additional restrictions on the reference set. These new restrictions regulate the freedom with which production units are compared and can participate in the set of efficient targets. The restrictions of the reference set effectively reduce the production possibility set, so the technical efficiency of production units cannot decrease. We anticipate that by restricting the production possibility set, we will obtain more realistic, effective targets for inefficient production units. In my opinion, the research on the restrictions of the reference set is likely the most promising field of research in the case of non-homogeneity in the DEA models. This opinion is evidenced by the articles published in the recent years, for example [1], [8], [11], [13], [15].

Acknowledgements

The study was supported by the project no. 19-08985S “Models for efficiency and performance evaluation in non-homogeneous economic environment”, funded by the Czech Science Foundation.

References

[1] Boďa, M., Dlouhý, M., Zimková, E. (2020). Modeling a shared hierarchical structure in data envelopment analysis: An application to bank branches. *Expert Systems with Applications*, article 162.
[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] Cook, W. D., Harrison, J., Imanirad, R., Rouse, P., Zhu, J. (2013). Data envelopment analysis with nonhomogeneous DMUs. *Operations Research*, 61(3), 666–676.
[4] Dlouhý, M. (2018). Measuring geographic inequalities: dealing with multiple health resources by data envelopment analysis. *Frontiers in Public Health*, 6, article 53.
[5] Dlouhý, M., Jablonský, J. & Zýková, P. (2018). *Analýza obalu dat* [Data envelopment analysis, in Czech]. Prague: Professional Publishing. ISBN 978-80-88260-12-7.
[6] 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(2), 245–259.
[7] Golany, B. & Thore, S. (1997). Restricted best practice selection in DEA: an overview with a case study evaluating the socio-economic performance of nations. *Annals of Operations Research*, 73, 117–140.
[8] Holý, V. (2022). The impact of operating environment on efficiency of public libraries. *Central European Journal of Operations Research*, 30(1), 395–414.
[9] Jablonský. J. (2019). Data envelopment analysis models in non-homogeneous environment. *Acta Universitatis Agriculturae et Silvilcuturae Mendelianae Brunensis*, 67(6), 1535–1540.
[10] Kumbhakar, S. C. & Lovell, C. A. K. (2000). *Stochastic frontier analysis*. Cambridge: Cambridge University Press.
[11] Mendelová, V. & Kráľ, P. (2021). An overview of methodological issues in data envelopment analysis: a primer for applied researchers. *Statistika*, 101(3), 329–350.
[12] Novosádová, I. & Dlouhý, M. (2007). Hodnotenie technickej efektívnosti nemocníc a odmeňovanie zdra votníkov [Evaluation of technical efficiency of acute hospitals and its relation to wages of personnel]. *Ekonomický časopis*, 55, 783–792.

[13] Razavi Hajiagha, S. H., Hashemi, S. S., & Amoozad Mahdiraji, H. (2016). Fuzzy C-means based data envelopment analysis for mitigating the impact of units' heterogeneity. *Kybernetes*, 45(3), 536–551.

[14] Sexton, T. R. & Lewis, H. F. (2003). Two-stage DEA: an application to major league baseball. *Journal of Productivity Analysis*, 19(2), 227–249.

[15] Zarrin, M., Schoenfelder, J. & Brunner, J.O. (2022). Homogeneity and best practice analyses in hospital performance management: An Analytical Framework. *Health Care Management Science*, online first.

[16] Zýková, P. & Jablonský, J. (2018). Analýza efektivnosti obchodních řetězců v České republice. *Trendy v podnikání*, 8(3), 47–54.

How Do Risk Appetite and Size Matter for Banking Credit Risk Management?

Xiaoshan Feng[1]

Abstract. A bank's ability to expand its business can be affected and limited to some extent by its size. In measuring the efficiency of a bank's credit risk management, it is critical to determine whether a bank's size has an impact, while the impact of a bank's risk appetite on risk management is also of concern. The objective of this paper is to examine how size and risk appetite affect the management of credit risk under the assumption of healthy efficiency scores are produced. To achieve this objective, this paper quantifies the management of credit risk, analyses the impact of bank size before and after modification of its efficiency, then includes risk appetite as one of the impact factors to investigate how size and risk appetite play out in the specific banking sector.

Keywords: Credit risk management; Size; Risk appetite; AHP; Data envelopment analysis; Logistic regression model.

JEL Classification: G21, C31, C67, C80, C61, C58
AMS Classification: 62M10, 91G40, 91G70, 90C05

1 Introduction

Size has always been a very controversial topic in the banking industry. In the context of a stricter regulation published by the EBA for banks with large assets, there are growing issues about whether bank size will become more important for credit risk management. In addition, determining a bank's risk appetite will also be critical to the study of credit risk management, as risk appetite indicates the executive's tolerance for risk management on the one hand, and reflects the bank's overall strategy on the other. To date, few studies have incorporated the risk appetite which is quantified using various subjective and objective characteristics of the banks and their executives into credit risk management research.

The purpose of this study is to investigate how size and risk appetite affect credit risk management under the assumption of healthy efficiency scores. In this paper, we selected 10 commercial banks in the Czech Republic. By employing data envelopment analysis (DEA), we further generate healthy efficiency scores after several modifications, then we will adapt the logistic regression model to access the significance of the selected determinants to the Czech banking industry, in which, an analytic hierarchy process (AHP) is utilized, integrating several subjective and objective characteristics to establish a risk appetite coefficient for each bank.

This paper is divided into five sections. The first section starts with the introduction and the last one ends with the conclusion. The second section includes the literature review. Section 3 presents a brief description of methodology and data collection. In the fourth section, the empirical results will be discussed.

2 Literature review

This paper is based on the author's previous study on the efficiency of banks' credit risk management, [6] measured the efficiency of credit risk management in five countries by using DEA then defined the significant determinants that play large roles in banks' credit risk management in different countries using a logistic regression model. However, the main weakness of this paper is that it does not eliminate the scale effect. Given the fact that banks are of different sizes, the efficiency will be biased if the analysis is performed using unmodified data with scale effects. There are rich studies that showed that size is significant for the profitability of the banking industry (Halkos and Salamouris [9]; Redmond and Bonhnsa [12]; Feng [5]). On top of that, in the selected banking sectors, larger banks will have a relatively higher capacity for credit risk management, both in terms of risk management systems and risk management methods (Hakenes and Schnabel [8]; Feng [6]). Studies that more frequently adopt a natural logarithmic in total assets when measuring bank size (Altunbas et al. [1]; Spathis et al [15]).

[1] VSB - Technical University of Ostrava, Department of Finance, Sokolská třída 33 702 00 Ostrava, Czech Republic, xiaoshan.feng@vsb.cz.

Several studies incorporated the non-performing loans ratio (NPLs) as a proxy of credit risk when measuring the efficiency of credit risk management in the banking sector. Undesirable outputs like NPLs may present in the banking sector which prefers to be minimized. Paradi and Zhu [10] mentioned three approaches when non-performing loans are incorporated in previous literature. The first is to leave the NPLs ratio as an output but use the inverse value. The second method is to treat this undesirable output as input, which is applied in other studies (Puri and Yadav [11]; Toloo and Hančlova [16]). The third one is to treat it as an undesirable output with an assumption of weak disposability, which requires that undesirable outputs can be reduced, but at a cost of fewer desirable outputs produced.

Wilson et al [19] emphasized that the risk appetite of financial institution executive has a significant impact on the risk and stability of the financial institution. The risk appetite of a bank varies on the regulatory mechanism and the institutional environment. [18] argued that executives' attitudes toward risk vary from country to country. They argued that the culture of different countries affects people's attitudes toward risk, but that people are risk-averse in each country. However, previous studies are limited to investigating the impact of a bank's credit risk management from the perspective of executive risk appetite. Therefore, in this study, we will first compare the efficiency result under the two solutions for eliminating the size effect in our sample, then assess the impact of the risk appetite of a bank to credit risk management, to diversify the choices of determinants, we also incorporate macroeconomic indicators and other bank relevant factors.

3 Methodology and Data collection

Following the previous literature, we apply DEA to measure the efficiency of credit risk management in selected 10 banks from the Czech Republic, during the period from 2012-2020. Moreover, we employ the logistic regression model to investigate the possible internal and external determinants of credit risk management efficiency.

3.1 Two Classic Models of Data Envelopment Analysis

DEA is a linear programming-based method, introduced by Charnes, Cooper, and Rhodes in 1978. DEA is used to evaluate the relative efficiency of a set of decision-making units (DMUs) with multiple inputs and multiple outputs. Then, Banker, Chames, and Cooper proposed a model in 1984, named the BCC, which is an extended version of the CCR model. The main difference between these two models is different returns to scale. The CCR model assumes all DMUs are operating at an optimal scale, that is, constant returns to scale (CRS); While the BCC model assumes variable returns to scale (VRS).

In DEA models, we measure the efficiency of each DMU. One of the most frequently used methods to measure efficiency is by the ratio. Suppose we have n DMUs in the population, each DMU produces s outputs while consuming m inputs. Consider DMU_j, j represents n DMUs, x_{ri} and y_{ri} are the matrixes of inputs and outputs respectively. The efficiency rate of such a unit can be expressed as:

$$\frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \tag{1}$$

The efficiency rate is the ratio of the weighted sum of outputs to weighted sum of inputs. The DEA model assumed inputs and outputs should be non-negative. Let DMU_j to be evaluated on any trial be designated as DMU_o, where $o = (1,2,\dots,n)$.
A ratio of two linear functions can construct the linear-fractional programming model as follows:

$$\max_{v,u} \theta = \frac{\sum_{r=1}^{s} u_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}} \tag{2}$$

$$\text{subject to } \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1, (j = 1,2,\dots,n) \tag{3}$$

$$u_1, u_2, \dots, u_r \geq 0, (r = 1,2,\dots,s) \tag{4}$$

$$v_1, v_2, \dots, v_i \geq 0, (i = 1,2,\dots,m) \tag{5}$$

Where θ is the technical efficiency of DMU_o to be estimated, v_i $(i = 1,2,\dots,m)$ is the optimized weight of input and the output u_r $(r = 1,2,\dots,s)$. $y_{rj} the$ is observed amount of output of the r-th type for the j-th DMU, x_{ij} is the observed amount of input of the i-th type for the j-th DMU.

Data Selection for Measuring the Efficiency by DEA

To make sure the results of applying DEA are accurate, the number of inputs and outputs, and DMUs must support the rule of thumb, which was proposed firstly by Golany and Roll [6], then developed by Bowlin [3], that is, it should have three times the number of DMUs as there are input and output variables if this condition will not be met, the results are not reliable (Toloo and Tichý[16]).

In this study, we will collect data from 10 representative commercial banks from the Czech Republic, which includes the large size bank, medium and small sizes. All data are from the annual report of each bank on a consolidated basis.

This study aims to investigate the efficiency of credit risk management in the banking industry. Therefore, we will apply the intermediation approach to measure the efficiency of credit risk management based on the DEA model. To assess credit risk modeling in the banking industry, Berg et al. [2] suggested using NPL as a proxy of credit risk in a nonparametric study of the bank production, Altunbas et al. [1] incorporated loan loss provisions (LLP) to analyze the efficiency of Japanese banks.

Generally, this paper developed two inputs and one output, with 10 DMUs which satisfies the rule of thumb. We selected the important indicators which can represent soundness: capital adequacy ratio (CAR); Asset quality: NPL; Capacity to withstand future credit risk: LLP; Total loan: TL.

	DEA_O	**DEA_1**	**DEA_2**
$Input\ x_1$	LLP	LLP	LLP
$Input\ x_2$	NPL	NPL	CAR
$Output\ y$	TL	ln (TL)	1-NPL

Table 1 Data selection for DEA

3.2 Logistic Regression Model

Furthermore, after we obtained efficiency scores based on the CCR model and the BCC model, we can estimate the determinants of banking credit risk management efficiency using the regression model, since the efficiency can be measured as binary outcomes, we can model the conditional probabilities of the response outcome, rather than give a binary result. Therefore, we apply the logistic regression model in this paper. The logistic model could be interpreted based on an underlying linear model, shown below:

$$Y_{i,t} = \beta_0 + X'_{i,t}\beta + \epsilon_{i,t}\ , i = 1, \dots, N, t = 1, \dots, T. \tag{7}$$

Where the subscripts i and t denote the cross-sectional (N) and time dimension (the T) of the panel data, respectively. There is k $(k = 1, \dots, K)$ regressor in $X_{i,t}$, not including a constant term. $X_{i,t}$ is explanatory variable value for i-th section at t-th dimension; β_0 is the incept; β is the slope coefficient of a $(k \times 1)$ vector. The variable $\epsilon_{i,t}$, can be called as the error term in the relationship, represents factors other than explanatory variables that affect dependent variables. Since we have a binary output variable $Y_{i,t}$, and we want to model the conditional probability $p(Y_{i,t} = 1|X'_{i,t} = x_{i,t})$ as a function of $x_{i,t}$:

$$\pi(x_{i,t}) = p(Y_{i,t} = 1|X'_{i,t} = x_{i,t}) \tag{8}$$

The logistic regression model can be constructed as follows:

$$log\frac{\pi(x_{i,t})}{1-\pi(x_{i,t})} = \beta_0 + X'_{i,t}\beta\ , i = 1, \dots, N, t = 1, \dots, T. \tag{9}$$

Data Selection for Logistic Regression Model

To investigate the determinants of the efficiency of banks' credit risk management, the dependent variable in the regression model is the efficiency score obtained from the previously mentioned DEA model, measured by 1 and 0, which represent DMU is efficient and inefficient, respectively. The independent variables are, respectively, GDP growth rate, which is the year-on-year annual GDP growth rate. Risk-weighted assets are calculated by the Standardized Approach (SA), which is calculated as the ratio of RWAs under SA to total RWAs. Size of bank is taking natural logarithmic of total assets of the bank. The risk appetite of the selected bank is measured through 7 criteria. In this paper, we adopted the AHP proposed by [14], that we designed the questionnaire for three experts who are general management in the Czech banking industry, through pair-wise comparison, the experts' experiences are utilized to estimate the relative magnitudes each pair of items, the larger the risk appetite coefficient of the commercial bank, indicating that the bank is less risk averse. To improve the accuracy of the logistic regression model, certain tests will be conducted to verify that the model meets the assumptions (i.e., binary dependent variable, little or no multicollinearity, etc.).

	CS	KB	CSOB	RB	SBER[2]	PPF	FIO	MMB	UNI	EQUA
Executive Nationality[3]	1	1	1	0	0	1	1	1	1	1
Education level[4]	7	7	6	7	7	7	7	7	7	7
Executive risk appetite	0.5	1	0.5	1	1	2	1	1	1	1
Length of tenure	6	3	6	5	2	7	19	5	1	1
C/I ratio (%)	47.60	50.50	54.60	60.40	71.10	22.70	31.60	45.70	55.28	64.70
CAR (%)	25.60	22.30	24.20	22.80	17.24	24.47	21.06	18.20	23.40	17.62
Loan growth rate (%)[5]	5.76	0.05	8.81	7.49	4.06	6.71	18.51	4.38	14.44	43.54

Table 2 Criteria of Risk Appetite Coefficient

4 Empirical Result

In this section, we employed DEA models to access the efficiency of the Czech banking credit risk management. In reference to the previous section, we applied two treatments that aim to eliminate the exited size effect, then compare the modified results to the original results, and point out the significant differences and improvements based on our modification to the data. In the second part of this section, we will analyze the significance of determinants based on the efficiency scores.

The original results showed that from 2012 to 2020, larger size banks have better credit risk management efficiency[6] than the smaller size, meanwhile, large asymmetry exhibited, see Table 3. Under the assumptions of CCR and BCC, the second largest bank ČSOB, which focuses on retail, SME, and mid-cap clients, is the only efficient DMU among 10 selected banks. Smaller size bank such as FIO and PPF showed insufficient capacity for credit risk management. As mentioned in the previous section, the result has strong size effect due to the selection of output TL, given the fact that in the Czech Republic, about 80% of the market share are dominated by 4 largest banks, the result turned to be biased by the size problem. Therefore, this study set out with an aim of eliminating the size effect and generate a healthy efficiency score.

After applying natural logarithmic on TL, the result from DEA_1 presents less asymmetry within the Czech banking industry than the DEA_O. The mid-size bank EQUA, mainly provides loans for retail and SME clients, is the most efficient DMU. While another mid-size MMB has the lowest score, whose main business strategy is to develop retail loans mortgages, commercial loans, and investment loans. Moreover, whose NPL is the highest one in the Czech banking industry due to the large amount of stage 3 loans from 2012-2016. In general, size of bank doesn't explain the efficiency scores under the DEA_1, this rather contradictory result compared to DEA_O could be attributed to the fact that natural logarithmic eliminated the size effect of the original data, which lead to a fair quality of the efficiency scores.

Turning now to the second treatment which we referred as DEA_2, is to take the inverse value of NPL and leave it as an output, this transformation will make the undesirable variable NPL becoming desirable, furthermore, we incorporate a new input CAR, represents the soundness of the bank. The result indicates the Czech-owned bank shows better efficiency than others and suggests smaller size banks have relatively higher efficiency. Comparing the DEA_2 and DEA_O, it can be seen that taking inverse value of NPL effectively solves the undesirable output

[2] From 30.4.2022, CNB revoked the banking license of Sberbank, due to a significant outflow of client deposits in response Russia's invasion of Ukraine
[3] Nationality is measured as 1- Czech Republic, 2- not Czech Republic.
[4] The education level is measured based on the classification from UNESCO.
[5] The loan growth rate is calculated from the average growth rate of the last 9 years.
[6] The efficiency scores shown in the table are the geometric mean of the banks' efficiency scores over the last nine years.

in a simple way, the concept of inverse value complies with the logic that the portion of performing loans which are not in default and have high certainty to receive future payments.

		CS	KB	CSOB	RB	SBER	PPF	FIO	MMB	UNI	EQUA
	OTE	0.98	0.88	1.00	0.34	0.14	0.11	0.07	0.19	0.54	0.12
DEA_O	PTE	0.99	0.92	1.00	0.78	0.88	0.84	0.90	0.62	0.80	0.92
	OTE	0.89	0.85	0.94	0.79	0.87	0.83	0.86	0.62	0.76	1.00
DEA_1	PTE	0.99	0.93	1.00	0.85	0.92	0.84	0.90	0.66	0.83	1.00
	OTE	0.78	0.85	0.85	0.84	0.91	0.92	0.99	0.72	0.83	0.92
DEA_2	PTE	0.93	0.94	0.96	0.95	0.98	0.95	0.98	0.93	0.93	1.00

Table 3 Results from DEA models

In the logistic regression model, under the assumptions of binary outcome and independent observations, no multicollinearity, etc., we interpreted the determinants with the sign and significant level. The probability of a bank being efficient increases with larger size and more conservative risk appetite based on the result from DEA_O. Moreover, the results from DEA_1 showed that only the proportion of using standardized approach to calculate RWAs was statistically significant in BCC model. That is, the probability of a bank being efficient increases with more banks granted the permission to use IRB approach to calculate RWAs for credit risk to meet capital requirement. As for the second treatment, where we took the inverse value of NPL and leave it as an output, in reference to DEA_2, the result suggests lower the bank's aversion to risk then higher the probability that bank can perform efficient credit risk management. Meanwhile, under a flourishing economy, the likelihood of banking credit risk management efficiency turns to be higher.

Variables	DEA_O	DEA_1	DEA_2
	13.21	-5.13	-3.43
C	[1.14]	[4.29]	[5.31]
	7.42***	0.21	-0.41
SIZE	[1.60]	[0.28]	[0.35]
	0.21	-0.12**	-0.06
SA	[0.16]	[0.05]	[0.06]
	-15.87**	2.07	5.97**
RA	[4.53]	[1.88]	[2.08]
	-0.23	0.04	0.46*
LAGGDPG	[0.46]	[0.12]	[0.17]

Table 4 Regression results

Comparing the three results, it can be seen that risk appetite is matter for the credit risk management in the Czech banking industry, based on the research, larger size bank turned to have lower risk appetite coefficient. Moreover, less use of standardized approach corresponds to [4] that using IRB provides better performance than using SA. Put differently, when more bank applied from CNB for the use of IRB approach to calculate RWAs, the industry will have better efficiency in credit risk management.

Taken together, these results suggest that when size effect exhibits in the input and output selections during employing DEA to measure credit risk management efficiency, the result will have strong size bias, then it will further influence the breakdown of determinants. Luckily, several treatments which can eliminate the size effect then lead to a fair efficiency score, the size bias is no longer the significant determinant on the credit risk management in the Czech banking industry. Conjointly, given the fact that only the two largest banks in selected banks of the Czech Republic hold conservative risk appetites, higher risk aversion matters for credit risk management efficiency when the size effect exists, while a moderate risk appetite is statistically significant when size effect is no longer matters. In general, it has strong evidence to prove that risk appetite is playing important role in the Czech banking industry.

5 Conclusion

The motivation of this paper is to further improve the efficiency result which was biased in the DEA models due to the size effect of the selected output. Furthermore, the main goal is to examine if size still matters for credit risk management in the selected banking industry and determine the impact of risk appetite and other determinants. Through certain treatments, the size effect gets eliminated. The result proves natural logarithmic can smooth size variance and reduce the asymmetry within the selected industry; And if takes the reverse value of undesirable output, needs to carefully consider the logic and reasonableness. The above-mentioned modifications for the size problem do not apply to every country. The impact of risk appetite on credit risk management in the Czech banking industry is significant in our results, quantify risk appetite from a broader perspective can help studying the impact of the risk appetite of commercial bank executives on certain risk, which will be an important policy reference for the financial institution.

Acknowledgements

The author was supported through the Czech Science Foundation (GACR) under project 20-25660Y and moreover by SP2022/4, an SGS research project of VSB-TU Ostrava. The support is greatly acknowledged.

References

[1] Altunbas, Y., Liu, M., Molyneux, P. & Seth, R. (2000). Efficiency and risk in Japanese banking. *Journal of Banking & Finance, Elsevier*, vol. 24(10) (pp. 1605-1628).
[2] Berg S.A., Forsund, F.R & Jansen, E.S. (1992). Malmquist Indexes of Productivity Growth During the Deregulation of Norwegian Banking, 1980-89. *The Scandinavian Journal of Economics* (pp. S211-S228).
[3] Bowlin, W.F. (1998). Measuring Performance: An Introduction to Data Envelopment Analysis (DEA). *Journal of Cost Analysis* 7 (pp. 3-27).
[4] Cucinelli, D., battista, M.L.D., Marchese, M. & Mieri, L. (2018). Credit risk in European banks: The bright side of the internal ratings-based approach. *Journal of Banking and Finance*, 93(C), 213–229.
[5] Feng, X. (2019). An Empirical Analysis of Macroeconomic and Bank Performance Factors Affecting Credit Risk in Banking for The Central European Countries. *37th International Conference on Mathematical Methods in Economics*, České Budějovice: University of South Bohemia in České Budějovice, 2019, 392-397.
[6] Feng, X. (2021). Efficiency of Credit Risk Management and Their Determinants in Central European Banking Industries. *39th International Conference on Mathematical Methods in Economics*, Faculty of Economics and Management, Czech University of Life Sciences Prague, 83-88.
[7] Golany, B. & Roll, Y. (1989). An Application Procedure for DEA. *Omega*, 17, (pp. 237-250).
[8] Hakenes, H. & Schnabel, I. (2011). Bank size and risk-taking under Basel II. *Journal of Banking & Finance*, 35 (6), 1436–1449.
[9] Halkos, G. & Salamouris, D. (2004). Efficiency measurement of the Greek commercial banks with the use of financial ratios: a data envelopment analysis approach. *Management Accounting Research*, 15(2), 201-224.
[10] Paradi, J. C. & Zhu, H. (2013). A survey on bank branch efficiency and performance research with data envelopment analysis. *Omega. Elsevier*, 41(1), 61–79.
[11] Puri, J. & Yadav, S. P. (2014). Expert Systems with Applications A fuzzy DEA model with undesirable fuzzy outputs and its application to the banking sector in India. *Expert Systems With Applications. Elsevier* Ltd, 41(14), 6419–6432.
[12] Redmond, G. & Bonhnsa, C. (2007). Bank size and profitability: one nation, one bank. *International of business research*. 8(1).
[13] Řepková, I. (2012). Measuring the efficiency in the Czech banking industry : Data Envelopment Analysis and Malmquist index, *Proceedings of 30th International Conference Mathematical Methods in Economics*, 2012, 781-786.
[14] Saaty, T.L. (1994). How to make a decision: the analytic hierarchy process. *Interfaces*, Vol. 24, No. 6, 19–43.
[15] Spathis, Ch., Kosmidou, K. & Doumpos, M. (2002). Assessing profitability factors in the Greek banking system: a multicriteria methodology. *International Transactions in Operational Research*, Vol. 9 No. 5, 517- 530.
[16] Toloo, M. & Hančlova J. (2019). Multi-valued measures in DEA in the presence of undesirable outputs. *OMEGA*, 94, 1-10.
[17] Toloo, M. & Tichý, T. (2015). Two alternative approaches for selecting performance measures in data envelopment analysis. *Measurement*, Vol. 65, 29-40.

[18] Weber, E.U. & Hsee, C. (1998). Cross-Cultural Differences in Risk Perception but Cross-Cultural Similarities in Attitudes Toward Perceived Risk. *Management Science* 44 (9), 1205－1217.
[19] Wilson, J., Casu, B., Girardone, C. & Molyneux, P. (2010). Emerging themes in banking: Recent literature and directions for future research. *The British Accounting Review*, 42(3), 153-169.

Closed-Loop Supply Chain Coordination by Contracts

Petr Fiala[1], Renata Majovská[2]

Abstract. A supply chain is a decentralized system where material, financial and information flows connect economic agents. There is much inefficiency in supply chain behaviour. Supply chain contracts are used to provide some incentives to adjust the relationship of supply chain partners to coordinate the supply chain, i.e., the total profit of the decentralized supply chain is equal to that achieved under a centralized system. When the supply chain integrates and coordinates the backward flows of goods along with the forward flows, it takes the form of a closed-loop supply chain. The integration of forward and reverse activities into a single system pursues environmental goals, creates new economic opportunities and provides competitive ad- vantages. The aim of this paper is to analyse and compare contracts for supply chain and closed-loop supply chain coordination.

Keywords: supply chain, closed-loop supply chain, coordination, contracts

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Supply chain (SC) is a decentralized system where material, financial, information and decision flows connect members. Recent years have seen a growing interest among researchers and practitioners in the field of supply chain management. Supply chain management is about matching supply and demand with inventory management. When one or more members of the supply chain try to optimize their own profits, system performance may be hurt. Among the solutions, supply chain contracts, which have drawn much attention from the researchers recently (for review [2], [15]), are used to provide some incentives to adjust the relationship of supply chain partners to coordinate the supply chain, i.e., the total profit of the decentralized supply chain is equal to that achieved under a centralized system. The format of supply chain contracts varies in and across industries. The particular contract adopted by the firms is the outcome of some negotiation process that could be also modeled (see [7]).

Traditional supply chain (SC) focuses on the management of forward flows going from upstream members (e.g., suppliers) to downstream units (e.g., consumers). The Closed-Loop Supply Chain (CLSC) also manages the backward flows from the downstream to the upstream suppliers. Product returns from consumers to producers or to another unit characterize the main difference between a classical supply chain SC, which focuses on forward flows of goods, and a CLSC.

Forward activities include new product development, product design and engineering, procurement and production, marketing, sales, distribution, and after-sale service (see [14]). Reverse activities refer to product acquisition, reverse logistics, points of use and of disposal, testing, sorting, refurbishing, recovery, recycling, remarketing, and reselling (see [9]).

There are three fundamental objectives for integrating forward and reverse activities into a single system:

- meet environmental objectives,
- create new economic opportunities, and
- provide competitive advantages.

There are economic reasons for focal companies to establish CLSCs:

- the backward activities imply that goods reaching their end-of-use or their end-of-life stage are returned to the focal company and used for remanufacturing or recycling purposes.
- the focal company offers a collection service to consumers who might have difficulties in getting rid of end-of-use/life goods.
- consumers who return goods are most likely interested in repurchasing as well a new good to continue to satisfy their needs.

[1] University/Institute, Department, Address, name1@university.edu. 1 Prague University of Economics and Business, Department of Econometrics, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic, pfiala@vse.cz
[2] University of Finance and Administration, Prague, Department of Computer Science and Mathematics, Estonská 500, 101 00 Praha 10, Czech Republic, renata.majovska@mail.vsfs.cz

There are also non-economic (environmental, social) motivations for a focal company to manage the return flows:

- the focal company makes sure that both end-of-use and their end-of-life goods are not disposed-off in the landfill;
- governments can also establish some specific collection targets by charging some fees when such targets are not achieved.
- the focal company needs to offer more jobs to its community, since the backward flows require the implementation of some atypical processes
- managing the collection process allows the focal company to avoid that competitors collect their products,
- use the returns to gain positions especially in the second-hand market.

Closed-loop supply chain management fulfil many activities. One of the most important activities is the coordination mechanisms to improve payoffs of the CLSC members. The aim of this paper is to analyze and compare contracts for the problem of supply chain (SC) and closed-loop supply chain (CLSC) coordination.

Contracts are evaluated by desirable features:

- coordination of the supply chain,
- flexibility to allow any division of the supply chain's profit,
- easy to use.

Contracts can be considered as specific models of game theory. We will analyze the game of two players, the manufacturer (M) and the retailer (R) with their profit functions.

2 Contracts in SC

The problem of double marginalization causes the inefficiency of the supply chain as a decentralized system. In order for the chain to achieve a centralized solution, coordination contracts can be used. Examples are a wholesale price contract and a buy back contract.

2.1 Double marginalization problem

Double marginalization (see [13]) is a well-known cause of supply chain inefficiency and the problem occurs whenever the supply chain's profits are divided among two or more members and at least one of the members influences price-dependent demand. Each firm only considers its own profit margin and does not consider the supply chain's margin.

We consider a supply chain with a manufacturer and a retailer that sells a product. The manufacturer produces each unit for a cost c and sells each unit to the retailer for a wholesale price w. The retailer chooses an order quantity q and sells q units at price $p(q)$, assuming that $p(q)$ is decreasing, concave and twice differentiable function.

Centralized solution assumes a single agent has complete information and controls the entire supply chain (this is referred as the first-best solution) to maximize supply chain profit

$$z(q) = (p(q) - c)\, q. \tag{1}$$

Since profit is strictly concave in quantity, the optimal quantity q^0 satisfies

$$\frac{dz(q)}{dq} = 0. \tag{2}$$

Decentralized solution assumes the firms have incomplete information and make choices with the objective of maximizing their own profits. The retailer's profit and the manufacturer's profit are

$$z_R(q) = (p(q) - w)\, q, \tag{3}$$

$$z_M(q) = (w - c)\, q. \tag{4}$$

Optimal solution of the problem we denote q^*.

If the centralized and decentralized solutions differ, investigate how to modify the firm's payoffs so that new decentralized solution corresponds to the centralized solution. It can be shown that the retailer orders less than the supply chain optimal quantity ($q^0 > q^*$) whenever the manufacturer earns a positive profit and it holds

$$z(q^0) > z_R(q^*) + z_M(q^*). \tag{5}$$

Marginal cost pricing ($w = c$) is one solution to double marginalization problem, but the manufacturer earns a zero profit. A better solution is a revenue-sharing contract, where the manufacturer earns $\lambda\, z(q)$ and the retailer earns $(1-\lambda)\, z(q)$, for $0 \leq \lambda \leq 1$. The wholesale price w is now irrelevant to each firm's profits and the supply chain earns the optimal profit.

2.2 Centralized solution for stochastic demand

When the demand is stochastic than the newsvendor model can be applied. The newsvendor model is not complex, but it is sufficiently rich to study important questions in supply chain coordination. In a standard newsvendor problem, the price is assumed to be fixed.

We consider a supply chain in one-period setting in which a manufacturer sells to a retailer facing stochastic demand from consumers. We assume that stochastic demand x has a continuous distribution $F(x)$ that is invertible. We define the following quantities:

q retailer's total order quantity;

c manufacturer's production cost;

p retail price.

The setting can be characterized as a newsvendor problem. Centralized solution is a benchmark for the decentralized supply chain. The centralized chain is considered as an integrated firm that controls manufacturing and sales to consumers. The profit of an integrated firm for stocking level q is

$$z(q) = (p - c)q - p \int_0^q F(x)\, dx. \tag{6}$$

The problem is concave in q and the optimal solution is given by

$$q^0 = F^{-1}\left(\frac{p-c}{p}\right). \tag{7}$$

The maximum system profit $z(q^0)$ is completely determined by the stocking level q^0. Decentralized solution can be improved by contracting. The contract coordinates the chain if it induces the choice of the centralized system's optimal stocking level q^0.

2.3 Wholesale price contracts for SC

With a wholesale price contract, the manufacturer charges the retailer w per unit purchased. The retailer faces a problem analogous to that of the integrated chain with. The principal difference is that the retailer must buy stock at the wholesale price w instead of producing it at cost c.

The retailer's profit is

$$z_R(q) = (p - w)q - p \int_0^q F(x)\, dx. \tag{8}$$

The retailer's problem is concave in q and the optimal solution is given by

$$q(w) = F^{-1}\left(\frac{p-w}{p}\right). \tag{9}$$

The manufacturer acts as a Stackelberg leader and anticipates how the retailer will order for any wholesale price. The manufacturer anticipates a demand curve $q(w)$ and the profit

$$z_M(w) = (w - c)q(w) = (w - c)F^{-1}\left(\frac{p-w}{p}\right). \tag{10}$$

The manufacturer knows exactly what retailer will order at every wholesale price and bears no responsibility for the product. All uncertainty regarding supply profits is foisted onto the retailer. The wholesale price contract co-ordinates the chain only if the manufacturer earns a non-positive profit. So, the manufacturer clearly prefers a higher wholesale price. As a result, the wholesale price contract is generally not considered a coordinating contract. The richer contracts differ from wholesale price contracts by allowing the manufacturer to assume some of the risk arising from stochastic demand. As an example, we introduce buy back contracts.

2.4 Buy back contracts for SC

With a buy back contract (see [12]) the manufacturer charges the retailer w per unit purchased, but pays the retailer b per unit remaining at the end of the season. A retailer should not profit from left over inventory, so assume $b \leq w$. There is assumed that a returns policy on the decentralized chain introduces no additional cost beyond that incurred by the centralized system.

The retailer's profit is

$$z_R(q) = (p-w)q - (p-b)\int_0^q F(x)\,dx. \tag{11}$$

The retailer still faces a newsvendor problem. The optimal solution is

$$q(w,b) = F^{-1}\left(\frac{p-w}{p-b}\right). \tag{12}$$

No returns or full returns are suboptimal. An intermediary policy results in chain coordination. The manufacturer offers a contract $(w(\varepsilon), b(\varepsilon))$ for $\varepsilon \in (0, p-c\rangle$ where

$$w(\varepsilon) = p - \varepsilon, b(\varepsilon) = p - \frac{\varepsilon p}{p-c}. \tag{13}$$

The retailer orders the integrated chain quantity $q(w(\varepsilon), b(\varepsilon)) = q^0$ and system profit is equal to the integrated chain profit $z(q^0)$.

Retailer's profit is increasing in ε: $\quad z_R(w(\varepsilon), b(\varepsilon)) = \frac{\varepsilon}{p-c} z(q^0). \qquad (14)$

Manufacturer's profit is decreasing in ε: $z_M(w(\varepsilon), b(\varepsilon)) = \left(1 - \frac{\varepsilon}{p-c}\right) z(q^0). \qquad (15)$

3 Contracts in CLSC

Several types of contracts, e.g., buyback, revenue-sharing, cooperative advertising, and two-part tariff contracts, have been proposed in the CLSC literature (see [6]). All these contracts were initially proposed in the context of supply chains. In the following sections, we will focus on some specific contracts. We will introduce the following notation:

$D(p)$ is the demand function depending on the retail price p and possibly on other decision variables;
A is the effort to attract returns;
$r \in (0, 1)$ is the return rate;
$r(A)$ is the return rate generated by the effort A;
Δ is the returns' residual value; and
I is the per-return incentive.

3.1 Buy-back contracts in CLSC

A buy-back contract allows the retailer to return unsold units to the manufacturer at the end of the selling season. A general CLSC game formulation is given below:

$$z_M = \max_{w,q,I}\{q\big(w + (\Delta - I)r(A)\big) - C_M(q) - w[q - D(\mathrm{p})]^+\}, \tag{16}$$

$$z_R = \max_{p,A}\{D(p)(p + Ir(A)) - wq - C_R(A)\}, \tag{17}$$

where q is the manufacturer's production quantity, $(q-D(p))^+$ is the unsold quantity that the retailer returns to the manufacturer, $C_M(q)$ the manufacturer's production cost, and $C_R(A)$ is the retailer's cost of the effort to collect end-of-life/use products.

A number of specific versions are proposed, but not all coordinate the CLSC. Differentiated buy-back contracts that distinguish between unsold and returned products do not coordinate the chain. Here are some coordinating procedures. A buy-back contract depending on the return deadline, τ, guarantees coordination, as it encourages the retailer to procure the optimal quantity and to offer a τ -dependent refund (see [18]).

Chen [3] also uses a buy-back contract to coordinate a CLSC managing stochastic returns. The manufacturer decides the wholesale price under an uncertain consumer willingness to return. The uncertainty can be resolved by the retailer sharing some consumers return information with the manufacturer.

Yang et al. [19] propose a refund incentive mechanism that a retailer can offer to two competing manufacturers based on a money-back guarantee mechanism. The retailer's convenience of providing an incentive fully depends on the returns' residual value and, in turns, is independent of the manufacturers' efficiency and performance. When the incentive is given to one manufacturer only, the competitor's profits and returns decrease. Finally, providing an incentive to both manufacturers is the best policy a retailer can implement, even if one manufacturer does not make any returns.

3.2 Revenue-sharing contracts

Revenue-sharing contracts were initially introduced in the context of supply chain coordination. They aim at mitigating the double marginalization effect by removing the marginalization at the wholesale price level. In a CLSC, a generic formulation of the two players' optimization problems is as follows:

$$z_M=\max_{I}\{D(p)(p\lambda + (\Delta - I)r(A))\}, \tag{17}$$

$$z_R=\max_{p,A}\{D(p)\big(p(1-\lambda) + Ir(A)\big) - C_R(A)\} \tag{18}$$

where $\lambda \in (0,1)$ is the revenue sharing parameter.

Govindan and Popiuc [8] propose a revenue-sharing contract (RSC) mechanism in which a manufacturer shares the revenues either with a retailer in a two-echelon CLSC, or with both a retailer and a distributor in a threeechelon CLSC. In both instances, coordination is achieved through an RSC with exogenous sharing parameters. Xie et al. [18] use a combination of revenue sharing and support programs to coordinate the CLSC. Under a revenuesharing contract, the retailer shares the revenues with the manufacturer, while in the support program, the manufacturer pays part of the retailer's service efforts. Both firms prefer the simultaneous adoption of both mechanisms, which makes it possible to reach coordination when both the sharing and support parameters are exogenous.

Xie et al. [16] model a coordination mechanism based on two sharing mechanisms: one based on forward flows, and one based on reverse flows. Both sharing parameters are determined by the manufacturer and, simultaneously, this configuration is complemented by an exogenous cooperative advertising program. Xie et al. [16] show that the presence of a dual sharing contract mechanism does not necessarily lead to coordination, because increasing one of the sharing parameters reduces the manufacturer's profits while it increases the retailer's profits. At the same time, the double marginalization effect still persists.

Han et al. [10] complement a per-return incentive with an RSC, showing that the RSC always allows the firms to coordinate a CLSC when the retailer collects. Later, they extend the model to a returns-disruption scenario, showing that the manufacturer's preferences change. Coordination becomes more challenging due to the penalty associated to return disruption.

3.3 Cooperative advertising contracts

Some models deal with cooperative advertising (or green-effort) contracts in which a manufacturer supports the efforts of a collector (for overview see [1]).

The players' profit functions are defined as follows:

$$z_M=\max_{w,I,B}\{D(p)[w+(\Delta-I)\, r] - BC_R(A)\}, \tag{19}$$

$$z_R=\max_{p,A}\{D(p)(p+Ir(A)) - wq-(1-B)\, C_R(a)\}, \tag{20}$$

where $B \in (0,1)$ is the support rate, that is, the percentage of the retailer's green efforts paid by the manufacturer.

De Giovanni and Zaccour [5] propose dynamic game that involves one manufacturer and one retailer. The aim is to achieve coordination by implementing a cost-revenue-sharing contract. The retailer supports the manufacturer's green efforts by paying at time t a share $B(t) \in (0,1)$ of the cost. The rationale for this support is that the manufacturer shares part of the remanufacturing advantages with the retailer, according to the exogenous sharing parameter, $\lambda \in (0,1)$. With a contract of the type $(B(t), \lambda)$, the authors characterize the conditions under which coordination is achieved.

De Giovanni [4] models a CLSC setting in which a manufacturer incentivizes a retailer to advertise more to increase the stock of green goodwill by sharing a part of its revenues. The transfer occurs through the implementation of an RSC, where the sharing rule can be either exogenous or endogenous. When sharing is endogenous, coordination is never achieved, because the manufacturer sets a low share and the retailer would then prefer a non-coordinated setting. Adopting an exogenous sharing parameter increases the chances of coordinating the chain. Coordination takes place when the sharing parameter is fixed within a specific range of values.

Hong et al. [11] propose a cooperative advertising program as well as a two-part tariff mechanism as coordination mechanisms. A two-part tariff mechanism coordinates an CLSC because it brings the retail price and the return rate to the same levels as a centralized chain.

Conclusions

There is a vast literature on supply chain (SC) and closed-loop supply chain (CLSC) contracts recently. The supply chain is fully coordinated, i.e., the total profit of the decentralized supply chain is equal to that achieved under a centralized system. The paper analyzes contracts that enable chain (SC) coordination. The aim of the paper was to compare contracts for SC chains and for CLSC chains. Many contracts for CLSC chains were initially proposed in the context of SC chains. The contracts in SC were an inspiration for CLSC, supplemented by factors motivating collection of end-of-use/end-of-life products. Some general models for CLSC contracts and specific modifications in the literature were investigated. The research of a general framework that synthesizes existing results for a variety of SC contract forms would be very desirable. The analysis of the simple cases of contracts gives recommendations for more complex real problem. Real problems in supply chains are solved by joint problem solving in supply chain partnership.

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] Aust, G. & Buscher, U. (2014). Cooperative advertising models in supply chain management: a review. *European Journal of Operational Research*, 234, 1–14.
[2] Cachon, G. (2003). Supply Chain Coordination with Contracts. In S. Graves & T. de Kok, (Eds.), *Handbooks in Operations Research and Management Science*: *Supply Chain Management*. North-Holland.
[3] Chen, J. (2011). The impact of sharing customer returns information in a supply chain with and without a buyback policy. *European Journal of Operational Research*, 213, 478–488.
[4] De Giovanni, P. (2017). Closed-loop supply chain coordination through incentivizes with asymmetric information. *Annals of Operations Research*, 253, 133–167.
[5] De Giovanni, P. & Zaccour, G. (2013). Cost–revenue sharing in a closed-loop supply chain. In V. Křivan & G. Zaccour (Eds.), *Advances in dynamic games* (pp. 395–421). Boston: Birkhäuser.
[6] De Giovanni, P. & Zaccour, G. (2019). A selective survey of game-theoretic models of closed-loop supply-chains. *4OR*, *17*, 1–44.
[7] Fiala, P. (1999). Modelování vyjednávacího procesu s více hodnotícími kritérii. *Politická ekonomie, 32*, 253-268.
[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*, 326–336.
[9] Guide, V. D. R. & Van Wassenhove, L. N. (2009). The evolution of closed-loop supply chain research. *Operations Research*, 57, 10–18.

[10] Han, X., Wu, H., Yang, Q. & Shang, J. (2017). Collection channel and production decisions in a closedloop supply chain with remanufacturing cost disruption. *International Journal* of *Production Research*, *55*, 1147– 116.

[11] Hong, X., Xu, L., Du, P. & Wang, W. (2015). Joint advertising, pricing and collection decisions in a closedloop supply chain. International Journal of Production Economics, 167, 12–22.

[12] Pasternack, B. (1985). Optimal pricing and returns policies for perishable commodities. *Marketing Science*, *4*, 166-176.

[13] Spengler, J. (1950). Vertical integration and antitrust policy. *Journal of Political Economy*, *58*, 347-352.

[14] Tabolt, S., Lefebvre, E. & Lefebvre, L. A. (2007). Closed-loop supply chain activities and derived benefits in manufacturing SMEs. Journal of Manufacturing Technology Management,18, 627–658.

[15] say, A. A., Nahmias, S., & Agrawal, N. (1999) Modeling Supply Chain Contracts: A Review. In S. Tayur, M. Magazine & R. Ganeshan (Eds.), *Quantitative Models for Supply Chain Management* (pp. 299-336). Bos- ton: Kluwer Academic Publishers.

[16] Xie, J., Liang, L., Liu, L. & Ieromonachou, P. (2017). Coordination contracts of dual-channel with coopera-tion advertising in closed-loop supply chains. *International Journal of Production Economics*, *183*, 528–538.

[17] Xie, J., Zhang, W., Liang, L., Xia, Y., Yin, J. & Yang, G. (2018). The revenue and cost sharing contract of pricing and servicing policies in a dual-channel closed-loop supply chain. *Journal of Cleaner Production*, *191*, 361–383.

[18] Xu, L., Li, Y., Govindan, K. & Xu, X. (2015). Consumer returns policies with endogenous deadline and supply chain coordination. *European Journal of Operational Research*, 242, 88–99

[19] Yang, H., Chen, J., Chen, X. & Chen, B. (2017). The impact of customer returns in a supply chain with a common retailer. *European Journal of Operational Research*, 256, 139–150.

Data Envelopment Analysis – Vehicle Routing Problem Optimization for the Inter-Municipal Cooperation in Public Security

Martin Flegl[1], Eva Selene Hernández Gress[2]

Abstract. Perception about the insecurity in Mexico remains high among its population and only 52.5% of the population identifies the performance of the State police as effective to eradicate the violence. That is why, it is of a high importance for the government to improve this situation. However, citizens consider the performance of the local governments as ineffective to solve important problems. This is mainly due to the limited resources and persistent corruption. In such situation, it is necessary to search for security strategies resulting in lower violence. In this article, we apply a combination of Data Envelopment Analysis and Vehicle Routing Problem to evaluate the technical efficiency of the public security and propose an inter-municipality cooperation. For this purpose, we use data related to the public security system in 125 municipalities in Jalisco state in Mexico. The results reveal a technical efficiency of .6944 with standard deviation of .2676. To improve the level of the technical efficiency, 78 routes between the municipalities were constructed, which lead to more efficient use of approximately 900 police units.

Keywords: Data Envelopment Analysis, Mexico, Optimization, Public Security, Vehicle Routing Problem.

JEL Classification: C44, C61, F52
AMS Classification: 90-08, 90C05, 91B32

1 Introduction

Mexican government spent almost 300 billion of Mexican pesos to improve the situation of violence in the country in 2020. Out of the total, 42.22 billion were allocated to the public security [19]. However, the insecurity remains alarming, as more than 23,000 reported cases per 100,000 citizens have been reported yearly since 2010 [17]. Citizens perceive as a main problem a low performance of the police, as only 52.5% of the population aged 18 and over identifies the performance of the State Police as effective in eradicating the crime, while this perception is 47.5% in the case of the Municipal Preventive Police. What is more, only 27.8% of them consider the government performance of their city to be effective in solving important problems of their citizens [18]. One of the reasons of such low performance is linked to a low budget and quality infrastructure destined in public security [4][21].

To improve this perception, it is necessary to elevate the institutional quality through the government's efficiency in decision-making [7][9]. It can be thought that if more personnel in the public security is hired, then the lower violence and better citizens' perception about its performance can be gained. The international standards indicate that a country (region) should have 2.8 police elements per 1,000 citizens, but Mexico reported only 1.2 police elements in 2020. In absolute numbers, the Mexican public security system lacks 101,458 security personnel [4]. As large differences in the development in the regions exist, some municipalities may have problems with low accessibility to services, lack of availability of resources or inability to use them [1]. That is why, the new allocation of the personnel must be carefully evaluated regarding the socio-economic and environmental conditions of each region, as well as to the size of the service operations [14].

Furthermore, the police work in rural areas depends on citizens' trust and respect for the police, which significantly affects the citizens' willingness to cooperate with the police [11]. Due to these conditions, it may be favorable improving the cooperation in the public security system, which would improve the resource allocation efficiency and the level of security. In fact, as Bel and Sebö [5] suggest, small municipalities should search an inter-municipal cooperation (IMC) to enhance the collaboration between governments, local councils, or agencies. Such

[1] 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.
[2] Tecnologico de Monterrey, School of Engineering and Sciences, Campus Hidalgo, Blvd. Felipe Ángeles 2003, 42083, Pachuca, Hidalgo, evahgress@tec.mx. ORCID: 0000-0001-8720-5997.

cooperation can increase the efficiency in combating the violence by optimizing the number of the public security personnel and their operability [22][25]. This is especially important when the resources are limited, and, in the case of Mexico, when there is a lack of police elements in the country. In this case, the IMC can be designed by using techniques for routing optimizations based on the Traveling Salesman Problem (TSP), in the case of smaller number of connections, or the Clustering TSP (CTSP) for larger number of connections, or using the Vehicle Routing Problem (VRP), when multi depots are considered [3].

The Data Envelopment Analysis (DEA) is widely used to assess the operational efficiency and performance [10][13]. In public security, the DEA analysis focuses either on a single period, performs the analysis over several years, or the two-stage DEA models are applied. For example, Wang, Le and Hou [26] applied the DEA to evaluate the efficiency of the security departments in 22 administrative districts in Taiwan to improve their management. Nepomuceno et al. [23] studied the public safety efficiency and effectiveness in 145 municipalities in Pernambuco, Brazil. Alda and Dammert [2] assessed police performance of 619 Peruvian municipalities considering environmental aspect in which police operate, whereas Benito, Martínez-Córdoba and Guillamón [6] analyzed the efficiency of police service in 99 municipalities in Spain. Hadad, Keren and Hanani [14] evaluated 13 police stations efficiency in the South of Israel. The outcome of these analyses provides overview about the policing efficiency and the understanding of factors that affect the violence and/or the efficiency. However, up to our knowledge, such analysis does not consider improvements through the IMC to enhance the efficiency of the security system.

That is why, the objective of the article is to assess the technical efficiency of the public security system in Mexico and propose the inter-municipal cooperation. For this purpose, the Data Envelopment Analysis is used in the first part of the analysis to obtain the municipalities' public security efficiency and the Vehicle Routing Problem is applied in the second part of the analysis to design the collaborations between the municipalities. The research question that arises is whether the efficiency to eradicate the violence in Mexico can be improved through the creation of the inter-municipal collaboration.

2 Materials and methods

2.1 Data

The analysis uses data related to the public security and the criminal incidence of 125 municipalities in Jalisco state, Mexico (Figure 1a). Jalisco is one of the most important Mexican states, as it is the 3rd biggest state according to the population (8,348,151), represented 6.9% of the total GDP of the country in constant prices in 2020 (4th in Mexico), and Jalisco contributed the most in the primary sector of the country with 12.2% of the corresponding GDP [15]. On the other hand, Jalisco also registered the 3rd most crimes in the country with 160,630 cases (7.926%), right behind Mexico City and Estado de México [24].

To evaluate the efficiency of the public security system, data from the National Census of Municipal Governments and Territorial Demarcations of Mexico City in 2019 (Census) [16] was used. The Census includes statistical and geographic information on the management and performance of the institutions that make up the Public Administration of each municipality. The DEA model uses as the inputs the number of personnel assigned to public security functions in the Municipal Public Administrations (PERSONNEL) and the average salaries according to income range of the public security functions (SALARY). Further, the model includes the number of infrastructure units to exercise the public security function that the Municipal Public Administrations have (UNITS), which encompasses the number of Commanderies, Security stations, Security modules, Police booths and others, in full operation in 2018. Finally, the information about the number of CCTV security camaras (CCTV) and the number of security panic buttons (BUTTON) in each municipality were also used. For the output part of the model, information related to the level of violence (VIOLENCE) obtained from the Executive Secretariat of the National Public Security System [24]. The violence is divided into seven juridical types of violence: Property crimes; Family crimes; Sexual freedom and safety crimes; Society crimes; Life and bodily integrity crimes; Personal freedom crimes; and Other affected legal assets (of common jurisdiction).

All the variables were recalculated per 1,000 citizens to eliminate the effect of population on the results [14][26]. In addition, the outputs have a character of the undesirable outputs. That is why, to satisfy the requirement of the DEA models for the outputs [2], each output was subtracted from a big M value as follows

$$M = \max\{a_{i2}\} + \min\{a_{i2}\},\ i = 1, 2, \dots, n.$$

and

$$\mathrm{Invert}_{i2} = M - a_{i2},\ i = 1, 2, \dots, n.$$

where n is the number of DMUs ($n = 125$). MaxDEA 7 Ultra software was used for the calculations.

2.2 Data Envelopment Analysis

The Data envelopment Analysis (DEA) evaluates decision-making units (DMU) regarding their capabilities to convert mm multiple inputs into ss multiple outputs [8]. If the model assumes constant returns to scale, the so-called CCR model can be used. The CCR input-oriented model for DMU_o is formulated as follows:

Maximize

$$q = \sum_{r=1}^{s} u_r y_{ro} \tag{1}$$

subject to

$$\begin{gathered} \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0, j = 1,2,\dots,n, \\ \sum_{i=1}^{m} v_i x_{io} = 1, \\ \mu_r, v_i \geq \varepsilon \text{ , } \varepsilon > 0, u_o \text{ free in sign.} \end{gathered} \tag{2}$$

where x_{ij} is the quantity of the input i of the DMU_j, y_{rj} is the amount of the output r of the DMU_j, and μ_r and v_i are the weights of the inputs and outputs $i = 1,2,\dots,m$, $j = 1,2,\dots,n$, $r = 1,2,\dots,s$ and ε is the so-called non-Archimedean element necessary to eliminate zero weights of the inputs and outputs. DMU is 100% efficient if $q = 1$, i.e., there is no other DMU that produces more outputs with the same combination of inputs, whereas DMU is inefficient if $q < 1$.

2.3 Vehicle Routing Problem

The Vehicle Routing Problem (VRP) is one of the most frequently encountered optimization problems in logistics, which aim is to design optimal delivery routes. In this case, each costumer is visited exactly once by one vehicle, each vehicle starts and ends its route at the depot and the capacity of the vehicle is not exceeded [3][12]. The analysis considers Pickup and Delivery Vehicle Routing Problem (PDVRP) and multiple depots with the following characteristics:

- Single vehicle is used in all municipalities. One of the most used vehicles by the Mexican municipal police is the Jeep Wrangler. That is why, the capacity of the vehicle is four persons. Each main municipality (depot) has eight vehicles available for the IMC operations.
- Each vehicle is loaded according to the required number of persons on a route. In each municipality, the loaded number of persons arrives, operates one hour, and leaves to next municipality on the route. The IMC usage of the police is recalculated regarding the number of municipalities on a route and the number of police loaded in the vehicle.
- We only consider cost per kilometer related to consumed gasoline (2.217 pesos, considering average consume of 10.3 liters of gasoline per 100km, and price of gasoline 21.52 pesos per liter). The other costs (such as insurance, depreciation of the vehicle and tires, etc.) are ignored.

For the calculations, the open source VRP Spreadsheet Solver was used. The complete description of the VRP algorithm can be found in [12].

3 Results

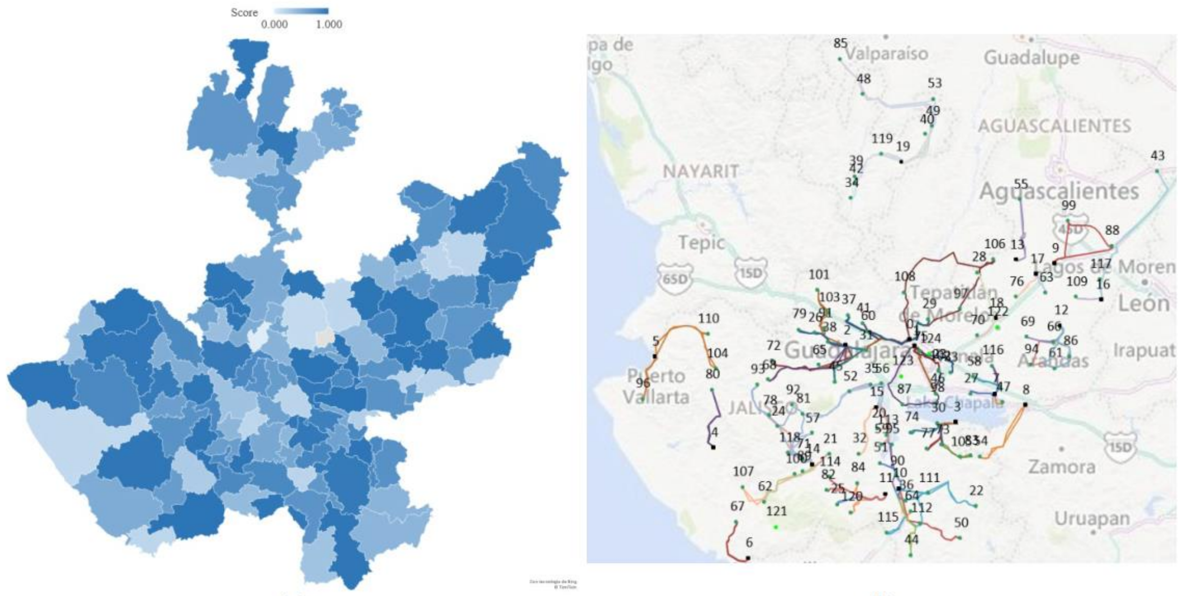
The results are divided into two parts. First, the overall technical efficiency of the public security system in Jalisco is evaluated; second, based on the efficiency results, the inter-municipal cooperation is proposed and its effect on the security system efficiency assessed.

The average technical efficiency to eradicate the violence of the whole state is 0.694 with standard deviation (SD) of 0.268 (Figure 1a, Table 1). The highest efficiency is observed in Costa-Sierra occidental region (0.811), Altos norte (0.774) and Sureste (0.732), whereas the lowest efficiency reported Valles (0.633), Ciénega (0.619) and Costa sur (0.590). Out of the 125 municipalities, only four reached the efficiency of 1.000: Atotonilco el Alto (Ciénega), Chapala (Sureste), Cuautitlán de García Barragán (Costa sur) and Magdalena (Valles). These municipalities are below the state's average considering the population (39,839 citizens compared to state's average of 66,785 citizens) and, what is more, these municipalities used only 0.765 of PERSONNEL per 1,000 citizens (state's average 4.333), 0.028 of UNITS per 1,000 citizens (0.205) and registered only 2.186 crime cases per 1,000 citizens (19.241).

Region	# of municipalities	Efficiency	SD
Alto sur	12	0.719	0.257
Altos norte	8	0.774	0.251
Centro	12	0.680	0.331
Ciénega	9	0.619	0.328
Costa sur	6	0.590	0.369
Costa-Sierra occidental	8	0.811	0.229
Lagunas	12	0.692	0.221
Norte	10	0.677	0.188
Sierra de Amula	14	0.699	0.277
Sur	12	0.705	0.277
Sureste	10	0.732	0.248
Valles	12	0.633	0.300
State	**125**	**0.694**	**0.268**

Table 1 Average technical efficiency by region.

The results indicate that, in average, the efficiency of all municipalities can be improved by 30.6%. As the CCR-input oriented model was used, this improvement can be achieved by optimizing the number of PERSONNEL used by each municipality, and correspondingly their salaries. The infrastructure (UNITS, CCTV and BUTTON) is already built or in function, therefore, its functionality will remain unchanged. In the Vehicle Routing Problem for the Inter-municipal cooperation (IMC), the construction of the depot (the main cities) is considered as follows (also due to the software limitations): 1) one depot is constructed in a municipality with the lowest efficiency (i.e., with the highest change of the PERSONNEL) in each region of the state; 2) eight more depots are constructed in the following municipalities with the lowest efficiency. In this case, the model will assure the biggest IMC leading to higher improvements in the efficiency of the public security system.



(a) (b)

Figure 1 (a) Efficiency of eradicating the violence in Jalisco, Mexico; (b) Routing for the IMC between the municipalities

The results of the VRP (Figure 1b) show that 78 routes were created to promote the IMC between the municipalities. As each municipality had a capacity of 8 vehicles (160 vehicles in total)), but only 48.75% of the capacity was used. This is mainly due to the location of the selected municipalities for the depots, as some depots are in the same region. In average, each route requires 119 kilometers and 3 hours and 26 minutes of the operational time (including 1 hour of the operation in each municipality on the route). The average driving time between two pair of municipalities takes 55 minutes and 51.65 kilometers. Both the needed time and the distance on each route indicate that the routes can be operated more than once every day or more vehicles can be allocated on these routes.

The total costs of the IMC routes are 20,577.44 pesos. On the other hand, the IMC resulted in reallocation of 6.77% PERSONNEL in average (approximately 900 police units in absolute numbers), which resulted in average salary savings of 326.15 pesos per person monthly in each municipality, i.e., 293,733.41 pesos in absolute numbers. Furthermore, these savings resulted in better efficiency in several municipalities (depending on number of operated routes and its size). For example, Totatiche in the region Norte operates 8 routes, which resulted in higher efficiency by +19.2% (from 44.8% to 64.0%), Autlán de Navarro in the Sierra de Amula region operates 8 routes with an improvement of 3.5% (27.7% to 31.2%) and Zapotlán el Grande in South Region with 8 routes and +3.1% of the efficiency improvement (20.4% to 23.6%). The less routes are designed, the lower the effect of the IMC on the efficiency improvements is.

One of the key motivations to create the IMC is to enhance the service quality, optimize the service-related transaction costs and improve governance [5]. In this case, the proposed IMC in the security system sounds reasonable as security is the common target of all municipalities, but not all of them may have enough resources to do so. In general, there is less crime in rural areas, so there are less police allocated, which may result in fewer security infrastructure [11][20]. The application of the VRP to design the routes looks appropriate, hence several limitations must be considered. First, each municipality is designed to only one route. So, the IMC does not allow a cooperation from another municipality, which may be desirable. Second, the number of depots and their location seem crucial for the efficiency improvements, as tight locations may restrict the number of designed routes by VRP.

4 Conclusions

Many ways to improve the level of security exist, however some of these can be impossible due to limiting resources. Therefore, it is necessary to search new ways of how to improve current security processes. In this article, the combination of DEA and VRP was applied to propose the Inter-municipal cooperation between 125 municipalities in Jalisco, Mexico. For this purpose, the technical efficiency was calculated in the first part of the model to evaluate the overall technical efficiency to eradicate the violence. In the second part, 20 municipalities with the lowest efficiency were identified. Those municipalities were considered as the depots in the VRP. The proposed IMC created 78 routes with a use of approximately 900 police units between the municipalities. What is more, considering the model settings, the proposed IMC leaded to an approximate monthly operational savings of 293,733.41 pesos in the whole state.

References

[1] Adebimpe Adekanbi, O. & Sánchez-Juárez, I. (2021). Socioeconomic Inequalities Among the Municipalities of Chihuahua, Mexico. *The Journal of Developing Areas*, 55(3), 229-244. https://dx.doi.org/10.1353/jda.2021.0063

[2] Alda, E. & Dammert, L. (2019). Weathering the Storm! The effects of the external environment on police efficiency in Peru. *Policing*: *An International Journal*, 42(6), 1124-1140. https://doi.org/10.1108/PIJPSM-03-2019-0033

[3] Anaya Fuentes, G.E., Hernandez-Gress, E.S., Seck Tuoh Mora, J.C. & Medina Marín, J. (2018). Solution to travelling salesman problem by clusters and a modified multi-restart iterated local search metaheuristic. *Plos One*, 13(8), e0201868. https://doi.org/10.1371/journal.pone.0201868

[4] Arellano, S. (2021). Faltan alrededor de 100 mil policías en México [About 100,000 police officers are missing in Mexico]. *México social*, [Online], Available: https://www.mexicosocial.org/faltante-de-policias/#:~:text=Desde%20esta%20perspectiva%20es%20fundamental,como%20est%C3%A1ndar%20a%20nivel%20planetario [18 Apr 2022].

[5] Bel, G. & Sebö, M. (2018). Does inter-municipal cooperation really reduce delivery costs?. An empirical evaluation of the role of scale economies, transaction costs, and governance arrangements. Research Institute of Applied Economics. *Working Paper* 2018/16, available: http://www.ub.edu/irea/working_papers/2018/201816.pdf

[6] Benito, B., Martínez-Córdoba, P.-J. & Guillamón, M.-D. (2021). Measurement and determinants of efficiency in the municipal police service. *Evaluation and Program Planning*, 85, 101904. https://doi.org/10.1016/j.evalprogplan.2020.101904

[7] Bjørnskov, C., Dreher, A. & Fischer, J.A.V. (2010). Formal institutions and subjective well-being: Revisiting the cross-country evidence. *European Journal of Political Economy*, 26(4), 419-430. https://doi.org/10.1016/j.ejpoleco.2010.03.001
[8] Cooper, W., Seiford, L. & Zhu, J. (2011). *Handbook on data envelopment analysis*. Nueva York: Springer.
[9] Danish, M.H. & Nawaz, S.M.N. (2022). Does institutional trust and governance matter for multidimensional well-being? Insights from Pakistan. *World Development Perspectives*, 25, 100369. https://doi.org/10.1016/j.wdp.2021.100369
[10] De La Hoz, E., Zuluaga, R. & Mendoza, A. (2021). Assessing and Classification of Academic Efficiency in Engineering Teaching Programs. *Journal on Efficiency and Responsibility in Education and Science*, 14(1), 41–52. https://doi.org/10.7160/eriesj.2021.140104
[11] Eman, K. & Bulovec, T. (2021). A case study of rural crime and policing in Pomurje region in Slovenia. *Journal of Rural Studies*, 85, 43-51. https://doi.org/10.1016/j.jrurstud.2021.05.012
[12] Erdoğan, G. (2017). An open source Spreadsheet Solver for Vehicle Routing Problems. *Computers & Operations Research*, 84, 62-72. https://doi.org/10.1016/j.cor.2017.02.022
[13] Flegl, M., Jiménez Bandala, C.A., Sánchez-Juárez, I. & Matus, E. (2021). Productivity analysis in the Mexican food industry. *Proceedings of the 39th International Conference on Mathematical Methods in Economics*: Prague, Czech Republic, 100-105.
[14] Hadad, Y., Keren, B. & Hanani, M.Z. (2015). Combining data envelopment analysis and Malmquist Index for evaluating police station efficiency and effectiveness. *Police Practice and Research*, 16(1), 5-21. https://doi.org/10.1080/15614263.2013.845945
[15] IIEG (2021). *Producto Interno Bruto de Jalisco en 2020 [Gross Domestic Product of Jalisco in 2020]*. Instituto de Información Estadística y Geográfica de Jalisco [Institute of Statistical and Geographic Information of Jalisco], [Online], Available: https://iieg.gob.mx/ns/wp-content/uploads/2021/12/Ficha-informativa- Producto-Interno-Bruto-de-Jalisco-2020-20211209.pdf [21 Apr 2022].
[16] INEGI (2020). *Censo Nacional de Gobiernos Municipales y Demarcaciones Territoriales de la Ciudad de México 2019 [National Census of Municipal Governments and Territorial Demarcations of Mexico City 2019]*. Instituto Nacional de Estadística y Geografía, available: https://www.inegi.org.mx/programas/cngmd/2019/ [20 Mar 2022].
[17] INEGI (2021). *Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública (ENVIPE) 2021 [National Survey of Victimization and Perception of Public Safety]*. Instituto Nacional de Estadística y Geografía [National Institute of Statistics and Geography], [Online], Available: https://www.inegi.org.mx/programas/envipe/2021/ [15 Apr 2022].
[18] INEGI (2022). *Encuesta Nacional de Seguridad Pública Urbana (ENSU (septiembre 2021) [National Survey of Urban Public Security (ENSU (September 2021)]*. Instituto Nacional de Estadística y Geografía [National Institute of Statistics and Geography], [Online], Available: https://inegi.org.mx/programas/ensu/ [15 Apr 2022].
[19] Institute for Economics & Peace (2021). *Mexico Peace Index 2021*: Identifying and Measuring the Factors That Drive Peace. Sydney, May 2021. Available from: http://visionofhumanity.org/resources [20 Apr 2022].
[20] Lindström, P. (2015). Police and crime in rural and small Swedish municipalities. *Journal of Rural Studies*, 39, 271-277. https://doi.org/10.1016/j.jrurstud.2014.12.004
[21] Monroy, J. (2020). Presencia criminal falta de policías, sin equipo… [Criminal presence, lack of police officers, without equipment…], *El Economista*, [Online], Available: https://www.eleconomista.com.mx/politica/Presencia-criminal-falta-de-policias-sin-equipo...-20200721-0142.html [18 Apr 2022].
[22] Moreira, G. & Ceccato, V. (2021). Testing theft transmission in and around São Paulo metro stations, Brazil. *Journal of Transport Geography*, 94, 103116. https://doi.org/10.1016/j.jtrangeo.2021.103116
[23] Nepomuceno, T.C.C., Santiago, K.T.M., Daraio, C. & Costa, A.P.C.S. (2020). Exogenous crimes and the assessment of public safety efficiency and effectiveness. *Annals of Operations Research*. https://doi.org/10.1007/s10479-020-03767-6
[24] SESNSP (2020). *Incidencia delictiva del Fuero Común, nueva metodología [Criminal incidence of the Com- mon Law, new methodology]*. Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública, available: https://www.gob.mx/sesnsp/acciones-y-programas/incidencia-delictiva-del-fuero-comun-nueva-metodologia?state=published [20 Mar 2022].
[25] Sestoft, D., Rasmussen, M.F., Vitus, K. & Kongsrud, L. (2014). The police, social services and psychiatry cooperation in Denmark—A new model of working practice between governmental sectors. A description of the concept, process, practice and experience. *International Journal of Law and Psychiatry*, 37(4), 370-375. https://doi.org/10.1016/j.ijlp.2014.02.007

[26] Wang, C.-N., Le, A.L. & Hou, C.-C. (2019). Applying Undesirable Output Model to Security Evaluation of Taiwan. *Mathematics*, 7(11), 1023. https://doi.org/10.3390/math7111023

Pedagogical Performance Modelling at the Faculty of Economics of the University of South Bohemia – Case Study

Ludvík Friebel[1], Jan Fiala[2]

Abstract.
Systemic steps that stem from a number of in-depth analyses and relations of various factors are proposed at the Faculty of Economics of the University of South Bohemia in České Budějovice following the accreditation of degree programmes to increase the efficiency of instruction and to perform optimisation of the internal structure. The aim of the faculty management is to continuously propose and implement measures that will lead to internal restructuring of the faculty and will take into account, e.g., the constantly changing numbers of students in particular degree programmes, the pass rate, the importance (weight) of particular courses, numbers of graduation theses supervised at particular departments, etc. The pedagogical performance of particular departments, or individual employees, can be inferred by taking all relevant factors into account, which may become a basis for making decisions concerning changes in the organisational structure of the faculty. The proposed calculation model delivers objective information concerning the performance of the faculty in relation to the curriculum of particular degree programmes.

Keywords: Probability model, Deterministic model, Operational research, Pedagogical performance of faculty, Efficiency, Optimisation.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Several researchers dealt generally with the chosen issue, using different methods and techniques. For example, Jablonský (2011) and (2016) used DEA for efficiency research for performance evaluation in Czech higher education. The same method was used by Johnes, Portela, Thanassoulis (2017) and Mikušová (2017). Comprehensive view of efficiency research in higher education institutions is provided by an extensive study of Worthington and Lee (2008).

In publication of Flégl and Vltavská (2013) is compared scientific performance of faculties of economy. Musil and Fischer (2015) dealt with the general performance measurement of educational services. However, unlike some previous articles, our contribution deals with measuring the performance of particular departments within one faculty. The chosen solution respects also the Act on Higher Education Institutions (1998).

Internal optimisation and subsequent organizational restructuring aiming at adapting to the changing external and internal conditions took place at a number of faculties of higher education institutions and universities in the Czech Republic and abroad through the selection of adequate organisational measures. The restructuring of departments at higher education institutions often occurred as the consequence of measures that had been adopted in the context of an accreditation process.

A study containing results of an analysis that focused on deriving an objective informative indicator concerning the performance of particular departments depending on degree programmes offered was prepared for the management of the Faculty of Economics of the University of South Bohemia as the basis for the submission of proposals for the optimisation of the internal structure.

The pedagogical performance of a department is understood as a quantitative assessment of the workload of all staff of the department, specifically the total absolute numbers of factors monitored.

[1] University of South Bohemia/Faculty of Economics, Department of Applied Mathematics and Informatics, Studentská 13, České Budějovice, friebel@ef.jcu.cz.
[2] University of South Bohemia/Faculty of Economics, Department of Applied Mathematics and Informatics, Studentská 13, České Budějovice, fiala@ef.jcu.cz.

In order to ensure the impartiality of the process of the assessment of the performance of departments, the following indicators were included in the stochastic calculation model: numbers of students in new and current bachelor's and master's degree programmes, course importance according to the profile, student success rate and the preference for a department from the perspective of the selection of the graduation thesis supervisor.

2 Data and Methods

All data presented below were sourced from the university information system and curricula of particular programmes and were utilised with the consent of the management of the University of South Bohemia. The analysis of these documents, the results of which led to the preparation of the calculation model, was utilised as the main research method. The calculation model produces a numerical assessment of the performance of particular departments while taking into account the abovementioned criteria and it is a practical example of the theory of operational research. Abbreviations for degrees of study and study programmes in our text are presented in table 1.

Abbr.	Degree of study/Study programme	Abbr.	Degree of study/Study programme
bc	Bachelor's degree programme	CR	Tourism
bcK	Bachelor's degree programme in the part-time form	EKINF	Economic informatics
n	Consecutive degree programme	PI	Business informatics
nK	Consecutive degree programme in a part-time form	UFRP	Accounting and financial management of the company
EM	Economics and management	AI	Applied informatics
MRR	Regional development management	OPCR	Commercial management - tourism
AEFP	Analysis in economic and financial practice	OPRM	Commercial management - retail management
FU	Finance and accounting	TD	Triple degree

Table 1 Abbreviations of degrees of study ana study programmes

Numbers of students in new degree programmes

For the purposes of the analysis, the number of students in new degree programmes that had already been opened in the academic year 2019/2020 constituted the basis. If the programmes had not been opened yet, the basis was the number of students in degree programmes that were replaced by the new degree programmes. An estimate (see tables 1 and 2) was used in the case of courses of study that were not preceded by another programme. An imbalance in the occupancy of particular programme is chiefly apparent in tables 2 and 3. The number of students on bachelor's study programmes depends on the capacity of individual programme. All those who apply are admitted to some study programmes. Entrance exams are then taken for master's degree programmes.

New degree programme	Original or parallel degree programme	Number of students
MRR-bc	MRR-bc, SPVS-bc	24
MRR-bcK	MRR-bcK	14
EM-bc	EM-bc, OPRM-bc, OPCR-bc,REP-bc	224
EM-bcK	EM-bcK, OPRM-bcK	52
AEFP-bc	AEFP-bc	2
PI-bc	EKINF-bc	23
PI-bcK	EKINF-bcK	18
FU-bc	UFRP-bc	58
FU-bcK	UFRP-bcK	34
CR-bc	New course of study	40 (estimate)
CR-bcK	New course of study	20 (estimate)

Table 2 Numbers of students who enrolled in bachelor's degree programmes in winter semmester 2019 (source: author's own calculations)

New degree programme	Original or parallel degree programme	Number of students
EM-n	EM-n, OP-n, REP-n	49
EM-nK	OP-nK, REP-nK	60
FU-n	UFRP-n	25
FU-nK	UFRP-nK	14
AI-n	AI-n	12
AEFP-n	AEFP-n	10
TD-n	TD-n	12

Table 3 Numbers of students who enrolled in master's degree programmes in winter semmester 2019, (source author's own calculations)

Curricula of accredited degree programmes

Curricula of accredited degree programmes constituted the basis for the purposes of the calculation of the workload of workplaces - departments. From the perspective of the calculation of contact hours, weights were used in the case of optional courses (table 4).

Profile	P	PZ	ZT	PV-A	PV-B	PV-C
Weight	1	1	1	0.500	0.333	0.250

Table 4 Weights of courses per profile (P – compulsory course, PZ – courses of a profiling core, ZT – basic theoretical courses, PV – elective course) (source: author's own calculations)

If the course exists in an English language version, the number of students enrolled in the course is divided equally between the corresponding options for the purposes of the calculation. The number of lectures, tutorials and consultations concerning courses involving multiple departments will be divided in the manner that corresponds to the manner in which respective departments are involved in their instruction.

Pass rate

The pass rate (transition coefficient) in particular year expresses the share of students accomplishing all necessary courses of given study programme against the number of all students in the study programme and is set for particular study programme on the basis of the Union Information from Students' Registers (SIMS). Therefore, the failure rate between winter and summer semesters of the same academic year is not taken into account. Likewise, due to the lack of more detailed data, we used the same value for all bachelor's programmes and analogously the same value for all master's programmes. Coefficients in table 5 are based on prior years and the succession of particular degree programmes (table 2 and 3) is taken into account.

	Year				Year	
Programme	1	2	3	Programme	1	2
bc	1	0.607	0.461	n	1	0.84
bcK	1	0.607	0.461	nK	1	0.84

Table 5 The pass rate of students of bachelor's and master's programmes without distinguishing the semester (source: author's own calculations)

Calculation of contact hours

The executed number of contact hours in a degree programme i for a course j is calculated in the following manner:

$$l_{ij} = as_i . ps_{it} . pc_{ij} . l_j \,, \tag{1}$$

where as_i stands for the number of students who are admitted to the first year of the degree programme i at the beginning of the year t, pc_{ij} stands for the pass rate of the programme i at the beginning of the year t, pc_{ij} stands for the probability of the selection of the specific course j within the degree programme i, l_{ij} stands for the number of lessons of the course j as it is outlined in the approved accreditation application. For the purposes of the calculation, it is necessary to distinguish lectures and tutorials, see the example of the calculation below.

Calculation example

There are 224 first-year students (see table 1) in the degree programme entitled EM-bc. The course entitled Probability theory and statistics 1 is in the winter semester of the second year of the degree programme and the course includes 14 hours of tutorials and 28 hours of lectures. There is also a version in English, therefore, 50 % of students are expected to enrol in the Czech version of the course. The pass rate for the first year is set to 60.7 % (see table 4).

Thus, the number of contact hours of lectures is:

$$224 \cdot 0.607 \cdot 0.5 \cdot 28 = 1902$$

and the number of contact hours of tutorials is:

$$224 \cdot 0.607 \cdot 0.5 \cdot 14 = 951.$$

3 Results

3.1 Version excluding final theses

Since the selection of the place of work where the student wishes to execute his/her final thesis is a matter of the student's choice, such courses were not included in the analysis. It is 6595 hours of consultations in total.

Department	SS	WS	Total	AE	THP	Ped. perf.
KMI	61751	101428	163179	12.65	1	12900
KEN	45141	53939	99083	6.85	1	14465
KUF	43888	44392	88280	13.20	1	6688
KŘE	47913	35011	82921	8.95	1	9265
KOD	37907	30493	68400	8,25	1	8291
KPH	24097	21557	45651	5.10	0.50	8951
KRM	5694	17621	23314	3.30	1	7065
KJE	2095	5288	7384	7.30	1.85	1012

Table 6 Numbers of contact hours in total excluding bachelor's and master's theses (SS – summer semester, WS – winter semester, AE – no. of academic employees, THP – no. of non-academic employees, Ped. perf. – no. of contact hours per academic employee) (source: author's own calculations)

The "efficiency" – pedagogical performance – of the corresponding department is calculated in the final column of table 6 for information purposes. This value is constructed as a share of the total number of contact hours and the number of academic staff members. Significant differences between departments are evident again.

Table 7 contains numbers of contact hours of lectures, tutorials and consultations while excluding graduation theses.

	Lectures			Tutorials			Consultations		
Dpt.	SS	WS	Total	SS	WS	Total	SS	WS	Total
KMI	27993	40613	68606	26105	46861	72966	7653	13954	21607
KEN	19421	23798	43220	19057	24419	43477	6663	5722	12386
KUF	19663	18782	38445	18526	19080	37606	5699	6530	12229
KŘE	21980	13576	35555	20986	16001	36986	4947	5434	10380
KOD	18359	8855	27214	16145	18794	34939	3403	2844	6247
KPH	3539	10097	13635	17959	8571	26529	2599	2889	5487
KRM	1238	5518	6756	3402	9464	12866	1054	2639	3692
KJE		674	674	1250	3479	4730	845	1135	1980

Table 7 Numbers of contact hours of lectures, of tutorials and of consultations excluding bachelor's and master's theses (source: author's own calculations)

3.2 Calculation including final theses.

Information in curricula provided the basis for determining the number of contact hours concerning graduation theses. Guaranties of study programmes in table 8 were determinative from the perspective of assigning graduation theses to specific departments.

Degree programme	BT1	BT2	DT1	DT2
MRR-bc, MRR-bcK	KRM	KRM		
EM-bc, EM-bcK	KEN	KŘE		
EM-n, EM-nK			KEN	KŘE
AEFP-bc, AEFP-n	KMI	KMI	KMI	KMI
PI-bc, PI-bcK	KMI	KMI		
FU-bc, FU-bcK	KUF	KUF		
FU-n, FU-nK			KUF	KUF
AI-n			UAI	KMI
TD-n			KRM	KRM

Table 8 Assigning final theses to departments (bc-bachelor's degree programme, n-consecutive degree programme, nK-consecutive degree programme in a distance form, BT-bachelor thesis, DT-diploma thesis)

It is the assignment of theses to individual workplaces that results in changes in the number of consultations provided by individual departments in table 9. The differences are, of course, greater for the departments they guarantee the degree programmes.

Dpt.	SS	WS	Total	Dpt.	SS	WS	Total
KMI	7888	14290	22178	KOD	3624	3065	6690
KEN	6111	8052	14163	KPH	1054	2639	3692
KUF	7330	6389	13720	KRM	985	1275	2260
KŘE	6880	5434	12314	KJE	2599	2889	5487

Table 9 Numbers of contact hours of consultations (source: author's own calculations)

Finally, table 10 indicates the total performance and per academic employee performance of departments covering all factors including final theses. What follows from the comparison of the total contact hours of relevant departments is primarily the unevenness in the performance of them. The KMI Department demonstrates the highest total and relative pedagogical performance regarding monitored indicators, which is almost fourteen times more than the lowest performance of the KJE Department. The situation is also similar regarding the number of contact hours per academic employee. Differences in the performance of departments between the winter and the summer semester are also clearly discernible in the same table. For instance, the KRM Department's performance is three times higher in the winter semester than in the summer semester. Number of contact hours between both semesters is balanced only for KUF and KPH departments.

Dpt.	Total	Ped. perf.	SS/WS	Dpt.	Total	Ped. perf.	SS/WS
KMI	185357	14653	60.2%	KOD	75090	8118	123.8%
KEN	113246	14426	82.7%	KPH	49343	14952	104.0%
KUF	102000	7727	100.9%	KRM	25574	2765	35.4%
KŘE	95235	9571	135.5%	KJE	12871	2524	57.4%

Table 10 Numbers of contact hours of all activities (Ped. perf. – no. of contact hours per academic employee, SS/WS – share btw. summer ana winter semester) (source: author's own calculations)

Besides the absolute and relative number of contact hours presented above, the number of different courses must also be taken into account see table 11, where the number of academic employees is also displayed. The highest total number of courses taught, 81, is posted by the KMI department, which represents almost 7 times the number of courses of the KPH Department. It is clear that the more courses, even if they are lectured for small groups of students, the more academic staff are needed to provide them. However, it is necessary to proceed with caution here, as some courses can be very similar in content.

Dpt.	SS	WS	Total	AE	Dpt.	SS	WS	Total	AE
KMI	34	47	81	12.65	KJE	8	15	23	5.1
KUF	22	23	45	13.2	KŘE	11	11	22	9.95
KOD	16	14	30	9.25	KRM	7	8	15	9.25
KEN	15	11	26	7.85	KPH	3	9	12	3.3

Table 11 Numbers of courses and academic employees (source: author's own calculations)

4 Discussion

The developed calculation model is a simple tool for the comparison of the pedagogical performance of individual departments. The real performance of staff members of the given places of work, and therefore their workload, can be deduced from the values obtained in the model and presented in table 10. The differentiation of semesters provides an effective argument in favour of the modification of a degree programme in the context of accreditations.

The model should be updated at the beginning of every academic year when the number of students admitted to individual degree programmes is clear as well as the failure rate in particular years of the programmes. In fact, these factors change every year and significantly influence the total workload of departments. The calculation model is very sensitive to the quality of data collected and it would be appropriate to connect it directly to the study agenda system. In the case of new degree programmes that are usually not integrated into the IS, adding information to the model may be a very laborious matter. The model must be perceived as a variable structure that must reflect new factors that may have a substantial influence on the performance of departments. If the significant deviations from the defined structure occur, it is necessary to adapt the personnel structure of the faculty to the trend.

Results of the analysis may serve well not only during the accreditation process but also for further information support of various internal measures of the faculty such as drafting budgets of departments or in the area of evaluation of academic staff members. The developed model is easily applicable to any faculty of a higher education institution.

4.1 Conclusion

A calculation model working with probability was introduced in the paper. The model is used for expressing the performance of departments of the Faculty of Economics of the University of South Bohemia in České Budějovice in numerical terms. Selected measurable factors, which change over time just as the structure of the model can change, were taken into account in the model. The analysis of the influence of degree programmes on the performance of departments contributed towards the successful accreditation of degree programmes and the execution of organisational changes at the faculty.

Acknowledgements

This paper has been produced with contribution of long-term institutional support of research activities by Faculty of Economics, University of South Bohemia, České Budějovice.

References

[1] Flégl, M. and Vltavská, K. (2013). Comparison of Different Approaches to Estimate Efficiency at Faculties of Economics in Public Higher Education Institutions. In *Proceedings of the 10th International Conference Efficiency and Responsibility in Education 2013* (pp. 136–144). Prague: Czech University of Life Sciences.

[2] Jablonský, J. (2011). Models for Efficiency Evaluation in Education. In *Proceedings of the 8th International Conference on Efficiency and Responsibility in Education 2011* (pp. 110-119). Prague: Czech University of Life Sciences.

[3] Jablonský, J. (2016) Efficiency analysis in multi-period systems: an application to performance evaluation in Czech higher education. *Central European Journal of Operational Research* vol. 24, 2, 283–296

[4] Johnes, J., Portela, M. and Thanassoulis, E. (2017). Efficiency in education. *Journal of the Operational Research Society*, vol. 68, 4, 331–338, doi:10.1057/s41274-016-0109-z

[5] Mikušová, P. (2017) Measuring the efficiency of the Czech Republic higher education institutions: An application of DEA. *ERIES Journal*, vol. 10, 2, 58–63, doi:10.7160/eriesj.2017.100204

[6] Musil, P. and Fischer, J. (2015). Measurement of output of educational services. *Politická ekonomie*, vol. 63, 2, 167–184.

[7] Worthington, AC. and Lee, BL. (2008) Efficiency, technology and productivity change in Australian universities 1998–2003. *Economics of Education Review*, vol. 27, 3, 285–298.

[8] Zákon č. 111/1998 Sb. Zákon o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách).

Evaluation of the Health Condition and Medical Resources in the Municipalities in the Czech Republic

Jakub Hanousek[1]

Abstract. We analyze data on the health condition of the population and medical resources in the 160 biggest municipalities of the Czech Republic. The method for evaluation is a data envelopment analysis. It is a method based on linear programming used to measure the efficiency of the production units. The production units in this article are the 160 biggest municipalities in the Czech Republic. The health condition of the population is considered of the most common fatal diseases per 1000 inhabitants in each municipality. Medical services are the number of several types of physicians and the number of pharmacies per 1000 inhabitants in each municipality. Municipalities that have low mortality from the most common fatal diseases and the large numbers of medical resources became efficient.

Keywords: Medical resources, Municipalities, Czech Republic, DEA

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

The goal of this article is to evaluate the health conditions and medical recourses in the 160 biggest municipalities in the Czech Republic. The motivation is finding the municipalities with low mortality from the five most known diseases and, at the same time, high level of medical resources. Such municipalities represent "good" place for living. The evaluation methods in this article are the data envelopment analysis models. Data envelopment models are used to evaluate the efficiency of several units based on the given inputs and outputs. The inputs in this article are deaths from diseases of the circulatory system, death from respiratory diseases, deaths from infectious and parasitic diseases, deaths of a tumor and deaths due to alcohol and drug use. The outputs are number of general practitioners, number of pharmacies, number of dentists, number of pediatricians and the number of special physicians.

The Czech Republic has 6529 municipalities. The biggest municipality in the Czech Republic has around 1.3 milion inhabitants. The smallest municipalities have only one inhabitant. That is a large variance in the number of inhabitants. We reduced the number of municipalities for this article to the 160 biggest. The variance in the 160 municipalities is from around 8.2 thousand to 1.3 million inhabitants. The average age in the municipalities is 42.8 years and the median is 43 years. The vast majority of municipalities have average age between 40 and 45 years. 92.5% municipalities fall in this interval. We can state that the average age in municipalities is homogenous.

2 Methods

DEA was developed by Charnes, Cooper and Rhodes in 1978 [3] and constructs the production frontier and evaulates the technical efficiency of production units. The production unit uses a number of inputs to produce outputs. The technical efficiency of the production unit is defined as the ratio of its total weighted output to its total weighned input or, vice versa, as the ratio of its total weighned input to its total weighned output. DEA model permits each production unit to choose its input and output weights to maximize its technical efficiency score. A technically efficient production unit is able to find such weights that the production unit lies on the production frontier [5]. The production frontier represents the maximum amounts of output that is produced by given amounts of input (the output maximization DEA model) or, alternatively, the minimum amounts of inputs required to produce the given amount of output (the input minimization DEA model). This article deals with two DEA models. The first model is input minimization model. The second model is output maximization model. Both models work with constant revenue from scale [1,3,4].

[1] Prague University of Economics and Business, W. Churchill Sq. 1938/4, 130 67 Prague 3 – Žižkov, Czech Republic, xhanj52@vse.cz

2.1 Input oriented model

$$
\begin{aligned}
&\text{Minimize: } \theta_{\mathbf{q}} - \boldsymbol{\varepsilon}\left(\mathbf{e}^{\mathbf{T}}\mathbf{s}^{+} + \mathbf{e}^{\mathbf{T}}\mathbf{s}^{-}\right), \\
&\qquad\qquad \mathbf{X}\boldsymbol{\lambda} + \mathbf{s}^{-} = \theta_q \mathbf{x}_{\mathbf{q}}, \\
&\text{Subject to: } \mathbf{Y}\boldsymbol{\lambda} - \mathbf{s}^{+} = \mathbf{y}_{\mathbf{q}}, \\
&\qquad\qquad \boldsymbol{\lambda}, \mathbf{s}^{+}, \mathbf{s}^{-} \geq \mathbf{0}.
\end{aligned}
\tag{1}
$$

θ_q is a variable which represents efficiency rate of a production unit. $\boldsymbol{\varepsilon}$ is an infinitesimal constant. The infinitesimal constant $\boldsymbol{\varepsilon} = 10^{-8}$. $\mathbf{e}^{\mathbf{T}} = (1,1,1...,1)$. $\mathbf{s}^{+}$ and $\mathbf{s}^{-}$ are vectors of additional variables. $\mathbf{X}$ is a matrix of inputs. $\mathbf{Y}$ is a matrix of outputs. $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_n)$ is a vector of weights which are assign to productions units. Weights are the variables in a model [4].

2.2 Output oriented model

$$
\begin{aligned}
&\text{Maximize: } \phi_q + \boldsymbol{\varepsilon}\left(\mathbf{e}^{\mathbf{T}}\mathbf{s}^{+} + \mathbf{e}^{\mathbf{T}}\mathbf{s}^{-}\right), \\
&\qquad\qquad \mathbf{X}\boldsymbol{\lambda} + \mathbf{s}^{-} = \mathbf{x}_{\mathbf{q}}, \\
&\text{Subject to: } \mathbf{Y}\boldsymbol{\lambda} - \mathbf{s}^{+} = \phi_q \mathbf{y}_{\mathbf{q}}, \\
&\qquad\qquad \boldsymbol{\lambda}, \mathbf{s}^{+}, \mathbf{s}^{-} \geq \mathbf{0}.
\end{aligned}
\tag{2}
$$

ϕ_q is a variable which represents efficiency rate of a production unit. Other part of model has same interpretation as a model 2. [4]

3 Application

We evaluate efficiency rate of 160 production units. The production units are the 160 biggest municipalities in the Czech Republic. All inputs and outputs are recalculated per 1000 inhabitants. The inputs and outputs are in table 1.

Inputs	Outputs
Deaths from diseases of the circulatory system	Number of general practitioners
Death from respiratory diseases	Number of pharmacies
Deaths from infectious and parasitic diseases	Number of dentists
Deaths of a tumor	Number of pediatricians
Deaths due to alcohol and drug use	Number of special physicians

Table 1: Inputs and outputs

3.1 Inputs

This section describes the inputs that are used in the models.

Death from diseases of the circulatory system

The indicator shows the number of deaths from diseases of the circulatory system as the number of these deaths per thousand living population. The data represents the average number of these deaths in 5 years. This data is available only at the district level, so the data describe the overall situation in the district and the same value is displayed in all municipalities in one district. We consider a lower number to be an indication of a better quality of life.

Death from respiratory diseases

The indicator shows the number of deaths from respiratory diseases as the number of these deaths per imaginary thousand living population. The data represents the average number of these deaths in 5 years. This data is available only at the district level, so the data describe the overall situation in the district and the same value is displayed in all municipalities in one district. We consider a lower number to be an indication of a better quality of life.

Death from infectious and parasitic diseases

The indicator shows the number of deaths from diseases of the circulatory system as the number of these deaths per imaginary thousand living population. The data represents the average number of these deaths in 5 years. This data is available only at the district level, so the data describe the overall situation in the district and the same value is displayed in all municipalities in one district. We consider a lower number to be an indication of a better quality of life.

Death of a tumor

The indicator shows the number of deaths from tumors as the number of these deaths per imaginary thousand living inhabitants. The data represents the average number of these deaths in 5 years. This data is available only at the district level, so the data describe the overall situation in the district and the same value is displayed in all municipalities in one district. We consider a lower number to be an indication of a better quality of life.

Death due to alcohol and drug use

The proportion of deaths related to alcohol and drug use indicates the average proportion of deaths associated with these causes in 5 years of the total number of deaths. This data is available only at the district level, so the data describe the overall situation in the district. We consider a lower proportion to be an indication of a better quality of life.

3.2 Outputs

This section describes the outputs that are used in the models.

Number of general practitioners

The number of practitioners indicates the number of general practitioners per imaginary thousand inhabitants in each municipality. We consider a higher number to be a factor increasing the quality of life.

Number of pharmacies

The number of pharmacies indicates the number of pharmacies per imaginary thousand inhabitants of a given municipality. We consider a higher number to be a factor increasing the quality of life.

Number of dentists

The number of dentists indicates the number of dentists per imaginary thousand inhabitants in each municipality. We consider a higher number to be a factor increasing the quality of life.

Number of pediatricians

The number of practitioners indicates the number of general practitioners per imaginary thousand inhabitants in each municipality. We consider a higher number to be a factor increasing the quality of life.

Number of special physicians

The number of specialists indicates the number of specialists per thousand inhabitants in each municipality. We consider a higher number to be a factor increasing the quality of life. [6]

4 Results

The results from model 2 are in table 2. The municipalities with high equipped medical resources and low mortality have a score 1. An efficient score in model 2 can be interpreted as a necessary proportional increase of medical recurses to make the municipality efficient. The results from model 1 are not shown, because we don´t have enough space in this article.

Municipality	Eff. score	Municipality	Eff. score	Municipality	Eff. score	Municipality	Eff. score
Třeboň	1.000	Louny	1.059	Velké Meziříčí	1.228	Svitavy	1.423
Hořice	1.000	Náchod	1.060	Nový Bor	1.229	Varnsdorf	1.451
Mohelnice	1.000	Poděbrady	1.062	Cheb	1.230	Liberec	1.464
Chotěboř	1.000	Rokycany	1.065	Písek	1.235	Přerov	1.470
Čáslav	1.000	Žďár n. Sáz.	1.067	Uničov	1.236	Vsetín	1.475
Prachatice	1.000	Vysoké Mýto	1.068	Sušice	1.240	Žatec	1.486
Rychnov n. K.	1.000	Beroun	1.076	Choceň	1.243	Dvůr Král. n. L.	1.522
Veselí n. M.	1.000	Lanškroun	1.078	Mělník	1.252	Aš	1.539
Domažlice	1.000	Litovel	1.080	Blansko	1.266	Sokolov	1.549
Kyjov	1.000	Slaný	1.083	Plzeň	1.268	Duchcov	1.553
Vlašim	1.000	Litoměřice	1.092	Kolín	1.270	Frýdek-Místek	1.560
Holešov	1.000	Mladá Bol.	1.104	Dobříš	1.277	Litvínov	1.580
Boskovice	1.000	Jesenice	1.105	Česká Třebová	1.291	Jablonec n. Nis.	1.602
Jaroměř	1.000	Ústí n. Orlicí	1.109	Břeclav	1.293	Hlučín	1.614
Mar. Lázně	1.000	Šternberk	1.111	Vrchlabí	1.296	Ústí nad Labem	1.617
Tachov	1.000	Nová Paka	1.119	Praha	1.297	Trutnov	1.633
Roud. n. L.	1.000	Olomouc	1.123	Česká Lípa	1.310	Mnichovo Hra.	1.646
Český Kruml.	1.000	Humpolec	1.124	Roztoky	1.317	Králův Dvůr	1.657
Turnov	1.000	Bruntál	1.136	Štětí	1.322	Krnov	1.664
Říčany	1.000	Polička	1.138	Otrokovice	1.330	Moravská Třeb.	1.691
Jičín	1.000	Klatovy	1.143	Příbram	1.338	Děčín	1.700
Benešov	1.000	Hodonín	1.143	Milevsko	1.338	Karviná	1.729
Kadaň	1.000	Tábor	1.150	Kuřim	1.347	Milovice	1.742
Uherské Hrad.	1.000	Nový Jičín	1.155	Bra. n. L.-St. Bol.	1.348	Bohumín	1.761
Šumperk	1.000	Tišnov	1.160	Nov. Měst. na M.	1.357	Kladno	1.782
Znojmo	1.000	Hradec Král.	1.161	Ostrov	1.363	Červený Kost.	1.812
Třebíč	1.000	Zlín	1.162	Přelouč	1.366	Chomutov	1.814
Karlovy Vary	1.000	Uherský Brod	1.166	Opava	1.377	Ostrava	1.822
Jihlava	1.000	Kutná Hora	1.176	Český Těšín	1.377	Kopřivnice	1.835
České Bud.	1.000	Hlinsko	1.178	Hostivice	1.378	Příbor	1.850
Valašské Mez.	1.010	Chrudim	1.180	Lysá nad Labem	1.378	Havířov	1.851
Havl. Brod	1.015	Kroměříž	1.181	Brno	1.382	Jirkov	1.854
Jindř. Hrad.	1.016	Pelhřimov	1.183	Čelákovice	1.392	Neratovice	1.869
Kralu. n. Vlt.	1.031	Semily	1.188	Rumburk	1.394	Most	1.908
Rakovník	1.032	N. M. n. Met.	1.193	Teplice	1.402	Chodov	1.982
Jeseník	1.035	Frýdl. n. Ostr.	1.197	Strakonice	1.407	Studénka	1.999
Nymburk	1.044	Zábřeh	1.200	Rož. p. Radhošt.	1.408	Orlová	2.128
Hranice	1.052	Vyškov	1.209	Pardubice	1.417	Bílina	2.198
Litomyšl	1.053	Prostějov	1.219	Fren. p. Radhošt.	1.417	Klášter. n. Ohří	2.238
Lovosice	1.055	Ivančice	1.226	Třinec	1.419	Krupka	2.334

Table 2: Results

30 municipalities are efficient. Other 130 municipalities need a proportional increase of outputs to become efficient. These 30 efficient municipalities are potential “good” place for living.

Region	Number of municipalities	Average eff. Score in the region input model	Average eff. Score in the region output model
Kraj Vysočina	9	1.108	0.912
Plzeňský kraj	6	1.119	0.901
Jihočeský kraj	9	1.127	0.902
Olomoucký kraj	11	1.139	0.889
Jihomoravský kraj	12	1.169	0.867
Zlínský kraj	9	1.193	0.855
Královéhradecký kraj	12	1.233	0.844
Pardubický kraj	13	1.249	0.815
Středočeský kraj	27	1.272	0.816
Hlavní město Praha	1	1.297	0.771
Liberecký kraj	6	1.299	0.788
Karlovarský kraj	7	1.380	0.764
Ústecký kraj	20	1.553	0.689
Moravskoslezský kraj	18	1.605	0.644

Table 3: Distribution of the 160 biggest municipalities in the regions and average eff. score

The average results in the regions of the Czech Republic are in table 3. There is the average result from municipalities in each region from model 1 and model 2. The three best regions are region Vysočina, region Plzeň and region Jihočeský. The three worst regions are region Moravskozlesky, region Ústecký and region Karlovarský.

5 Conclusion

This paper focuses on measuring health conditions and medical recourses in the 160 biggest municipalities in the Czech Republic. The variations were measured by two data envelopment models. The inputs are 5 deadly diseases (diseases of the circulatory system, respiratory diseases, infectious and parasitic diseases, tumors and diseases from alcohol and drug use). The outputs are 5 medical resources (general practitioners, pharmacies, dentists, pediatricians and special physicians). 30 municipalities are efficient other 130 municipalities are inefficient. We have assigned municipalities to the 14 regions of the Czech Republic. We calculated the average efficient score for each region. The best three regions are region Vysočina, Plzeňský and Jihočeský. The worst three regions are Moravskoslezský, Ústecký and Karlovarský. The health condition of the population in a country, region or municipality is one of the important parameters for measuring life quality. Results from this study could be useful as one of the parameters for measuring life quality in region or municipalities in the Czech Republic.

Acknowledgements

This work was supported by the project no. F4/42/2021 of the Internal Grant Agency, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Dlouhý, M. & Hanousek, J. (2019). An assesment of Regional Variations: AnApplication to Polish Regions. In: *37 th International Conference on Mathematical Methods in Economics*, České Budejovice: University of South Bohemia in ČeskéBudějovice, Faculty of Economics, pp. 281-286

[2] Dlouhý, M. (2018). Measuring Geographic Inequalities: Dealing with multiple Health Resources by Data Envelopment Analysis. *Frontiers in Public health*, 6(53), pp. 1-6.

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

[4] Jablonský, J. & Dlouhý, M.(2004): *Modely hodnocení efektivnosti produkčních jednotek*. Praha: Professional Publishing, pp. 183.

[5] Kumbhakar, S. C &, Lovell, C. A. K (2000). *Stochastic frontier analysis*. Cambridge: Cambridge University Press, pp. 355.

[6] DATLAB, STEM [online] Available at https:kdejedobre.cz

Model Design for Team Roles in Agile IT Projects

Ondřej Havazík [1], Petra Pavlíčková[2], Jan Rydval [3]

Abstract. Scrum is one of the most widely used agile methodologies in project management, especially in the field of information technology. Due to its wide use, it also brings a large number of problems that can occur when using it. One such issue may be, for example, the filling of suitable candidates (employees) for individual roles in the team, whether they are leaders or management positions or technological, business, design, etc. It is necessary to estimate the importance of individual team roles. Therefore, it's crucial to find an appropriate mathematical model to determine the preferences of each role. The Analytic Network Process (ANP) is used to create a mathematical model. The individual clusters of the mathematical model are described. The preferences of individual Scrum roles are derived both based on the structure of the model and based on the subjective evaluation of project management students.

Keywords: Agile, Analytic Network Process (AHP), Scrum, Team Roles

JEL Classification: C44
AMS Classification: 90B50

1 Introduction

The agile approach to project implementation has been applied in many companies recently, not only in software development. Agile methods reflect that software development projects are often too complex to define their full scope and all requirements in the planning phase at the beginning of the project ([2], [3]). Companies are also forced to come up with new or improved products in a much shorter time than before, and agile methodologies help them to do this. Furthermore, the agile approach encourages close collaboration between the development team and the customer, adapting to change and working in rapid, iterative development cycles, making them different from the waterfall way of project management.

However, despite the many successes of introducing agile methodologies into how projects are managed, most organizations struggle to implement agile into their project management fully. While agile methods are claimed to be easy to understand, they are difficult to follow in practice ([7], [15]). One of the reasons for the lack of adherence to the basic principles of agile methodologies is, that they may not be taught sufficiently.

When teaching agile software development, it is important to consider the following factors: student collaboration in teams, communication with the customer, and a well-defined scope and role of the agile team. Based on experiments with these methods, Kropp and Meier [6] found that using agile methodologies in teaching has a positive impact on student learning outcomes. Yilmaz and O'Connor [18] introduced the gamification of different parts of the Scrum methodology in real projects for better learning of agile project management tools. This gamification requires a clear definition of the roles of the agile team, but then it has a positive impact on the work efficiency and software developers. Thus, it is clear, that teaching and practicing agile approaches improve teamwork and effectiveness, and a crucial part of this is the definition of agile team roles and estimation of their importance.

The issue of assessing the importance of project team roles is very complex. Therefore, it is necessary to represent this decision problem using a decision network where the dependence of the elements (nodes) of the same level can be assumed. Thus, to obtain higher quality results in complex decision problems, Ziemba [19] suggests that the Analytic Network Process (ANP) can be used. For decision problems containing dependencies between elements of the same hierarchy level, ANP is a more appropriate approach than AHP (Analytic Hierarchical Process) Ziemba [19]. A decision network is used to represent the decision problem. The decision network consists of clusters and nodes, representing the elements of the decision problem, and with respect to the network structure of the decision problem, the ANP is used to determine cardinal quantitative information about alternatives (in this case, project roles). The ANP model for estimating the importance of project team roles is also used and considered

[1] Czech University of Life Sciences Prague, Faculty of Economics and Management, Department of Systems Engineering, Kamýcká 129, 165 21 Praha 6 – Suchdol, havazik@pef.czu.cz
[2] Czech University of Life Sciences Prague, Faculty of Economics and Management, Department of Systems Engineering, Kamýcká 129, 165 21 Praha 6 – Suchdol, pavlickovap@pef.czu.cz
[3] Czech University of Life Sciences Prague, Faculty of Economics and Management, Department of Systems Engineering, Kamýcká 129, 165 21 Praha 6 – Suchdol, rydval@pef.czu.cz

as a suitable tool by Rydval [9]. The decision model's real dependencies between criteria and alternatives (nodes of the decision network) reflect the actual decision problem situation. Considering this, the ANP model allows more accurate results [14]. Saaty ([13], [14]) describes this problem in more detail and explains the dependencies between decision elements.

The aim of this paper is to propose a simplified decision model for assessing the importance of individual team roles in an agile-led project. The decision model is represented in the form of a decision network with defined elements affecting the project team roles.

2 Materials and methods

2.1 Agile approach

The agile family of development methods were born out of a belief that an approach more grounded in human reality – and the product development reality of learning, innovation, and change – would yield better results. Agile principles emphasize building working software that people can get hands on quickly, versus spending a lot of time writing specifications up front. Agile development focuses on cross-functional teams empowered to make decisions, versus big hierarchies and compartmentalization by function. And it focuses on rapid iteration, with continuous customer input along the way [17].

2.2 Agile team roles

Scrum team [15] consists of one Scrum Master, one Product Owner, and Team. Within a Scrum team, there are no subteams or hierarchies. The Scrum team should be a small team, typically 10 or fewer people. Scrum team should share the same Product Goal, Product Backlog, and Product Owner. The Scrum team is responsible for all product related activities from stakeholder collaboration, verification, maintenance, research and development, and anything else. The Scrum Team is accountable for creating a valuable, useful increment every Sprint. In Scrum, there are three roles: The Product Owner, The Team, and The Scrum Master. Together these are known as The Scrum Team. The Product Owner is responsible for maximizing return on investment (ROI) by identifying product features, translating these into a prioritized list, deciding which should be at the top of the list for the next Sprint, and continually re-prioritizing and refining the list. The Scrum Master helps the product group learn and apply Scrum to achieve business value. The Scrum Master is not the manager of the Team or a project manager; instead, the Scrum Master serves the Team, protects them from outside interference, and educates and guides the Product Owner and the Team in the skillful use of Scrum.

2.3 ANP

Many decision problems cannot be structured hierarchically without inner dependencies of the same-level elements. And because these decision problems may also involve many interactions and dependences of higher-level elements in a hierarchy on lower-level elements they cannot be structured into an Analytic Hierarchy Process (AHP) model ([12], [14]). The Analytic Network Process (ANP) is a generalization of the AHP. In ANP, where the decision problem is structured like a network, the structure of the problem may include dependences between the all-level elements of the decomposition structure. The ANP model can reflect the increasing complexity of a network structure, where the network can be created from different groups of elements. Each group of elements (cluster) includes a relative homogeneous set of elements. Connections can exist between clusters and between the elements i.e. between the elements inside the cluster as well as between the elements from different clusters.

In AHP for hierarchical trees, synthesized global priorities are calculated by multiplying the local priorities, which are determined via pairwise comparisons of the priority of the parent element. In ANP, this process is replaced by the Limit Matrix calculation ([13], [14]). ANP can be used in various industries where it is necessary to implement important evaluations, such as marketing, politics, military, etc. The basic steps of the ANP:

(i) Creation of a decision problem network which describes dependency among decision elements. The ANP allows the inner and outer dependence within and among a sets of elements;

(ii) Pairwise comparisons of the elements within the clusters and among the clusters ([13], [14], [10]). The inconsistency of these comparisons has to be controlled. The inconsistency is measured by the consistency ratio CR:

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

where RI represents Saaty's random index. A pairwise comparison matrix is considered inconsistent for $CR > 0.1$. And CI is consystenci index and is calculate as:

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

where l_{max} is the largest eigenvalue of Saaty's matrix and n is the number of criteria. Saaty's matrix is considered to be sufficiently consistent if I_S<0.1;

(iii) The Supermatrix construction. The priorities derived from the pairwise comparisons are inputed into Unweighted Supermatrix. This Supermatrix has to be normalized using clusters weights, the Weighted Supermatrix is calculated

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

where each block $\boldsymbol{Wij}$ of the Supermatrix consists of:

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

where:

$$\textstyle\sum_i^n w_{ij} = 1, j \in \langle 1, n \rangle, \tag{5}$$

(iv) and the last step is computation of the Limit Supermatrix and global preferences of decision elements. The Limit Matrix is used to obtain stable weights from Weighted Supermatrix. Raising the Weighted Supermatrix to powers generates the Limit Matrix, the powers will converge to a given matrix (the Limit Supermatrix), or the powers will converge to a cycle of matrices (the Limit Supermatrix is the average of these matrices). From Limit Matrix the final global preferences decision-making elements are extracted. See Adams [1] for more information about algorithms for computing the Limit Matrix.

3 Results

As mentioned above, we used the ANP to find out the preferences of the Scrum roles. This is the main problem that our model (**Figure 1**) will deal with. In order to build this model, it is necessary to select individual Clusters. Four primary clusters were chosen for this model: Team Roles, Activities, Hard Skills, and Soft Skills.

3.1 Clusters description

Cluster – Team Roles (R) - The roles include management / managerial roles, namely Scrum Master (also known as Agile coach) and Product Owner. Next will be the professional roles that have been used in the game, such as junior and senior designer and UX designer. In each Scrum team, professional roles can vary, depending on the team's focus or the type of project the team is working on. In contrast, each Scrum team must have the two management team roles mentioned above. This Cluster contains several relations to other clusters.

Cluster – Activities (A) – The cluster of activities is significant because activities are essential for successfully completing a project. However, as with Scrum's roles in the team, the activities here will vary depending on the type of project the team is working on and will also vary depending on the role in the Scrum team. Product Owner has fixed activities, for example, creating and prioritizing individual tasks in the product backlog, communication with the customer and stakeholders, communication with team members, etc. Activities such as training individual team members in self-management, supervising, or managing all Scrum events are essential to the Scrum Master.

Cluster – Hard Skills (HS) - Another significant cluster, which must not be missing in the model, is a subset of competencies related to, i.e., hard skills. These skills are called hard because one has to learn and master them through hard work and practice. A typical example of hard skills in the IT world can be: a person can program in a specific programming language, controls particular computer programs in which can be done various activities, knowledge around specific Cisco networks, ability to automate tests, work with specific technologies, etc.

Cluster – Soft Skills (SS) - The last cluster involved in the model is the second subset of competencies, which is the opposite of hard skills, namely soft skills. These skills are very important in every profession. Different skills stand out for each individual according to determine in which professional direction he could go. Typical examples of these skills include verbal and nonverbal communication, problem-solving skills, public speaking.

3.2 Relationships between clusters

Each cluster (**Figure 1**) has several relations: input, output, and one self-loop. The input relation in these subchapters will not be analyzed, but only the output relation from all clusters will be described.

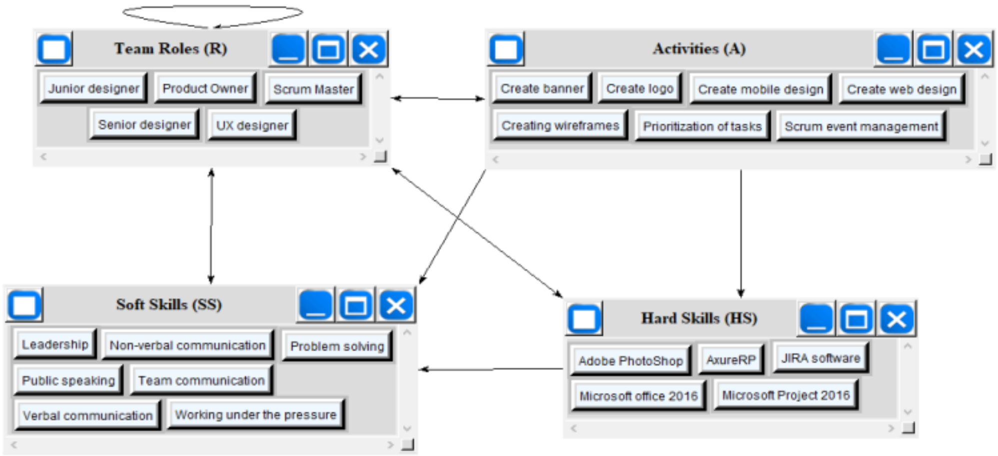


Figure 1 ANP model

Cluster – Team Roles (R) - The first output relation leads to the cluster of activities. The relation describes which activities are important for the particular role. Regarding the role, the next relations show what kind of hard and soft skills are important for particular roles. The self-loop goes back to the same role cluster. It models a situation where we have an absence of a specific competence in the team, which must be covered by someone else from the current Scrum team members according to their abilities (primarily hard and soft skills).

Cluster – Activities (A) - There are also three links from the cluster of activities, as with roles. The first leads to hard skills. The relation describes what hard skills are important for specific activities. The other connection is almost the same, but it only concerns to the soft skills. And the last link is the link back to the roles. That indicates the importance of the roles from the point of view of the activities of the Scrum team.

Cluster – Hard Skills (HS) - Compared to the previous two clusters, hard skills have only two output links to soft skills and back to roles. The first is the link to soft skills, which tells us how important they are in order to acquire hard skills. It is possible to master them from several different points of view, and it can be efficiently, improving in new ones, etc. The second link to the roles means, what role in the team is crucial to include specific hard skills.

Cluster – Soft Skills (SS) - The last cluster has relationship to the roles. This connection tell us, which role in the team is important to have the appropriate soft skills. As can be seen in the figure above (Figure 1), the individual nodes (24 nodes in total) are interconnected according to the relations between the clusters. Not every node must have relation to all nodes in a given cluster.

3.3 Pilot study

In this paper, we present a pilot study using the above-proposed decision model to determine the preferences of its elements. This pilot study is a preliminary small-scale study before the main research in defining the different agile project team roles and their importance. The pilot study is conducted to determine whether the clusters of the above-mentioned model are defined appropriately and whether respondents can understand and make pairwise comparisons between the elements. The group of respondents represents 13 students and 3 educators in the field of teaching agile approaches in project management at FEM, CZU Prague. After conducting a pilot study, the significance values of each role of the agile team and the significance values of the other elements in the decision model are obtained. It is also determined how easy or complicated it is for the respondents to complete the questionnaire and make pairwise comparisons and whether they are able to maintain consistency in their responses.

The pilot study was conducted by having each respondent complete a questionnaire consisting of 21 pairwise comparison matrices. The matrices contained varying numbers of elements (nods) for pairwise comparison (minimum 2, maximum 7). In each pairwise comparison, the respondent assigned a preference value of 1 to 9 (according to the Saaty scale) to the preferred nod (the non-preferred nod was thus given an inverted value). The respondent performed all pairwise comparisons in the questionnaire in this way, and his/her comparison was then con- verted to the SuperDecisions software (ANP is in this paper carried out by the SuperDecisions software [16]). Thus, the synthesized priorities of all nods in the ANP model were obtained from the Limit Matrix of each respondent. The overall preferences for agile team roles (**Table 1**) were then obtained by averaging the obtained respondents' results. The raw values for each element (Raw column) are drawn directly from the Limit Matrix for our network. The normalized values (N- columns) are normalized by adding the elements in individual clusters (A, R, HS, SS) and dividing each element by the sum of the cluster to yield the normalized vector. The sum of the numbers in the normalized vector is 1. Column N-A represents the values obtained from all respondents; column N-M represents the values from men; column N-M represents the values from women. According to the results, the essential roles are Scrum Master and Product Owner. Verbal communication, Team communication and, after them, Problem solving are almost equally considered the most important Soft Skills.

Decision Nod	Raw	N-A	N-M	N-W	Decision Nod	Raw	N-A	N-M	N-W
A - Create banner	0.0092	0.0767	0.0762	0.0770	HS - Adobe PhotoShop	0.0257	0.1459	0.1396	0.1511
A - Create logo	0.0092	0.0767	0.0762	0.0770	HS - AxureRP	0.0244	0.1386	0.1317	0.1443
A - Create mobile design	0.0151	0.1255	0.1262	0.1250	HS - JIRA software	0.0528	0.3001	0.2898	0.3088
A - Create web design	0.0151	0.1255	0.1262	0.1250	HS - Microsoft office	0.0448	0.2544	0.2553	0.2536
A - Creating wireframes	0.0151	0.1255	0.1262	0.1250	HS - Microsoft Project	0.0283	0.1610	0.1835	0.1422
A - Prioritization of tasks	0.0282	0.2351	0.2346	0.2355	SS - Leadership	0.0265	0.1078	0.1315	0.0911
A - Scrum event management	0.0282	0.2351	0.2346	0.2355	SS - Non-verbal communication	0.0316	0.1284	0.1412	0.1193
R - Junior designer	0.0828	0.1808	0.1787	0.1821	SS - Problem solving	0.0411	0.1673	0.1548	0.1761
R - Product Owner	0.1062	0.2318	0.2287	0.2336	SS - Public speaking	0.0231	0.0941	0.0923	0.0955
R - Scrum Master	0.1090	0.2380	0.2390	0.2374	SS - Team communication	0.0424	0.1725	0.1737	0.1717
R - Senior designer	0.0921	0.2010	0.1974	0.2032	SS - Verbal communication	0.0427	0.1736	0.1656	0.1792
R - UX designer	0.0680	0.1484	0.1562	0.1438	SS - Working under the pressure	0.0384	0.1562	0.1409	0.1671

Table 1 Overall Preferences

However, looking more closely at the results (**Figure 2**), it is interesting that the preferences of men and women were essentially the same within the Activities cluster (A) and Team Roles cluster (R), while not within the Soft Skills cluster (SS). For example, when assessing the soft skill of leadership, women placed 15% less importance on this soft skill than the average, while men placed 20% more importance on it.

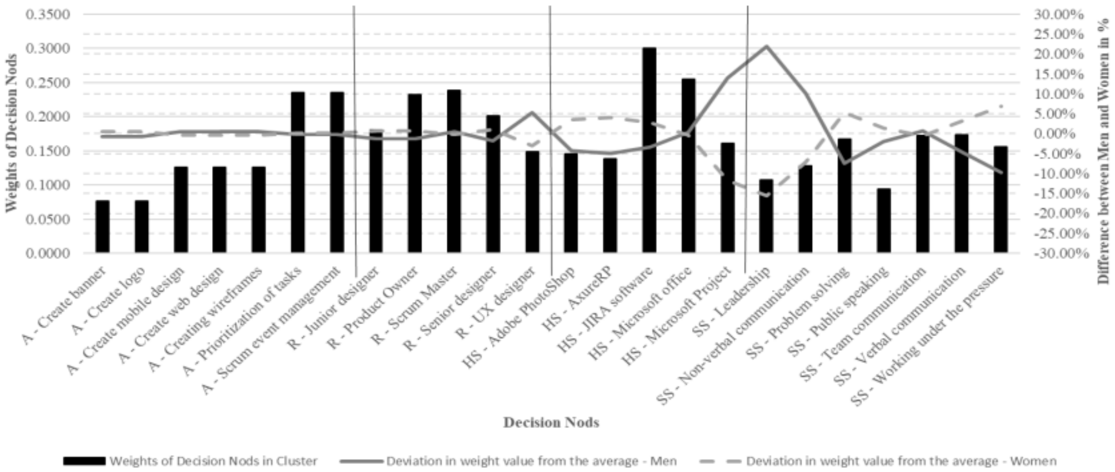


Figure 2 Weights of Decisions Nods

Saaty decided the threshold for CR of 0.10. If CR is greater than this threshold, then it questions the credibility of the decision maker's judgements. To maintain consistency, it is possible that the decision-maker re-evaluates his comparisons. In our pilot study, more than half of the respondents did not maintain consistency in their pairwise comparisons in our case study. Therefore, we went through their responses again with them and adjusted their responses that had a CR > 0.2 at their option to preserve consistency. But this is not always possible. Saaty [11] suggested that for the matrices of order 3 and 4, the CR thresholds can be taken as 0.5.

4 Discussion

The higher complexity of the definition and evaluation of project team roles' preferences shows that the application of the ANP is appropriate. Ziemba [19] pointed out the higher quality of ANP results. For the decision problems containing the dependence between the elements of the same hierarchy level, is ANP a more suitable approach than AHP. Furthermore, consideration of the real dependencies between criteria in the decision model makes the model precisely reflect the real decision problem, and it allows to obtain more precise results ([4], [8]). We found in our pilot study that the clusters of our model were defined appropriately, and respondents understood and made pairwise comparisons between the elements. But it was very problematic to maintain consistency in the decision-maker's judgements. When the decision-maker was asked to make pairwise decisions in matrices of order 3, 4 or larger, decisions were often inconsistent. It is not always expected of the decision-maker to re-evaluate his comparisons; consistency must be managed in another way. Saaty [11] suggested for the matrices of order 3 and 4 the thresholds for CR as 0.5. A new approach to ensuring consistency introduce Hlavatý and Brožová [5]. The approach is based on a nonlinear optimisation model. They show how to adjust the matrix values to maintain the original information contained in the pairwise comparison with achieving an acceptable inconsistency.

5 Conclusion

We proposed a simplified model for defining and obtaining the importance of individual roles in the agile team. Our decision model is represented by the ANP decision network. The decision model consists of four clusters (A, R, HS, SS) with corresponding decision nodes. In a pilot study, we have demonstrated the suitability of the design of both clusters and decision nodes. The selected respondents were able to perform a pairwise assessment of each element of the decision ANP network. However, guaranteeing the consistency of their decisions was problematic. Inconsistent results were treated by reevaluating with the appropriate respondents. In the future, we will conduct a major investigation of the relevance of the different roles of the agile team and treat consistency differently.

Acknowledgements

This paper was supported by IGA of the FEM, Czech University of Life Sciences in Prague, grant no. 2019A0008.

References

[1] Adams, B. (2011). SuperDecisions Limit Matrix Calculations. *Decision Lens Inc*., DeLand FL, USA.
[2] Boehm, B., Grunbacher, P. & Briggs, R.O. (2001) Developing groupware for requirements negotiation: lessons learned. In: *IEEE Software*, 18(3). pp. 46–55.
[3] Campbell, J., Kurkovsky, S., Liew, Ch.W. & Tafliovich, A. (2016) Scrum and Agile Methods in Software Engineering Courses. In: *47th ACM Technical Symposium on Computing Science Education*, TN, USA.
[4] Golcuk, I. & Baykasoglu, A. (2016). An analysis of DEMATEL approaches for criteria interaction handling within ANP. *Expert Syst. Appl*, 46, 346–366.
[5] Hlavatý, R. & Brožová, H. (2022) Fixing Inconsistent Saaty's Matrix using Optimisation Model. In: *Proceedings of the 40th International Conference Mathematical Methods in Economics*, Jihlava: College of Polytechnics Jihlava. (in print)
[6] Kropp, M. & Meier, A. (2013) Teaching agile software development at university level: Values, management, and craftsmanship, *Software Engineering Education Conference*, Proceedings, pp. 179–188. doi: 10.1109/CSEET.2013.6595249.
[7] López-Martínez, J., Juárez-Ramírez, R., Huertas, C., Jiménez, S. & Guerra-García, C. (2016) Problems in the adoption of agile-scrum methodologies: A systematic literature review. In: *4th International conference in software engineering research and innovation*, doi:10.1109/conisoft.2016.30.
[8] Mandic, K., Bobar, V. & Delibasi´c, B. (2015). Modeling Interactions Among Criteria in MCDM Methods: *A Review. Lect. Notes Bus. Inf. Process*, 216, 98–109.
[9] Rydval, J., Bartoška, J. & Jedlanová, T. (2019). Sensitivity Analysis of Priorities of Project Team Roles Using the ANP Model. In: *Proceedings of the 37th International Conference Mathematical Methods in Economics*, České Budějovice: University of South Bohemia. ISBN 978-80-7394-760-6
[10] Saaty, T.L (2008). *Relative measurement and its generalisation in decision making*. RACSAM, 102(2).
[11] Saaty, T.L. (1980) *The Analytic Hierarchy Process*. McGraw-Hill, New York.
[12] Saaty, T.L. (2000). *Fundamentals of the Analytic Hierarchy Process*. RWS Publications, Pittsburgh, 2000.

[13] Saaty, T.L. (2001). *Decision Making with Dependence and Feedback*: *The Analytic Network Process, The Analytic Hierarchy Process Series*. Pittsburgh: IX, RWS Publications.

[14] Saaty, T.L. (2003). *The Analytic Hierarchy Process (AHP) for Decision Making and the Analytic Network Process (ANP) for Decision Making with Dependence and Feedback,* Creative Decisions Foundation.

[15] Schwaber, K. (2004) *Agile Project Management with scrum*. Microsoft Press.

[16] SuperDecisions Software for Decision-Making. (2018). URL <http://www.superdecisions.com/>.

[17] Sutherland, J. (2022) *The Scrum papers*, http://jeffsutherland.com/scrumpapers.pdf

[18] Yilmaz, M. & O'Connor, R.V. (2016) A scrumban integrated gamification approach to guide software process improvement: A Turkish case study, *Tehnicki Vjesnik*, 23(1), pp. 237-245. https://doi.org/10.17559/TV- 20140922220409

[19] Ziemba, P. (2019). Inter-Criteria Dependencies-Based Decision Support in the Sustainable wind Energy Management. *Energies 2019*, 12, 749; doi:10.3390/en12040749.

Is There Any Dependence Between a Football Club's Financial Health and Its First League Performance?

Jana Heckenbergerová[1], Irena Honková[2]

Abstract. The presented contribution aims to analyze the financial health of Czech football clubs. Financial results of selected first league football clubs are evaluated using the summary indicator IN05. Satisfactory results are achieved only by three clubs out of twelve, one club corresponds with the so-called grey zone and the remaining clubs are in the bankruptcy zone. Furthermore, the stated hypothesis verifies whether there is no correlation between club performance and its financial health. At the first sight, it seems logical that the league's performance depends on the financial situation, and one can assume that the stated hypothesis has to be rejected. Nevertheless, provided statistical analysis utilizing the Spearman's correlation coefficient proves the opposite.

Keywords: Football club, Financial health, Index IN05, Correlation coefficient

JEL Classification: C12, G32
AMS Classification: 62G10, 62P20, 91B24

1 Introduction

We often ask ourselves, whether the financial results of companies reflect their performance. Is there any direct dependence of accounting data on performance? At first sight, it seems that yes as it should not be possible to survive with permanently negative economic results. Companies shall necessarily reach a situation in which they are not able to cover their operational activity with capital. On the contrary, a company with good financial results is sufficiently strong in the capital so that it can increase its performance as it is not limited by insufficient sources.

This contribution deals with the relationship between performance and financial health. It is focused on Czech football clubs in the first league, doing business in CZ NACE 9312 Sport club activities. With obtained league points, they are suitable for investigation of mentioned dependence as there is a direct measurement of the performance quality. In most other sectors quality of performance is hard to measure.

1.1 Literature Review

The football club economics has been mapped at the turn of the twentieth century. Vamplew in [27] presents the economics of Scottish football clubs in 1890-1914. Dobson & Goddard in [7], [8] have analyzed the performance, revenue, and cross subsidization in the Football League in 1927-1994. They believed that the unusual nature of the product of sporting leagues, and the effort for competitive balance on the field of play, provide an economic justification for redistributing financial resources between football clubs. They examined the relation- ship between changes in the distribution of playing. Access between different types of clubs and changes in gate revenues and attendances, and considers how flows of revenues within the football industry have been modified by various type of cross subsidy. The authors are relative optimistic about the potential for professional football to remain viable financially at all levels within the existing league structure.

In the last decade, several authors have dealt with the economics of football clubs. Demil & Lecocq in [6] designed a dynamic model for their financial stability as they believe that the environment of this sport is also dynamic. Their business model concept generally revers to the articulation between different areas of a firm´s activity designed to produce a proposition of value to customers. Two different uses of the term were noted. The first is the static approach – as a blueprint for the coherence between core business model components. The second was a more transformational approach using the concept as a tool to address change and innovation in the organization. They built the framework to try to reconcile these two approaches to consider business model evolution, looking particularly at the dynamic created by interactions between its business model´s components.

They illustrated the framework with the case of the English football club Arsenal FC over the decade. They viewed business model evolution as a fine tuning process involving voluntary and emergent changes in and between

[1] University of Pardubice, Faculty of Economics and Administration, Studentská 95, 532 10 Pardubice, jana.heckenbergerova@upce.cz
[2] University of Pardubice, Faculty of Economics and Administration, Studentská 95, 532 10 Pardubice, irena.honkova@upce.cz

permanently linked core components, and find that firm sustainability depends on anticipating and reacting to sequences of voluntary and emerging change, giving the label „dynamic consistency" to this firm capability to build and sustain its performance while changing its business model. Dobson, in his book named „The economics of football"[9], states economic issues and specificities of English football clubs. He presents a detailed economic analysis of professional football at club level. He reflects the development of the economics of professional football over the past ten years using a combination of economic reasoning and statistical and econometric analysis. Topics covered include some of the most hotly debated issues currently surrounding professional football, including player salaries, the effects of management on team performance, betting on football, racial discrimination and the performance of football referees. Dabscheck in [5] presented a collection of analytical contributions by internationally regarded scholars in the field, which extensively examine the many economic challenges facing the world´s most popular team sport. Littlewood in [20] explains the football financing, whereas Szymanski in [25] deals with football club bankruptcies. He states that football clubs go bankrupt following a series of negative shocks. Tabakovic & Wollmann in [26] deal with the money effect on the football of university environment.

It results from the above that many authors deal with football finances but do not address the effectiveness. This can be divided in a technical and a financial effectiveness. The technical effectiveness is focused on the match victory [14] and deals with factors that affect it, for example the pass strategy [10]. The financial effectiveness is addressed by Miragaia et al in [21] who states that in monitored period 2009-2014 only 10 professional clubs of 15 were financial effective. Skrinjaric & Barisic in [24] dealt with the effect of success/failure of the Croatian national team on the Croatian stock exchange in 2014-2018 and found that there was no dependence.

The correlation between victories of the specific team Juventus with shares traded on a stock exchange and its shares was examined by Botoc et al in [4]. Their positive effect was found. In the same way, the study of Godinho & Cerqueira [13] confirms for most of clubs with shares traded on a stock exchange that the victory has positive effect on their shares price. Floros states in his research [11] that from four analysed football clubs two clubs showed the positive effect, one club showed the negative effect and one club the neutral effect. Bell et al in [2] states that the match importance matters and Scholtens & Peenstra in [23] state that the stock exchange is rather affected by losses than by victories. Aglietta et al in [1] state that losses/victories are not related to the stock exchange, a random is rather concerned. Leach & Szymanski in [19] state that when the football clubs are traded in the stock exchange, they are usually focused on profit.

The research of the effect of financial results on the position in national or international football competitions is focused only on the dependence of the stock exchange index on this position. Football clubs that do not trade their stock on the public stock exchange have not been analyzed in this way yet. Due to this reason, our contribution deals with the positions of the Czech football clubs in the 1st league with dependence on their financial health. The hypothesis that the position of Czech football clubs in the basic part of the 1st league does not depend on their financial health is verified.

2 Methodology

Results from the basic part of 1st league 2019/2020 were found from the source Fortuna Liga [12]. In addition, accounting data for 2019 were found in the Magnus Web database [3].

The ROE indicator and synthetic indicators (IN99, IN05) were considered as financial health evaluation methods. According to [18], ROE indicator was proven to be not too much reliable predictive indicator. This study demonstrated no statistically significant linkage between the value of bankruptcy model and the prosperity of the business in the following year expressed by ROE values.

The accuracy of synthetic model ranges from 45-95%. In the Czech environment, accuracy of the IN05 indicator overcomes others commonly used synthetic indicators [17]. For this reason, the IN05 index was chosen to define the financial situation of analyzed football clubs. It is defined as:

$$IN05 = 0.13 * \frac{Assets}{Liabilities} + 0.04 * \frac{EBIT}{Paid\ interest} + \\ + 3.97 * \frac{EBIT}{Assets} + 0.21 * \frac{Revenues}{Assets} + 0.09 * \frac{Current\ assets}{Short\text{-}term\ liabilities} \tag{1}$$

The IN05 index shows the highest success rate for middle companies bankruptcy risk identification, for other companies the mention index is overall successful. The IN05 index has two limits: 0.90 as the lower limit and 1.60 as the upper limit. Based on the survey [22], companies with the IN05 index below the minimum value do not create any value and they have 97% chance to reach their bankruptcy. Companies above the upper limit have 92% chance not to bankrupt. Companies of which the IN05 index is found in the so-called grey zone 0.90 – 1.60 have similar chances of bankruptcy or not; nevertheless, the mentioned classification is not suitable for the predictions.

Relation between two datasets can be identified using correlation tests. Pearson's correlation coefficient (R) shows the power of linear dependence and requires normality of input datasets. Monotony can be verified by Spearman test. The Spearman´s coefficient (Rs) can be defined as the regular Pearson correlation coefficient in term of the proportion of variability accounted for, except that Spearman Rs is computed from ranks. It assumes that the variables under consideration were measured at least an ordinal (rank order) scale and the individual observations (cases) can be ranked into two ordered series (p_i, q_i).

$$Rs = 1 - \frac{6\sum_i^n (p_i - q_i)^2}{n(n^2-1)} \tag{2}$$

The Spearman´s coefficient is dimensionless number and as it does not require any information about distribution, it is part of non-parametric statistics. [28]

Both coefficients range from -1 to 1. Value -1 shows decreasing dependence and value 1 shows increasing dependence. If correlation coefficient is equal to 0 then there is no statistically detectable linear dependency between variables. It should be noted that variables can be dependent even if correlation coefficient is equal to 0. [15] Critical values for Pearson and Spearman test are summarized in the Statistical tables [16].

3 Results and Discussion

The number of clubs in the 1st league in 2019/2020, in its basic part, was sixteen. Due to the fact that four clubs have not published their final accounts for given period, only twelve selected clubs were considered, see the table 1.

Position	Football club
1.	Slavia
2.	Plzeň
3.	Sparta
4.	Jablonec
5.	Liberec
6.	Baník
7.	Slovácko
8.	Bohemians
9.	Mladá Boleslav
10.	Olomouc
11.	Teplice
12.	Zlín

Table 1 Position of football clubs after basic part of the 1st league in the season 2019/2020

Source: own processing from [12]

For selected football clubs, the summary index IN05 was evaluated using (1). Results are summarized in following Table 2. When the value of the IN05 index exceeds the limit of 1.6, the company is financially healthy. This is valid only in three cases – Sparta, Plzeň and Mladá Boleslav. On the contrary, IN05 values under 0.9 indicate serious financial issues and the bankruptcy zone. Surprisingly, most of the selected football clubs have financial troubles. Therefore we recommend that deeper financial analyses of such sports clubs should be performed. Generally, it seems that the football clubs are over-financed by foreign capital and their own capital is negative. There must be financial sources that are permanently lending to sports clubs. But their economical motivation is questionable as these clubs are not able to pay back the borrowed capital.

Football club	IN05 index	Position
Sparta	1.691232	3.
Plzeň	1.665086	2.
Mladá Boleslav	1.664807	9.
Liberec	1.161018	5.
Olomouc	0.476984	10.
Slavia	-1.51992	1.
Jablonec	-1.86418	4.
Baník	-2.27616	6.
Zlín	-2.49145	12.
Teplice	-4.87937	11.
Bohemians	-9.79435	8.
Slovácko	-11.4299	7.

Table 2 IN05 index values for selected clubs Source: own processing from [12]

Let us state our hypothesis,

H0: There is no correlation between IN05 index and league position.

The Pearson´s coefficient R = -0.29281 and its critical value Rcrit = 0.576 (n=12, α=0.05). As |R| < Rcrit H0 cannot be rejected on significance level 0.05. Therefore, this test confirms that there is no linear dependence between the position of Czech football clubs in the basic part of the 1st league in 2019/2020 and its financial health.

The value of Spearman´s coefficient was calculated and Rs = - 0.48252. Tabular critical value for n=12 and α=0.05 corresponds to Rscrit = 0.5804. This results in the inequality |Rs| < Rscrit and therefore H0 cannot be rejected again. Spearman's test shows that there is no monotony (ranked dependence) between the bankruptcy risk and 1st league position.

Values of both Pearson´s and Spearman's coefficients are negative. This shows a nonsignificant decreasing effect between analyzed variables. It can be explained in a natural expected way, football clubs with higher financial health perform better in 1st league than clubs with financial troubles.

As mentioned above in the literature review, most of the published studies are focused on the success rate of football clubs in competitions that positively affect the value of their market value through the growth of their share price. This study aims at smaller clubs, namely first-league football clubs in the Czech Republic, where the effect of success on the share price growth cannot be tested. Nevertheless, it can be analogically assumed that there is a dependence of success on the summary financial results. In our case, football club financial health is expressed by the IN05 synthetic indicator.

From performed financial analysis using the IN05 summary index, we can conclude that only three clubs of twelve were above the limit of 1.6 indicating their financial health. It is important to remark that rather than above the limit, they were directly on it. The remaining football clubs showed absolutely alarming financial results and it is strange how they can survive from a long-term point of view.

Performed statistical tests surprisingly confirm that there is no significant dependence between football clubs' financial situation and their 1st league results. However, negative values of evaluated correlation coefficients show that a higher resulting position is connected with a lower summary index value. This confirms the natural behavior of the Czech football environment.

4 Conclusion

The company's financial health should be the keystone of its effectiveness. Football clubs same as other companies have been established in order to generate profit and the same economic principles should be valid for them as for others. Even football clubs need to have a positive cash flow for activities and proper financing. Surprisingly, performed financial analyses of Czech first-league football clubs show that most of them are over- financed by foreign capital. Their summary index IN05 indicates serious financial troubles and high bankruptcy chances. Only three clubs (Sparta, Plzeň and Mladá Boleslav) of the selected twelve have satisfactory financial results. In addition, another surprising feature of the Czech football environment has been found. There is neither significant correlation nor ranked correlation between the IN05 index and club 1st league position. Only insignificant negative dependence between a football club's financial health and its first league performance has been found in the season 2019/2020.

Acknowledgements

This contribution has been supported by the institutional support of the University of Pardubice, Czech Republic.

References

[1] Aglietta, M., Andreff, W. & Drut, B. (2018). Stock market and football. *Revue d economie politique*, 118(2), 255-296.

[2] Bell, AR., Brooks, C., Matthews, D. & Sutcliffe, C. (2012) Over the moon or sick as a parrot? The effects of football results on a club´s share price. *Applied Economics*, 44 (26), 3435-3452.

[3] Bisnode. MagnusWeb [software]. Bisnode Česká republika, a. s. [2020-08-30] Available from https:// magnusweb.bisnode.cz

[4] Botoc, C., Mihancea, E. & Molcut, A. (2019). Football and Stock Market Performance Correlation: Evidence from Italy. *Scientific Annals of Economics and Business*, 66 (4), 525-539.

[5] Dabscheck, B. (2015). Handbook on the Economics of Professional Football. *Economic Record*, 91 (295), 539-542.

[6] Demil, B. & Lecocq, X. (2010). Business Model Evolution: In Search of Dynamic Consistency. *Long Range Planning*, 43(2-3), 227-246.

[7] Dobson, SM. & Goddard, JB. (1998). Performance and revenue in professional league football: evidence from Granger causality tests. *Applied Economics*, 30 (12), 1641-1651.

[8] [8] Dobson, SM. & Goddard, JB. (1998). Performance, revenue, and cross subsidization in the Football League, 1927-1994. *Economic History Review*, 51 (4), 763-785.

[9] Dobson, SM. & Goddard, JB. (2011). *The Economics of football*, ISBN: 9780521517140

[10] Emara, N., Owens, D., Smith, J. & Wilmer, L. (2017). Serial correlation in National Football League play calling and its effects on outcomes. *Journal of Behavioral and Experimental Economics*, 69, 125-132.

[11] Floros, C. (2014) Football and Stock Returns: New Evidence. In *International Conference on Applied Economics (ICOAE 2014)* (201-209). Greece: Mediterranean Agron Inst Chania, Chania.

[12] Fortuna Liga. 1. Liga 2019/2020*: Základní část*. [2022-05-10] Available from: https://www.sport.cz/fotbal/ceska-1-liga/#section-list-top

[13] Godinho, P. & Cerqueira, P. (2018). The Impact of Expectations, Match Importance, and Results in the Stock Prices of European Football Teams. *Journal of Sports Economics*, 19 (2), 230-278.

[14] Keskin, HI. & Ondes, H. (2020). Measuring the Effeciency of Selected European Football Clubs: DEA and Panel Tobit Model. *Sosyoekonomi*, 28(43), 153-174.

[15] Kozel, M., Mynářová, L. & Svobodová, H. (2011*). Moderní metody a techniky marketingového výzkumu*. Praha: Grada Publishing, a. s.

[16] Kubanová, J. & Linda, B. (2007) *Kritické hodnoty a kvantily vybraných rozdělení pravděpodobnosti*, Pardubice, ISBN: 80-7194-852-7

[17] Kuběnka, M. & Myšková, R. (2019). Obvious and Hidden Features of Corporate Default in Bankruptcy Models. *Journal of Business Economics and Management*, 20(2), 368-383.

[18] Kuběnka, M. & Slavíček, O. (2016). Relationship between level of prosperity and failure prediction. In *Political Sciences, Law, Finance, Economics and Tourism Conference Proceedings* (505-512). Sofie: STEF92 Technology Ltd. ISBN 978-619-7105-74-2

[19] Leach, S. & Szymanski, S. (2015). Making Money out of football. Scottish *Journal of Political Economy*, 62(1), 25-50.

[20] Littlewood, M. (2015). Money and Football: A Soccernomics Guide. *Economic Affairs*, 35 (3), 497-498.

[21] Miragaia, D., Ferreira, J., Carvalho, A. & Ratten, V. (2019). Interactions between financial efficiency and sports performance Data for a sustainable entrepreneurial approach of European professional football clubs. *Journal of Entrepreneurship and Public Policy*, 8(1), 84-102.

[22] Neumaierová, I. & Neumaier, I. (2005). Index IN 05. In *Evropské finanční systémy: Sborník příspěvků z mezinárodní vědecké conference* (143-146). Brno: Masarykova univerzita v Brně.

[23] Scholtens, B. & Peenstra, W. (2009). Scoring on the stock exchange? The effect of Football Matches on stock market returns: an event study. *Applied Economics*, 41 (5), 3231-3237.

[24] Skrinjaric, T. & Barisic, P. (2019). Effects of Football Match Results of Croatian National Team on Stock Returns: Evidence from Zagreb Stock Exchange. *Zagreb International Review of Economics & Business*, 22 (1), 13-45.
[25] Szymanski, S. (2017,). Entry into Exit: Insolvency in English Professional Football. *Scottish Journal of Political Economy*, 64 (4), 419-444.
[26] Tabakovic, H. & Wollmann, TG. (2019). The impact of money on science: Evidence from unexpected NCAA football outcomes. *Journal of Public Economics*, 178.
[27] Vamplew, W. (1982). The Economics of a Sports Industry – Scottish Gate – Money Football, 1890-1914. *Economic History Review*, 35 (4), 549-567.
[28] Walker, I. (2013). *Výzkumné metody a statistika*. Praha: Grada Publishing, a. s.

Optimisation Approach to Dealing with Saaty's Inconsistency

Robert Hlavatý[1], Helena Brožová[2]

Abstract. We revisit the issue of the AHP method, wherein it often happens that an evaluation of a matrix based on Saaty's scale is inconsistent. The inconsistency of the matrix originates from ill-defined pairwise comparisons provided by a decision-maker. The measurement of the inconsistency is very well known. The less discussed issue is fixing the original data to achieve the required consistency. It is not expected from the decision-maker to re-evaluate the comparisons because: 1) the decision-maker is not available anymore, 2) the decision-maker is not capable of making such adjustments that would lead to inconsistency improvement. We propose our own approach based on a nonlinear optimisation model. We show how to adjust the matrix values to preserve the original information contained in the pairwise comparison while achieving acceptable inconsistency. We compare our approach with the earlier ones in the end.

Keywords: AHP, inconsistency, nonlinear optimization model, Saaty's matrix

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

1 Introduction

The Analytic hierarchy process (AHP) developed by Saaty [9] has been a standard in quantitative decision-making. The AHP is based on pairwise comparisons done by a decision-maker who seeks to determine the relative importance of criteria in a given decision-making problem. It was shown by Saaty [10]that one of the drawbacks of the AHP may be the inconsistency of pairwise comparisons. A decision-maker is required to provide a degree of preference between two criteria using a given scale. When the number of pairwise comparisons is too large, it is a common phenomenon that the transitivity of preference among criteria is not satisfied to a required level. The cause of such misevaluating originates in the way the people build preferences. Slovic [12] shows the background of preference construction and how the preferences can vary depending on the decision-maker's approach. It is no wonder that if a decision-maker uses common sense or intuition for preference evaluation, the inconsistency of the preferences can appear unintentionally. Consequently, the resulting weights of criteria originate in ill-defined preferences, and the entire process becomes invalid and can lead to wrong conclusions.

It is practically impossible to achieve perfect consistency or transitivity of the pairwise comparisons, especially for a large number of criteria [8]. There are different means of measuring inconsistency, each with its own thresholds, for which the inconsistency of the pairwise comparisons is still acceptable. Blankmeyer [1] discusses the differences between Saaty's eigenvalue approach and the least-square approach. Bozóki and Rapcsák [2] present their overview of different inconsistency measures and acceptable inconsistency levels. This issue is further developed by Ramík and Vlach [8], who generalise the inconsistency concepts and compare them. Our contribution will focus on Saaty's original approach to inconsistency measuring. This approach is presented further in chapter 2.1. Our motivation to study inconsistency problems draws from practical issues that we encounter in our university teaching practice. Students often get in touch with the AHP in their theses. They have also learned to properly test if the pairwise comparison matrices (PCM) are inconsistent to an acceptable level, thus justifying the validity of their calculation. The problem arises when a PCM is found inconsistent, and the natural question is how to deal with this issue properly. We intended to develop a methodology tool that would be accessible to students and help them overcome this issue.

The literature offers different approaches to dealing with the inconsistency of PCM. We have tackled this issue in our previous work [5] to show how a particular value of PCM should be changed in order to make the PCM consistent provided that there exists such a feasible change. The approach was based on deriving the exact formulae for PCM based on the characteristic equation of PCM. However, this approach requires choosing a particular pairwise comparison that should be modified to change PCM to acceptable inconsistency. This approach can be

[1] CZU Prague, Department of Systems Engineering, hlavaty@pef.czu.cz.
[2] CZU Prague, Department of Systems Engineering, brozova@pef.czu.cz.

helpful when a decision-maker is asked to provide reevaluation because our method [5] computes feasible inconsistency bounds on each pairwise comparison in PCM. In reality, though, the usual problem is that the decision-maker cannot be reached anymore to provide a new evaluation. Then it leaves one with a question about which pairwise comparison out of all should be changed. This is why we consider our approach useful for PCM with possibly missing data. The issue is also solved by Shiraishi, Obata and Daigo [11] or Ramík [7] using crisp or fuzzy approaches to the problem, respectively. In our current contribution, we set the following assumptions for a given inconsistent PCM:

1. the PCM contains all pairwise comparisons provided by a decision-maker
2. the decision-maker is not available for the PCM reevaluation, OR the decision-maker is unable to do so due to his inability to identify where the inconsistency originates from
3. all pairwise comparisons in PCM are of equal importance, and none of these pairwise comparisons is seen as ill-defined

We further justify these assumptions; 1. The assumption is natural when a decision-maker is asked to state his preference, and there is no reason for not providing a complete evaluation; 2. This is often seen in practice. A decision-maker (as a stakeholder) is asked to provide his preferences, but the further calculations of AHP are carried out by someone else. Even if the decision-maker is available for reevaluation, it is impossible for him to decide which values should be fixed to achieve acceptable inconsistency; 3. All pairwise comparisons are made naturally with a decision-maker's goodwill. Even if mutually inconsistent, it is impossible to tell which pairwise comparisons should be redefined. This is why we aim to find an algorithm which numerically changes the original PCM without any judgment, such that most of the original information is preserved. Several approaches have been developed for this sake. Xu and Wei [13] propose an algorithm that gradually improves original PCM values by a given formula until the acceptable consistency is reached. Cao, Leung and Law [3] propose a heuristic that performs better than the earlier approach under given criteria. Later, Zhang et al. [14] propose another algorithm outperforming the earlier two under the same criteria. Their algorithm adopts a segment tree to gradually approach the greatest lower bound of the distance with the original PCM. In the following sections, we overview the necessary basis for measuring inconsistency, propose a new approach to fix an inconsistent matrix, show a numerical example, and compare it with the other approaches. Finally, we discuss how well our method performs compared to the others under the same criteria and present our own additional criterion of the result quality.

2 Methods

In this chapter, we summarise the necessary basics of PCM and inconsistency measuring. Because the inconsistency involves calculating matrix eigenvalues, we make a short note about the process. Next, we introduce our methodology of PCM adjustments in order to make it consistent.

2.1 Pairwise comparison matrix and inconsistency

The AHP is a broader concept which allows making complex decisions by structuring it into smaller parts. This process involves determining the relative importance of criteria or alternatives. According to Saaty [9], this is done using a pairwise comparison matrix (PCM). Evaluation carried out by PCM is sometimes referred to as Saaty's method. The principle of the method is to express preferences between criteria or alternatives using a given scale. The evaluation should be carried out by a decision-maker who acts as a stakeholder in the given problem. We assume there is only one decision-maker for our purposes here. The usual scale proposed by Saaty uses preference values from 1 (equal importance) to 9 (strong preference) and their reciprocal values, thus ranging in $\left[\frac{1}{9}, 9\right]$. The PCM $A \in \mathbb{R}^{nxn}$ is defined as follows:

$$A = \begin{pmatrix} 1 & a_{12} & \dots & a_{1n} \\ a_{21} & 1 & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & 1 \end{pmatrix} \tag{1}$$

for which it holds true that $a_{ij} = {}^{1}/_{a_{ji}}, i, j = 1,2, \dots, n, a_{ij} \in \left[\frac{1}{9}, 9\right]$. Once evaluated, the matrix is used to determine the relative importance vector $\boldsymbol{w}^{\boldsymbol{T}} = (w_1, \dots, w_n)$ of criteria $1, \dots, n$ by *Eigenvalue method*, which solution is given by

$$\boldsymbol{w} = \lim_{k\to\infty}\left(\frac{A^k e}{e^T A^k e}\right) \tag{2}$$

Where $e^T = (1,1,\dots,1)$. The vector $\boldsymbol{w}$ is the desired outcome of the process and acts as a basis for further decisions for whichever decision-making problem. According to Saaty's approach [10], the inconsistency of PCM A is measured by the consistency ratio CR given as

$$CR = \frac{CI}{RI_n} \tag{3}$$

where RI_n represents random index. A given PCM is considered inconsistent for $CR > 0.1$, according to Saaty [10]. The values of RI_n differs for particular n and was proposed by Saaty [10] based on simulating random PCMs. The CI stands for consistency index and is calculated as

$$CI = \frac{\lambda_{\max} - n}{n-1} = \frac{\rho(A) - n}{n-1} \tag{4}$$

and λ_{max} is the maximum eigenvalue of A, and $\rho(A)$ is the spectral radius of A . Equation (4) holds true by the Perron-Frobenius theorem since A is strictly positive and irreducible being a PCM.

2.2 Spectral radius calculation issues

We make a short note on the practicality of spectral radius calculation in relation to our methodology. Calculation of spectral radius is non-trivial, and various methods can be used. Numerical methods for calculation of $\rho(A)$ are known to have cubic complexity and $\alpha n^3, \alpha > 0$ operations is asymptotically needed to calculate the spectrum of a matrix [4]. In the practical sense, these methods cannot be performed by hand calculation and neither there is an explicit formula to obtain $\rho(A)$. As a first-choice approach, one can consider using Matlab (if in possession) or various online tools such as WolframAlpha. Apart from such automated tools, there is a way to calculate $\rho(A)$ using the MS Excel solver add-on. It is enough to solve the following equation

$$\det(A - \lambda I) = 0 \tag{5}$$

for a given A, I is the identity matrix, and λ is the variable. This equation is true for the entire spectrum of a matrix. To find the specific eigenvalue $\lambda = \rho(A)$ it is necessary to use a gradient approach starting with a sufficiently large λ. It is assured by the nature of the gradient approach that $\lambda = \rho(A)$ will be found by solving equation (5). The starting value of λ for the gradient algorithm can be determined by the upper bound of $\rho(A)$ which is for any PCM easily expressed as

$$\rho(A) \leq \max_i \sum_j a_{ij} \tag{6}$$

It is also possible to give an estimation of the value of $\rho(A)$ by using a chosen matrix norm $\|\cdot\|$

$$\rho(A) \leq \|A^k\|^{\frac{1}{k}} \tag{7}$$

for some sufficiently large $k \in \mathbb{Z}$. From our experience, this may not work properly for larger k in MS Excel due to rounding errors.

2.3 Fixing inconsistent PCM

In this section, we introduce our approach to fixing inconsistent PCM. Assume there is an inconsistent pairwise comparison matrix A such that $CR(A) > 0.1$. We introduce the algorithm that produces an adjusted matrix denoted as B: $b_{ij} \in \mathbb{R}^{nxn}$ such that $CR(B) \leq 0.1$. This can be achieved only if some $a_{ij} \in A, i \neq j$ are adjusted. We propose two different optimisation models, M1 and M2, based on a similar methodology but with slightly different objectives. The goal of both optimisation models is to change the original matrix as least as possible while achieving acceptable inconsistency.

Model M1

In the first case, we propose an assumption that a relative difference between original elements $a_{ij}, i < j$ and adjusted elements b_{ij}, $i < j$ is expressed by a set of deviations d_{ij}, $i < j$ in the following manner

$$\frac{|b_{ij} - a_{ij}|}{a_{ij}} \leq d_{ij}, i < j \tag{8}$$

Note that this process is only done for $i < j$ as we only need to change the upper triangular matrix without the diagonal. Inequality (8) states that an absolute difference between original and adjusted elements $|b_{ij} - a_{ij}|$ relative to the original value a_{ij} would be at most d_{ij}. To overcome the issue of using modulus, the term (8) can be rewritten as

$$\frac{b_{ij}-a_{ij}}{a_{ij}} \le d_{ij}, i < j \wedge \frac{b_{ij}-a_{ij}}{a_{ij}} \ge -d_{ij}, i < j \tag{9}$$

This way, there is a possibly different deviation d_{ij} for each adjustment. In the first model, we seek to minimise the average deviation from the original data. The optimisation model is following

$$\min_{i<j} \frac{\sum d_{ij}}{(n^2-n)/2} \tag{10}$$

s.t.

$$\frac{b_{ij}-a_{ij}}{a_{ij}} \le d_{ij}, i < j, \frac{b_{ij}-a_{ij}}{a_{ij}} \ge -d_{ij}, i < j \tag{11}$$

$$\det(B - \rho(B)I) = 0 \tag{12}$$

$$\frac{\rho(B) - n}{(n-1).RI_n} \le 0.1 \tag{13}$$

$$\frac{1}{9} \le b_{ij} \le 9 \tag{14}$$

The model minimises the average deviation (10), which is expressed by (11). We seek such a constellation of all b_{ij} such that the determinant (12) produces such a spectral radius $\rho(B)$ that makes the consistency ratio acceptable (13). All adjusted values $b_{ij}, i < j$ must remain within the Saaty's scale extrema. The number $\rho(B)$ can act as a variable or be calculated a priori to satisfy (13) as an equation.

Model M2

In the second case, we slightly change our approach from the model M1. Instead of setting individual deviations d_{ij}, we introduce a single deviation variable D. We impose the equivalent constraints on the problem:

$$\frac{b_{ij}-a_{ij}}{a_{ij}} \le D, i < j, \frac{b_{ij}-a_{ij}}{a_{ij}} \ge -D, i < j \tag{15}$$

The D then acts as the maximum possible deviation between any pair of original elements $a_{ij}, i < j$ and adjusted elements b_{ij}, $i < j$. The optimisation model M2 is following

$$\min (D) \tag{16}$$

s.t.

$$\frac{b_{ij}-a_{ij}}{a_{ij}} \le D, i < j, \frac{b_{ij}-a_{ij}}{a_{ij}} \ge -D, i < j \tag{17}$$

$$\det(B - \rho(B)I) = 0 \tag{18}$$

$$\frac{\rho(B) - n}{(n-1).RI_n} \le 0.1 \tag{19}$$

$$\frac{1}{9} \le b_{ij} \le 9 \tag{20}$$

The model minimises the maximum deviation D to produce a set of $b_{ij}, i < j$ such that the adjusted matrix B is consistent. The interpretation of the constraints (17-20) matches model M1.

3 Results

As an illustration of our approach, we provide an example of inconsistent PCM, and we compute adjusted PCMs using Model 1 and Model 2, respectively. We adopt the same testing PCM as Xu and Wei [13], Cao, Leung and Law [3] and Zhang et al. [14] to be able to compare our results with their methodologies. The testing PCM is following:

$$A = \begin{pmatrix} 1 & 5 & 3 & 7 & 6 & 6 & 1/3 & 1/4 \\ 1/5 & 1 & 1/3 & 5 & 3 & 3 & 1/5 & 1/7 \\ 1/3 & 3 & 1 & 6 & 3 & 4 & 6 & 1/5 \\ 1/7 & 1/5 & 1/6 & 1 & 1/3 & 1/4 & 1/7 & 1/8 \\ 1/6 & 1/3 & 1/3 & 3 & 1 & 1/2 & 1/5 & 1/6 \\ 1/6 & 1/3 & 1/4 & 4 & 2 & 1 & 1/5 & 1/6 \\ 3 & 5 & 1/6 & 7 & 5 & 5 & 1 & 1/2 \\ 4 & 7 & 5 & 8 & 6 & 6 & 2 & 1 \end{pmatrix} \tag{21}$$

The matrix generates the weight vector (rounded) $\boldsymbol{w}^T = (0.17, 0.05, 0.19, 0.02, 0.03, 0.04, 0.17, 0.33)$ and $CR(A) = 0.17$. The matrix is considered inconsistent according to the value of CR. Applying our model M1, the following adjusted matrix is computed:

$$B_{M1} = \begin{pmatrix} 1 & 5 & 3 & 7 & 6 & 6 & 1/3 & 1/4 \\ 1/5 & 1 & 1/3 & 5 & 3 & 3 & 1/5 & 1/7 \\ 1/3 & 3 & 1 & 6 & 3 & 4 & \mathbf{1.67} & 1/5 \\ 1/7 & 1/5 & 1/6 & 1 & 1/3 & 1/4 & 1/7 & 1/8 \\ 1/6 & 1/3 & 1/3 & 3 & 1 & 1/2 & 1/5 & 1/6 \\ 1/6 & 1/3 & 1/4 & 4 & 2 & 1 & 1/5 & 1/6 \\ 3 & 5 & \mathbf{0.59} & 7 & 5 & 5 & 1 & 1/2 \\ 4 & 7 & 5 & 8 & 6 & 6 & 2 & 1 \end{pmatrix} \tag{22}$$

Note that A and B_{M1} only differs in one element, that is $a_{37} \neq b_{37}$ and consequently $a_{73} \neq b_{73}$. All other preferences have been preserved and $CR(B_{M1}) = 0.1$. If we wanted to achieve even better consistency, i.e. $CR \ll 0.1$, then as a consequence, the adjusted matrix (22) would have to be even farther from the original matrix (21). Since our goal is to preserve as much as possible of the original information, we only achieve the inconsistency threshold itself. This adjusted matrix (22) produces the weight vector (rounded) $\boldsymbol{w}_{M1}^T = (0.18, 0.06, 0.13, 0.02, 0.03, 0.04, 0.2, 0.35)$. The vector $\boldsymbol{w}_{M1}^T$ differs from the original $\boldsymbol{w}^T$ in all values but in five values, it differs with some significance (hundredths), as shown in red. Next, we apply model M2 and compute the adjusted matrix:

$$B_{M2} = \begin{pmatrix} 1 & 4.042 & 2.425 & 8.341 & 4.851 & 4.851 & 0.397 & 0.298 \\ 0.247 & 1 & 0.359 & 4.042 & 2.425 & 2.425 & 0.238 & 0.165 \\ 0.412 & 2.785 & 1 & 7.149 & 3.575 & 4.035 & 4.851 & 0.238 \\ 0.120 & 0.247 & 0.140 & 1 & 0.397 & 0.298 & 0.137 & 0.111 \\ 0.206 & 0.412 & 0.280 & 2.518 & 1 & 0.596 & 0.238 & 0.135 \\ 0.206 & 0.412 & 0.248 & 3.357 & 1.679 & 1 & 0.238 & 0.135 \\ 2.518 & 4.196 & 0.206 & 7.309 & 4.196 & 4.196 & 1 & 0.404 \\ 3.357 & 6.050 & 4.196 & 9 & 7.421 & 7.421 & 2.474 & 1 \end{pmatrix} \tag{23}$$

The adjusted matrix B_{M2} differs from the original in all values. The adjusted matrix B_{M2} generates the weight vector (rounded) $\boldsymbol{w}_{M2}^T = (0.17, 0.06, 0.18, 0.02, 0.03, 0.04, 0.16, 0.35)$. Once again, all values are actually different from the original. However, if rounded to two decimals, four values change with some significance.

4 Discussion

We have applied two different models to compute adjusted matrices B_{M1} and B_{M2}. The main difference between these matrices is how they differ from the original matrix A. The B_{M1} only adjusts one value (and the reciprocal as a consequence) but with a rather significant change in the preference scale (6 to 1.67). The preference direction is preserved but not as strong as the original statement. The B_{M2} on the other hand, adjusts all values of the original but not with such significant changes. All values of B_{M2} preserve the original preference direction of A, and at the same time, the largest difference is $\max_{i,j}\{|a_{ij} - b_{M2\_ij}|\} = 1.42$. The principal question is how to measure the quality of an adjustment to be able to compare different approaches. We adopt the metrics δ and σ of Xu and Wei [13], and we also introduce another metric $\bar{\delta}$ to compare the approaches:

$$\delta = \max_{i,j}\{|a_{ij} - b_{ij}|\} \;, \qquad \sigma = \frac{\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}(a_{ij} - b_{ij})^2}}{n}, \qquad \bar{\delta} = \frac{\sum_{i \neq j}|a_{ij} - b_{ij}|}{n^2 - n} \tag{24}$$

The authors claim that acceptable thresholds are $0 < \delta < 2$ and $0 < \sigma < 1$. These thresholds are originally based on Ma [6], who justifies this as the necessity of adjustment being at most one degree on the Saaty's scale

(1,3,5,7,9). Although we accept these measures as plausible, we do not see too much reason for such thresholds as we consider the main aim to preserve the original preference direction above all. Our metric $\bar{\delta}$ expresses the average deviation when the diagonal elements are not counted in, and we do not set any threshold for it and use it only for comparison. In Table 1, we compare the different methodologies by three different metrics.

	Xu and Wei [13]	**Cao, Leung and Law [3]**	**Zhang et al. [14]**	**Model M1**	**Model M2**
$\boldsymbol{\delta}$	10.465	11.031	1.157	4.329	1.421
$\boldsymbol{\sigma}$	1.907	2.071	0.471	0.544	0.567
$\bar{\boldsymbol{\delta}}$	1.132	1.233	0.321	0.09	0.411

Table 1 Comparison of different methods according to δ, σ and $\bar{\delta}$

The data in Table 1 show that both of our approaches outperform in all indicators Xu and Wei [13] and Cao, Leung and Law [3]. We do not match the model of Zhang et al. [14] in the original metrics, but we consider these a bit questionable. In the average deviation, our model M1 shows the best characteristics. Another criterion, which is not quantitative but possibly significant, is the user-related complexity of approaches. We dare say that skilled users can easily use our approach performing just one computation in MS Excel, unlike the earlier approaches.

5 Conclusion

We have introduced our way of adjusting for an inconsistent Saaty's matrix. We compared our approach with the other literature using different metrics. Our approach performs better or worse depending on the chosen metrics. Our further intention is to do computational experiments to reveal the average quality of our approach and possibly develop changes in our approach to achieve even better results.

Acknowledgements

This paper was supported by Czech Science Foundation (GAČR), project No. P403-22-11117S.

References

[1] Blankmeyer, E.(1987). Approaches to consistency adjustment. *Journal of Optimization Theory and Applications*, 54(3), 479-488.

[2] Bozóki, S., & Rapcsák, T. (2008). On Saaty's and Koczkodaj's inconsistencies of pairwise comparison matrix. *Journal of Global Optimization*, 42, 157–175.

[3] Cao, D., Leung, L.C. & Law, J.S. (2008). Modifying inconsistent comparison matrix in analytic hierarchy process: A heuristic approach. *Decision Support Systems*, (44), 944-953.

[4] Hladík, M. (2019). *Lineární algebra (nejen) pro informatiky*. Praha: Matfyzpress.

[5] Hlavatý, R. (2014). Saaty's matrix revisited: Securing the consistency of pairwise comparisons. In J.Talašová, J.Stoklasa & T.Talášek (Eds.), *Proceedings of 32nd International Conference Mathematical Meth- ods in Economics (MME 2014)* (pp.287-292). Olomouc: Palacký University.

[6] Ma, W. (1994). An approach to modifying pairwise comparison matrices and two criteria of modificatory effectiveness. *Systems Engineering Theory and Practice*, 12, 51-55.

[7] Ramík, J. (2014). Incomplete fuzzy preference matrix and its application to ranking of alternatives. International *Journal of Intelligent Systems*, 29, 787-806.

[8] Ramík, J. & Vlach, M. (2013). Measuring consistency and inconsistency of pair comparison systems. *Kybernetika*, 49, 465-486.

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

[10] Saaty, T.L (2008). Relative measurement and its generalisation in decision making. *RACSAM*, 102(2), 251–318.

[11] Shiraishi, S., Obata, T. & Daigo, M. (1998). Properties of a positive reciprocal matrix and their application to AHP. *Journal of the operations research*, 41, 404–414.

[12] Slovic, P. (1995). The construction of preference. *American Psychologist*, 50(5), 364–371.

[13] Xu, Z. & Wei, C. (1999) A consistency improving method in the analytic hierarchy process. *European Journal of Operational Research*, 116, 443-449.

Zhang, H., Sekhari, A., Ouzrout, Y. & Bouras, A. (2014). Optimal inconsistency repairing of pairwise comparison matrices using integrated linear programming and eigenvector methods. *Mathematical Problems in Engineering*, 2014.

Modeling the Influence of Opinion Leaders in E-commerce Networks

Jiří Homan[1], Ladislav Beránek[2 *], Radim Remeš[3]

Abstract. The development of Internet technologies and Web 2.0 in recent years, also in connection with the current pandemic, has stimulated the growth of e-commerce. Increased attention is also paid to business applications, strategies, and user behavior. Users in the e-commerce environment have access to share their experiences. They have access to the knowledge and understanding of other users. In this environment, opinions and decisions can profoundly affect each other.
In this paper, we examine the interaction mechanism of a group of users (autonomous agents) in an e-commerce social network. We focus mainly on the power of opinion leaders in forming the opinion of other users. Through simulations, we investigate the mechanism of influencing users' opinions by influential opinion leaders. The simulation results show that it is crucial to increase the integrity and credibility of opinion leaders. It's the only option how to improve the influence of various e-commerce applications and business models,

Keywords: Opinion dynamics, e-commerce, network, simulation

JEL Classification: C63
AMS Classification: 90B10

1 Introduction

In recent years, the transition to Internet technologies has accelerated in connection with the epidemic of recent years. At the same time, the interest in buying goods through various online stores has accelerated. The number of online stores has increased, and competition has intensified. Therefore, more and more attention is being paid to examining user behavior [17]. Users can exchange their experiences orally, through social networks, and other online tools in their shopping behavior. Users can express here their opinions and requirements. The specific features of communication in the online environment are as follows: the group effect, opinions, and demands spread rapidly, and the development of opinions and opinions can be complicated and have different outcomes. It is clear that in such a cooperative exchange of experiences and ideas, users interact. [3]. Opinion dynamics as a scientific discipline describes the process of forming opinions in user groups. A social network can represent such a group in which nodes and links symbolize users and their interactions. Users keep their opinions up to date, which can lead to either consensus or disagreement at the final stage.

Various studies on collective behavior have shown that several members, called leaders, can strongly influence group decision-making and action. These influential users can initiate new actions (publish their views) that other group members, called followers, easily follow. [9, 12]. Scientific studies suggest that leadership is essential in human coordination [11]. Leadership can occur due to temperament, dominance, or knowledge [11] and can benefit the group. These leaders often greatly influence the opinions of other members of a group. It is also interesting that views spread faster in a social network with opinion leaders than in a social network without these leaders [2]. Therefore, various authors discuss ways to influence and lead emerging views to reach a consensus or a given consensual value. In e-commerce, according to the authors [5], opinion leaders can directly influence the decision-making results of consumers, who often consider the opinion of other influential users. According to the authors [16], opinion leaders can advise other consumers on purchasing decisions, influencing their attitudes, beliefs, and behaviors. In their paper, the authors [8] try to understand the impact of opinion leaders on opinion followers. They analyze the mechanism of opinion interaction between opinion leaders and opinion followers.

People are much more willing to accept or influence others when differences of opinion are small. Conversely, when their differences of opinion are relatively significant, they may not be ready to change. And the degree of convergence depends on the difference of opinion and other parameters, which may rely on users' characteristics

[1] University of South Bohemia/Faculty of Economics, Studentska 13, Ceske Budejovice, homanj00@ef.jcu.cz.
[2] * University of South Bohemia/Faculty of Economics, Studentska 13, Ceske Budejovice, beranek@ef.jcu.cz.
[3] University of South Bohemia/Faculty of Economics, Studentska 13, Ceske Budejovice, inrem@ef.jcu.cz.

and opinion leaders. However, our results show that introducing a degree of convergence for different users could affect the likelihood of consensus and the opinion configuration in the population.

The rest of this document is arranged as follows. The following is a brief overview of the literature. In Part 3, we present a modified Deffuant model [7], including a heterogeneous degree of convergence. The main results are described in section 4. We study a modified Deffuant model with different population sizes using numerical simulations. Conclusions and discussions are in section 5.

2 Literature background

Models of opinion and decision dynamics focus on the mechanism of interaction between opinions. They are based on the assumption that users will decide based on the opinions of their neighbors in the network. Therefore, a model based on the study of social media is suitable for studying the mechanism of dissemination of opinion. Many popular models of opinion dynamics have been introduced using various evolutionary rules, such as the DeGroot model [6], the voter model [13], the majority rule [7], the Friedkin and Johnsen model [13], the limited confidence model [10], and opinion dynamics on social networks [7]. Authors Afshar and Asadpour [1] extended the traditional Deffuant - Weisbuch model and proposed a model in which agents gradually change their views through deliberate interactions. Other models include the Weisbuch-Deffuant model [14], which assumes that differences in attitudes between two individuals will have a direct effect. The Hegselmann-Krause model [15] is similar to the Deffuant model. Their common feature is that only nodes with similar views can communicate. The difference is that individuals in the Weisbuch-Deffuant model always communicate with only one neighbor, while in the Hegselmann-Krause model, individuals communicate with more than one individual. It is assumed that network users communicate with each other within a specific range of trust of individual users. It should be noted that various other models can be used to examine the dynamics of opinions [4]. However, little consideration is given to individual heterogeneity and dynamic conformity, and qualitative analysis is the primary method. This article proposes a modified model for the actions of opinion leaders and dynamic conformity.

3 Methods

In the e-commerce environment, the assumption is usually accepted that opinion values can be considered continuous rather than discrete. In addition, because potential consumers do not generally know each other, they tend to trust opinions that are similar to their own but do not trust opinions that are very different. Dynamics models of opinion, such as the Deffuant model and the Hegselmann-Krause model, are therefore applicable because they define opinion as a continuous value instead of a discrete value and further include some limited model reliability. That is, they have the parameter δ as a threshold. In these models, opinions with differences above δ cannot influence each other according to these models.

As for the update mechanism, the Deffuant model demonstrates a process based on compromise. In each step, agent i is randomly selected to communicate with neighboring agent j (also chosen randomly). Under certain conditions given by Equation 1, these two agents update their position [7]:

$$\begin{cases} x_i(t+1) = \begin{cases} x_i(t) + \gamma\big(x_j(t) - x_i(t)\big), & if \ \big|x_i(t) - x_j(t)\big| < \delta \\ x_i(t), & otherwise, \end{cases} \\ x_j(t+1) = \begin{cases} x_j(t) + \gamma\big(x_i(t) - x_j(t)\big), & if \ \big|x_j(t) - x_i(t)\big| < \delta \\ x_j(t), & otherwise, \end{cases} \end{cases} \tag{1}$$

The real number $x_i(t)$ represents the user's opinion at each discrete time t.

As another model, we present the Hegselmann – Krause model. In this model, people can respond to everyone else's opinions and update their opinions accordingly (averaged over all other opinions). The equation for updating the opinion in the Hegselmann – Krause model is as follows [15]:

$$x_i(t+1) = \frac{\sum_{j:|x_i(t)-x_j(t)|<\mu} w_{ij} x_j(t)}{\sum_{j:|x_i(t)-x_j(t)|<\mu} w_{ij}} \tag{2}$$

where w_{ij} is the corresponding value of the weights of the neighborhood matrix of user i, it can be seen that the agents are entirely dependent on the opinions of others when updating their own opinion.

In our model, we consider a network consisting of n users. We consider a situation where there are opinion leaders in the network. These leaders (labeled l) have an opinion that they can change, and this action affects the acceptance of followers (index i). The mathematical description is based on the described previous models. The values expressing the opinions of $x(t)$ followers and leaders are updated according to the relation:

$$\begin{cases} x_i(t+1) = x_i(t) + \dfrac{1}{|\Omega_i(t)|} \displaystyle\sum_{j\in\Omega_i(t)} \gamma_{ij}(t)\left(x_j(t) - x_i(t)\right), & if\ |x_i(t) - x_{j\neq l}(t)| < \delta\ and\ i \neq l \\ x_l(t+1) = x_l(t) + u(t) \end{cases} \tag{3}$$

where $\Omega_i(t)$ is the set of agents' neighbors at time t. $|\Omega_i(t)|$ is the number of neighbors of the user i at time t, $u(t)$ is a specific change in the opinion of the leader l.

The convergence parameter $\gamma_{ij}(t)$ represents the degree of interacting factors converging their views towards each other. In the models of previous authors [15], the degree of convergence is a particular constant value in the range [0, 1/2]. However, this would indicate that users' opinions are rapidly converging, even though they have significant differences of opinion. Two interacting users (unless one is an opinion leader) with a high opinion difference $S_{ij}(t)$ just don't like to adapt their opinions. On the other hand, if $S_{ij}(t)$ is low, it is not a problem for the user to accept the other's views. To take this into account, we express the convergence parameter $\gamma_{ij}(t)$ in our model as follows:

$$\begin{cases} \gamma_{ij\neq l}(t) = \dfrac{1}{1 + e^{k_j S_{ij}(t)}} \\ \gamma_{il}(t) = \dfrac{1}{1 + e^{k_l S_{ij}(t)}} \end{cases} \tag{4}$$

The convergence parameter can therefore be individual for each pair of users and is also different for opinion leaders. This parameter is the degree of trust in another user (leader) or the degree of influence.

As part of the simulations, we consider a system consisting of N users on a random graph, where users can communicate with each other. Each user i has their opinion xi(t) ∈ [0, 1] in each time step t. In addition, opinion leaders are randomly distributed in the network. These leaders have a specific opinion, which is fixed, or changes according to the function u(t) (see equation 3). Initially, each agent (including leaders) is assigned an opinion randomly selected in the interval [0, 1].

One elementary generation of opinion dynamics can be described as the following steps.

(i) The agent i is chosen randomly. Then we go through the neighbors j of user i, and their difference of opinion $S_{ij}(t) = |\, x_i(t) - x_j(t)|$ is calculated.

(ii) If their difference of opinion between users exceeds certain reliability, i.e., if $S_{ij}(t) > \delta$, they will leave their opinion unchanged. However, suppose an opinion leader appears among the neighbors of the user i. In that case, the difference between the opinion of the leader l and the user-follower i may be more significant than the constraint given by the reliability of the neighbors δ.

iii) User's opinion is calculated according to Eq. (3).

The above steps are repeated until a steady-state is reached.

The Monte Carlo method is used for the simulation. Each step of this method will have N/2 generations. The steps described by the dynamics of opinion dynamics will be repeated until a steady-state is reached. Based on our model, a computer simulation method is adopted, which examines the strength of the influence of opinion leaders and the development of collective opinions. The simulation result will be clusters of users with "identical opinions." We chose the following assumptions for the experiments: the proportion of opinion leaders will vary in different simulations. For the initial setup, we chose a network of 200 nodes with one negative opinion leader and one positive opinion leader (Fig. 2). We calculated the degrees of trust (individual convergence parameter) of opinion leaders towards positive and negative opinion leaders in this order for $k_l = 10$, and the levels among followers' views are 2. Without losing generality, we change the share of positive leaders to monitor the impact of these changes on the development of collective opinions and the influence on the power of opinion leaders. For steady states, we observe four quantities: the number of n_c clusters and the normalized size of the largest c_l, defined as the fraction of agents in the most extensive set.

4 Simulation and analysis results

Initially, we observed the convergence parameter, which depends on the opinion difference $S_{ij}(t)$ between users i and j (resp. i and l), and the tunable parameter k, which characterizes the agent's sensitivity to the opinion difference.

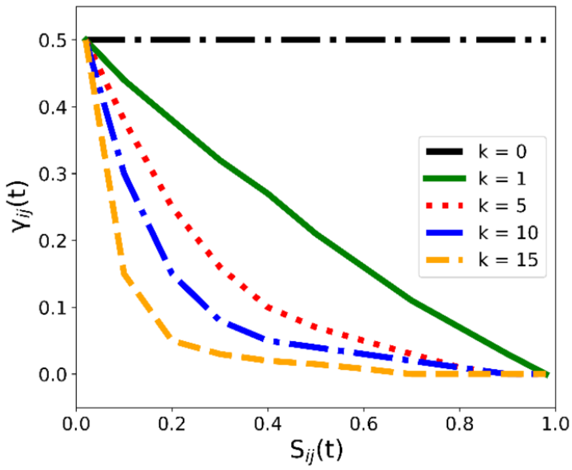
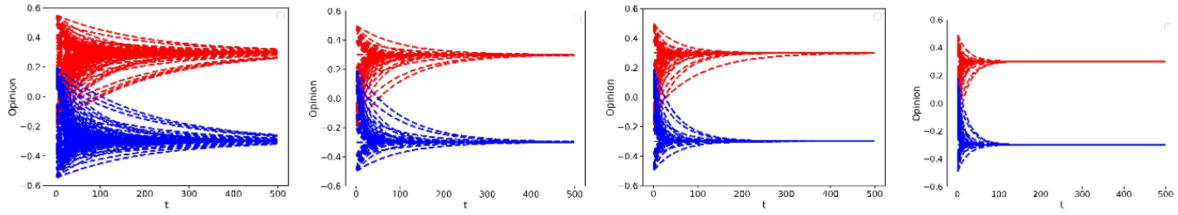


Figure 1 The convergence rate $\gamma_{ij}(t)$

Fig. 1 shows $\gamma_{ij}(t)$ as a function of $S_{ij}(t)$ for different k. To ensure $\gamma_{ij}(t) \in [0, 1/2]$, it is required that $k \geqslant 0$ (or k_i, where $i = j, l$, see equation 4). As shown in Fig. 1, when $k = 0$, the degree of convergence is a constant value, and the model degenerates to the original Deffuant model with $\gamma = 1/2$. For non-zero k, both a large difference of opinion and a large k will slow down the convergence of opinions between the interacting agents. In this Fig. 1, we do not distinguish between followers and opinion leaders. In simulations, we use low k_l values for opinion leaders.

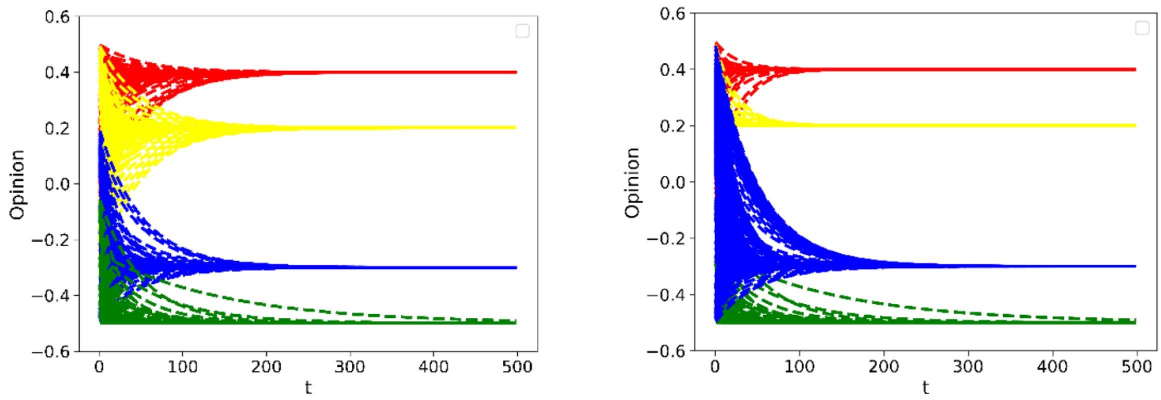
The results of simulations of the development of opinion dynamics are shown in Figure 2. The simulations were performed for a relatively small population of $N = 200$. Initially, we chose the same numbers of positive and negative opinion leaders. The followers in this picture have the same degree of confidence (convergence parameter) as positive and negative opinion leaders. Figure 2 shows how the range of confidence levels of opinion followers decreases. At the same time, the level of confidence increases, and both positive and negative opinion leaders quickly reach their target opinions. In addition, Figure 2 shows the development of the following followers' opinions. First, the speed of convergence of opinion leaders is accelerating as their level of confidence increases. Second, regardless of whether the final opinions of the followers are divided into one of three clusters, these clusters are distributed symmetrically within the opinion interval; no apparent bias against any of the subgroups of opinion leaders is observed. Fourth, increasing self-confidence cannot significantly improve the power of influence.



a) b) c) d)

Figure 2 Time evolution of opinions with N = 200. (a) $k_i = 5$, $\delta = 0.05$, (b) $k_i = 5$, $\delta = 0.10$, (c) $k_i = 5$, $\delta = 0.25$, and (d) $k_i = 5$, $\delta = 0.50$.

The results also suggest that in a network with the same confidence level in positive and negative opinion leaders, the confidence level of opinion followers contributes to the more significant influence of opinion leaders. However, when the confidence level of opinion followers exceeds a certain level, the strength of opinion leaders may not increase. In this situation, the views of opinion followers are not entirely dominated by any of the groups of opinion leaders. When the confidence level of opinion followers is high enough, they seem to have more difficulty deciding between the views of positive and negative leaders. As a result, the influence of positive and negative opinion leaders compensates each other.



a) b)

Figure 3 Evolution of the collective opinions with heterogeneous confidence levels. (a) $k_i = 5$, $k_l \in [1, 5]$, $\delta = 0.25$; (b) $k_i \in [1, 5]$, $k_l = 5$, $\delta = 0.25$.

To examine the impact of the level of confidence of opinion followers on the influence of opinion leaders, provided that the two subgroups of leaders are not well aligned, we changed the convergence parameter of opinion followers towards η_{ij} leaders via the k_l parameter (Figure 3). Then we repeat the process of the previous simulation with the other parameters without change. Thus, opinion leaders have varying degrees of confidence in the opinion leader. From Fig. 3, four types of opinion configurations can be found, for $k_i = 5$ and $k_l \in [1, 5]$. In Fig. 3b), there are different values of k_i. In steady states, several clusters of opinions are formed, which are so-called fragmentations. It turns out that the parameters concerning users-followers have a decisive influence on forming clusters of opinions around the opinion leaders. Due to the herd mentality, more agents tend to leave the opinion leaders they initially followed and join the largest opinion group. As can be seen, the results of the simulation for finite reliability show that the distribution of opinions is evolving towards clusters of certain opinions. This clustering is a convergence process in which opinions within a single group will converge indefinitely to a consensual state. On the one hand, dynamics always end with gathering opinions in clusters. At a steady-state, the agents in the various clusters are separated and will no longer exchange views.

5 Conclusion and discussion

In this paper, we use the Deffuant model, which we modified by introducing an individually adjustable parameter *k* for the degree of convergence. In our model, the degree of convergence depends on this individually adjustable parameter k and the difference of opinion of the two interacting agents within the limited reliability. This parameter *k* may differ in our model for various user-follower leader pairs (k_i or k_l). When two agents interact, a larger *k* and a more significant difference of opinion lead to a slower rate of convergence of the interacting agents. It means that an appropriate reduction in the degree of convergence in each interaction is most favorable for the consensus. In addition, there is an optimal limited level of confidence k_{ij}, which leads to a local maximum probability of reaching complete agreement when k is not too large. The simulation results also show that instead of increasing the relaxation time of the dynamics, introducing a heterogeneous convergence rate can reduce the relaxation time when the reliability is limited to a certain extent.

The results of our paper may be useful for quantitative analysis of social group collective decision-making in e-commerce networks. Our results show that the degree of the opinion of users-followers has a more critical effect than the influence of opinion leaders in opinion clusters formation. Therefore, strengthening the credibility of opinion leaders is a crucial prerequisite for maximizing the impact of e-commerce promotion.

Acknowledgements

This study was supported by project grant no. IGS02/2021, Faculty of Economics, University of South Bohemia, Czech Republic.

References

[1] Afshar, M. & Asadpour, M. (2010). Opinion formation by informed agents. *J. Artif. Soc. Social Simul*., 13 (4), 1–16.

[2] Altafini, C. (2012). Dynamics of opinion-forming in structurally balanced social networks, *PLoS One*, 7, e38135.

[3] Balazs, J.A. & Velasquez, J.D. (2016). Opinion mining and information fusion: a survey. *Inf. Fusion*, 27 95–110 .

[4] Beranek, L., Tlusty, P. & Remes, R. (2010). The Usage of Belief Functions for an Online Auction Reputation Model. In: M. Houda, J. Friebelova (Eds.) *Proceedings of the 28th Conference on Mathematical Methods in Economy 2010* (pp. 49–54). Faculty of Economics, University of South Bohemia, Ceske Budejovice, Czech Republic.

[5] Chaudhry, S. & Irshad, W. (2013). Opinion leadership and its role in buyer decision making, *Acad. Contem. Res. J*., 7(1), 7–14.

[6] DeGroot, M. (1974). Reaching a consensus. *Journal of the American Statistical Association*, 69(345), 118–121.

[7] Hegselmann, R., Krause, U. et al. (2002). Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of artificial societies and social simulation*, 5(3), 2-34.

[8] Huang, Ch., Dai, Q., Han, W., Feng, Y., Cheng, H., Li, H. (2018). Effects of heterogeneous convergence rate on consensus in opinion dynamics. *Physica A: Statistical Mechanics and its Applications*, 499, 2018, 428-435.

[9] Katz, E. (1957). The two-step flow of communication: An up-to-date report on a hypothesis. *Public Opin. Q*., 21(1), 61–78.

[10] Mirtabatabaei, A. & Bullo, F. (2012). Opinion dynamics in heterogeneous networks: Convergence conjectures and theorems. *SIAM Journal on Control and Optimization*. 50(5), 2763–2785.

[11] Nowak, A., Szamrej, J. & Latane, B. (1990). From private attitude to public opinion: a dynamic theory of social impact. *Psychol. Rev*. 97, 362–376.

[12] Solomon, M.R. (2014). *Consumer behavior: buying, having and being*. Boston: Prentice-Hall.

[13] Su, W., Chen, G. & Yu, Y. (2017). Finite-time elimination of disagreement of opinion dynamics via covert noise. *IET Control Theory & Applications*, 12(4), 563-570.

[14] Tejavibulya, P. & Eiamkanchanalai, P. (2011). The impacts of opinion leaders towards purchase decision engineering under different types of product involvement. *Complex. Syst. Eng. Manage*. 2, 12–22.

[15] Turcotte, J., York, C., Irving, J. et al. (2015). News recommendations from social media opinion leaders: Effects on media trust and information seeking. *J. Comput.-Med. Commun*., 20(5), 520–535.

[16] Villanueva, J., Yoo, S. et al. (2008). The Impact of marketing-induced versus word-of-mouth customer acquisition on customer equity growth. *J. Market. Res*., 45(1), 48–59.

[17] Yu, S. (2012). The dynamic competitive recommendation algorithm in social network services. *Inf. Sci*., 187, 1–14.

Mathematical Model for Air Carrier Irregularity Operation Management Under Conditions of Partial Uncertainty

Jiří Hořínka[1], Dušan Teichmann[2], Lenka Kontriková[3]

Abstract. The business model of air transport operators (air carriers) operating scheduled passenger transport is based on periodically recurring passenger transport on selected routes. As in any real process, so-called operational irregularities occur in the case of planned operation. Some of them may affect air operations to the extent that it will not be possible to operate all scheduled flights. In such cases, the air carrier's operational control centre staff must decide which flights will be operated and which will not be served. The article aims to present a mathematical model that can help decision-making staff of the operational control department, who have the solution of these irregularities in their competence. The article will present a mathematical model for selecting flights maximizing the number of passengers carried, which will partially reflect the conditions of uncertainty. The input quantities representing the maximum time shifts of the connections will be burdened with uncertainty. Fuzzy numbers and fuzzy sets will be used to model uncertainty. Computational experiments with the proposed model will be performed on model data in the optimization software Xpress-IVE.

Keywords: Mathematical model, Optimization, Airline transport operation, Carried passengers, uncertainty

JEL Classification: C44
AMS Classification: 90C70

1 Motivation and analysis of the current state of affairs

The current state of the airline operational control is characterized by the so-called predictive operation control or predictive steering. Despite the best planning of all activities essential for the realization of flights according to the flight schedule, it is a common reality that the realization of the flight schedule does not proceed exactly according to plan. These unexpected changes are called irregularities, abbreviated IRROP. Furthermore, in addition to unexpected changes in the values of factors affecting the plan, these values are subject to uncertainty. The presented article shows one of the possible approaches to work with uncertainty in the case of operational control steering.

2 Analysis of the current state of affairs in the field of operation control during irregularity operations

Several works devoted to increasing the efficiency of air carriers in the establishment of IRROP (Irregularity operations) have been found in the literature. A global view of the IRROP solution is provided by the publication [1]. Collaborative decision making (CDM) is considered in the publication to be the main tool for solving IRROP. Publication [2] provides partial insights into the issue of minimization of impacts. The work deals with the issue of minimizing the total tactical costs caused by departures delays at the airport. Computational experiments were performed at Frankfurt Airport for 20 departures at the peak of the morning wave departures.

Also, other work [3] deals with the issue of crew movement control at IRROP in the operation of air carriers. The optimization approach in the published article makes it possible to decide on the number of exchanged crews in the event of an irregularity and the total number of crews required to cover the scope of operation.

[1] VSB-Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15/2172, 708 33 Ostrava-Poruba, Czech Republic, horinka@centrum.cz
[2] VSB-Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15/2172, 708 33 Ostrava-Poruba, Czech Republic, dusan.teichmann@vsb.cz
[3] VSB-Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15/2172, 708 33 Ostrava-Poruba, Czech Republic, lenka.kontrikova@vsb.cz

3 Problem formulation and mathematical model

A set of flights I is given. For each flight $i \in I$, the earliest scheduled pre-flight commencement time for flight operations is defined – t_i, flight duration including times required to perform all pre-flight and flight operations T_i, maximum pre-flight commencement time delay for the flight (if pre-flight operations are commenced no later than at the end of this period, there will be no flight delays) – a_i and an indication of the number of passengers who have purchased a ticket c_i for the flight $i \in I$. For each pair of flights $i \in I$ and $j \in I$, the "repositioning (empty) flight time" is further defined (including cases where the destination is also the departure airport of the next flight) after the end of the flight operation $i \in I$ until the time of the start of the flight $j \in I - \tau_{ij}$. There is also a total number of *N* aircraft (excluding the aircraft whose flight originated IRROP) that can be used to operate flights.

Assume that the maximum delay time for the start of pre-flight activities for flight $i \in I$ is burdened with uncertainty. The reasons for uncertainty in the case of coefficients on the right sides of the group of limiting conditions are, for example, longer planned ground service time (turnaround time) than is actually needed or the possibility of shorter actual time of the next flight due to favorable meteorological conditions.

The task is to optimize the daily use of the aircraft fleet, ie. decide on the sequence of flight operations by individual aircraft so that the transport is offered to a sufficiently large number of passengers at the highest possible level of trust, see eg [4].

In the general formulation of the task of optimizing the daily utilization of the aircraft fleet, let us assume the conditions of a homogeneous aircraft fleet, i.e., the possibility that all aircraft in the fleet are interchangeable.

In order to model the decision in the optimization problem, we introduce a group of variables x_{ij}, for $i \in I \cup \{0\}$, $j \in I \cup \{0\}$ and $i \neq j$ with domains containing values 0 and 1 (these will be bivalent variables). The bivalent variables x_{ij} will model the decision that the same aircraft will be deployed to operate flight $j \in I$ after flight control $i \in I$. The value $x_{ij} = 1$ after the end of the optimization calculation will represent the decision that the flight $i \in I$ and the flight $j \in I$ will be included in the daily operation plan of the aircraft in the presence of an immediate precedence $i \prec\prec j$, i.e. that among the flight operator $i \in I$ and $j \in I$ it will not be operated in any other flight. The value $x_{ij} = 0$ after the end of the optimization calculation will represent the decision that there will be no immediate precedence between flights $i \in I$ and $j \in I$ no other flight will be operated in that order. The value $x_{ij} = 0$ after the end of the optimization calculation will represent the decision that there will be no immediate precedence between flights $i \in I$ and $j \in I$ in the daily operation plan of the aircraft.

The variables x_{0j} for $j \in I$ and x_{i0} for $i \in I$ will have a specific position between the variables x_{ij}, for $i \in I \cup \{0\}$ and $j \in I \cup \{0\}$. If $x_{0j} = 1$ applies to the bivalent variable x_{0j} after the optimization calculation, then the flight operator $j \in I$ will not be preceded by the operation of another flight. If $x_{0j} = 0$ applies to the bivalent variable x_{0j} after the end of the optimization calculation, then in the plan of daily operation of the aircraft, the flight operator $j \in I$ will be preceded by the operation of another flight. The variables x_{i0} will have analogous meanings. If $x_{i0} = 1$ applies to the bivalent variable x_{i0} after the end of the optimization calculation, then in the plan of daily operation of the aircraft, the operation of another flight will not follow the service $i \in I$. If $x_{i0} = 0$ applies to the bivalent variable $x_{i0} = 0$ after the end of the optimization calculation, then the flight $i \in I$ will be followed by the operation of another flight.

One of the ways to work in mathematical models with uncertainty is fuzzy numbers, in the presented article it will be triangular fuzzy numbers.

The main task remains - it is still necessary to decide on the sequence of flight operations by individual aircraft, but to take into account their uncertainty in their operational rescheduling.

The mathematical model has the form:

$$\max f(x, z) = \sum_{j \in I} c_j \sum_{i \in I \cup \{0\}} x_{ij} \tag{1}$$

subject to:

$$\sum_{i \in I \cup \{0\}} x_{ij} \leq 1 \qquad \text{for } j \in I \tag{2}$$

$$\sum_{i \in I} x_{ji} \leq \sum_{i \in I \cup \{0\}} x_{ij} \qquad \text{for } j \in I \qquad (3)$$

$$\sum_{j \in I} x_{0j} \leq N \qquad (4)$$

$$t_i + T_i + z_i + \tau_{ij} \leq t_j + z_j + M \cdot (1 - x_{ij}) \qquad \text{for } i \in I \text{ and } j \in I \qquad (5)$$

$$z_i \leq \widetilde{\boldsymbol{a}}_{\boldsymbol{\iota}} \qquad \text{for } i \in I \qquad (6)$$

$$x_{ij} \in \{0; 1\} \qquad \text{for } i \in I \cup \{0\} \text{ and } j \in I \qquad (7)$$

$$z_i \in R_0^+ \qquad \text{for } i \in I \qquad (8)$$

Function (1) represents the optimization criterion - the total number of passengers whose flights the aircraft will be dispatched. The group of restrictive conditions (2) shall ensure that a maximum of one aircraft is dispatched to operate each flight. The group of restrictive conditions (3) ensures that if the flight control $j \in I$ does not take place (situation when $\sum_{i \in I \cup \{0\}} x_{ij} = 0$), then the flight control $i \in I$ cannot occur after it either. If flight $j \in I$ occurs (situation where $\sum_{i \in I \cup \{0\}} x_{ij} = 1$), then the group of conditions ensures that it will be followed by a maximum of one flight $i \in I$. Restriction (4) will ensure that no more aircraft are used to operate the flights than the carrier has available. The group of restrictive conditions (5) will ensure that if the flight operator $j \in I$ after the flight $i \in I$ is not time-permissible, then it will not be scheduled. The group of restrictive conditions (6) shall ensure that the delay in the time of commencement of pre-flight operations at the aerodrome of departure of the flight does not exceed the time that would cause the flight to be delayed. The groups of constraints (7) and (8) define the domains of the variables used in the proposed model.

The occurrence of uncertainty on the right-hand sides in the group of limiting conditions (6) will be modeled by fuzzy sets of values approximately smaller than the values modeled by triangular fuzzy numbers ($\widetilde{a}_\iota = (a_{i1}, a_{i2}, a_{i3})$, $i \in I$, replacing constants modeling the maximum time periods shifts of individual flights. The causes of uncertainty in the case of the coefficients on the right sides of the group of limiting conditions (6) are, for example, a longer planned ground handling time (turnaround time) than is realistically needed or the possibility of a shorter actual flight time of the next flight due to favorable meteorological conditions (tail wind). If the value of the left side in a specific condition becomes less than or equal to the value of a_{i1} or a_{i2} (the choice of this value is made by the solver according to the specific situation), then the condition will be fulfilled at the level of credibility 1. If the value of the left side in a specific condition exceeds the value of a_{i2} or a_{i3} (its choice again depends on the solver according to the specific real conditions of the problem being solved), then the condition will be fulfilled at the level of credibility 0.

In order to be able to solve the model (1) - (8) well, it is necessary to perform its transformation into a linear model. The general procedure of transformation is given in the literature [4]. The linearized version of the fuzzy linear model of the optimization problem (1) - (8) has the form:

$$max\, f(x, z, h) = h \qquad (9)$$

subject to:

$$\sum_{i \in I \cup \{0\}} x_{ij} \leq 1 \qquad \text{for } j \in I \qquad (10)$$

$$\sum_{i \in I} x_{ji} \leq \sum_{i \in I \cup \{0\}} x_{ij} \qquad \text{for } j \in I \qquad (11)$$

$$\sum_{j \in I} x_{0j} \leq N \qquad (12)$$

$$t_i + T_i + z_i + \tau_{ij} \leq t_j + z_j + M \cdot (1 - x_{ij}) \qquad \text{for } i \in I \text{ and } j \in I \qquad (13)$$

$$z_i \leq h \cdot a_i + (1 - h) \cdot (a_i + p_i) \qquad \text{for } i \in I \qquad (14)$$

$$\sum_{j \in I} c_j \cdot \sum_{i \in I \cup \{0\}} x_{ij} \geq h \cdot F^{max} + (1 - h) \cdot (F^{max} - r) \qquad (15)$$

$$x_{ij} \in \{0; 1\} \qquad \text{for } i \in I \cup \{0\} \text{ and } j \in I \qquad (16)$$

$$z_i \in R_0^+ \qquad \text{for } i \in I \qquad (17)$$

$$h \in R_0^+ \qquad (18)$$

Function (9) represents the optimization criterion - the level of confidence to which a sufficiently large number of passengers whose flight has been dispatched is achieved (even at the cost of an acceptable delay). The comments on the groups of restrictive conditions (10) - (13) are the same as the comments on the groups of restrictive conditions (2) - (5). The function of the constraint group (14) ensuring the acceptability of the time shift of the planned pre-flight commencement time for the purpose of $i \in I$ flight operations will be explained by the interpretation of the individual values achieved on the right-hand side when the confidence level changes. At the level of confidence $h = 1$, a time shift of the planned time of commencement of pre-flight activities for the purpose of $i \in I$ flight operation is achieved only in cases where the planned time reserve according to the current flight schedule is used. If the value of the variable z_i exceeds the value a_i but does not exceed the value $a_i + p_i$, the confidence level of the obtained time shift (and consequently the obtained solution) will decrease and at the moment when the value of the time shift of the planned time of pre-flight operations flight $i \in I$ reaches the value $a_i + p_i$, there is a state when the level of trust, which is achieved by a sufficiently large number of passengers whose flight was dispatched (even at the cost of acceptable delay), is $h = 0$. Restriction (15) creates a relationship between the level of trust achieved and the total number of passengers whose flights have been dispatched (even at the cost of an acceptable delay). Constraint groups (16) - (18) define the domains of variables used in the model.

4 Computational experiments

Consider the following optimization problem. The air carrier has a homogeneous aircraft fleet to operate 10 flights on a specific day. The operational data required to solve the optimization problem are summarized in Table 1 and Table 2. Table 1 contains data on flights, Table 2 contains data on the time required for aircraft flights between airports (airports landing of flight $i \in I$ and airport of departure $j \in I$). An airport with index $i = j = 0$ represents an airport with a carrier base station. Flight times t_i are expressed in the number of minutes that elapse from the selected point on the timeline. If the selected point is e.g. 5:00, then the value of the pre-flight start time 100 for flight 1 corresponds to the real time 6:40.

Flight i	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
t_i	100	130	140	190	250	280	360	395	495	600
T_i	50	70	40	130	95	100	70	200	90	130

Table 1 Values t_i and T_i

Flights	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	130	125	80	100	40	50	60	75	60	120
2	125	135	45	55	40	130	80	70	65	40
3	100	200	135	130	140	60	75	50	60	70
4	130	140	200	120	240	50	80	75	60	80
5	160	75	150	100	170	200	80	90	100	90
6	115	140	65	85	60	130	90	70	75	60
7	125	120	105	65	60	120	110	90	65	70
8	65	100	115	55	70	110	120	100	95	130
9	135	130	105	85	60	120	130	140	55	100
10	95	140	125	110	40	80	70	110	155	105

Table 2 Values τ_{ij}

In the first phase of the computational experiment, the minimum number of aircraft needed to operate the scheduled 10 flights will be calculated. The calculated value will serve as one of the input data for the next phases of the calculation, in which the proposed model (9) - (18) will be experimented. However, to calculate the minimum number of aircraft, it is necessary to use another model, namely a model in the form of:

$$min\, f(x) = \sum_{j \in I} x_{0j} \tag{19}$$

subject to:

$$\sum_{i \in I \cup \{0\}} x_{ij} = 1 \qquad \text{for } j \in I \tag{20}$$

$$\sum_{j \in I \cup \{0\}} x_{ij} = 1 \qquad \text{for } j \in I \tag{21}$$

$$t_i + T_i + \tau_{ij} \le t_j + M \cdot (1 - x_{ij}) \qquad \text{for } i \in I \text{ a } j \in I \tag{22}$$

$$x_{ij} \in \{0; 1\} \qquad \text{for } i \in I \cup \{0\} \text{ a } j \in I \cup \{0\} \tag{23}$$

Function (19) represents the optimization criterion - the minimum number of aircraft required to operate 10 flights. The group of restrictive conditions (20) will ensure that each flight is serviced. The group of restrictive conditions (21) will ensure just one transfer after the flight service $i \in I$. The group of restrictive conditions (22) will ensure that time-inadmissible transfers between flights do not occur. The constraint group (23) defines the domains of the variables used in the model.

After completing the optimization calculation, a minimum number of 4 aircraft was calculated with the following flight sequences: $0 \to 1 \to 6 \to 9 \to 0$, $0 \to 2 \to 5 \to 0$, $0 \to 3 \to 7 \to 10 \to 0$ and $0 \to 4 \to 8 \to 0$. This completes the first phase of the computational experiment.

In the second phase of the computational experiment, assume that 1 aircraft fails. In addition, in this phase of the computational experiment, we determine all the input data for the solution of the model (9) - (18).

So, we have only 3 planes available. Since 4 aircraft are the minimum number of aircraft, it is clear that all 10 flights will not be dispatched. This assumption can be verified by adding a condition to the model (19) - (23) in the form:

$$\sum_{j \in I} x_{0j} \le 3 \tag{24}$$

After the optimization calculation is completed, the optimization software reports that the task does not have an acceptable solution.

In the next procedure, we determine the maximum possible delay times of the start times of pre-flight activities for individual flights a_i, where $i \in I$. We determine these times, for example, according to whether the given flight is directed to a hub-type destination, and it is necessary to maintain connections to connecting flights there. If the start time of pre-flight activities cannot be changed over time, the value of the maximum allowable shift for a given flight is 0. For the purposes of the computational experiment, it will also apply that for the last operated flights (flights 5, 8, 9, 10), a_i, where $i = 5, 8, 9, 10$, set uniformly, namely $a_5 = a_8 = a_9 = a_{10} = 0$. The results of the calculations are summarized in Table 3. Table 3 also shows the numbers of passengers with a purchased ticket for the given flights (it is assumed that the numbers of passengers do not change during the given day).

Flight *i*	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$\boldsymbol{a_i}$	80	10	105	0	0	50	100	0	0	0
$\boldsymbol{c_i}$	100	120	128	150	135	140	165	100	110	95

Table 3 Values a_i and c_i

Other input data needed to verify the functionality of the model are the left, main and right values of triangular fuzzy numbers $\tilde{\boldsymbol{a}}_i = (a_{i1}, a_{i2}, a_{i3})$. For further calculations we will choose one of the possible approaches mentioned in the work [4]. It will be an approach in which the key values are a_{i1} and a_{i3}. For key values of fuzzy numbers $\tilde{\boldsymbol{a}}_i$ it will apply that $a_{i1} = a_i$ (value from the second row in table no. 3) and $a_{i3} = a_{i1} + p_i$, where $p_i = 2 \cdot a_{i1}$. This means that the maximum acceptable flight delay rate is equal to twice the time reserve a_i.

The last input data to the model (9) - (18) are the values F^{max} and r. These values are determined in accordance with the procedure given in the work [4] The value F^{max} corresponds to the maximum number of passengers whose flight will be departed on a given day under the most favorable operating conditions (the most favorable operating conditions correspond to the maximum use of time values p_i). We calculate the value of r as the difference $r = F^{max} - F^{min}$, where F^{min} corresponds to the maximum number of passengers whose flight will be dispatched on a given day under the least favorable operating conditions (least favorable operating conditions correspond to the minimum use of time values p_i, ie the case when $p_i = 0$).

In order to calculate the value F^{max}, we solve the model (1) - (8), in which we substitute the values a_{i3} for $i = 1, \ldots, 10$ on the right-hand side of the group of conditions (6).

In order to calculate the value F^{min}, we solve the model (1) - (8), in which we substitute the values a_{i1} for $i = 1, \ldots, 10$ on the right-hand side of the group of conditions (6). By solving the model (1) - (8) for the values a_{i3} we get the total number of passengers with the flight on a given day $F^{max} = 1\ 148$ and the flight plans served by individual aircraft $0 \to 1 \to 6 \to 3 \to 7 \to 0$, $0 \to 2 \to 5 \to 9 \to 0$ and $0 \to 4 \to 8 \to 0$. Therefore, the flight 10 is not dispatched. By solving the model (1) - (8) for the values a_{i1} we get the total number of passengers with the flight on a given day $F^{min} = 1\ 143$ and flight plans served by individual aircraft $0 \to 2 \to 5 \to 7 \to 10 \to 0$, $0 \to 3 \to 6 \to 9 \to 0$ and $0 \to 4 \to 8 \to 0$. Therefore, flight 1 is not dispatched.

The third phase of the solution follows, where we solve the model (9) - (18). After substituting and completing the optimization calculation, we obtain the value h = 0.7h. Based on the obtained value, we can therefore conclude that a sufficiently large number of passengers with a departed flight will have a confidence level of 0.7 in the event of a failure of 1 aircraft.

5 Conclusion

The presented paper deals with the operational control issue of an air carrier operating regular passenger transport with an insufficient number of aircraft in conditions of uncertainty. Insufficient number of aircraft can be caused by several problems, which can be related to technical defects on the aircraft, complications caused during the flight or problems caused by the absence of a suitable crew. Uncertainty is modeled through fuzzy mathematics, specifically using triangular fuzzy numbers. For the solution itself, a fuzzy linear model is introduced, enabling to maximize the number of passengers with the departure flight during the planning period. The model fuzzificates the maximum times by which the start times of pre-flight activities can be postponed. Computational experiments are performed with the proposed model, which document the functionality of the proposed approach. Further activities will be directed to the fuzzification of other quantities appearing in the model, including the coefficients of the purpose function.

Acknowledgements

The paper was supported by internal project from the Faculty of Mechanical Engineering Faculty of VSB – Technical University of Ostrava, SP 2022/62 Development and research in transport and logistics.

References

[1] Barnhart, C.; Laporte, G.; Ball, M.; Nemhauser, G.; Odoni, A. *Handbooks in Operations Research & Management Science: Transportation: Chapter 1 Air Transportation: Irregular Operations and Control*. 14; North Holland, 2006; ISBN 9780444513465.

[2] Evler, J.; Asadi, E.; Preis, H.; Fricke, H. Airline Ground Operations: Schedule Recovery Optimization Approach with Constrained Resources. *Transportation Research Part C: Emerging Technologies* 2021, 128, doi:10.1016/j.trc.2021.103129.

[3] Wei, G.; Yu, G.; Song, M. Optimization Model and Algorithm for Crew Management During Airline Irregular Operations. *Journal of Combinatorial Optimization* 1997, 1, doi:10.1023/A:1009780410798.

[4] Teodorović, D.; Vukadinović, K. *Traffic Control and Transport Planning*; Springer Netherlands: Dordrecht, 1998; Vol. 13; ISBN 978-94-010-5892-6.

Numerical Valuation of the Investment Project Flexibility Based on the PDE Approach: An Option to Contract

Jiří Hozman [1], Tomáš Tichý [2]

Abstract. The solution to the optimal investment decision, which captures the value of a flexibility embedded in a project, plays an important role in the decision-making process. In this paper we focus on a real options approach interpreting the flexibility value as the option premium and we extend our previous research to an option to contract operating scale according to market conditions.
Following a contingent claim analysis the values of both the project and the embedded flexibility, expressed as functions of time and underlying output price (following a stochastic process), can be identified as solutions of relevant PDE systems of the Black-Scholes type. More precisely, the link between project and flexibility values is realized through a payoff function, which can be enforced with respect to the flexibility type at any time prior to or at expiration date.
Due to the presence of the American constraint the real option pricing problem is not solvable analytically in general, and therefore appropriate numerical methods have to be employed. Analogously to pricing of financial options and in line with our results achieved in this field of financial engineering, the discontinuous Galerkin method is applied to solve the relevant governing equations. The capabilities of the numerical scheme resulted are illustrated on a simple contraction decision problem.

Keywords: real option pricing, project value, option to contract, Black-Scholes inequality; American option, discontinuous Galerkin method

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

1 Introduction

Traditional valuation approaches, based on discounted cash flow methodology, do not recognize the important qualitative and quantitative characteristic of some of the intrinsic attributes of the investment opportunities, namely, irreversibility of investments, choice of timing and last but not least uncertainty over the future rewards from investments, see [3]. Accordingly, in a such stochastic world, the classical net present value rule significantly underestimates the true value of an investment project, as it is not able to capture the flexibility value of such project, especially in a long time horizon.

Therefore, to provide a more realistic model of investment behaviour, the modern investment theory, known as real options approach, was built on the interpretation of the flexibility value as the option premium, see [11]. There exists a large number of various valuation techniques related to real options approach, see, e.g., [10] for a brief overview. Among them, the contingent claims analysis enjoys greater interest, because the formulation via a partial differential equation (PDE) provides comprehensive information on modern investment issues in terms of a wide spectrum of input data and possibilities of further post-processing.

In this short contribution we extend our recent results focused on the numerical valuation of options to extend the operating scale at a fixed time in the future, see the conference paper [8], where the discontinuous Galerkin (DG) method is applied to solve the relevant governing equations. We proceed as follows — in Section 2 the relevant PDE models are formulated, while in Section 3 a numerical valuation scheme is presented. Finally, in Section 4 a simple numerical experiment related to reference data is provided.

2 PDE Model for Pricing of Options to Contract

In contrast to expansion options from [8], we concentrate here on valuing the flexibility of an investment project that adopts the embedded option to contract the operating scale with respect to current market conditions. In order

[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.

to describe the flexibility value (of the project), it is first necessary to describe the value of the project itself. Subsequently, we are able to find the real option value by solving the relevant PDEs that link both option and project values, see inspiring ideas in [9]. In line with [8] we assume that project values as well as real option prices can be expressed as functions of the actual time t and the commodity (output in general) price P following a geometric Brownian motion, as stated in [4], i.e.,

$$\mathrm{d}P(t) = (r - \delta)P(t)\mathrm{d}t + \sigma P(t)\mathrm{d}W(t), \quad P(0) > 0, \tag{1}$$

where $r > 0$ is the risk-free interest rate, $\delta > 0$ is the mean convenience yield on holding one unit of the commodity, $W(t)$ is a standard Brownian motion and $\sigma > 0$ is the volatility of the commodity price.

Further, we denote by $V_0(P,t)$ the value of the project, which does not have any options on contraction. In contrast, the function $V_1(P,t)$ stands for the value of an investment project with the embedded option to contract the production rate requiring the amount $C > 0$. Since the possibility of this single decision-making (i.e., scaling down the production rate) is related to any time prior to or at prespecified time $T > 0$, options on contraction correspond to the conventional American put options with the payoff function defined as follows

$$\Pi(V_0, V_1) \equiv \Pi(V_0(P,T), V_1(P,T)) = \max(V_1(P,T) - V_0(P,T) - C, 0), \quad P \geq 0. \tag{2}$$

Note that it is also possible to introduce the payoff function (2) with term $+C$ instead of $-C$ and in this case we speak of the disinvestment costs C that we will save by a contraction of the operating scale, cf. [2].

Let $T^* > T$ be the maximum life-time of both projects and $\varphi_0(P,t)$ and $\varphi_1(P,t)$ represent (after-tax) cash flow rates associated with the given project. Following the contingent claims approach [1], involving no arbitrage opportunities for trading in the underlying assets together with the delta-hedging techniques, one can characterize value functions V_0 and V_1 between expiry date T and project life-time T^* as solutions of a couple of deterministic backward PDEs (see [3]):

$$\frac{\partial V_i}{\partial t} + \underbrace{\frac{1}{2}\sigma^2 P^2 \frac{\partial^2 V_i}{\partial P^2} + (r-\delta)P\frac{\partial V_i}{\partial P} - rV_i}_{\mathcal{L}_{BS}(V_i)} = -\varphi_i, \quad P \in (0,\infty),\ t \in [T, T^*),\ i = 0,1, \tag{3}$$

with the homogeneous terminal conditions $V_i(P,T^*) = 0$, $P \in (0,\infty)$, $i = 0,1$.

In what follows we present the governing equation for the embedded flexibility representing the value added to the project function, i.e., $V_1(P,t) \geq V_0(P,t)$ for all $P \geq 0$ and $t \in [0,T)$. More precisely, we set $F(P,t) = V_1(P,t) - V_0(P,t)$ as the option value at the current price P and actual time $t \in [0,T)$. Further, taking into account an equivalence of cash flow rates of both projects during the option life-time (i.e., $\varphi_1(P,t) = \varphi_0(P,t)$ on $[0,T)$) and encompassing the early exercise constraint of American options (i.e., $F(P,t) \geq \Pi(V_0(P,T), V_1(P,T))$ at any time $t \in [0,T)$), the value function F satisfies the so-called moving-boundary problem. For this kind of problems it is also necessary (apart from solving the governing equation) to determine two regions separated by a free boundary $\mathcal{E}$ driven by the optimal exercise price $P^*(t)$, see [15]. The presence of the American constraint thus represents a quite new and essential extension of the methodological concept of real options valuation from [8] and the treatment of this early exercise feature is listed on the following lines and in Section 3, respectively.

Let $\Omega_{\mathrm{E}} \subset (0,\infty)$ denote the exercise region, where it is optimal to exercise the option early, in other words, scale down the operating rate, thus we solve the following problem

$$\frac{\partial F}{\partial t} + \mathcal{L}_{BS}(F) < 0, \quad P \in \Omega_{\mathrm{E}},\ t \in [0,T), \quad \text{with } F(P,t) = \Pi(V_0(P,T), V_1(P,T)), \tag{4}$$

where $\mathcal{L}_{BS}$ is the second order linear differential operator of the Black-Scholes (BS) type defined in (3). While in the continuation region, it is not optimal to exercise early and we solve the problem

$$\frac{\partial F}{\partial t} + \mathcal{L}_{BS}(F) = 0, \quad P \in (0,\infty)\backslash\overline{\Omega_{\mathrm{E}}},\ t \in [0,T), \quad \text{with } F(P,t) > \Pi(V_0(P,T), V_1(P,T)). \tag{5}$$

Note that to guarantee the well-posedness of (4)–(5), it is enforced a continuity of the option value F and the partial derivative $\partial F/\partial P$ on the free boundary $\mathcal{E}$, known as the high-contact condition for American options, see [14].

There are several approaches how to handle the early exercise feature, among the widely used ones, just penalty techniques [15] allow us to reformulate both problems (4) and (5) into one equation valid everywhere in both regions, i.e.,

$$\frac{\partial F}{\partial t} + \frac{1}{2}\sigma^2 P^2 \frac{\partial^2 F}{\partial P^2} + (r-\delta)P\frac{\partial F}{\partial P} - rF + q_F = 0, \quad P \in (0,\infty),\ t \in [0,T), \tag{6}$$

where q_F is defined to ensure American constraint $F(P,t) \geq \Pi(V_0(P,T), V_1(P,T))$ and satisfy the conditions:

$$q_F(P,t) = 0, \text{ if } F(P,t) > \Pi(V_0(P,T), V_1(P,T)), \qquad q_F(P,t) > 0, \text{ if } F(P,t) = \Pi(V_0(P,T), V_1(P,T)). \quad (7)$$

The quantity q_F can be viewed as an additional nonlinear source term in the governing equation and its essential role is to guarantee that the value of an option under American constraint cannot fall below its payoff function at any time $t \in [0,T)$. Note that the penalty approach can be unified for both European and American exercise features, if we put $q_F(P,t) = 0$ in (6) for all $P > 0$ and $t \in [0,T)$ in the case of a European exercise right.

3 Numerical Solution

According to Section 2 to determine the present value of the flexibility of an investment project one have to solve two consecutive problems. First, a pair of PDEs (3) with homogeneous terminal conditions is solved to construct payoff function (2). Consequently, the solution of the problem (6) with penalty (7) represents the desired real option value. Since there are no analytical formulae for finite maturity American options in general, the valuation should rely on numerical approaches. In our study, we employ the DG method, successfully used also in the field of financial options pricing (see, e.g., [6] and [7]), to improve the numerical valuation process. We proceed as follows. At first, we localize the governing equations to a bounded spatial domain and discuss the choice of suitable boundary conditions. Next, we recall the variational form of the penalty term for the American constraint. Finally, we mention the standard discretization steps and present the numerical scheme.

3.1 Localization to a computational domain

The spatial localization of terminal-value problems (3) and (6) is necessary to the subsequent numerical treatment. For this purpose let $P_{\max} \gg 0$ denote the maximal sufficient value of the commodity price satisfying $P_{\max} > P^*(t)$ for all $t \in [0,T)$. Then, we define the computational domain $\Omega = (0, P_{\max})$ and formally restrict the governing equations and the relevant terminal conditions to the bounded domain Ω. Therefore, we have to impose project as well as option values at both endpoints $P = 0$ and $P = P_{\max}$.

First, the project values V_i, $i = 0,1$, for all $t \in [T, T^*)$, on boundary $\{0, P_{\max}\}$ can be simply estimated by the net present value approach, see [9]. Specifically, under the given cash flow rates φ_i, $i = 0,1$, the present values of both projects at endpoints of Ω are defined as

$$V_i(0,t) = \int_t^{T^*} \varphi_i(0,\xi) e^{-r(\xi - t)} \mathrm{d}\xi, \qquad V_i(P_{\max},t) = \int_t^{T^*} \varphi_i(P_{\max},\xi) e^{-r(\xi - t)} \mathrm{d}\xi, \quad t \in [T,T^*), \ i = 0,1. \quad (8)$$

Secondly, for the real option value, the choice of boundary conditions has to reflect the type of flexibility that this option provides. Obviously, for a sufficiently large value of $P_{\max}$, the option to contract the operating scale is of no value at $P = P_{\max}$. On the other hand, when $P = 0$ it is optimal to exercise (i.e., to contract) early and thus option values at the left boundary point of Ω are set in the accordance with American constraints given by payoff function (2). In accordance with the above, we prescribe Dirichlet boundary conditions in the form

$$F(0,t) = \Pi(V_0(0,T), V_1(0,T)), \qquad F(P_{\max},t) = 0, \quad t \in [0,T). \quad (9)$$

3.2 Penalty method

In order to handle the American early exercise feature and force the solution of (6) to be equal to the payoff in the exercise region Ω_{E}, we were inspired by [15] and introduce, for a sufficiently regular function v, the variational form of penalty term q_F as

$$\langle q_F(t), v\rangle = c_p \int_\Omega \chi_{\mathrm{exe}}(t)\Big(\Pi(V_0,V_1) - F(t)\Big) v \,\mathrm{d}P = \underbrace{c_p \int_\Omega \chi_{\mathrm{exe}}(t)\Pi(V_0,V_1) v \,\mathrm{d}P}_{Q_{\mathrm{R}}(v)} - \underbrace{c_p \int_\Omega \chi_{\mathrm{exe}}(t) F(t) v \,\mathrm{d}P}_{Q_{\mathrm{L}}(F,v)}, \quad (10)$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product in $L^2(\Omega)$. The function $\chi_{\mathrm{exe}}(t)$ in (10) is defined as an indicator function of the region Ω_{E} at time instant t and $c_p > 0$ represents a weight to enforce the early exercise. In line with [6], we set c_p proportional to $1/\tau$, where τ is the time step introduced in (13). The form (10) can be split into linear functional Q_{R} and bilinear form Q_{L}, and we place their discrete variants on opposite sides of the fully discrete formulation of (6), see (13).

3.3 Numerical valuation scheme

Since the governing equations (3) and (6) are closely related to the class of convection-diffusion equations and exhibit a hyperbolic behaviour as $|r - \delta| \gg \sigma^2$, the DG method (see [12] for a complete overview) is applied in order to cope with these challenging attributes. We modify the numerical scheme proposed in [8] to pricing real options under the American constraint. Accordingly, the approximate solutions representing project as well as real option values are constructed as the piecewise polynomial, generally discontinuous, functions of the p-th order defined over the partition of the domain Ω with the assigned mesh size h, i.e., with partition nodes $0 = P_0 < P_1 < \ldots < P_N = P_{\max}$. In the rest of the paper, we denote the space of such functions by S_h^p.

The development of the whole numerical scheme consists of two consecutive phases — spatial semi-discretization and temporal discretization. Within the first phase the time variable is left continuous and semi-discrete solutions are introduced using the variational formulation as solutions of the particular systems of the ordinary differential equations, see [6] and [7]. Secondly, we realize the discretization with respect to the time coordinate using an implicit Euler scheme that results into a sequence of linear algebraic problems related to a time partition $T^* = t_0 > t_1 > \cdots > t_R = T > t_{R+1} > \cdots > t_M = 0$ with fixed time step $\tau = T^*/M$. Further, denote $u_{h,m}^{(i)} \in S_h^p$, $i = 0, 1$, the approximation of the corresponding project value functions V_i from (3) at time level $t_m \in [T, T^*]$, $m = 0, \ldots, R$. Similarly, we define the DG approximate solution of problem (6) as functions $w_h^m \approx F(\cdot, t_m)$, $t_m \in [0, T]$, $m = R, \ldots, M$. Starting from zero initial project values $u_{h,0}^{(0)}$ and $u_{h,0}^{(1)}$, the desired value of flexibility $w_h^M \approx F(\cdot, 0)$ is computed in the following three steps (note $t_{m+1} - t_m = -\tau$):

$$\left\langle u_{h,m+1}^{(i)}, v_h \right\rangle - \tau \mathcal{A}_h\left(u_{h,m+1}^{(i)}, v_h\right) = \left\langle u_{h,m}^{(i)}, v_h \right\rangle - \tau \ell_h^{(i)}(v_h)(t_{m+1}) + \tau \left\langle \varphi_i(t_{m+1}), v_h \right\rangle \quad \forall\, v_h \in S_h^p,\ m = 0, 1, \ldots, R-1,\ i = 0, 1, \tag{11}$$

$$\left\langle w_h^R, v_h \right\rangle = \left\langle \Pi\left(u_{h,R}^{(0)}, u_{h,R}^{(1)}\right), v_h \right\rangle \quad \forall\, v_h \in S_h^p, \tag{12}$$

$$\left\langle w_h^{m+1}, v_h \right\rangle - \tau \mathcal{A}_h\left(w_h^{m+1}, v_h\right) + \tau Q_h\left(w_h^{m+1}, v_h\right) = \left\langle w_h^m, v_h \right\rangle - \tau \ell_h(v_h)(t_{m+1}) + \tau q_h(v_h)(t_{m+1}) \quad \forall\, v_h \in S_h^p,\ m = R, \ldots, M-1, \tag{13}$$

where the bilinear form $\mathcal{A}_h(\cdot,\cdot)$ stands for the discrete variant of the spatial partial differential operator $\mathcal{L}_{BS}$ from (3). Further, the linear forms $\ell_h^{(i)}(\cdot)(t)$ and $\ell_h(\cdot)(t)$ contain terms arising from boundary conditions (8) corresponding to the particular project value V_i and from boundary conditions (9) related to the option value F, respectively. Concerning the American constraint treated by (10) in scheme (13), we introduce new forms $Q_h(\cdot,\cdot)$ and $q_h(\cdot)(t)$ as discrete variants of Q_{L} and Q_{R}. For a detailed derivation of all above-mentioned forms, we refer the interested reader to [7]. Moreover, for practical purpose, to evaluate forms Q_h and q_h we use an element-wise approximation of the early exercise region as

$$\chi_{\mathrm{exe}}(t_m)\big|_{[P_l, P_{l+1}]} \approx \widetilde{\chi_{\mathrm{exe}}}(t_m)\big|_{[P_l, P_{l+1}]} := \begin{cases} 1, & \text{if } w_h^{m-1}(P_c) < w_h^R(P_c) \\ 0, & \text{if } w_h^{m-1}(P_c) \geq w_h^R(P_c) \end{cases}, \quad t_m \in [0, T],\ 0 \leq l \leq N-1, \tag{14}$$

where P_c is the midpoint of the interval $[P_l, P_{l+1}]$ and w_h^R is given as S_h^p-approximation of the payoff function Π depending on states $u_{h,R}^{(i)}$, $i = 0, 1$, see (12). Finally, note that the equations (11) and (13) result into a sequence of systems of linear algebraic equations with sparse matrices that uniquely determine the relevant solutions on the corresponding time levels, see [6].

4 Illustrative Case Study: Option to Contract Mining Production

In this section, we briefly illustrate capabilities of the numerical scheme introduced above. The presented numerical experiments arise from practical issues in the iron ore mining industry and evaluate American as well as European contraction options under various scenarios. The three-step valuation scheme (11)–(13) is implemented in the solver Freefem++, incorporating GMRES as a solver for non-symmetric sparse systems, for more details, see [5].

As in [9] we consider an iron ore mine, value of which is given by the project value $V_0(P,t)$, and a mining project of value $V_1(P,t)$, adopting the embedded option $F(P,t)$ to scale down the production rate any time $t \in [0,T]$ with respect to current market conditions. Further, let Q denote the total reserve of the iron ore mine in thousands of million dry metric tonnes (dmt) and $q_i(t) \geq 0$, $i = 0,1$, be the iron ore production rates (in thousands of million dmt per year) associated with projects V_i, $i = 0,1$. The life-times $T_0^* > 0$ and $T_1^* > 0$ (in years) of both projects depend on operating rates q_i and are defined as minimal admissible values that satisfy the relationship $Q = \int_0^{T_0^*} q_0(\xi)\mathrm{d}\xi = \int_0^{T_1^*} q_1(\xi)\mathrm{d}\xi$. The contraction feature of the option implies that $T^* = T_1^* > T_0^* \gg T$ and therefore we assume that production rates satisfy

$$q_0(t) = \begin{cases} s(t), & \text{if } t \in [0, T_0^*), \\ 0, & \text{if } t \in [T_0^*, T^*], \end{cases} \qquad q_1(t) = \begin{cases} s(t), & \text{if } t \in [0, T), \\ \kappa \cdot s(t), & \text{if } t \in [T, T^*), \end{cases} \tag{15}$$

where the function $s(t)$ corresponds to the production rate related to the project having no embedded options and the factor $\kappa \in (0,1)$ represents the contracted mining production rate. Further, in line with [4], we define the after-tax cash flow rates of relevant projects as follows

$$\varphi_i(P,t) = q_i(t)\Big((1-D)P - c(t)\Big)(1-B), \quad P \in [0, P_{\max}],\ t \in [0, T^*], \qquad i = 0,1, \tag{16}$$

where $c(t)$ is the average cash cost rate of iron ore production per dmt, D is the rate of state royalties and B is the income tax rate.

The numerical experiments are performed on the reference data from [9], where European and American real put options are priced using a fitted finite volume method. In all cases, we consider the following project and market data:

$$Q = 10,\ \ s(t) = \frac{e^{0.007t}}{10},\ \ D = 0.05,\ \ B = 0.3,\ \ c(t) = 25e^{0.005t},\ \ T = 1,\ \ C = 1-\kappa,\ \ \sigma = 0.3,\ \ r = 0.06,\ \ \delta = 0.02, \tag{17}$$

which are the representatives of parameter values of practical significance. More precisely, the value of C is given in thousands of million USD and $c(0) = 25$ USD per dmt corresponds to prices for the calendar year 1988, see [13].

In order to ensure a relevant comparison with the reference results from [9], we employ a piecewise linear approximation (i.e., $p = 1$) on domain Ω with $P_{\max} = 60$ USD and uniform partition with mesh size $h = 0.6$. In parallel with this, the time step is set as $\tau = 0.01$ and the American early exercise feature is handled with $c_p = 10/\tau$ in (10). Further, using (15) and (17), easy calculation leads to $T_0^* \doteq 75.8$ and we determine T^* in a similar way for various factors κ.

Within the simple case study, we investigate the behaviour of the option values for various contraction rates $\kappa \in \{0.25, 0.50, 0.75\}$ and illustrate a general relationship between European and American options to contract under the same market conditions. Figure 1 (left) records the DG approximations of flexibility values (in thousands of million USD) at present time $t = 0$ for all three scenarios under the European as well as American constraints. At first glance, plots are similar to the conventional financial put options with the relevant BS model parameters and illustrate an intuitive expectation that the value of flexibility F is a decreasing function of the rate κ in the region of low output prices (i.e., for smaller ones than some critical value). Moreover, it is apparent for all cases that American options cost more than their European counterparts, i.e., early exercise feature increases value of the project flexibility. This distinctive feature of American options is well resolved for particular scenario ($\kappa = 0.5$) in more detail in Figure 1 (right), where differences (in thousands of million USD) between American and European option prices (at $t = 0$) are depicted together with the comparative results from [9]. Without surprise, one can easily observe that piecewise linear DG approximations match well the reference values and give fairly the same results as the fitted finite volume method. In brief the numerical approach performed shows its promising potential by producing financially meaningful results, easily interpreted and subsequently used in the decision-making process.

5 Conclusion

The real options approach and especially related valuations techniques pose a very challenging part of capital budgeting. In this paper we have recalled PDE models to valuation of investment projects and options to contract the production rate under the current market conditions. Concurrently, we have presented a numerical scheme based on the DG method to solve particular governing equations. The elaborated simple case study from the iron ore mining industry provides very good similarity to reference results and suitability of the DG method for real option pricing issues. One possible future research objective seems to be a multi-stage sequential investment decision that combines various options to change operating scale (expansion or contraction) into one compound option that is able to capture changing investment strategies in a long time horizon within the whole decision-making process.

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 SP2022/4, an SGS research project of VSB-TU Ostrava.

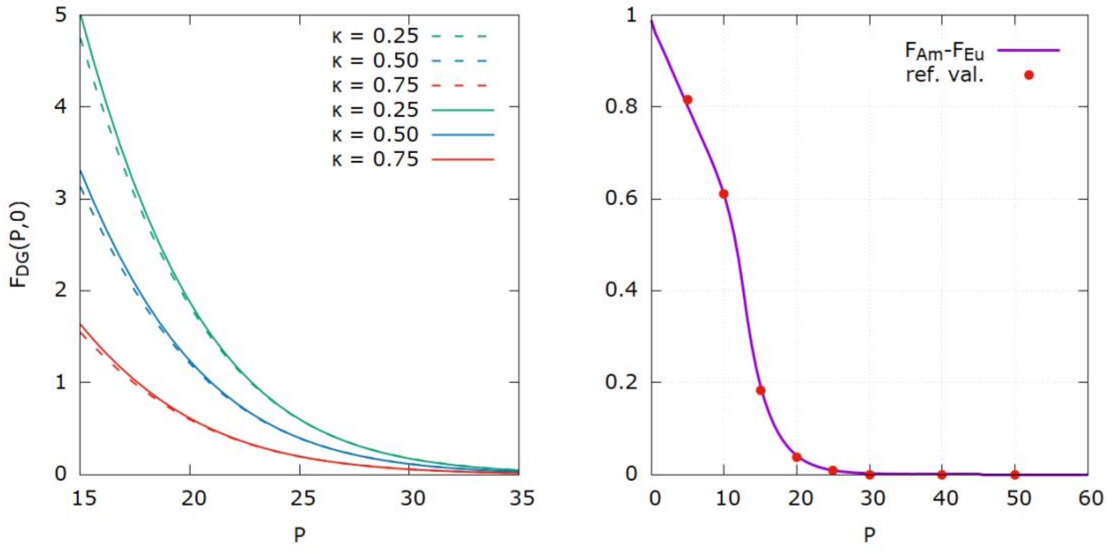


Figure 1 Approximate embedded option values for different contraction rates (left) under American (solid line) and European (dashed line) constraints; the difference between American and European contraction option prices at $t = 0$ for $\kappa = 0.50$ (right)

References

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

[2] Čulík, M. (2010). Flexibility and Project Value: Interactions and Multiple Real Options. *AIP Conference Proceedings*, 1239, 326–336.

[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. (2018). DG framework for pricing European options under one-factor stochastic volatility models. *Journal of Computational and Applied Mathematics*, 344, 585–600.

[7] Hozman, J. & Tichý, T. (2020). The discontinuous Galerkin method for discretely observed Asian options. *Mathematical Methods in the Applied Sciences*, 43, 7726–7746.

[8] 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.

[9] Li, N. & Wang, S. & Zhang, S. (2020). Pricing options on investment project contraction and ownership transfer using a finite volume scheme and an interior penalty method. *Journal of Industrial & Management Optimization*, 16, 1349–1368.

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

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

[12] Rivière, B. (2008). *Discontinuous Galerkin Methods for Solving Elliptic and Parabolic Equations: Theory and Implementation.* SIAM, Philadelphia.

[13] World Bank (1992). *World Development Report 1992. Development and the Environment*. New York: Oxford University Press.
[14] Wong, H.Y. & Zhao, J. (2008). An artificial boundary method for American option pricing under the CEV model. SIAM *Journal on Numerical Analysis*, 46, 2183–2209.
[15] Zvan, R. & Forsyth, P.A. & Vetzal, K.R. (1998). Penalty methods for American options with stochastic volatility. *Journal of Computational and Applied Mathematics*, 91, 199–218.

Location of Capacity Completion Centers in Distribution Systems with Heterogeneous Vehicle Fleet

Andrea Hrníčková[1], Dušan Teichmann[2], Denisa Mocková[3]

Abstract. An important group of operations taking place in distribution parts of logistics chains is the completion of consignments. It is usually necessary to decide in which warehouse the completion of consignments will take place if distribution systems contain several warehouses. The decision of the place of completion is influenced by location of the required items in individual warehouses and by quantity of the required items in the warehouses. The presented paper deals with the problem of location of multiple completion centres with limited capacity in the conditions of a distribution system containing several warehouses, with stored items divided into several size categories. It is possible to store items of the same size in several warehouses and a vehicle fleet is heterogeneous. The problem is conceived as an optimization problem, which is solved by the methods of mathematical programming. When creating a mathematical model, the knowledge from the field of transport and location tasks is used. A computational experiment will be performed with the proposed model, and it will be performed for several numbers of warehouses to be completion.

Keywords: Mathematical Model, Optimization, Distribution, Warehouse, Completion, Location

JEL Classification: C610
AMS Classification: 90B05, 90C08

1 Introduction

The article freely follows and expands the presented article at the QUANTITATIVE METHODS IN ECONOMICS (MULTIPLE CRITERIA DECISION MAKING XXI) 2022 „Mathematical Model for Localization of the Final Shipment Distribution Warehouse at a Distributor Operating in Virtual Reality". The article [2] dealt with the issue of the location of one dispatch centre with the completion of multi-products orders (further only the completion centre) into one of the existing warehouses so that the total costs of the completion and the dispatch of required warehouse products are minimal. The distribution network contained model data of 4 – 8 warehouses, distance matrix was known, the handling costs of one normalized product in each warehouse were known and costs per 1 kilometre driven. The distribution was realized using a homogeneous vehicle fleet, capacity of warehouses was unlimited, and therefore every warehouse was a candidate for the location of a dispatch centre with the completion of multi-product orders (further only the candidate).

The presented article deals with the location of p completion centres within the existing distribution network. Furthermore, the available capacity of the warehouses will be limited, and the distribution will be realized using heterogeneous vehicle fleet. As in the original article the minimization of the total costs of the completion and the dispatch of required warehouse products will remain the optimisation criterion. The handling costs of one normalized product (SPO) are also known. The distribution network contains 19 warehouses. The experiments will be performed for different values of p.

2 Motivation to solve the problem and analysis of the current state of knowledge

The optimization of the distribution is demanding processes influenced by demand of customers. It is good to have large distribution networks with a lot of warehouses for better customer satisfaction and to have tactical located completion centres at the same time.

[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, teichdus@fd.cvut.cz.
[3] Czech Technical University in Prague, Department of Logistics and Management of Transport, Horská 3. 128 03 Praha 2, mockova@fd.cvut.cz.

Modifications of the location routing problem (LRP) or vehicle routing problem (VRP) deal with the issue of the process optimization between final completion and dispatch with limited capacity and end customers. Approaches based on mathematical programming, heuristic or metaheuristic methods are especially suitable for working with these problems.

In [5] the authors dealt with Capacitated Location-Arc Routing Problem with Deadlines (CLARPD) and a fleet of capacitated heterogeneous vehicles. They proposed a mixed integer programming model, that determined which depots should be opened. It also determined the routes to be served with deadlines and assigned vehicles to each open depot. Moreover, efficient route plans were designed thereby to minimize total travel costs. Since CLARPD is an NP-hard problem, a genetic algorithm (GA) with constructive heuristic to generate initial solutions were designed. Furthermore, the authors investigated the usage of simulated annealing (SA) to determine GA performance. The results of the experiments showed the promising potential of GA. According to the authors, the proposed methods could be used for the appropriate location of depots with road salt and planning routes for the salting of roads during the winter months.

In [6] the authors dealt with the issues related to reduce emissions produced in logistics. They studied low-carbon LRP with a heterogeneous vehicle fleet, simultaneous pickup and delivery and time windows. A two-phased hybrid heuristic algorithm. Firstly, they used a GA to group the customer points according to temporal-spatial distance to construct the initial path, and then they used variable neighbourhood search algorithm to optimize location of the depots and to determine distribution routes. To improve the result achieved, elements of SA were implemented into the second algorithm and a vehicle modification strategy was also added.

In [1] the authors dealt with a similar problem, they focused on the issue of VRP with respect to the electric heterogeneous vehicle fleet for a two-echelon recycling network (recycling stations and recycling centres) within reverse logistics. The authors solved this problem by using two models and at the same time a hybrid GA in combination with large-scale neighbourhood search algorithm is proposed to achieve better results. Numerical experiments showed besides that the usage of a heterogeneous vehicle fleet could reduce costs. According to the authors, the proposed procedure could be used by recycling companies in the construction of recycling centres and route planning. The procedure also could further provide a basis for companies to solve the relationship between the economy, the environment and resources.

Another possible approach to solve heterogeneous capacitated vehicle routing problem (HCVRP) was dealt in [3], where the authors used Deep Reinforcement Learning (DRL), which is a combination of Deep Neural Networks and Reinforcement Learning. In solving the problem, they considered both min-max and min-sum objectives for HCVRP, which aimed to minimize the longest or total vehicle driving time. In this case DRL method was based on the attention mechanism with a vehicle selection decoder accounting for the heterogeneous fleet constrains and a node selection decoder accounting for the route construction. The mechanism learnt to generate a solution by automatic selection of the vehicle, as well as a node that was assigned to it in every step. Numerical experiments showed that the method is faster than the other modern methods (i.e., SISR), brings better solutions and generalized problems with a large number of customers for MM-HCVRP i pro MS-HCVRP.

3 Problem formulation and its mathematical model

In this chapter, the solved problem is presented and a mathematical, which is used to perform the experiments described in the following chapter is presented.

3.1 Formulation of optimisation problem

The set of warehouses I of distributor is defined in the distribution system. For each warehouse $i \in I$, the requirement a_i for warehouse products is defined (a SPO type of product is chose as the basic product, BPO and XPO type products are normalized to SPO type products). The transport between warehouses is ensured by a set of vehicles C with different capacities (heterogeneous vehicle fleet). Vehicle capacity $k \in C$ is c_k. The distance matrix D is known, where its elements d_{ij} represent distances between warehouses. Furthermore, the costs per 1 kilometre driven by vehicle with capacity c_k marked as l_k are known. For each warehouse $i \in I$, its available capacity f_i is defined by the amount of the normalized products SPO, and its handling costs are marked as e_i. The handling costs are only included for transported products from other warehouse to completion centres. Each warehouse without completion centre could be assigned to the only one warehouse with opened completion centre. The goal is to choose a location of a maximum p completion centres providing completion and dispatch so that the total costs of completion and dispatch of warehouse products are minimal.

3.2 Design of optimisation model

In order to model decision in a mathematical model, we implement three groups of variables into the optimization problem.

The first group of variables consists of bivalent variables x_i modelling the decision on the opening of the completion centre in the warehouse $i \in I$. If after the optimization, $x_i = 1$, it means that the completion centre will be located in the warehouse $i \in I$. If so $x_i = 0$, then it means that the completion centre will not be located in the warehouse $i \in I$.

The second group of variables consists of non-negative integer variables y_{ijk} representing the number of trips made between warehouse $j \in I$ without opened completion centre and warehouse $i \in I$ with open completion centre by the vehicle with capacity c_k, where $k \in C$.

The last group of constrains of bivalent variables z_{ij} modelling the assignment of warehouse $j \in I$ without opened completion centre to warehouse $i \in I$ with open completion centre. If after the optimization, $z_{ij} = 1$, it means that the warehouse $j \in I$ without open completion centre is assigned to warehouse $i \in I$ with open completion centre. If so $z_{ij} = 0$, it means that the warehouse $j \in I$ without open completion centre is not assigned to warehouse $i \in I$ with open completion centre.

The symbol M represents prohibitive constant. If the value of the prohibitive constant was less than the maximum possible volume of trips in all sessions, the binding conditions would malfunction and thus the optimal solution would be negatively affected.

The mathematical model of the problem has the form:

$$\min f(x,y,z) = \sum_{i \in I} \sum_{j \in I} \sum_{k \in C} d_{ij} \cdot y_{ijk} \cdot l_k + \sum_{i \in I} \sum_{j \in I, j \neq i} e_i \cdot a_j \cdot x_i \tag{1}$$

subject to:

$$a_j \cdot (1 - x_j) \leq \sum_{i \in I} \sum_{k \in C} c_k \cdot y_{ijk} \qquad \text{pro } j \in I \tag{2}$$

$$y_{ijk} \leq M \cdot (1 - x_i) \qquad \text{pro } i \in I, j \in I, \text{a } k \in C \tag{3}$$

$$y_{ijk} \leq x_i \cdot M \qquad \text{pro } i \in I, j \in I \text{ a } k \in C \tag{4}$$

$$\sum_{i \in I} x_i \leq p \tag{5}$$

$$\sum_{k \in C} y_{ijk} \leq z_{ij} \cdot M \qquad \text{pro } i \in I \text{ a } j \in I \tag{6}$$

$$z_{ij} \leq \sum_{k \in C} y_{ijk} \qquad \text{pro } i \in I \text{ a } j \in I \tag{7}$$

$$\sum_{j \in I} a_j \cdot z_{ij} \leq f_i \qquad \text{pro } i \in I \tag{8}$$

$$x_i \in \{0,1\} \qquad \text{pro } i \in I \tag{9}$$

$$y_{ijk} \in Z_0^+ \qquad \text{pro } i \in I, j \in I \text{ a } k \in C \tag{10}$$

$$z_{ij} \in \{0,1\} \qquad \text{pro } i \in I \text{ a } j \in I \tag{11}$$

Function (1) represents optimal criterion – the total costs of the completion and dispatch of the required warehouse products. The first part of the criterion represents total transport costs, second part of it represents the handling costs of the completion of consignment in the open completion centres. The group of constrains (2) will ensure the conversion of required products from warehouses without opened completion centre to warehouses with opened completion centre for the volume of trips. The group of constrains (3) will ensure that if the warehouse is the completion centre at the same time, products will not be transported. The group of constrains (4) will ensure bond between values of variables x_i and y_{ijk}. The condition (5) will ensure that the completion centres will be a maximum of p. The groups of conditions (6) and (7) will ensure bond between values of variables z_{ij} for $i \in I$ and $j \in I$ and y_{ijk} for $i \in I, j \in I$ and $k \in C$. The group of the constrains (8) will ensure acceptance of the completion centres capacities. The groups of obligatory conditions (9) – (11) define domains of the variables used in the mathematical model.

4 Calculation experiments with the mathematical model

We will demonstrate the functionality of the model in the following example. The goal is to assess the completion of multi-product shipments in completion centres located in a maximum of p warehouses which are part of the distribution network. The following model data were used in the computational experiments. The distribution network included a total of 19 warehouses, namely Kladno, Kutná Hora, Mělník, Mladá Boleslav, Beroun, Kolín, Benešov, Nymburk, Příbram, Rokycany, Kralupy, Jičín, Poděbrady, Říčany, Čáslav, Přelouč, Louny, Dobříš and Jirny. The distances between the individual warehouses were determined according to [4]. Table 1 shows the weekly customer requirements, handling costs, and available capacity for individual warehouses. For experimental purposes, the products were normalized as the SPO product, which was chosen as the base.

Warehouse	a_i $[ks\ SPO]$	e_i $[CZK \cdot ks\ SPO^{-1}]$	f_i $[ks\ SPO]$	Warehouse	a_i $[ks\ SPO]$	e_i $[CZK \cdot ks\ SPO^{-1}]$	f_i $[ks\ SPO]$
Kladno	5 367	1,89	42 000	Kralupy	1 452	1,6	29 400
Kutná Hora	2 784	2,05	45 000	Jičín	3 687	2,03	30 500
Mělník	3 579	1,92	56 000	Poděbrady	5 784	1,87	30 200
Mladá Boleslav	3 679	1,73	20 000	Říčany	8 795	1,94	35 000
Beroun	1 489	1,84	20 600	Čáslav	2 014	1,67	36 420
Kolín	2 680	1,94	21 540	Přelouč	4 523	1,99	30 600
Benešov	1 754	2	23 500	Louny	9 687	1,84	30 210
Nymburk	7 425	1,8	24 530	Dobříš	5 320	1,91	30 800
Příbram	8 005	1,75	27 000	Jirny	3 840	1,82	20 000
Rokycany	6 500	2,15	28 000				

Table 1 Weekly customer requirements, handling costs, and available capacity of each warehouse

The value of handling costs depends on the technology, technical equipment, and location of the warehouse. There is a heterogeneous vehicle fleet available. Its capacity and cost per 1 km driven are shown in Table 2.

Vehicle k	c_k $[ks\ SPO]$	$l_k [CZK \cdot km^{-1}]$
1	2 000	26
2	4 500	28,5
3	6 000	29
4	3 800	27
5	7 900	30,5

Table 2 Capacity and cost per 1 km driven for each vehicle

The Xpress-IVE optimization software was used to obtain the optimal solution. The experiments were performed on a personal computer equipped with Intel Celeron with parameters of 1.8 GHz and 4.0 GB of RAM.

They were performed for $p = 2, \dots, 19$. Since the solution was not changed in any of the tested cases, only the results of selected optimization calculations are given, summarized in Table 3, and Table 4.

p | warehouse i	3	9	Value OV [CZK]
2	1	1	334 539
3	1	1	334 539
4	1	1	334 539
5	1	1	334 539

Table 3 Optimal location of the completion centre and the value of the optimization criterion

As an optimal solution out of 19 model warehouses with the variant choice p, the model always choses only 2 completion centres located in warehouses 3 (Mělník) and 9 (Příbram). In this case, the value of total costs was CZK 334,539.

Table 4 shows the determined values of the variables $y(i, j, k)$, where i represents the completion centre, j represents the warehouse without the opened completion centre and k represents the assigned vehicle. The values of the variables correspond to the number of trips made from warehouse j to completion centre i. Warehouses 1, 2, 4, 6, 8, 11, 12, 13, 15, 16, 17, and 19 are assigned to the completion centre placed in warehouse 3. The completion centre for remaining warehouses 5, 7, 10, 14, and 18 is placed in warehouse 9. The solution is again the same for all cases.

Trips | *p*	**2**	**3**	**4**	**5**	**Trips | *p***	**2**	**3**	**4**	**5**
y(3,1,3)	1	1	1	1	y(3,17,3)	1	1	1	1
y (3,2,4)	1	1	1	1	y(3,17,4)	1	1	1	1
y (3,4,4)	1	1	1	1	y(3,19,2)	1	1	1	1
y (3,6,4)	1	1	1	1	y(9,5,1)	1	1	1	1
y(3,8,5)	1	1	1	1	y(9,7,1)	1	1	1	1
y(3,11,1)	1	1	1	1	y(9,10,5)	1	1	1	1
y(3,12,4)	1	1	1	1	y(9,14,3)	1	1	1	1
y(3,13,3)	1	1	1	1	y(9,14,4)	1	1	1	1
y(3,15,4)	1	1	1	1	y(9,18,3)	1	1	1	1
y(3,16,3)	1	1	1	1	**Summary**	**19**	**19**	**19**	**19**

Table 4 Number of trips made from warehouse j to completion centre i assigned to individual vehicle k for selected p and their total number

5 Conclusion

The presented article deals with the issue of the location of the completion centre within the distribution network. The aim was to present a mathematical model minimizing the costs of internal distribution between warehouses and completion centres located in the network, including handling and completion costs. The functionality of the model was tested on tasks with a different maximum value of the number of operated completions centres located in the network.

The obtained results clearly show that the available capacity of candidates for the completion centre, the distance of candidates from other warehouses located in the network, and handling costs play a significant role when locating the completion centre. The performed experiments further proved that not only the numbers of warehouses and candidates but also different values of input data within the same number of warehouses and candidates are decisive for computational complexity. For example, in experiments with 19 warehouses and 4 candidates, the calculation time ranged from 189.4 s to 6 hours. The calculation came to an end there without reaching the optimal solution.

It is also necessary to pay attention to the differences in computing time when changing the values of the maximum number of completions centres in a task with the same number of warehouses and candidates. As there is no completion centre capable of completing all requirements, computational experiments were started with completion centres $p = 2$, the maximum number. With the maximum number of completions centres $p = 2$ was found that in order to achieve the minimum costs of 335,538 monetary units, it was necessary to build completion centres in warehouses 3 and 9. Thereupon, the maximum number of completions centres was gradually increased from a value of $p = 2$ to $p = 19$. The values of total costs did not decrease and there were no changes in the location of the centers. However, the computational time has changed, as it is shown in Table 5.

Maximum number of completions centres	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Calculation Time [s]	802,6	358,6	189,4	180,5	140	150,4	154,3	151,6	153,2
Maximum number of completions centres	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**
Calculation Time [s]	151,3	154,2	285,3	280,8	284,5	285,5	285,7	288,5	456,6

Table 5 Change of computation times depending on the complexity of the task

The results of computational experiments showed that in the proposed model, it is more appropriate to start conducting experiments on higher values of the maximum number of candidates. In case of higher values of maximum numbers, the results were achieved in a shorter candidate calculation time. Future research could deal with the influence of one-time fixed costs on the location of the completion centre (cost for cleaning, maintenance, and preparation of unused space), the influence of an asymmetric distance matrix, or the introduction of uncertainties into the model. It would also be interesting to deal with the comparison of the use of homogeneous and heterogeneous vehicle fleets in terms of costs incurred.

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] Cao, S., Liao, W. & Huang, Y. (2021). Heterogeneous fleet recyclables collection routing optimization in a two-echelon collaborative reverse logistics network from circular economic and environmental perspective. *Science of the Total Environment*, 758, 1-15.

[2] Hrníčková, A.; Kováč, M.; Teichmann, D. & Mocková, D. (2022). Mathematical Model for Localization of the Final Shipment Distribution Warehouse at a Distributor Operating in Virtual Reality. *Proceeedings of 21th International Conference on Quantitative Methods in Economics*, 76-81.

[3] Li, J., Ma, Y., Gao, R., Cao, Z., Lim, A., Song, W. & Zhang, J. (2021). Deep reinforcement learning for solving the heterogeneous capacitated vehicle routing problem. *IEEE Transactions on Cybernetics*, Orlando, Florida, USA, 1-14.

[4] Mapy (2022) Available at: https://www.mapy.cz (Accessed: 5.4.2022)

[5] Mirzaei-Khafri, S., Bashiri, M., Soltani, R. & Khalilzadeh, M. (2019). A mathematical model for the capacitated location-arc routing problem with deadlines and heterogeneous fleet. *Transport*, 34(6), 692-707.

[6] Wang, X. & Li, X. (2017). Carbon reduction in the location routing problem with heterogeneous fleet, simultaneous pickup-delivery and time windows. *Procedia computer science*, 112, 1131-1140.

Volatility of Corn Futures with Markov Regime Switching GARCH Model

Michaela Chocholatá[1]

Abstract. This paper deals with the analysis of the volatility of corn futures based on the daily values from January 3, 2018 to March 30, 2022. Both the univariate GARCH model and Markov switching GARCH (MS GARCH) model were estimated to illustrate the switching behaviour of analysed series during the period under consideration. The estimation results of GARCH model proved high volatility persistence of the corn market and the two-regime MS GARCH model enabled to capture various volatility switches during the analysed period.

Keywords: volatility, corn, GARCH model, regime switching, MS GARCH model

JEL Classification: C58, D53
AMS Classification: 62M05, 62M10, 91B84

1 Introduction

To analyse the volatility of agricultural futures has become a challenging issue especially during the last two decades. As pointed out by [6], „corn futures are the most liquid and active market in grains, with 350 000 contracts traded per day". The largest corn growers are the United States. In addition to the United States, the largest corn exporters are Argentina, Brazil, Ukraine and France [19]. The price of corn is related to the prices of other commodities like meat and oil. Higher prices of corn imply expensive livestock feed which, on the other hand, leads to the increases of meat prices. Furthermore, the rise of energy prices has led to using corn for production of e.g., ethanol and biofuel. Futures enable to consider all the outside factors determining the prices (supply, demand, weather, stocks, transportation, political events, etc.), so the prices can thus remain relatively predictable. Nowadays, due to the impacts of COVID-19 pandemic, the tense world stocks and uncertainty caused by the conflict in Ukraine, price volatility (instability) for commodities on the world market is likely to continue [23].

Since the end of 2019 the world has been facing the global COVID–19 crisis caused by the coronavirus SARS-CoV-2. In March 2020 the World Health Organization declared the new coronavirus pandemic. The ongoing COVID-19 pandemic has heavily influenced the economic environment including the financial and commodity markets all over the world. There have been published plenty of studies documenting the impacts of this pandemic. Kotyza et al. [14] analysing the sugar prices, pointed out the substantial increase in the stock market volatility implied by the COVID-19 pandemic. The unprecedented reaction of equity and commodity markets to the novel coronavirus was pointed out by Amar et al. [1]. Analysis of daily prices of soybean futures was published by Yin and Wang [22] who accented that futures price of agricultural products plays a significant role in resource allocation.

Silvennoinen and Thorp [18] studied the bi-variate conditional volatility and correlation dynamics for individual commodity futures (including corn futures) and financial assets. They presented evidence favouring closer commodity and financial market integration during the period 1990 – 2009 using the generalized autoregressive conditional heteroscedasticity (GARCH) class models. GARCH-class framework by forecasting of corn futures volatility was used by Wang and García [20] and Musunuru et al. [15]. Da Silveira et al. [7] categorized the empirical studies into different groups according to the approaches used for investigation of the volatility dynamics in agricultural markets. The focus on the volatility persistence of agricultural commodities can be found e.g., by Balcombe [3]. Živkov et al. [24] used the Markov switching generalized autoregressive conditional heteroscedasticity (MS GARCH) model to analyse the volatility spillover effect between the four agricultural futures – corn, wheat, soybean, and rise. De la Torre-Torres et al. [8] applied the MS GARCH model in the analysis of oil and natural gas futures.

The aim of the paper is to analyse the volatility of corn futures comprising daily data from January 3, 2018 to March 30, 2022 based on the use of univariate GARCH model and MS GARCH model, respectively. Firstly, to explore the volatility persistence of traditional GARCH(1,1) model and that of the two-regime MS GARCH(1,1)

[1] University of Economics in Bratislava, Faculty of Economic Informatics, Department of Operations Research and Econometrics, Dolnozemská cesta 1, 852 35 Bratislava, michaela.chocholata@euba.sk.

model. Secondly, the paper aims to investigate if the timing of volatility switches in MS GARCH(1,1) model is in line with the commonly known turbulent issues like e.g., COVID-19 and outbreak of the war in Ukraine.

The rest of the paper is organized as follows. Section 2 is devoted to methodology issues including the GARCH and MS GARCH framework. Data and empirical estimation results are outlined in section 3 and section 4 concludes.

2 Methodology

In general, the futures prices returns of agricultural commodities (i.e., log returns data yt) do not generally follow a normal distribution, have a nonzero skewness and kurtosis greater than three [7]. The conditional mean of the log returns data is often assumed to be zero and it is supposed that the series is not serially correlated. In case of dynamics in the conditional mean, the Box-Jenkins ARMA models can be used. To capture the time-varying volatility of log returns, the ARCH class models have become very popular during the last three decades. The ARCH model was introduced by Engle [9] and extended to the GARCH model by Bollerslev [4]. Mathematical formulation of the GARCH(1,1) model is as follows:

$$h_t = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1} \tag{1}$$

where the symbols α_0, α_1 and β_1 denote the unknown parameters. To ensure the positivity of conditional variance h_t, it must hold that $\alpha_0 > 0, \alpha_1 > 0$ and $\beta_1 \geq 0$.

Since the financial markets face to calm and turbulent periods, many researchers (see, e.g., [2], [5], [24]) have pointed out that the traditional GARCH class models fail to capture the true variation in volatility in case of structural breaks in empirical time-series. To overcome the overestimation of the volatility persistence and thus the GARCH model misspecification, the traditional GARCH class model can be merged with the Markov switching (MS) model [12]. This combined model, called the MS GARCH model, enables to capture the structural breaks endogenously.

The two-regime MS GARCH specification [11] is as follows [2]:

$$y_t|(s_t = k, I_{t-1}) \sim D\left(0, h_{k,t}, \boldsymbol{\epsilon_k}\right), \tag{2}$$

where I_{t-1} is the information set at time $t-1$, symbol $D\left(0, h_{k,t}, \boldsymbol{\epsilon_k}\right)$ denotes the continuous distribution with zero mean, time-varying conditional variance $h_{k,t}$ and additional shape parameters included in the vector $\boldsymbol{\epsilon_k}$. Regime variable s_t assumes only integer values $k = \{1,2\}$ and is governed by a first-order Markov process. A transition probability matrix **P** of dimension 2×2 contains transition probabilities $p_{ij} = P\{s_t = j|s_{t-1} = i\}$ giving the probability that the regime i (at time $t-1$) will be followed by the regime j (at time t). Since at any time, the variable should be in one of the two considered regimes, it should hold that

$$p_{i1} + p_{i2} = 1, \qquad i = 1,2 \tag{3}$$

The average length of being in a specific regime, i.e. expected duration $E(D_i)$ for $i \in \{1,2\}$ is as follows – see e.g., [5]:

$$E(D_i) = \frac{1}{1-p_{ii}}. \tag{4}$$

The stable probabilities π_i of being in a particular regime ($i = 1,2$) can be calculated as follows:

$$\pi_i = (1-p_{ii})/(2-p_{ii}-p_{jj}). \tag{5}$$

Instead of the normal distribution, e.g., the Student's t-distribution can be used to model the fat-tailed distribution. To estimate both the GARCH and MS GARCH models, the R software packages "rugarch" [10] and "MSGARCH" [2] can be used.

3 Data and Empirical Results

The paper deals with the daily closing prices of corn futures (US cent/1 bushel)[2] traded on Chicago Board of Trade (CBoT) spanning from January 3, 2018 to March 30, 2022. The data were retrieved through the R package "quantmod" [17] from the web-page of Yahoo! [21]. The whole analysis was carried out with the use of softwares EViews and R.

The focus of the paper is on the log returns data, i.e. continuously compounded returns yt, which are graphically depicted in Figure 1 together with the closing prices of the corn futures. Figure 1 includes the basic descriptive statistics of log returns as well. As expected, in case of the financial time series, the mean of daily log returns is close to zero. The returns' daily volatility is of 1.69 %, the distribution of log returns is negatively skewed and highly leptokurtic. From the beginning of 2018 till the first half of 2020 the behaviour of the daily closing prices (left axis) was quite stable, significantly rising tendency was detected during the COVID-19 pandemic in the second half of 2020 and first half of 2021. The third quarter of 2021 was marked by a fall in prices, while in the following winter months the rising trend of prices is clearly visible. This tendency lasted essentially until the end of the analysed period also marked by the outbreak of the war in Ukraine. Furthermore, based on the visual assessment of log returns (right axis) it is possible to identify the volatility clustering, i.e., that periods of low volatility are followed by high volatility periods. The highest volatilities were detected during the first half of 2021.

Descriptive statistics – log returns

Mean	0.0007
Median	0.0007
Maximum	0.0621
Minimum	-0.1910
Std. Dev.	0.0169
Skewness	-1.2984
Kurtosis	19.7426

Figure 1 Daily closing prices (US cents per 1 bushel) and log returns of corn futures; descriptive statistics of log returns

In the next step, the conditional volatility GARCH(1,1) model of log returns was estimated supposing the zero conditional mean and no serial correlation of returns. The estimation is based on the Student's *t*-distributed error assumption with parameter υ indicating the degrees of freedom. The estimation results together with the unconditional volatility σ , degrees of freedom υ of the standardized Student's *t*-distribution and the log-likelihood (LL) are gathered in Table 1. All the parameters were statistically significant at 5% significance level, there was neither autocorrelation in the standardized residuals nor in the squared standardized residuals till the lag 300 at the 1% significance level. The ARCH-LM test results confirmed no remaining heteroscedasticity. The assumption of covariance-stationarity for the GARCH(1,1) model, i.e., $\alpha_1 + \beta_1 < 1$, was fulfilled. The choice of the Student's *t*-distribution was confirmed to be appropriate, since the relatively small degrees of freedom parameter υ implied significant departure from normality [5]. Calculated unconditional volatility σ was quite close to its sample counterpart outlined in Figure 1. The time-varying behaviour of the conditional volatility is depicted in Figure 2 indicating extremely high values in the second half of 2021.

[2] Corn: 1 bushel = approx. 25,4016 kg, corn contract is for 5 000 bushels.

α_0	α_1	β_1	$\alpha_1+\beta_1$	σ	υ	LL
8.10^{-6}	0.0884	0.8910	0.9804	2.0204 %	4.2169	2986.953
**	***	***	-	-	***	-
		Diagnostic test results of standardized residuals				
	Q(300)	$Q^2(300)$	ARCH-LM			
	245.52	278.69	0.0170			

Table 1 GARCH(1,1): Estimation results and diagnostic test results

Note: In all tables the symbols ***, ** and * indicate the rejection of the null hypothesis at 1%, 5% and 10% level of significance, respectively.

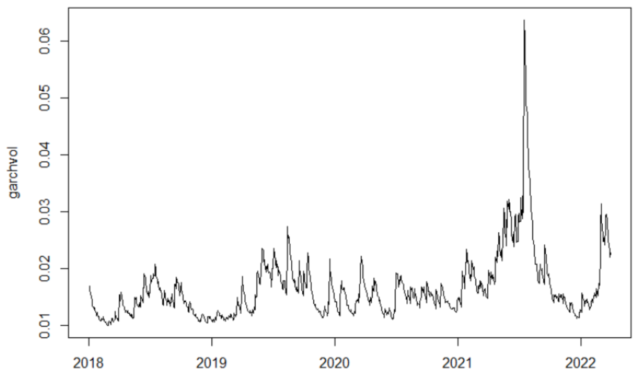


Figure 2 GARCH(1,1): Conditional volatility

Table 2 provides the estimation results of the MS GARCH(1,1) model based on the assumption of Student's *t*-distributed innovations and two regimes – regime 1 (low volatility regime) and regime 2 (high volatility regime). The degrees of freedom parameters υ of the *t*-distribution were fixed across both regimes [5]. Since the estimated value of υ was higher than 2, it was confirmed that the modelled distribution had finite variance and fatter tails than the normal distribution. Almost all estimated parameters proved to be statistically significant (with exception of constant in regime 1). The heterogeneous character across the two regimes is clearly observable concerning the unconditional volatility values. On the other hand, there was almost no difference in within-regime volatility persistence of the MS GARCH(1,1) model across the regimes ($i = 1, 2$). Regime 1, which can be denoted as a low volatility regime, was characterized by only a little bit higher within-regime volatility persistence in comparison to high volatility regime 2. The persistence of regimes (i.e., second source of volatility persistence), is given by transition probabilities p_{11} and p_{22}, i.e., by probabilities of staying in regime 1 and regime 2, respectively (Table 3). The results proved no substantial difference in persistence of regimes. With regard to the stable probabilities (π_1 and π_2), the results proved slightly higher stable probability of the high volatility regime 2 of around 0.5606 and expected duration of 2.2758 days, while the probability of low volatility regime 1 was of around 0.4394 and expected duration of 1.7838 days. As pointed out by Klaassen [12], the LL values can offer an initial view to assess whether the regime persistence is a considerable source of volatility persistence. However, based on results presented in Table 1 and Table 2, respectively, the log-likelihoods corresponding to MS GARCH(1,1) model were lower than their GARCH(1,1) model counterpart, i.e., the results did not confirm that the consideration of regimes could help to capture the volatility persistence.

	α_{01}	α_{11}	β_1	$\alpha_{11}+\beta_1$	σ_1	υ	**LL**
Regime 1	8.10^{-7}	0.0034	0.9869	0.9903	0.8838 %	5.0950	2879.2547
		*	***	-	-	***	-
	α_{02}	α_{12}	β_2	$\alpha_{12}+\beta_2$	σ_2	υ	**LL**
Regime 2	9.10^{-6}	0.0710	0.9191	0.9901	2.9787 %	5.0950	2960.2648
	*	**	***	-	-	***	-

Table 2 MS-GARCH(1,1): Estimation results

p_{11}	p_{21}	π_1	π_2	$E(D_1)$	$E(D_2)$
0.9859	0.0110	0.4394	0.5606	1.7838	2.2758
***	***	-	-	-	-

Table 3 Transition results

Figure 3 shows the smoothed probabilities for both volatility regimes (i.e., states) of the MS GARCH(1,1) model indicating that the smoothed probabilities did not move too often between the two regimes. Furthermore, since the smoothed probabilities were close to 0 or to 1 this pointed out the low level of uncertainty concerning the occurrence of a particular regime in concrete time.

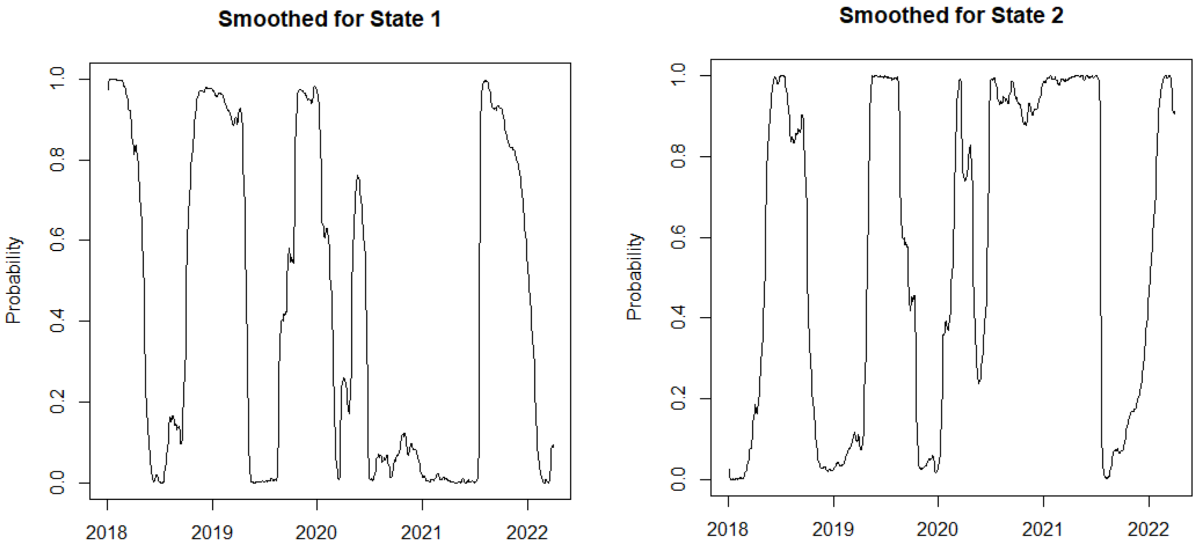


Figure 3 Smoothed probabilities for individual regimes

4 Conclusion

As it is commonly known, corn belongs to one of the most important commodities for humanity and thus the price of corn has been closely monitored by investors. The state of supply and demand has the main impact on the final stock exchange price of corn, the price is also affected by the appreciation or depreciation of the US dollar. Furthermore, the price of corn and its volatility on the stock exchange are affected by climate change and weather changes. Another important factors to be mentioned, concerning the corn prices, are the seasonal cyclical movements and geopolitical situation in the world. The most used financial instruments for corn trading are the socalled futures [16]. The aim of this paper was to analyse the corn futures volatility based both on the traditional GARCH and MS GARCH models. The traditional GARCH(1,1) model indicated high volatility persistence and clearly extremely high volatility values in the second half of 2021. Although the two-regime MS GARCH(1,1) model enabled to capture various volatility switches during the analysed period, the substantial differences neither in within-regime volatility persistence nor in persistence of regimes, was confirmed. Overall, the results did not confirm that the consideration of regimes could help to capture the volatility persistence of corn futures.

Acknowledgements

This work was supported by the Grant Agency of Slovak Republic – VEGA grant no. 1/0193/20 “Impact of spatial spillover effects on innovation activities and development of EU regions” and VEGA 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".

References

[1] Amar, A. B., Belaid, F., Youssef, A. B., Chiao, B. & Guesmi,K. (2021). The unprecedented reaction of equity and commodity markets to COVID-19. *Finance Research Letters*, 38, 101853.

[2] Ardia, D., Bluteau, K., Boudt, K., Catania, L. & Trottier, D.-A. (2019). Markov-switching GARCH models in R: The MSGARCH package. *Journal of Statistical Software*, 91(4), 1–38.
[3] Balcombe K. (2009). The Nature and Determinants of Volatility in Agricultural Prices. https://mpra.ub.uni-muenchen.de/24819/
[4] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327.
[5] Chocholatá, M. (2022). Volatility regimes of selected central European stock returns: a Markov switching GARCH approach. *Journal of Business Economics and Management* (article in press), 1–19.
[6] CME Group. (2022). *Corn Futures and Options*. [Online]. Available at: https://www.cme-group.com/makets/agriculture/grains/corn.html [cited 2022-05-01].
[7] da Silveira, R. L. F., Maciel, L. dos S., Mattos, F. L. & Ballini, R. (2017). Volatility persistence and inventory effect in grain futures markets: evidence from a recursive model. *Revista de Administração*, 52(4), 403–418.
[8] De la Torre-Torres, O.V., Galeana-Figueroa, E. & Álvarez-García, J. A. (2020). Test of Using Markov-Switching GARCH Models in Oil and Natural Gas Trading. *Energies*, 13, 129.
[9] Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987–1007.
[10] Ghalanos, A. (2020*). rugarch: Univariate GARCH models*. R package version 1.4-4. [Online]. Available at: https://CRAN.R-project.org/package=rugarch [cited 2022-04-01].
[11] Haas, M., Mittnik, S. & Paolella, M. S. (2004). A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics*, 2(4), 493–530.
[12] Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2), 357–384.
[13] Klaassen, F. (2002). Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics*, 27(2), 363–394.
[14] Kotyza, P., Czech, K., Wielechowski, M., Smutka, L. & Procházka, P. (2021). Sugar Prices vs. Financial Market Uncertainty in the Time of Crisis: Does COVID-19 Induce Structural Changes in the Relationship? *Agriculture*, 11(2), 93.
[15] Musunuru, N., Yu, M. & Larson, A. (2016). Forecasting Volatility of Returns for Corn Using GARCH Models. *Texas Journal of Agriculture and Natural Resources*, 26, 42–55.
[16] Plecháč, P. (2020). *Kukuřice – komodita pro obchodování na burze, online graf a kurz kukuřice*. [Online]. Available at: https://www.tradecz.cz/kukurice-komodita-pro-obchodovani-na-burze-online-graf-a-kurz-kukurice/ [cited 2022-05-10].
[17] Ryan, J. A., Ulrich, J. M., Thielen, W., Teetor, P. & Bronder, S. (2020). Package 'quantmod'. [Online]. Available at: https://cran.r-project.org/web/packages/quantmod/quantmod.pdf [cited 2022-04-01].
[18] Silvennoinen, A. Thorp, S. (2013). Financialization, crisis and commodity correlation dynamics. *Journal of International Financial Markets, Institutions and Money*, 24 (C), 42–65.
[19] *Trading Economics*. (2022). Corn. [Online]. Available at: https://tradingeconomics.com/commodity/corn [cited 2022-05-01].
[20] Wang, X. & Garcia, P. (2011). Forecasting Corn Futures Volatility in the Presence of Long Memory, Seasonality and Structural Change. *2011 Annual Meeting*, July 24-26, 2011, Pittsburgh, Pennsylvania 103749, Agricultural and Applied Economics Association.
[21] Yahoo! (2022). *Yahoo! Finance*. [Online]. Available at: https://finance.yahoo.com/ [cited 2022-04-09]
[22] Yin, T. & Wang, Y. (2021). Nonlinear analysis and prediction of soybean futures. Agricultural Economics – Czech, 67(5), 200–207.
[23] Združenie pestovateľov obilnín. (2022). Ceny. [Online]. Available at: https://www.obilninari.sk/ceny/ [cited 2022-05-01].
[24] Živkov, D., Kuzman, B. & Subić, J. (2020). What Bayesian quantiles can tell about volatility transmission between the major agricultural futures? *Agricultural Economics* – Czech, 66(5), 215–225.

Pareto Front Approximation using Restricted Neighborhood Search

Jaroslav Janáček[1], Marek Kvet[2]

Abstract. The final decision on deployment of emergency service centers includes balancing between conflicting objectives. Usually, system and fair criteria are considered. The system criterion expresses average response time of the system to randomly emerging demand for service of a system user. The fair criterion takes into account the worst situated users and it can be expressed as a number of users' demands outside a given radius from the nearest located center. To support the balancing process made by the responsible decision makers, a set of non-dominated system designs is to be determined to enable the final decisions under knowledge of consequences. The further presented research has been evoked by successfully applied "sandwich" exact method, where the neighboring members of the Pareto front were determined using a couple of exact optimization methods. Due to big computational time demanded by these methods, the whole process of Pareto front determination needed unacceptable time. We concentrate on speeding up the approach by employing an incrementing heuristic based on swap operation with respect to a given restriction. We study an impact of the heuristic usage on the method efficiency.

Keywords: location problems, conflicting criteria, Pareto front approximation, restricted neighborhood search

JEL Classification: C44, C61
AMS Classification: 90C05, 90C06, 90C10, 90C27

1 Introduction

Designing large-scale service systems in public sector is not easy at all. The combinatorial nature of such problems often requires the application of different methods of operational analysis, mathematical modeling, but also the development of specialized algorithms or even complex software tools to solve them effectively [1, 4, 6, 11, 13]. In this paper, we address a group of discrete network location problems that take into account two conflicting quality criteria.

Various interest groups with different ideas about how the system should work effectively usually accompany making strategic decisions with long-term impacts. This fact can significantly complicate the decision-making process. Consider the following example: If we optimized the average availability of a service for users by minimizing the shortest time in which the service could be provided (so-called system criterion), there would certainly be a small group of people too far away from the nearest source of provided service. Such a system design would be considered unfair [2, 3, 7]. Although they contribute as much to the functioning of the system as others, they have significantly worse access to the service. On the other hand, if we minimized only the number of those who are disadvantaged by their location (so-called fair criterion), we would worsen the value of the former system criterion. Thus, these two mentioned simple objectives are contradictory [5, 7, 12].

Simultaneous optimization of the system and fair criteria is not possible. Improving one criterion entails worsening the other and vice versa. Therefore, it is not appropriate for the output of the search for the optimal location of service centers to be only one solution.

Regarding the output of bi-criteria optimization, it must be noted that one solution is not enough and the set of all feasible solutions of a given problem can be extremely large. For this reason, providing a small subset of solutions in which a specific feature applies seems to be a good way. We call such a subset the *Pareto front* [5, 7, 8, 9]. More details about the Pareto front and its features will be discussed in the following section.

[1] University of Žilina, Faculty of Management Science and Informatics, Univerzitná 8215/1, 010 26 Žilina, Slovakia, jaroslav.janacek@fri.uniza.sk
[2] University of Žilina, Faculty of Management Science and Informatics, Univerzitná 8215/1, 010 26 Žilina, Slovakia, marek.kvet@fri.uniza.sk

Research published in [5, 7] has shown that obtaining the exact Pareto front is a very demanding task requiring enormous computational time. From a practical point of view, it is therefore difficult to imagine. Therefore, experts in the field of applied informatics and operations research are moving towards the development of efficient and fast heuristic methods that would produce a Non-Dominated Solutions Set (*NDSS*) as an approximation of the exact Pareto front. The main scientific content of this paper is aimed at speeding up the former developed approach by employing an incrementing heuristic based on swap operation with respect to a given restriction. We study here an impact of the heuristic usage on the suggested method efficiency.

2 Non-Dominated Solution Set and its Quality

All kinds of the *p*-location problems, which form the background of public service system designing, have one common property - combinatorial nature. Let us focus now on the output of bi-criteria optimization following from the set of all feasible solutions.

Since almost all public service systems have limited human, technical or any other resources, the problem often turns into the selection of exactly p elements from the set of m candidates in order to optimize given criterion. Thus, the set of all feasible solutions of the problem can be formulated by the expression (1).

$$\left\{P \subseteq \{1, \ldots, m\}, |P| = p\right\} \tag{1}$$

It must be noted that the cardinality of this set can be extremely large to process. Nevertheless, the original large set can be presented by a special subset called *Pareto front*. It can be understood as a small set of solutions, in which the property of so-called *non-dominance* holds for each pair of its elements. Let us note that in the bi-criteria optimization, each feasible solution P can be evaluated by two criteria $f_1(P)$ and $f_2(P)$ no matter what form they take. The *non-dominance* can be explained in the following way: A solution P from (1) is called a non-dominated solution if every other solution R from (1) for which $[f_1(P), f_2(P)] \neq [f_1(R), f_2(R)]$ satisfies the following clause $f_1(P) < f_1(R)$ or $f_2(P) < f_2(R)$. The set of all non-dominated solutions is called a *Pareto front* or a *Pareto set* [5, 7, 8, 12]. Thus, the Pareto front is formed by non-dominated solutions, which cannot be improved in one objective unless the other objective is worsened. As the process of the complete Pareto front obtaining is very demanding, we focus on heuristic search of a good approximation of the Pareto front for the bi-criterial problem. The notion of a Pareto front can be easily understood from the left part of Figure 1, in which the black circles correspond to the members of the exact Pareto front and the white ones represent such feasible solutions, which are dominated by the Pareto front members. In the figure, the symbols *MLM* and *MRM* are used to denote the most left and the most right member of the Pareto front respectively.

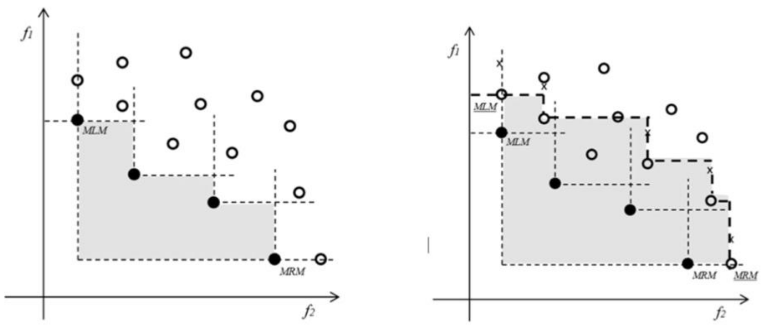


Figure 1 Pareto front and its approximation quality

Another important aspect needed to be solved when working with a Pareto front and its approximation by a set of non-dominated solutions *NDSS*, consists in comparison of two sets (usually the exact and approximate Pareto fronts). Obviously, each of them may have different cardinality. As each member of the approximate set must be either a member of the Pareto set or it must be dominated by some member, we will measure the accuracy by the difference of areas under curves determined by the Pareto front and *NDSS* as shown in the right part of Figure 1. When comparing the sizes of two grey polygons, the bordering solution denoted as *MLM* and *MRM* must either be the same or very near. Otherwise, the comparison becomes useless [9, 12].

If we need to study the dynamics of a heuristic method designed for *NDSS* completion, we can follow these assumptions: The area A_{NDSS} for the approximate set *NDSS* with n elements can be computed according to (2)

$$A_{NDSS} = \sum_{k=1}^{n-1} f_1\left(\boldsymbol{y}^k\right) * \left(f_2\left(\boldsymbol{y}^{k+1}\right) - f_2\left(\boldsymbol{y}^k\right)\right) \tag{2}$$

The area A_{NDSS} or the difference $A_{NDSS} - A_{PF}$, where *PF* denotes the exact Pareto front, may express the quality of the *NDSS* as approximation of the Pareto front. Based on the A_{NDSS}, we suggest a dynamic characteristic of any heuristic. We exploit the fact that any incrementing heuristic inspects the set $\boldsymbol{Y}$ solution by solution and only occurrence of successful candidate corresponds with the instant, when *NDSS* is changed and, simultaneously, the area A_{NDSS} decreases. Let us consider T as a time horizon or testing a heuristic used as a generator of solutions-candidates for *NDSS* update. Let the sequence $\{t_0, t_1, \ldots, t_s\}$ be a set of instants, when the heuristic produced a successful candidate. Furthermore, let $NDSS(t_r)$ denote the *NDSS* after the r-th candidate insertion and let $NDSS(t_0)$ correspond to a starting *NDSS*. Then, the dynamic characteristic of the heuristic can be described by a volume V_{Heur} computed according to (3), or it can be described by the mean value of area $\underline{A}$ computed as $\underline{A} = V_{Heur}/T$. In the case study presented at the end of this paper, we report the mean value of area $\underline{A}$.

$$V_{Heur} = \sum_{r=0}^{s-1} A_{NDSS(t_r)} * \left(t_{r+1} - t_r\right) + A_{NDSS(t_s)} * \left(T - t_s\right) \tag{3}$$

3 Repeated Refinement of Non-Dominated Solution Set

The suggested approach is based on repeated refinement of a current set *NDSS* consisted only of two solutions, where the first one is obtained by minimization of the objective f_2 and the second one is a solution, for which the objective f_1 is as small as possible. During one run of so called outer cycle, the current *NDSS* with *noNDSS* solutions ordered according to increasing values of the objective f_2, the solutions $\boldsymbol{y}^1, \ldots, \boldsymbol{y}^{noNDSS-1}$ are gradually fathomed using a local search and the current *NDSS* is updated. A run of the outer cycle starts with $\boldsymbol{y}^1$ (so called the most left bordering solution). If solution $\boldsymbol{y}^k$ is processed, a local search is applied, which starts with initial solution $\boldsymbol{y}^k$ and improves expression $w_1 f_1(\boldsymbol{y}) + w_2 f_2(\boldsymbol{y})$. During the local search performance, the inspected solutions are evaluated according to objectives f_1 and f_2 and used as candidates for immediate *NDSS* update. When the local search is finished, then the k-th solution of the current *NDSS* is compared to the starting solution (former k-th solution) and if the k-th solution of *NDSS* has not been changed, the refinement process continues with k+1st solution. In the opposite case, the local search starts with the changed k-th solution of *NDSS*. The run of the outer cycle ends with completing of local search of $\boldsymbol{y}^{noNDSS-1}$ unless the solution has been changed during the search. The repeated refinement process continues with next run of outer cycle until the determined computational time T elapses. The repeated refinement process can be described by the following steps.

0. Construct the initial *NDSS* with *noNDSS* solutions. Set t = *currentTime,* set *noOuter* = 0.
1. If *currentTime* - $t \le T$ then set *noOuter* = *noOuter* + 1, set k = 1 and go to Step 2, otherwise terminate and output *NDSS, noNDSS* and *noOuter*.
2. Copy the k-th solution of *NDSS* to $\boldsymbol{y}$ and apply procedure *LocalSearch*(*noNDSS, NDSS*) to the starting solution $\boldsymbol{y}^k$ and then go to step 3.
3. If $\boldsymbol{y}^k \neq \boldsymbol{y}$, then go to step 2, otherwise continue with step 4.
4. If k < *noNDSS* – 1, then set $k = k + 1$ and go to step 2, otherwise go to step 1.

It must be noted that the procedure *LocalSearch* updates both the set *NDSS* and the number *noNDSS* of its solutions. The used procedure *LocalSearch* is based on a neighborhood search according to the best-admissible strategy. The neighborhood of a current solution $\boldsymbol{y}^c$ is defined as a set of feasible solutions, which can be obtained by performing one permitted operation on the current solution. These neighboring solutions are inspected solution by solution. The inspection of a neighboring solution $\boldsymbol{y}$ involves computation $f_1(\boldsymbol{y})$ and $f_2(\boldsymbol{y})$, possible update of *NDSS* by $\boldsymbol{y}$, and update of the best-found solution $\boldsymbol{y}^b$ of the neighborhood from the point of the minimal value of expression $w_1 f_1(\boldsymbol{y}) + w_2 f_2(\boldsymbol{y})$. If the inspected neighborhood contains better solution than $\boldsymbol{y}^c$, then the current solution $\boldsymbol{y}^c$ is updated by $\boldsymbol{y}^b$ and the search is continued with the new neighborhood, otherwise the local search finishes. This searching process may be restricted so that $\boldsymbol{y}^b$ is updated only by such solutions, which satisfy $a_1 f_1(\boldsymbol{y}) + a_2 f_2(\boldsymbol{y}) \le A$.

The procedure *LocalSearch* performs according to the following steps, where *GetNext*($\boldsymbol{y}^c$) gives the next solution of the neighborhood of the current solution $\boldsymbol{y}^c$ and if there is no next solution, it gives dummy solution $\boldsymbol{d}$.

0. Set $\boldsymbol{y}^c$ to the input solution, initialize $\boldsymbol{y}^b = \boldsymbol{y}^c$ and $f^b = w_1 f_1(\boldsymbol{y}^c) + w_2 f_2(\boldsymbol{y}^c)$.
1. Set $\boldsymbol{y} = GetNext(\boldsymbol{y}^c)$.
   If $\boldsymbol{y} = \boldsymbol{d}$, go to step 3, otherwise go to step 2.
2. *UpdateNDSS*($\mathbf{y}$).
   If $w_1 f_1(\boldsymbol{y}) + w_2 f_2(\boldsymbol{y}) < f^b$, then set $\boldsymbol{y}^b = \boldsymbol{y}$ and $f^b = w_1 f_1(\boldsymbol{y}) + w_2 f_2(\boldsymbol{y})$.
   Go to step 1.
3. If $\boldsymbol{y}^b \neq \boldsymbol{y}^c$, then set $\boldsymbol{y}^c = \boldsymbol{y}^b$ and go to step 1, otherwise terminate the local search.

Comment: If the restriction should be used, then the close $w_1 f_1(\boldsymbol{y}) + w_2 f_2(\boldsymbol{y}) < f^b$, in step 2 should be enlarged this way: $(w_1 f_1(\boldsymbol{y}) + w_2 f_2(\boldsymbol{y}) < f^b)$ and $(a_1 f_1(\boldsymbol{y}) + a_2 f_2(\boldsymbol{y}) \leq A)$.

4 P-Location Problem with Conflicting Criteria

The p-location problem with contradictory system and fair criteria studied in this paper can be formulated making use of the following notations. Let I be the set of m candidates, i.e. $I = \{1, 2, \ldots, m\}$, from which exactly p elements are to be chosen to establish a service center. Thus, a simple solution of any p-location problem can be described in two different ways: either by a list of indexes from I containing exactly p elements or by a vector of location variables. The first way was used to define the complete feasible solutions set by (1). The second approach via a vector is more suitable for implementation and practical usage, because it directly corresponds to the mathematical model. The decision on service center locating at a location $i \in I$ is usually modelled by a binary variable $y_i \in \{0, 1\}$, which takes the value of one if a center should be located at i and it takes the value of zero in the opposite case. This way, the following expression (4) can define the set $\boldsymbol{Y}$ of all feasible solutions.

$$\boldsymbol{Y} = \left\{ \boldsymbol{y} \in \{0,1\}^m : \sum_{i=1}^{m} y_i = p \right\} \tag{4}$$

As far as the system and fair criteria are concerned, their mathematical formulations need several new symbols to be introduced. Let the symbol J be used to denote the set of n network nodes, i.e. $J = \{1, 2, \ldots, n\}$, in which the system users are aggregated. Obviously, the sets I and J may be the same and it is no problem if $m = n$. Each element $j \in J$ is connected with a weight coefficient b_j. The value of b_j may have many different interpretations, i.e. the number of system users sharing the location j or expected frequency of demands for service at this node. Furthermore, let the constant t_{ij} represent the traversing time from the service center location i to the user's location j. Since the system criterion follows the concept of so-called *generalized disutility* explained in [8, 10, 12], it is assumed that r nearest located centers can participate in the provision of the service for clients. In the criterion description, q_k stands for probability of the case that the k-th nearest center is the closest one, which is accessible. Operation $\min_k\{\}$ returns the k-th minimal value of the specified list of values. After these assumptions, the system criterion can be formulated according to the expression (5).

$$f_1(\boldsymbol{y}) = \sum_{j \in J} b_j \sum_{k=1}^{r} q_k \min_k \left\{ t_{ij} : i \in I;\ y_i = 1 \right\} \tag{5}$$

The fair criterion (6) expresses the number of users, whose distance from the nearest located service center expressed by time exceeds the given limit T.

$$f_2(\boldsymbol{y}) = \sum_{j \in J} b_j \max \left\{ 0,\ sign\left(\min \left\{ t_{ij} : i = 1, \ldots, m;\ y_i = 1 \right\} - T \right) \right\} \tag{6}$$

5 Numerical Experiments

Presented experiments were aimed at revealing the impact of restriction of the local search on efficiency of the repeated refinement process. Two series of experiments were suggested and performed. The first series minimizes in the local search linear combination of the objectives so that $w_1 = 1$ and $w_2 = 0$ and no restriction was used. The second series uses the same coefficients $\boldsymbol{w}$ in the local search, but the restriction $a_1 f_1(\boldsymbol{y}) + a_2 f_2(\boldsymbol{y}) \leq A$ was imposed on the search. The coefficients were determined so that $a_1 = 0$ and $a_2 = 1$ and $A = f_2(\boldsymbol{y}^{k+1})$.

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 making use of the NetBeans IDE 8.2 environment.

As far as the dataset is concerned, we have built on previous research activities published in [5, 7, 8, 9, 12], in which the data of road network of Slovak self-governing regions were taken into account. Used benchmarks are denoted by the abbreviations of particular regions as depicted in Figure 2.

Figure 2 Administrative organization of Slovakia

In the problem instances, all inhabited network nodes represent both the set of candidates for a service center and the set of system users' locations. The parameter r used in the system criterion (5) was set to 3. The associated coefficients q_k for k=1, …, r were set in percentage in the following way: q_1 = 77.063, q_2 = 16.476 and q_3 = 100 - q_1 - q_2. These values were obtained from a simulation model of existing Emergency Medical Service system in Slovakia [10]. Parameter T used in the fair criterion (6) was set to the value of 10 minutes [5, 7, 8, 9, 12].
Basic characteristics of used problem instances are summarized in Table 1. For each benchmark corresponding to one row of the table, the column denoted by m contains the number of possible service center locations. The parameter p expresses the number of facilities to be located. The middle part of the table contains the number of solutions *NoS* forming the Pareto front. In the column denoted by *Area* we provide the readers with the size of A_{PF}. The right part of the table contains additional information about the bordering points of the Pareto front.

Table 1 Benchmarks characteristics and the exact Pareto fronts features

Region	m	p	*NoS*	*Area*	$f_1(MLM)$	$f_2(MLM)$	$f_1(MRM)$	$f_2(MRM)$
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

It must be noted that an individual run of the whole solving process was limited by 5 minutes threshold. The obtained results are reported in the following two tables. While Table 2 contains the results for the algorithm without any restriction, Table 3 summarizes the obtained results with applying suggested restriction. Both tables have the same structure. Each column represents one instance. The row *CT* contains the computational time in seconds. The objective functions f_1 and f_2 of the most left and the most right member of *NDSS* are not reported due to the fact that they are almost the same as in the exact Pareto front. The number of *NDSS* members is reported in the row denoted by *noNDSS*. The size of area formed by the *NDSS* members is given in the row denoted by *FinArea*. The row denoted by *MeanArea* contains the mean value of area. The notation *noOS* expresses the number of outer cycles performed during the solving procedure.

Table 2 Results of numerical experiments with no restriction

Region	BA	BB	KE	NR	PO	TN	TT	ZA
CT	300.0	488.9	538.3	329.2	642.0	302.8	306.5	317.4
noNDSS	30	160	193	100	178	85	62	89
FinArea	600880	978508	1214873	738046	846817	865033	824330	411994
MeanArea	601010.7	1045510.9	1404536.1	754208.4	916246.2	868713.2	828293.3	422549.4
noOS	1093	1	2	5	1	17	38	11

Table 3 Results of numerical experiments with suggested restriction

Region	BA	BB	KE	NR	PO	TN	TT	ZA
CT	300.1	303.0	345.5	303.4	323.6	303.1	301.9	303.9
noNDSS	31	205	231	99	249	83	64	89
FinArea	581857	955901	1258817	740919	828837	865163	818920	410206
MeanArea	581918.9	998618.3	1299127.9	746261.4	868151.8	867047.3	820978.8	414136.2
noOS	3289	3	5	24	2	82	125	38

6 Conclusions

This paper was aimed at such public service system designing, in which two conflicting criteria are to be taken into account. To overcome high computational demands for obtaining the Pareto front, a heuristic approach based on a simple swap algorithm was suggested. To accelerate the associated solving process, a restriction of the local search was developed. Presented results of experiments using real small and middle-sized benchmarks have confirmed that the restriction enables to make the algorithm significantly more effective by higher number of performed outer cycles. The adjustment helped to obtain better results as concerns the Pareto front approximation quality measured by the value of area and the cardinality of the non-dominated solutions set.

Future research could be aimed at developing several new strategies or even whole heuristics, which could produce a good Pareto front approximation in an acceptably short computational time.

Acknowledgements

This work was supported by the grants VEGA 1/0216/21 “Design of emergency systems with conflicting criteria using artificial intelligence tools” 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 work was supported also 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 Confer- ence 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] 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.

[4] 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.

[5] 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.

[6] Ingolfsson, A., Budge, S. & Erkut, E. (2008). Optimal ambulance location with random delays and travel times. *Health care management science*, 11(3), pp. 262-274.

[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. (2021). Swap Heuristics for Emergency System Design with Multiple Facility Location. In: *Mathematical Methods in Economics*, 2021, pp. 226-231.

[9] Janáček, J. & Kvet, M. (2022). Quality Measure of Pareto Front Approximation. In *Quantitative Methods in Economics 2022*, in print.

[10] 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.

[11] Jánošíková, Ľ. (2007). Emergency Medical Service Planning. In: *Communications* 9(2), pp. 64-68.

[12] Kvet, M. & Janáček, J. (2021). Incrementing Heuristic for Non-Dominated Designs of Emergency Medical System. In: *SOR 2021 Proceedings*, pp. 429-474.

[13] Marianov, V. & Serra, D. (2002). Location problems in the public sector, *Facility location – Applications and theory* (Z. Drezner ed.), Berlin, Springer, pp 119-150.

Comparison of Bond Yields to Maturity Using Hawawini-Vora and IRR Methods

Petr Jiříček[1], Stanislava Dvořáková[2]

Abstract. The aim of the paper is to compare yield rates to maturity for selected bonds. Yield rates to maturity will be calculated using two methods: the classic Hawawini-Vora method and the internal rate of return (IRR) method. The IRR method is generally used to evaluate investments in the financial market. The paper will compare both methods for different types of bonds depending on the yield, maturity and market value of the bonds.

Keywords: bond yield, internal rate of return, Hawawini-Vora method

JEL Classification: C6, G23
AMS Classification: 65H04, 68R10

1 Introduction

The aim of the contribution is the comparison of calculating the yield rate to maturity of bonds by means of two basic methods. The first method is the approximate Hawawini-Vora method [5]. The other method is the internal rate of return method (IRR), originally outlined by Böhm-Bawerk [1]. The recommendation for calculating yield to maturity of bonds by the method of IRR reflects the fact that the IRR method has several significant advantages. These are connected namely with the so-called conventional cash-flows generated by investment to bonds, when any change of the cash-flow polarity occurs. Problems of the IRR method were completely dealt with in the past e.g. by Teichroew et al. [14] in compliance with solving the polynomial by the so-called Descartes' rule. The definition of the conventional type of investment is newly mentioned by Kulakov & Kulakova [9]. Hazen [7] solved the problem of multiple real IRRs. Interpreting complex roots of the IRR method is dealt with by Pierru [12]. Pressacco et al. [13] define the so-called quasi-IRR which may not have real-valued IRR. The topic of using IRRs was summarised by Osborne [11] and Magni [10]. Dhavale & Sarkis [2] dealt with multiple roots analysis of stochastic type of internal rate of return, which can be applicable if the yield rate to maturity of the bond is variable (in the case of inflation-linked bonds).

Unlike this method, other approaches to the determination of the bond yield rate to maturity use approximate methods. Here, four methods of approximation can be historically classified [6]:

i) methods consisting of approximation by direct expansion of the present value of an annuity in powers of the unknown true yield rate;

ii) methods involving approximations by expansion of the present value of an annuity after replacing the unknown true yield rate by the sum of known trial rate near the true yield rate;

iii) methods based on interpolation between two or more trial yield rates giving nearly correct value;

iv) methods providing empirical approximations.

The last methods ignore the time value of money (e.g. discounting of the bond yields). From among these methods, that commonly employed by financial analysts is the YTM method [8] that used the ratio of the yield per period to the average market price of the bond. There is a more accurate method which is widely used, using weight functions instead of a simple arithmetic mean expressing the relationship between the market price and the face value of the bond [5]. After the authors, this is referred to as the Hawawini-Vora method.

[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.

2 Methods, assumption and model formulations

For constructing and analysing the IRR model [3] which methodically comes from the net present value (NPV) of financial investment, the following notation and assumptions are used.

The market price of the bond is defined as the equation:

$$P = \sum_{t=1}^{n} \frac{C}{(1+i)^t} + \frac{F}{(1+i)^n}, \tag{1}$$

where P is the market price of the bond,

C is the coupon value,

F is the face (nominal) value,

n is the maturity of the bond,

i is the exact yield rate to maturity of the bond on financial market.

The general calculation of the yield rate of a bond (i) of n years requires solving a rational fractional function of degree n:

$$NPV = \sum_{t=1}^{n} \frac{C}{(1+IRR)^t} + \frac{F}{(1+IRR)^n} - P = 0, \tag{2}$$

where IRR is the internal rate on return (the numerically solved exact yield rate to maturity of the bond on financial market); it is yield rate to maturity (i) for $NPV = 0$.

There are no general algebraic solutions to equation (1) for n higher than four. Therefore, we must use approximation methods or numerical solutions.

When solving the calculation of the IRR from (2), it is, however, preferable to convert the rational fractional function (1) by the substitution $x = \frac{1}{1+i}$ to a polynomial function

$$f(x) = a_1 x + a_2 x^2 + a_3 x^3 + \cdots + a_n x^n. \tag{3}$$

By using suitable mathematical software (e.g. Maple or Matlab), we can calculate the roots of this polynomial and by reversing $i = \frac{1}{x} - 1$ transform the real-valued roots back to the i variable. The only problem with the conversion is when $x = 0$. Then $i = \infty$.

We do not consider imaginary roots yet because their importance is ambiguous in economic practice in financial markets. For solutions of the yield rate to maturity of the bonds, it is sufficient to find only real-valued positive roots of the polynomial function. This procedure is preferable because there are more algorithms for seeking the root of a polynomial and they are simpler than algorithms for seeking the root of a rational fractional function. To calculate the roots, it is necessary to use numerical methods because, as well known, there are no formulas for analytical calculation of roots of polynomials of a degree higher than 3.

Descartes rule of signs: By means of this rule, it is possible to determine how many positive roots there are for a polynomial function. Assuming that there is a polynomial function $f(x)$, then the number of positive roots is equal to the number of variations of the sign between individual terms of the polynomial. This method finds even the solutions which are not from the set of real numbers, but from the set of complex numbers. The rule says that the number of positive real-valued roots of a polynomial is equal to the maximum of the number of the sign variations.

Several advances commonly employed by financial analysts include the use of an approximate yield rate to maturity of the bond.

1. This approximate yield rate (a) is computed according to Kaplan [4]:

$$a = \frac{C + \frac{F-P}{n}}{\frac{F+P}{2}}. \tag{4}$$

Currently, the abbreviation YTM (yield to maturity) is used for this term.

2. Based on Henderson's approximation of yield rate to maturity of the bond, Hawawini and Vora state for yield rate values in the range 0 < i ≤ 1 (which responds 100% p.a.) and for maturity of the bond in the range n ∈ ⟨1, ∞) new weighting function $\theta\ (n, i)$ in the shape [5]:

$$\theta(n,i) = \frac{(1+i)^n}{(1+i)^n - 1} - \frac{1}{i \cdot n}. \tag{5}$$

Formula (5) represents the new distribution of weights between the face value and the market value of the bond in equation (4). To recalculate the yield rate to maturity approximation value, Hawawini and Vora then assign a weight of 0.4 to the face value of the bond (F) and a weight of 0.6 to the market value of the bond (P).

The Hawawini-Vora derived approximation equation is in the form:

$$AYTM = \frac{C + \frac{F-P}{n}}{0.4\ F + 0.6\ P}. \tag{6}$$

Currently, the abbreviation AYTM (average yield to maturity) is used for this term.

3 Results

A comparison of the two IRR and Hawawini methods mentioned above is made for two different nominal interest rates on the bond, namely 5% p.a. and 12% p.a.

a) As basic parameters for comparing the yield rate to maturity calculation using the IRR and AYTM method (Hawawini-Vora), assume the expected annual coupon yield of the bond (C) of CZK 50. This corresponds to the nominal interest rate on the bond of 5% p.a. from face value of the bond (F) in the amount of CZK 1,000. Assume that the market price of the bond (P) reaches CZK 1,100.

   Figure 1 shows us the course of the dependence of the yield rate on the maturity of the bond (IRR, AYTM) depending on the number of years until the maturity of the bond (n) – under the above assumed parameters.

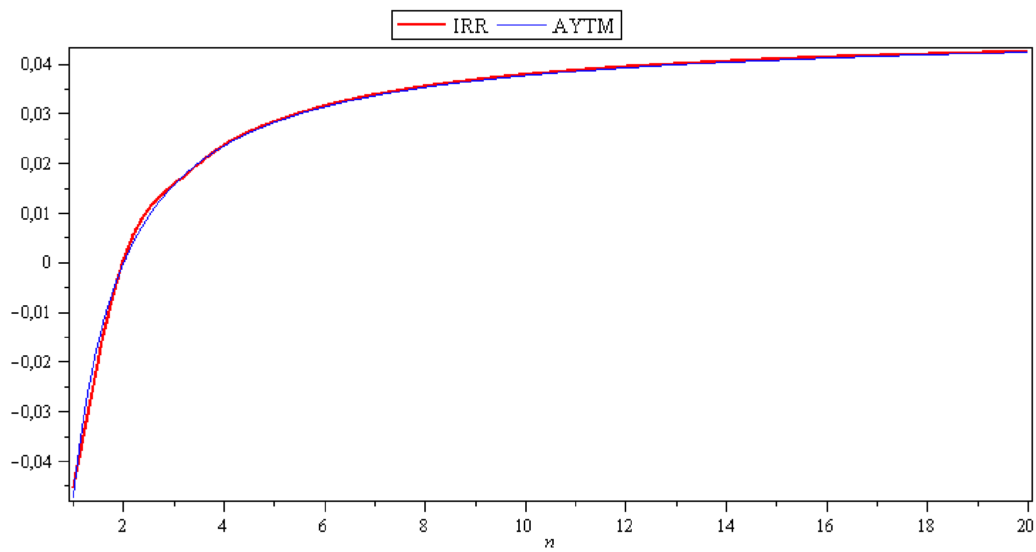


Figure 1 The course of IRR and AYTM for C = CZK 50, F = CZK 1,000; P = CZK 1,100

Now we define for the difference between the approximation of the bond yield to maturity by the Hawawini-Vora method and the value determined by the IRR method in the form:

$$d(C; n, i) = AYTM(C; n, i) - IRR(C; n, i) \tag{7}$$

for each the constant market price of the bond ($P = const$).

Figure 2 shows the difference between the calculated yield rate values to maturity using the IRR and AYTM methods (Hawawini-Vora) depending on the number of years to maturity of the bond at lower inflation rates. We see that the minimum difference is reached from the value of $n = 2$ and $n = 6$.

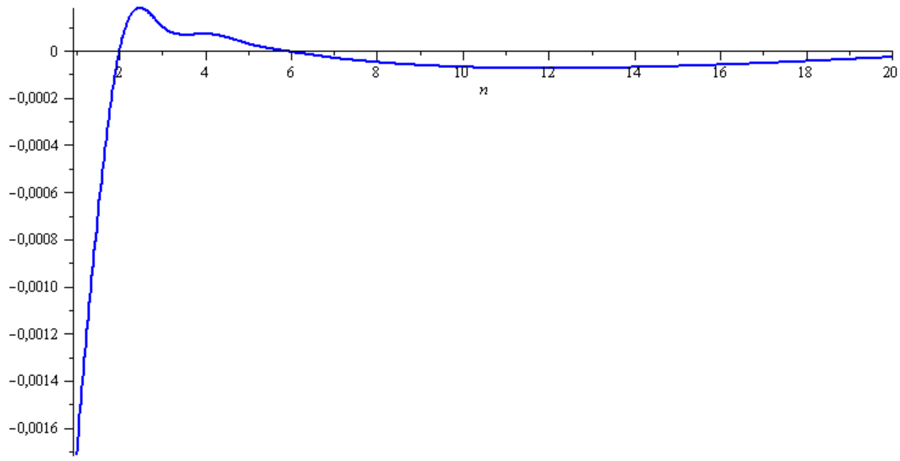


Figure 2 Difference curve (7) between Hawawini-Vora (AYTM) and IRR for C = CZK 50, F = CZK 1,000; P = CZK 1,100

b) Now, for the comparison of the yield rate to maturity calculation using the IRR and AYTM methods (Hawawini-Vora), let us assume the expected annual coupon yield of the bond (C) of CZK 120, which corresponds to the nominal interest rate on the bond of 12% p.a. from face value of the bond (F) in the amount of CZK 1,000. Suppose again that the market price of the bond (P) reaches CZK 1,100.

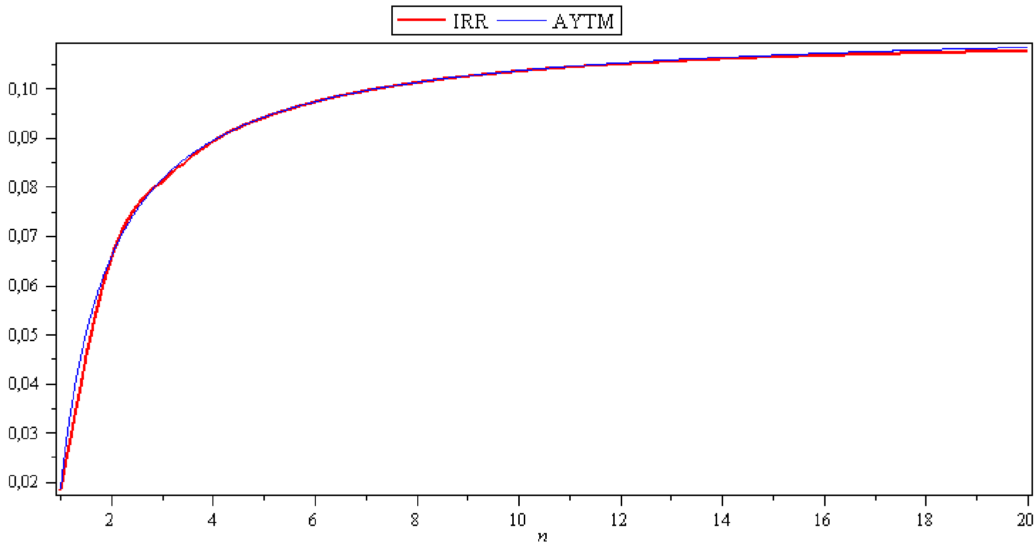


Figure 3 The course of IRR and AYTM for: C = CZK 120; F = CZK 1,000; P = CZK 1,100

Figure 4 shows the difference between the calculated yield rate to maturity values using the IRR and AYTM (Hawawini-Vora) methods, depending on the number of years to maturity of the bond at higher inflation rates. We see that the minimum difference is reached at n = 7.5.

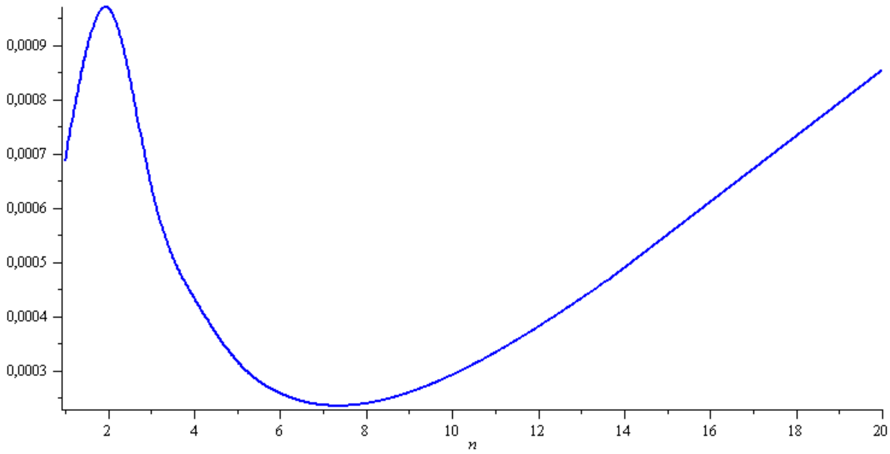


Figure 4 Difference curve (7) between Hawawini-Vora (AYTM) and IRR for C = CZK 120; F = CZK 1,000; P = CZK 1,100

4 Conclusion

The YTM and AYTM are approximation methods commonly used by financial market analysts to calculate the yield to maturity of bonds (i). The AYTM method was newly derived from the original YTM method by Hawawini and Vora using the weight function $\theta(n, i)$. This function reflects the influence of two basic parameters of the bonds, the nominal value stated on the bond at the time of its issue and the market value achieved at the time of its trading on the financial market. The stated aim of the paper was to compare the yield to maturity of the bond (i) using the Hawawini and Vora (AYTM) method with the calculation of the yield to maturity of the bond using the numerical solution of the IRR (internal rate on return) method.

The results show that the difference of the curves $AYTM = \varphi(n)$ and $IRR = \rho(n)$ is not significant, as shown by the course recorded in Figure 2 and Figure 4. However, the minimum of the difference curve $d(C;\ n, i)$ is different for different values of the coupon yield on the bond (C), reflecting the setting of the interest rate of the bond under different conditions of the inflation rate at the issue of the bond. The minimum of this difference under conditions of low inflation occurs for the values of time to maturity $n = 2$ and $n = 6$. Under the conditions of a higher inflation rate, if it is necessary to pay a higher coupon yield on the bond (in the current situation in the economy), this difference is minimal for the values of the time to maturity $n = 7.5$.

Acknowledgements

This research was supported by the College of Polytechnics Jihlava, under Grant no. INT/2022/0005 “Modeling the effectiveness of projects”.

References

[1] Böhm-Bawerk, E. V. (1891). *The positive theory of capital*. Available from internet http:// mises.org/books/positivetheory.pdf.

[2] Dhavale, D. G., & Sarkis, J. (2018). Stochastic internal rate of return on investment in sustainable assets generating carbon credits. *Computers* & *Operation Research*, 89, 324–336. DOI: 10.1016/j.cor.2017.02.014.

[3] Dvořáková, S., & Jiříček, P. (2019). Analysis of multiple economic internal rate of return. In *37th International Conference on Mathematical Methods in Economics 2019* (398–402). České Budějovice: University of South Bohemia in České Budějovice, Faculty of Economics.

[4] Hawawini, G. A., & Vora, A. (1979). *On the Theoretic and Numeric Problems of Approximating the Bond Yield to Maturity*. MPRA Paper 44889, University Library of Munich, Germany.

[5] Hawawini, G. A., & Vora, A. (1981). *On the History of Yield Approx*imation. Available from internet https://flora.insead.edu/fichiersti_wp/inseadwp1981/81-14.pdf.

[6] Hawawini, G. A., & Vora, A. (1982). Yield Approximations: A Historical Perspective. *The Journal of Finance*, 37(1), 145–156. DOI: 10.1111/j.1540-6261.1982.tb01100.

[7] Hazen, G. B. (2003). A new perspective on multiple internal rates on return. *The Engineering Economist,* 48(1), 31–51. DOI: 10.1080/00137910308965050.

[8] Kaplan, S. (1967). Computer Algorithms for Finding Exact Rates of Return. *The Journal of Business,* 48, 389–392.

[9] Kulakov, N. Y., & Kulakova, A. N. (2013). Evaluation of Nonconventional Projects. *The Engineering Economist,* 58(2), 137–148. DOI: 10.1080/0013791X.2012.763079.

[10] Magni, C. A. (2011). Average Internal Rate of Return and investment decisions: a new perspective. *The Engineering Economist* 55(2), 150–180. DOI: 10.1080/00137911003791856

[11] Osborne, M. (2010). A resolution to the NPV-IRR debate? *The Quarterly Review of Economics and Finance* 50(2), 234–239.

[12] Pierru, A. (2010). The simple meaning of complex rates of return. *The Engineering Economics* 55(2), 105–117. DOI: 10.1080/00137911003775081.

[13] Pressacco, F., Magni, C. A., & Stucchi, P. (2011). *A quasi-IRR for a project without IRR*. Available from internet http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1800348.

[14] Teichroew, D., Robichek, A. A., & Montalbano, M. (1965). An analysis of criteria for investment and financing decisions under certainty. *Management Science* 12(3), 151–179.

Performance Evaluation of Lithuanian II Pillar Pension Funds Using Rolling Window Technique

Audrius Kabašinskas[1], Miloš Kopa[2], Kristina Šutienė[3], Aušrinė Lakštutienė[4], Aidas Malakauskas[5]

Abstract. This paper presents results of performance evaluation of Lithuanian II pillar pension funds using rolling window technique. The Lithuanian pension system has three pillars: mandatory (Ist, social insurance system), quasi-optional (IInd, life-cycle pension funds) and optional (IIIrd, any kind of pension funds or insurance). Investments in II pillar from standard funds were changed to life-cycle funds in 2019. To reveal different behavior of market risk and performance of funds, we used 120 days windows (rolled by 1 day). Risk-adjusted performance of funds was measured by employing mean return, average recovery and Sharpe-based ratios such Calmar ratio, Sortino ratio, adjusted Sharpe ratio, VaR Sharpe ratio. However, to describe market risk we only focused on 5 special time windows related to COVID-19.

Keywords: pension funds, rolling window, performance measurement, risk assessment

JEL Classification: C44, D81, G32, G11, H55, J32
AMS Classification: 62P05, 62P20 , 91G15, 91G70

1 Introduction and literature review

As the pension fund sector is developing rapidly around the globe, regulators are constantly exposed to new challenges, which have to be managed to maintain financial stability. The ever-changing environment and the global recession only complicate the work of these institutions, which must respond quickly to emerging risks and make every effort to eliminate or at least reduce those risks. For these reasons, traditional supervision is increasingly being replaced by new requirements that focus on a comprehensive risk assessment. As pension funds outlook should be long-term, it is important that fund managers are able to increase the value of their portfolio by successfully investing in selected asset classes.

Significant differences between European and US pension funds exists, and some countries choose not only to restrict the investment freedom of fund managers but full-out ban investing in certain assets. OECD countries set limits or a total ban on investing in real estate, private equity, or loans [1]. Boon et al. [2] elaborate that stricter funding requirements lead to a decline in risky investment in assets, which is most pronounced during the financial crisis. It is observed that different pension funds managed by companies operating under the same market conditions show different performance results [3-9].

The global financial crisis, ongoing Covid-19 pandemic, both raise scientific debates and questions about how and to what extent regulators and supervisors should influence pension fund investment strategies or allow professional pension fund managers to make short-term decisions to protect fund assets from significant losses. Is it precisely the supervisory authorities that need to pursue only the goals of a long-term strategy?

European Union does not regulate member state pension fund systems, though specific aspects are governed by European regulation on fund management. Each country locally decides upon its pension system's set-up, which most commonly consists of 3 pillars [1]. In Lithuania there are Ist being mandatory, ensures base pension level and is purely managed by local authority. IInd pillar is quasi-optional and regulated by local laws, which could differ vastly between countries: life cycle funds are mandatory in Lithuania [10,11]. Use of funds in IInd pillar is generally limited - the beneficiary can receive a payout only when a certain age is reached in a form of one-off payment or annuity (if the balance is more than a set threshold). Participation in IIIrd pillar is completely optional and is mostly supported indirectly via tax cuts. IIIrd pillar fund set-up or management is not regulated as heavily as IInd pillar

[1] Kaunas University of Technology, Dept. of Mathematical Modeling, Studentu 50-144, Kaunas, Lithuania, audkaba@ktu.lt
[2] Dept. of Mathematical Modeling, Studentu 50-144, Kaunas, Lithuania; Charles University, Dept. of Probability and Mathematical Statistics, Sokolovská 83, Prague, Czech Republic, kopa@karlin.mff.cuni.cz
[3] Kaunas University of Technology, Dept. of Mathematical Modeling, Studentu 50-144, Kaunas, Lithuania, krisuti@ktu.lt
[4] Kaunas University of Technology, School of Economics and Business, Gedimino 50 - 401; Kaunas, Lithuania, ausrine.lakstutiene@ktu.lt
[5] Kaunas University of Technology, School of Economics and Business, Gedimino 50 - 401; Kaunas, Lithuania, aidasmal@gmail.com

(e.g. does not have to follow life-cycle investment strategy) and assets can be freely allocated. Key document describing pension fund management is the prospect which is approved by each country governing body (LT - Bank of Lithuania). Prospect defines all the key characteristics of a fund including pricing, investment areas and strategies.

Adjustments in fund asset allocation are defined by two strategies: strategic asset allocation, which defines key markets and is normally reviewed once per year, and tactic asset allocation, which defines particular sectors and is normally reviewed once per month. Tactic asset allocation is the only point where will be operating in terms of investments. But normally it would not make any drastic adjustments as pension funds follow long-term investment strategy, which ignores short-term adjustments. The investment strategy of the pension fund must be based on a strategic allocation of pension assets which, according to the pension company, aims to ensure an optimal ratio of risky to less risky asset classes throughout the accumulation period, taking into account typical average participant factors such as risk tolerance. The amount of contributions, the accumulated amount, the remaining duration of participation in pension accumulation, the most commonly chosen type of pension benefits, longevity, and important assumptions set by another pension accumulation company. Changes in the investment strategy are considered significant when the share of risky and less risky assets in the pension fund strategy is changed by more than 5 percentage points, and the investment strategy of the pension fund, its implementation and suitability shall be reviewed and evaluated at least once every three years. Pension funds are divided into four groups according to their investment strategy. Most of the second pillar pension funds are “mixed”: assets of these funds are invested into high investment risk asset classes (e.g. equities) and into less risky asset classes (e.g. government bonds) [12]. According to the data of the Bank of Lithuania [13] and the recommendations of the Association of Financial Analysts [14], pension funds in Lithuania are classified into several categories based on the investment strategy in equities: 1) - conservative pension funds (assets under management (hereinafter, AUM) are not invested into equities); 2) - pension funds investing a small part into equities (up to 30 per cent of AUM are invested into equities); 3) - pension funds investing a medium part into equities (up to 70 per cent of AUM are invested into equities); and 4) - pure equity pension funds (up to 100 per cent of AUM are invested into equities). The review by the Bank of Lithuania shows [15] that the asset value accumulated in 2nd pillar has amounted to EUR 5.91 million at 31 December 2021; the number of participants has reached 1.388 million. They are managed by 5 companies: "Aviva Lietuva" (AVIVA), "INVL Asset Management" (INVL), "Luminor investiciju valdymas" (LUMINOR), "SEB investiciju valdymas" (SEB), "Swedbank investiciju valdymas" (SWED). Since the transition of IInd pillar pension funds to life cycle asset allocation approach, the returns have not been homogeneous: robust first year returns were followed by a rapid fall in the first quarter of 2020. Based on data provided by Bank of Lithuania, during the outbreak of Covid-19, pension fund returns fell by almost 15%, while cumulative gains still remained positive at 1.8% [16]. Throughout 2021, pension fund performance has remained positive, while markets remain cautious of on-going COVID-19 pandemic, rising geopolitical tensions due to Russo-Ukrainian war and hindrance of supply value chain on world economy. The decline in global stock indices and bond prices suggests that changes in global equity markets are affecting the performance of pension funds.

In further sections of this paper, we provide idea how to explore risk and performance of pension funds market in rapidly changing economical environment. To understand how market risk changes due to COVID-19 we have estimated correlation in five time windows (120 days each). These correlations were compared to long-term correlation. Moreover, rolling window technique was used to show how performance of pension funds changes over time. Such analysis allow us to understand short-term behavior of Lithuanian pension system during uncertain times. It must be noted that Bank of Lithuania does not regulate short-term fluctuations as they only focus on long-term stability of pension system. However, for participant of pension system short-term fluctuations are psychologically stressful.

2 Methodology

The performance of pension fund is appraised by combining a measure of risk and a measure of reward that are estimated for ex-post daily returns ri, i= 1, $\cdots$, n. In the literature, there is a broad consensus how to estimate the reward, but this is not true for the risk quantification, as the ambiguity surrounding the risk concept could be clearly seen [17,18]. This led to a derivation of numerous risk indices suggested by researchers and practitioners, which implies that we don’t have a unique way to objectively measure risk. However, on the other hand, this could be seen as some sort of advantage, as the risk could be studied from different angles. Therefore, in the paper we focus on one of the most commonly used approaches to measure the performance of some investment by estimating the Sharpe ratio [19]. More specifically, the several distinct ratios that were built on the concept of appraising the reward to variability are quantified by considering different risk measures. Particularly, the greater value is, the better trade-off of risk and return is observed.

We begin with conventional Sharpe ratio. However, as in most cases, we observe non-normally distributed returns, it makes sense to consider the adjusted Sharpe ratio introduced in [20]. This ratio adjusts for skewness and kurtosis by incorporating a reward for a positive skewness and kurtosis less than 3, or a penalty for a negative skewness and kurtosis greater than 3. In order to take into account extreme values possibly observed during Covid-19 crisis, the downside or extreme risk is considered by including risk measures for this purpose in the dominator. For example, Sortino ratio [21], which is another extension of Sharpe ratio, however, in denominator semi-standard deviation is used. Similarly, Calmar ratio [22] uses the maximum drawdown rather than semi-standard deviation in the denominator. Value-at-Risk (VaR) [23,24] is an industry standard for measuring extreme risk, therefore we include Reward-to-VaR measure in the study by computing Sharpe type ratio, with $VaR1{-}\alpha$ in denominator($\alpha=$ 0.05). Additionally, the recovery time or drawdown duration, denoted as Average Recovery, is computed as the time taken to recover. All mentioned ratios are calculated using 120 days rolling window technique when window is rolled by 1 day. Finally, the impact of Covid-19 on the performance of pension funds is estimated by dividing the entire period into smaller sub-periods, and the reallocation of pension funds among clusters is observed.

3 Results

In Lithuania, pension funds are managed by five pension accumulation companies such as AVIVA, INVL, LMNR, SEB, and SWED. Since 2019, the life-cycle concept has been adopted and life-cycle funds have been established. As such, the investment risk of pension funds changes regarding the participants' age. At the beginning of 2020, all financial markets have been hit by Covid-19 crisis, which resulted in a detrimental impact on pension funds results. The historical net asset values observed on daily basis were collected from the websites of pension accumulation companies (PACs), namely AVIVA, INVL, LUMINOR, SEB, and SWED, starting from PFs launch date at the beginning of January of 2019. As each manager operates seven pension funds of different age groups, a data set of 35 PFs was composed. All PFs demonstrated a growth with a varying slope till Covid-19 crisis, during which the funds have experienced a maximum drawdown ranging from 8.7% to 31.01%. The recovery period was long enough, but it seems that all funds achieved their value observed before crisis or even higher. In each age group we can observe that AVIVA funds were outperformed by other funds, while the performance of the funds managed by other PACs varies depending on age group. Therefore, in further analysis the whole period was split into five non-overlapping periods of length 120 days such as 2019/03/19–2019/09/06 (Period A), 2019/09/09–2020/02/28 (Period B), 2020/03/02–2020/08/24 (Period C), 2020/08/25–2021/02/12 (Period D) and 2021/02/15–2021/08/06 (Period E).

On the basis of observed net asset value, the daily simple returns of PFs have been calculated as the main variable to analyse funds. The graphical illustration of returns is depicted in Figure 1.

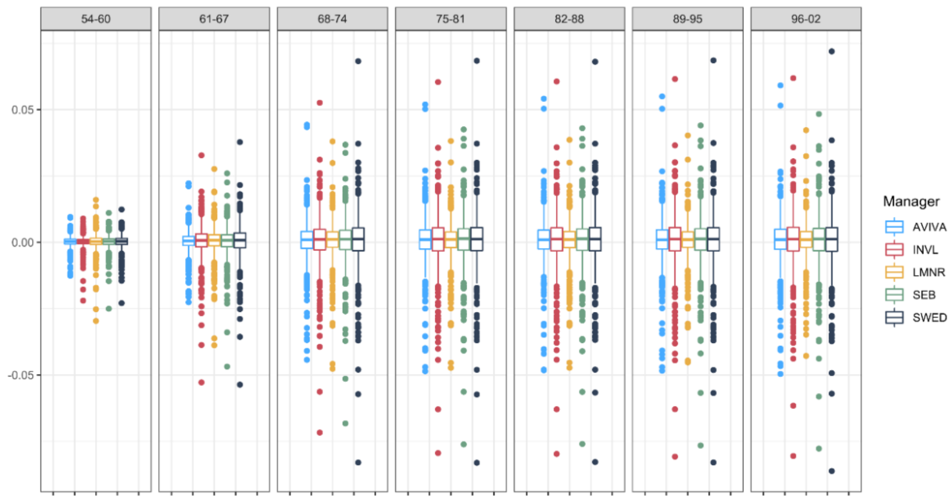


Figure 1 Boxplots of daily simple returns

The visual representation of daily returns in Figure 1 reveals that the deviation of returns is much larger for pension funds of age groups from 68-74 to 96-02, which is a quite expected result because of dominating investments in stocks. Comparing the managers in between, we can observe that LMNR funds are more consistent, except age group 54-60. The largest uncertainty is observed for the funds managed by SWED and INVL, with many observations distributed in the left side of distribution. Comparatively, AVIVA funds are slightly less extreme than others, especially on the negative side of distribution, but still experienced a long negative tail.

To quantify the expected reward and risk of PFs, the Sharpe-based performance measures described in the section "Methodology" have been estimated for each considered period A-E.

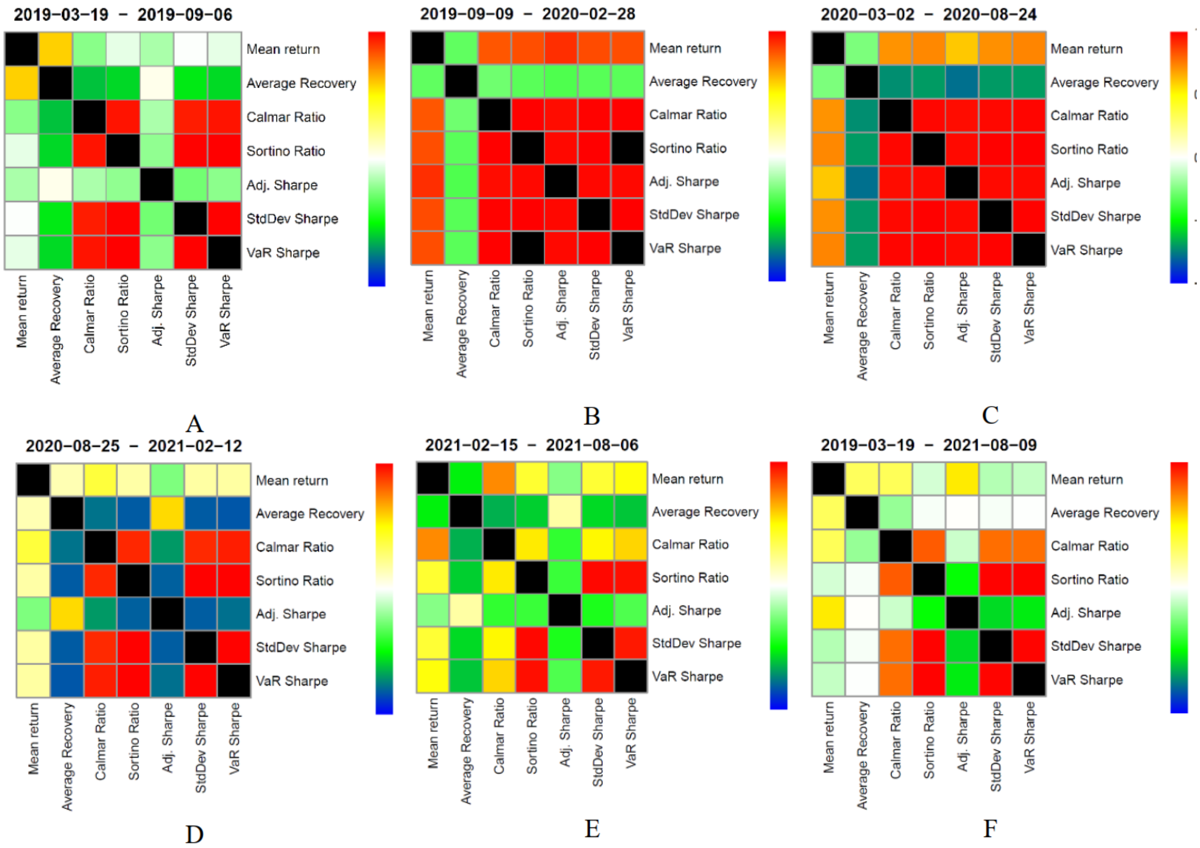


Figure 2 Correlation between measures in periods A–F.

In period A, a strong correlation (see Figure 2) was observed only between Calmar, Sortino and Sharpe ratios. Average recovery and adjusted Sharpe ratios correlate negatively to all other ratios, while mean return correlates slightly (positively only to average recovery). Significant changes in correlations are observed in period B. Mean return begins to correlate strongly positively to nearly all ratios, except average recovery, while the later one remains the only negatively correlated ratio. In period C (just after COVID-19 outbreak) situation is quite similar to period B, however, average recovery correlates more negatively (strong negative correlation) and correlation of mean return becomes weaker. Period D is more similar to period A than to any other period analyzed, however, average recovery and adjusted Sharpe ratios correlate to other ratios more negatively (very strong negative correlation). Finally, in period E the only strong positive correlation remains between Sortino and Sharpe ratios, while correlation of average recovery and adjusted Sharpe ratios increased comparing to period before. Such configuration of correlations means that after one year market situation returns to pre-COVID status.

Figure 3 represents evolution of estimated ratios (mean return, average recovery, Calmar ratio, Sortino ratio, adjusted Sharpe ratio, VaR Sharpe ratio) over entire period by rolling 120 days window and the window rolls by 1 day ahead.

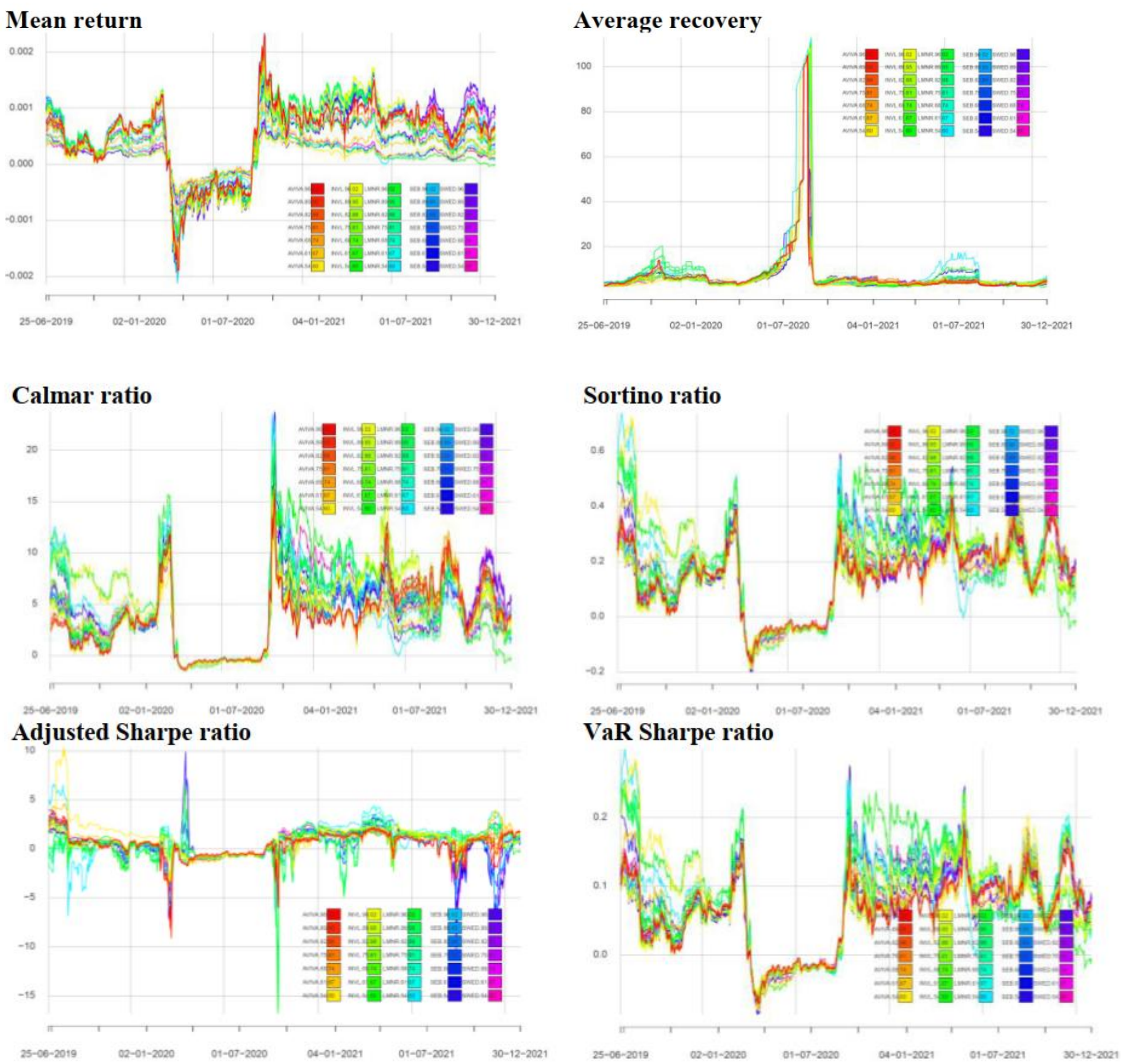


Figure 3 Evolution of estimated ratios (mean return, average recovery, Calmar ratio, Sortino ratio, adjusted Sharpe ratio, VaR Sharpe ratio) over entire period by rolling 120 days window

From Figure 3 we see that Calmar, Sortino and different Sharpe ratios behave very similar under different market conditions. Hence, there is no surprise that their evolution over time (see Figure 3) is quite similar, too: Sharpe ratio decline at the end of March 2020 was observed. The negative level of mentioned ratios was typical for them until September 2020. Finally, the Sharpe ratio increase was followed by high volatility period. Calculation of all these ratios is based on mean return that is why its evolution may look similar to other ratios, especially during COVID decline. Furthermore, evolution of adjusted Sharpe ratio in off-COVID periods (A, D and E) is quite different comparing to previously mentioned ratios. It is interesting to note that adj. Sharpe ratio exhibits large deviations before Sharpe ratio increase or decrease of other ratio and could be used as early warning indicator. Very different evolution of average recovery ratio coincides with observation from Figure 2 where it correlated highly negatively to other ratios. It is interesting that Sharpe ratio decline of average recovery ratio was observed earlier than other ratios reacted to changes in the market that is why it could be used to indicate end of crisis (COVID-19 in this case).

4 Conclusions

In Lithuania, pension funds are managed by five pension accumulation companies, which are AVIVA, INVL, LMNR, SEB, and SWED. Since 2019, the life-cycle concept has been adopted and life-cycle funds have been established. As such, the investment risk of pension funds changes regarding the participants' age. At the beginning of 2020, all financial markets have been hit by Covid-19 crisis, which resulted in a detrimental impact on pension funds results. Therefore, in the study we aimed to compare the performance of pension funds during this crisis by splitting the whole period into five sub-periods of length 120 days. For measuring risk-adjusted performance of funds, the Sharpe-based ratios have been computed to assess the risk in different terms such as standard deviation, semi-standard deviation, value-at-risk, drawdown, skewness and kurtosis. Additionally, the mean return and average recovery have been determined. The results have shown that depending on the time periods, the estimated Sharpe-based ratios may correlate to the different extent. To process and summarize all this analysis, the clustering analysis has been employed, where the optimal number of clusters in each sub-period was determined based on the Silhouette validity index. From this analysis, several findings have been revealed. First, before the crisis, the non-conservative funds belonging to AVIVA, INVL, SEB and SWED demonstrated very similar performance, where the exception has been determined for LMNR funds. Not surprisingly, the most conservative funds that are proposed for the age group of 54-60, were split into different clusters. Second, in subsequent periods, the migration of pension funds from one cluster to another has been observed whatmay suggest that it has been the impact of few waves observed for COVID-19 crisis. Third, no differences over time have been observed among funds for age groups of 68-74, 75-81, 82-88, 89-95, 96-02 managed by the same manager whatmay raise a question about the uniqueness of investment strategies they apply.

All PFs demonstrated a growth with a varying slope till Covid-19 crisis, during which the funds have experienced a maximum drawdown ranging from 8.7% to 31.01%. The recovery period was long enough, but it seems that all funds achieved their value observed before crisis or even higher. In each age group we can observe that AVIVA funds were outperformed by other funds, while the performance of the funds managed by other PACs varies depending on age group, and we can clearly observe the impact of COVID-19 crisis on PFs performance. In both crisis and non-crisis periods, life cycle fund performance, across all pension fund managers was homogeneous. This suggests that different PFs invest in uniform financial instruments and markets. Since all PFs reacted similarly to COVID-19 it can be argued that the regulatory impact on funds is uniform across all fund managers, and though fund strategies are formulated individually, differences were not evident.

Acknowledgements

This project has received funding from the Research Council of Lithuania (LMTLT), agreement No S-MIP-21-32.

References

[1] OECD (2021). https://www.oecd.org/finance/private-pensions/globalpensionstatistics.htm .
[2] Boon, L.N. & Briere, M. & Rigot, S. (2018). Regulation and pension fund risk-taking. *Journal of International Money and Finance*, 84(C), 23–41.
[3] Strumskis, M. & Balkevicius, A. (2016). Pension fund participants and fund managing company shareholder relations in Lithuania second pillar pension funds. *Intellectual Economics*, 10(1), 1–12.
[4] Bula, R. (2020). Transition matrix and stochastic kernel for repeatability assessment of performance of polish open pension funds. *Journal of Entrepreneurship and Sustainability Issues*, 8, 984–1005, 12.
[5] Kabasinskas, A. & Sutiene, K. & Kopa, M. & Valakevicius, E. (2017). The risk and return profile of Lithuanian private pension funds. Economic Research-*Ekonomska Istraživanja*, 30(1), 1611–1630.
[6] Kabasinskas, A. & Sutiene, K. & Kopa, M. & Luksys, K. & Bagdonas, K. (2020). Dominance-based decision rules for pension fund selection under different distributional assumptions. *Mathematics*, 8(5).
[7] Artiga González, T. & van Lelyveld, I. & Lučivjanská, K. (2020). Pension fund equity performance: Patience, activity or both? *Journal of Banking and Finance*, 115 (C):S0378426620300790.
[8] Dopierala, L. & Mosionek-Schweda, M. (2021). Pension fund management, investment performance, and herding in the context of regulatory changes: new evidence from the polish pension system. *Risks*, 9(1).
[9] Kopa, M. & Kabašinskas, A. & Šutienė, K. (2022). A stochastic dominance approach to pension fund selection. *IMA journal of management mathematics* 33(1), 139–160.
[10] Parliament of the Republic of Lithuania (2018). *Pension accumulation law in the Republic of Lithuania* (Lietuvos Respublikos pensijų kaupimo įstatymas).
[11] Bank of Lithuania (2018). *Regulation on pension fund management*, 2018.
[12] Medaiskis, T. & Gudaitis, T. (2017). Evaluation of second pillar pension funds supply and investment strategies in baltics. *Journal of Business Economics and Management*, 18(6), 1174–1192.
[13] Bank of Lithuania (2012). *Rules on usage of benchmark indices* (Lyginamųjų indeksų naudojimo taisyklės). Decision nr. 03-155 (2012-07-12).
[14] Association of Financial Analysts (2004). *Recommendations*. https://www.cfainstitute.org/about.
[15] Bank of Lithuania (2021). *Results of pension funds*. https://www.lb.lt/lt/pf-veiklos-rodikliai.

[16] Bank of Lithuania (2020). *Analysis and forecast of economics*. https://www.lb.lt/lt/mv-ekonomikos-analize-ir-prognozes.
[17] Levy, H. (2016). *Stochastic Dominance - Investment Decision Making under Uncertainty*. Springer.
[18] Bacon, C.R. (2019). *Performance Attribution: History and Progress.* CFA Institute Research Foundation Publications.
[19] Bacon, C.R. (2021). *Practical Risk-Adjusted Performance Measurement*. John Wiley & Sons.
[20] Pezier, J. & White, A. (2006). *The Relative Merits of Investable Hedge Fund Indices and of Funds of Hedge Funds in Optimal Passive Portfolios.* ICMA Centre Discussion Papers in Finance, icma-dp2006-10, Henley Business School, Reading University.
[21] Sortino, F. & van der Meer, R. (1991). Downside Risk. *Journal of Portfolio Management* 17(4), 27–31.
[22] Young, T.W. (1991). Calmar Ratio: A Smoother Tool. *Futures Magazine*, October.
[23] Dowd, K. (1998). Beyond Value at Risk: *The New Science of Risk Management*. John Wiley & Sons. Alexander, G. J. & Baptista, A. M. (2003). Portfolio Performance Evaluation Using Value at Risk. *Journal of Portfolio Management* 24(4), 93–102.

A Bootstrap Comparison of Robust Regression Estimators

Jan Kalina[1], Patrik Janáček [2]

Abstract. The ordinary least squares estimator in linear regression is well known to be highly vulnerable to the presence of outliers in the data and available robust statistical estimators represent more preferable alternatives. It has been repeatedly recommended to use the least squares together with a robust estimator, where the latter is understood as a diagnostic tool for the former. In other words, only if the robust estimator yields a very different result, the user should investigate the dataset closer and search for explanations. For this purpose, a hypothesis test of equality of the means of two alternative linear regression estimators is proposed here based on nonparametric bootstrap. The performance of the test is presented on three real economic datasets with small samples. Robust estimates turn out not to be significantly different from non-robust estimates in the selected datasets. Still, robust estimation is beneficial in these datasets and the experiments illustrate one of possible ways of exploiting the bootstrap methodology in regression modeling. The bootstrap test could be easily extended to nonlinear regression models.

Keywords: linear regression, robust estimation, nonparametric bootstrap, bootstrap hypothesis testing

JEL Classification: C14
AMS Classification: 62F40

1 Introduction

As the linear regression represents the most fundamental model in current econometrics [5], it is crucial to estimate its parameters without being excessively influenced by the presence of outliers in the data [11]. Robust regression estimators started to become established alternatives to the least squares since late 1960s [18]. The practical analysis of economic data however lags behind the current trends in mathematical statistics, although new robust estimators have recently been developed and investigated. The robust estimates are more variable (less efficient) compared to the least squares for non-contaminated models and choosing the robust fit in every situation is not necessarily optimal. The robust procedures are either intended to replace the least squares as self-standing estimators, or they have the potential to accompany the least squares as a sort of diagnostic procedures [1]. In the second situation, the user does not have to decide whether the least squares estimator is reliable or whether the robust fit is preferable. In any case, users of robust statistics should also have the ambition to compare the performance of several methods and to decide for the method that is able to outperform other methods.

The approach based on understanding robust regression estimators as diagnostic tools for the least squares has been developed from the very dawn of robust estimation [6]. Such approach is still topical in current data analysis, as documented e.g. by the application of the least weighted squares estimator [24] in the study of [9], where robust analysis is presented primarily as a tool revealing non-robustness of a standard data analysis. An example of such a recent standing-alone methodology is the robust regression by means of the method of moments of [2], which is reliable under contaminated as well as non-contaminated models. In addition, robust estimators start to obtain their own diagnostic tools (cf. [25]). In general, if a robust estimator yields a (sufficiently) different result from the least squares, the user should investigate the dataset closer and search for explanations [17]; in this context, a formal hypothesis test would be very useful. Let us however proceed with formulating the test problem carefully.

In standard terminology, a coefficient estimate is a realization of an estimator obtained for the sample at hand. With the sample data in place, two different estimates are either different or not. Still, it may be useful to ask whether the expectations of two alternative estimates are equal or not. Thus, the question is how to perform a formal hypothesis test of equality of the means of two alternative estimators, especially for small sample sizes. If the two estimators are consistent, they are already asymptotically unbiased and any such test is redundant. Still, the test may be meaningful for situations with finite samples when assuming consistency is not desirable. In fact,

[1] The Czech Academy of Sciences, Institute of Information Theory and Automation, Pod Vodárenskou věží 4, Prague 8, Czech Republic & The Czech Academy of Sciences, Institute of Computer Science, Pod Vodárenskou věží 2, Prague 8, Czech Republic, kalina@cs.cas.cz
[2] The Czech Academy of Sciences, Institute of Information Theory and Automation, Pod Vodárenskou věží 4, Prague 8, Czech Republic, janacekpatrik@gmail.com

some of the robust estimators such as the least trimmed squares or least weighted squares require to assume a lengthy list of technical assumptions in order to achieve consistency; also, each of the estimators has its own set of specific assumptions [23, 24]. Not relying on the consistency has also the advantage that a possible extension of the test to nonlinear regression is straightforward; consistency properties remain unknown for some robust nonlinear regression esteimators (including common types of robust neural networks [20]). This motivates our aim to propose a test of equality of the means of two alternative linear regression estimators based on nonparametric bootstrap. We recall that nonparametric bootstrap represents a popular methodology for estimating variability (i.e. the covariance matrix) of various robust regression estimates [13].

Testing equality of the means of two regression estimates, which seems not be mentioned in recent robustness literature [6], was discussed as an important topic in financial applications in the recent paper [12]. There, a Hausman-type test was proposed to compare the mean of the MM-estimate with the mean of the least squares fit. We recall that the Hausman test based on the difference of the two-stage least squares and the ordinary least squares is based on the asymptotic covariance matrix of the difference; the test is meaningful in econometric models with endogenic variables, where the least squares estimator is not consistent [5]. Naturally, deriving an asymptotic test of H_0 requires to derive the asymptotic covariance matrix of $\hat{\beta}^A - \hat{\beta}^B$ and cannot be obtained only as a combination of two individual covariance matrices for $\hat{\beta}^A$ and $\hat{\beta}^B$. Nevertheless, the work [12] compared two estimators that are consistent and asymptotically normal, exploiting known formulas for their asymptotic covariance matrices. Because we do not want to assume asymptotic normality to be available, we resort to a nonparametric bootstrap procedure. In Section 2, a nonparametric bootstrap test of equality of two means of two alternative linear regression estimates is proposed. Its performance over three real economic datasets is presented in Section 3. Section 4 brings conclusions.

2 Comparing two regression estimates

Throughout the paper, we consider the standard linear regression model

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

which may be expressed in the matrix notation as $Y = X\beta + e$. Here, we consider p fixed regressors (predictors) available for the total number of n observations (measurements). In our notation, $\beta = (\beta_1, \ldots, \beta_p)^T$ is the vector of parameters, where the i-th row of X will be denoted as $X_i = (X_{i1}, \ldots, X_{ip})^T$ for $i = 1, \ldots, n$. We assume the errors to be independent identically distributed; such random sampling assumption justifies using nonparametric bootstrap in (1), and we need the assumptions of the so-called classical linear regression model [5], i.e. $\mathsf{E}\, e = 0$, $\mathsf{var}\, e = \sigma^2 \mathcal{I}$ for a $\sigma > 0$, $\mathsf{E}\, X^T e = 0$, and $h(X) = p$. While the consistency of the least squares is already ensured in the classical linear regression under these assumptions, we do not assume additional assumptions required to achieve consistency of robust estimators in (1).

Our aim is to test equality of two regression estimates of β in (1). These two estimates will be denoted as

$$\hat{\beta}^A = (\hat{\beta}_1^A, \ldots, \hat{\beta}_p^A)^T \quad \text{and} \quad \hat{\beta}^B = (\hat{\beta}_1^B, \ldots, \hat{\beta}_p^B)^T, \tag{2}$$

while we typically take one to be the least squares estimator and the other to be one of available robust estimators. We are interested in testing equality of the expectations of these two estimators. Formally, the null hypothesis can be expressed as

$$H_0: \ \mathsf{E}\hat{\beta}_j^A = \mathsf{E}\hat{\beta}_j^B \tag{3}$$

for any of the indexes $j = 1, \ldots, p$, against the corresponding two-sided alternative hypothesis. The true but infeasible expectations naturally depend on the distribution of errors, while we omit any distributional assumptions here, and on the particular contamination of the data (severity of contamination, type of outliers).

Principles of bootstrap (resampling) applicable to confidence intervals as well as hypothesis tests have been well known in econometrics and were thoroughly discussed e.g. in [4] and references presented therein. While the range of commonly used bootstrapping approaches is quite broad, we rely here on the nonparametric bootstrap, which is known to perform well in regression modeling tasks [4]. A confidence interval for $(\hat{\beta}^A - \hat{\beta}^B)$ is simply obtained from the bootstrap distribution for the random variable $(\hat{\beta}^A - \hat{\beta}^B)$. The approach for bootstrap-based testing of equality of the means of two alternative regression estimates is formally proposed in Algorithm 1. Algorithm 1 exploits the effective idea to select the bootstrap samples once for all j and only afterwards to construct the confidence intervals separately for every particular $j = 1, \ldots, p$. We recall that consistency of both $\hat{\beta}^A$ and $\hat{\beta}^B$ already ensures (3) to hold. The presented nonparametric bootstrap test is inspired by [8], where nonparametric bootstrap estimation for robust regression was discussed and presented with an algorithm.

3 Experiments

We consider three real publicly available datasets. These were selected as datasets, where it is meaningful to explain a continuous response by a linear regression model using several regressors. From the original data, we however keep only continuous regressors, omitting all discrete ones. Let us first describe these datasets, which all have an economic background and do not contain any missing values.

Algorithm 1 Nonparametric bootstrap test of H_0 (3) of equality of means of $\hat{\beta}_j^A$ and $\hat{\beta}_j^B$ for a given $j \in \{1, \dots, p\}$

Input: Data rows $(X_{i1}, \dots, X_{ip}, Y_i)$, $i = 1, \dots, n$
Input: $j \in \{1, \dots, p\}$
Input: $S > 0$
Output: Decision function of the test of H_0 in (3)
1: **for** $s = 1$ to S **do**
2: Generate n bootstrap samples

$$({}_{(s)}X_{j1}^*, \dots, {}_{(s)}X_{jp}^*, {}_{(s)}Y_j^*), \quad j = 1, \dots, n, \tag{4}$$

by sampling with replacement from $(X_{i1}, \dots, X_{ip}, Y_i)$, $i = 1, \dots, n$.
3: Consider a linear regression model in the form

$${}_{(s)}Y_j^* = {}_{(s)}\gamma_0 + {}_{(s)}\gamma_1 {}_{(s)}X_{j1}^* + \dots + {}_{(s)}\gamma_p {}_{(s)}X_{jp}^* + {}_{(s)}v_j, \quad j = 1, \dots, n, \tag{5}$$

with random errors ${}_{(s)}v_1, \dots, {}_{(s)}v_n$.
4: Compute the estimator $\hat{\beta}_A$ of β in (5).
5: Compute the estimator $\hat{\beta}_B$ of β in (5).
6: $\hat{\theta}_s^j := (\hat{\beta}_j^A - \hat{\beta}_j^B)$.
7: **end for**
8: Arrange the values $\hat{\theta}_1^j, \dots, \hat{\theta}_S^j$ in ascending order as

$$\hat{\theta}_{(1)} \leq \dots \leq \hat{\theta}_{(S)}. \tag{6}$$

8: Construct the 95 % confidence interval as

$$\left[\hat{\theta}^j_{(\lfloor h \rfloor)}, \hat{\theta}^j_{(\lfloor n-h \rfloor)}\right], \tag{7}$$

where $h = \lfloor 0.025n \rfloor$ and $\lfloor x \rfloor$ denotes the greatest integer smaller or equal to x.
9: Reject H_0 (3) if and only if the confidence interval (7) does not cover 0.

The datasets represent important bench- marking data well known in robust statistics, as they contain outliers and robustness is known to be meaningful and beneficial for their modeling.

- Cirrhosis dataset with $n = 46$ and $p = 4$ available e.g. in [21]. The death rate from cirrhosis is considered in individual U.S. states. This response is explained by
  - X_1 = percentage of urban population,
  - X_2 = number of late births,
  - X_3 = wine consumption per capita, and
  - X_4 = consumption of hard liquor per capita (X_4).
- Education dataset with $n = 50$ and $p = 3$ contained e.g. in the package [22]. Per capita expenditures on public education are considered in the 50 U.S. states. This response is explained by
  - X_1 = number of residents in urban areas,
  - X_2 = per capita personal income, and
  - X_3 = percentage of individuals below 18 years of age in the population.
- Pasture dataset with $n = 67$ and $p = 3$ available e.g. in [21]. The rental price of pastures in different places in Minnesota is considered. This response is explained by
  - X_1 = rent of arable land,
  - X_2 = number of milk cows per square mile, and
  - X_3 = difference between pasturage and arable land.

3.1 Robust estimators

The least squares estimator, which is used as the reference estimator here for comparing with robust estimates, is computed by the function lm of R software. In the computations, we use the following robust estimators.

1. Least trimmed squares (LTS) [18]. In the computations, we use the function ltsReg of [22] with the trimming constant h = ⌊3n/4⌋. Properties of the LTS were derived in [23].
2. LTS-RLS, which denotes the LTS estimator accompanied by the reweighted version (reweighted least squares) described in [18]. We use again h = ⌊3n/4⌋.
3. MM-estimator with breakdown point equal to 0.5 and with efficiency equal to 0.95. Properties of MM-estimators were derived in [15]. For the computation, we use the function lmrob of [22].
4. LWS-lin, defined as the least weighted squares (LWS) estimator [24] with linearly decreasing weights [8]; properties of the LWS estimator bLWS of β were derived in [24].
5. LWS-log, defined with weights generated by the logistic function [8].
6. LWS-trim, defined with trimmed linear weights [8].
7. LWS-err, defined with weights exploiting the (so-called) error function [8].

Except for the LWS, which remains much less known in the econometric community, the considered robust estimators can be characterized as well established tools. We do not present results of S-estimators on the given data because of numerical instability of their implementation in [22].

3.2 Results

We perform all computations in R software [16] exploiting the robustbase package [22]. Point estimates of the differences $\hat{\beta} - \hat{\beta}^{LS}$ evaluated for the robust estimates of Section 3.1 are presented in Table 1. We do not present results for the intercept, because we understand the test to be meaningful only for the slopes. As revealed in the table, highly robust methods yield quite different results from non-robust methods. Particularly, LTS-RLS seems yield the estimates most different from those of the least squares, while MM-estimators and all versions of the LWS represent more or less a compromise between the least squares and LTS-RLS.

The table presents also nonparametric bootstrap confidence intervals for the differences $\hat{\beta} - \hat{\beta}^{LS}$ obtained always with $S = 1000$ bootstrap samples. Using the bootstrap confidence intervals for hypothesis testing, all obtained intervals cover the value 0. Thus, the bootstrap test yields no significant result on the usual level of 5 %. In other words, we do not find any significant difference between any two estimates. This is true in spite of the already mentioned differences among the point estimates corresponding to different estimation procedures.

The regression estimates that have narrower confidence intervals should be preferable for practical applications. In the cirrhosis dataset, the narrowest confidence intervals are those comparing the mean slopes of LTS and LWS-trim with the slopes of the least squares. In the education dataset, MM-estimator is the best and LWS-log remains only slightly behind. In the pasture dataset, LTS and MM-estimator have the narrowest confidence intervals and all versions of the LWS fall behind.

4 Conclusions

Robust estimators for the linear regression model have already established their position in the analysis of econometric data, although some promising estimators with a high breakdown point remain to be almost unknown to the econometric community. A bootstrap test of equality of the means of two regression estimates is developed in this paper based on a bootstrap confidence interval for the difference between the two estimators. We are particularly interested in the LWS estimator which can be characterized as an estimator with only rare applications; see [24] for the LWS in linear regression or [7] for the LWS in the location model. Still, the proposed bootstrap test can be used to compare any two robust estimators so that its usage is not limited to the LWS estimator.

The numerical study is performed here for three real economic datasets. Rejecting the null hypotheses would require more observations in our datasets, because of relatively large values of variances of individual regression estimates. Actually, significance remains achievable only for a very heavy contamination for $n < 70$, although point estimates subjectively seem very different from each other. This is an interesting result as such: on one hand, robust estimation is typically applied to handle small samples [10], but on the other hand, the variability of robust estimators (e.g. of the LWS estimator) has not been sufficiently investigated (see [8]). The results also reveal the difficulty of reliable regression modeling under small samples.

Estimator	Intercept	Mean difference for the regressor X_1	X_2	X_3	X_4
Cirrhosis dataset ($n = 46$)					
LTS	17.8	0.617	-1.56	0.38	1.44
		$[-1.233; 2.477]$	$[-3.30; 0.18]$	$[-2.49; 3.25]$	$[-0.48; 3.36]$
LTS-RLS	-3.3	0.438	-1.36	0.04	1.38
		$[-1.321; 2.209]$	$[-3.10; 0.38]$	$[-2.68; 2.76]$	$[-0.48; 3.24]$
MM	1.9	0.053	-0.13	0.01	0.09
		$[-2.095; 2.195]$	$[-2.22; 1.96]$	$[-2.94; 2.96]$	$[-1.90; 2.08]$
LWS-lin	19.4	0.619	-1.54	-0.27	-0.03
		$[-1.106; 2.344]$	$[-3.02; 0.34]$	$[-2.72; 2.25]$	$[-1.79; 1.73]$
LWS-log	3.5	0.349	-0.46	-0.47	-1.38
		$[-1.418; 2.113]$	$[-2.18; 1.26]$	$[-3.10; 2.16]$	$[-3.19; 0.43]$
LWS-trim	0.4	0.297	-0.25	-0.53	-1.51
		$[-1.303; 1.907]$	$[-1.77; 1.27]$	$[-2.89; 1.83]$	$[-3.32; 0.30]$
LWS-err	5.0	0.196	-0.47	-0.20	0.23
		$[-1.616; 2.015]$	$[-2.23; 1.29]$	$[-3.01; 2.61]$	$[-1.65; 2.11]$
Education dataset ($n = 50$)					
LTS	264	0.126	-0.04	-0.56	-
		$[-0.372; 0.624]$	$[-0.16; 0.08]$	$[-1.93; 0.81]$	-
LTS-RLS	309	0.070	-0.02	-0.73	-
		$[-0.380; 0.520]$	$[-0.12; 0.08]$	$[-1.83; 0.37]$	-
MM	278	0.068	-0.02	-0.66	-
		$[-0.232; 0.368]$	$[-0.09; 0.05]$	$[-1.36; 0.04]$	-
LWS-lin	275	0.072	-0.03	-0.64	-
		$[-0.278; 0.422]$	$[-0.11; 0.05]$	$[-1.44; 0.16]$	-
LWS-log	322	0.066	-0.03	-0.78	-
		$[-0.264; 0.396]$	$[-0.11; 0.05]$	$[-1.58; 0.02]$	-
LWS-trim	345	0.075	-0.03	-0.80	-
		$[-0.255; 0.405]$	$[-0.11; 0.05]$	$[-1.63; 0.03]$	-
LWS-err	127	0.028	-0.01	-0.74	-
		$[-0.342; 0.398]$	$[-0.10; 0.08]$	$[-1.64; 0.16]$	-
Pasture dataset ($n = 67$)					
LTS	3.77	-0.102	-0.175	3.91	-
		$[-0.278; 0.074]$	$[-0.478; 0.128]$	$[-0.80; 8.60]$	-
LTS-RLS	4.67	-0.113	-0.104	1.60	-
		$[-0.269; 0.043]$	$[-0.387; 0.179]$	$[-1.50; 4.70]$	-
MM	1.57	-0.056	-0.018	2.03	-
		$[-0.162; 0.050]$	$[-0.242; 0.206]$	$[-1.07; 5.13]$	-
LWS-lin	1.77	-0.086	0.029	0.04	-
		$[-0.196; 0.024]$	$[-0.201; 0.259]$	$[-3.06; 3.14]$	-
LWS-log	3.77	-0.152	-0.048	-1.37	-
		$[-1.102; 0.798]$	$[-0.252; 0.156]$	$[-4.17; 1.43]$	-
LWS-trim	4.40	0.155	0.138	1.69	-
		$[-0.805; 1.115]$	$[-0.072; 0.348]$	$[-1.21; 4.59]$	-
LWS-err	0.77	0.036	0.008	-0.45	-
		$[-0.084; 0.156]$	$[-0.232; 0.248]$	$[-3.65; 2.75]$	-

Table 1 Results of the experiments over three datasets. For each dataset and estimate $\hat{\beta}$, point estimates of $\hat{\beta} - \hat{\beta}^{LS}$ are given, where $\hat{\beta}^{LS}$ is the least square estimate of the true β. Bootstrap confidence intervals for the estimated differences $\hat{\beta} - \hat{\beta}^{LS}$ are also given.

We can make a general conclusion that robust statistics has shifted since its origins in direction to self-standing efficient methods. A comparison of point estimates (without any hypothesis test) based entirely on a visual inspection of the presented tables was common in early books on robustness (such as [18]), but such approach becomes outdated and we recommend to consider point estimates to be always accompanied by bootstrap estimates of their variability. Such estimates were presented e.g. in [8] however only for data with $p = 1$. It is also necessary to mention that bootstrap as a computational technique, helpful in solving various practical questions (e.g. estimating the covariance matrices of regression estimators), has not been so much acknowledged in theoretical approaches to robust statistics [6].

Higher attention of robust statisticians should be paid to methods for high-dimensional data [3]. The robustbase package [22] for robust statistical methods contains mainly datasets that are even smaller than the datasets analyzed here. We intend to perform further computations on larger data as well as to extend the bootstrap test procedure to robust multivariate (or high-dimensional) estimators [14] or robust neural networks [20], for which there are no available results on consistency.

Acknowledgements

This work was supported by the Czech Science Foundation grant GA21-05325S ("Modern nonparametric methods in econometrics"). The authors would like to thank Barbora Peštová for discussion about the nonparametric bootstrap procedure. Jan Tichavský´ helped with the implementation and a reviewer provided helpful comments.

References

[1] Cerioli, A., Riani, M., Atkinson, A.C., & Corbellini, A. (2018). The power of monitoring: How to make the most of a contaminated multivariate sample. *Statistical Methods & Application*, 27, 559–587.
[2] Čížek, P. (2016). Generalized method of trimmed moments. *Inference*, 171, 63–78.
[3] Filzmoser, P. & Todorov, V. (2011). Review of robust multivariate statistical methods in high dimension. *Analytica Chimica Acta*, 705, 2–14.
[4] Godfrey, L. (2009). *Bootstrap tests for regression models*. London: Palgrave Macmillan.
[5] Greeene, W.H. (2018). *Econometric analysis*. 8th edn. Harlow: Pearson Education Limited.
[6] Jurečková, J., Picek, J. & Schindler, M. (2019). *Robust statistical methods with R*. 2nd edn. Boca Raton: CRC Press.
[7] Kalina, J. (2018). A robust pre-processing of BeadChip microarray images. *Biocybernetics and Biomedical Engineering*, 38, 556–563.
[8] Kalina, J., Tobišková, N. & Tichavský, J. (2019). A nonparametric bootstrap comparison of variances of robust regression estimators. In *Conference Proceedings MME 2019, 37th International Conference Mathematical Methods in Economics*. University of South Bohemia, České Budějovice, 168–173.
[9] Kalina, J., Vašaničová, P. & Litavcová, E. (2019). Regression quantiles under heteroscedasticity and multicollinearity: Analysis of travel and tourism competitiveness. *Ekonomický časopis/Journal of Economics*, 67, 69–85.
[10] Kitchenham, B., Madeyski, L., Budgen, D., Keung, J., Brereton, P., Charters, S., Gibbs, S. & Pohthong, A. (2017). Robust statistical methods for empirical software engineering. *Empirical Software Engineering*, 22, 579–630.
[11] [Kreinovich, V., Sriboonchitta, S. & Huynh, V.N., Eds. (2017). *Robustness in Econometrics*. Cham: Springer.
[12] Maravina, T.A., Martin, R.D. (2022). A Hausman type test for differences between least squares and robust time series factor model betas. *Journal of Mathematical Finance*, 12, 411–434.
[13] Maronna, R.A., Martin, R.D. & Yohai, V.J. (2019). *Robust statistics. Theory and methods*. 2nd edn. Hoboken: Wiley.
[14] Marozzi, M., Mukherjee, A. & Kalina, J. (2020). Interpoint distance tests for high-dimensional comparison studies. *Journal of Applied Statistics*, 47, 653–665.
[15] Omelka, M. & Salibiáan-Barrera, M. (2010). Uniform asymptotics for S- and MM-regression estimates. *Annals of the Institute of Statistical Mathematics*, 62, 897–927.
[16] R Core Team (2019). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna. URL http://www.R-project.org/.
[17] Rousseeuw, P.J. & Hubert, M. (2018). Anomaly detection by robust statistics. *WIREs Data Mining and Knowledge Discovery*, 8, e1236.
[18] Rousseeuw, P.J. & Leroy, A.M. (1987). *Robust regression and outlier detection*. New York: Wiley.

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

[20] Rusiecki, A. (2013). Robust learning algorithm based on LTA estimator. *Neurocomputing*, 120, 624–632.

[21] Späth, H. (1992). *Mathematical algorithms for linear regression*. San Diego: Academic Press.

[22] Todorov, V. & Filzmoser, P. (2009). An object-oriented framework for robust multivariate analysis. *Journal of Statistical Software*, 32, 1–47.

[23] Víšek, J. Á. (2006). The least trimmed squares. Part III: Asymptotic normality. *Kybernetika*, 42, 203–224.

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

[25] Víšek, J. Á. (2019). S-weighted instrumental variables. In Skiadas, C.H. & Bozeman, J.R., Eds.*: Data analysis and applications 1: Clustering and regression, modeling-estimating, forecasting and data mining*. Vol. 2. London: Wiley, 53–72.

Single Sampling LTPD Plans for Inspection by Variables with Known Standard Deviation

Nikola Kaspříková[1]

Abstract. The paper addresses the Lot Tolerance Proportion Defective plans. The Lot Tolerance Proportion Defective plans which minimize the mean inspection cost per lot of the average process quality were originally designed by Dodge and Romig for sampling by attributes. The plans for the inspection by variables have been then proposed for a quality characteristic which follows normal distribution. The variables sampling plans generally allow to achieve significant savings in the mean inspection cost.

The variables inspection plans available so far include single sampling plans for the case that the standard deviation of the quality characteristic is not known. This paper addresses the case of the known standard deviation. The tables of plans for single sampling by variables when the standard deviation is known are provided in this paper for several values of the input parameters. The software implementation of the calculation of plans which may be useful for the situations not covered in the tables is discussed.

Keywords: acceptance sampling, single sampling plans, inspection by variables, LTPD

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

The paper addresses the Lot Tolerance Proportion Defective plans. The lot tolerance proportion defective (LTPD) acceptance sampling plans were designed by Dodge and Romig to minimize the mean number of items inspected per lot of the process average quality when the remainder of rejected lots is inspected. Such plans are called rectifying plans. The plans were first designed by Dodge and Romig for the inspection by attributes, see [2]. Plans for the inspection by variables and for the inspection by variables and attributes (all items from the sample are inspected by variables, the remainder of rejected lots is inspected by attributes) were then proposed and it was shown that these plans are in many situations more economical than the corresponding Dodge-Romig attribute sampling plans. The LTPD plans for inspection by variables and attributes have been introduced in [7], using approximate calculation of the plans. Exact operating characteristic, using non-central t distribution, has been later implemented for the calculation of the plans in the LTPDvar package [6]. The operating characteristics used for these plans are discussed by Jennett and Welch in [3] and by Johnson and Welch in [4]. It has been shown that these plans may perform better than the original attribute sampling plans, the economic analysis is published in [8]. The calculation of the LTPD variables sampling plans is implemented in the R extension package [6], covering both operating characteristics shown in [3] and [4]. The package also covers the new LTPD variables plans which are using the exponentially weighted moving average (EWMA) statistic in the inspection procedure to reflect the recent development in acceptance sampling plans design, for details and references see [6].

The variables inspection plans available so far include single sampling plans for the case that the standard deviation of the quality characteristic is not known. This paper addresses the case of the known standard deviation.

The structure of this paper is as follows: first, the design of the original Dodge-Romig LTPD sampling plans for the inspection by attributes (see [1]) is recalled. Then we recall the design of the the LTPD variables sampling plans as shown in [8] and introduce the plans for the known standard deviation case. The tables of plans for single sampling by variables when the standard deviation is known are provided in this paper for several values of the input parameter. The calculation and economic evaluation of the plans is done using the free [6] software which has been published on the Comprehensive **R** Archive Network.

[1] Prague University of Business and Economics, Department of Mathematics, Nám. W. Churchilla 4, Praha, Czech Republic

2 Dodge-Romig LTPD plans

In this section, let us recall the LTPD attribute inspection plans first. For the inspection procedures in which each inspected item is classified as either good or defective (the acceptance sampling by attributes), Dodge and Romig (see [1]) consider sampling plans (n, c) which minimize the mean number of items inspected per lot of process average quality Is, assuming that the remainder of the rejected lots is inspected

$$I_s = N - (N-n) \cdot L(\bar{p};n;c) \tag{1}$$

under the condition

$$L(p_t;n;c) \leq \beta, \tag{2}$$

where $L(p,n,c)$ is the operating characteristic (the probability of accepting a submitted lot with the proportion defective p when using the plan (n,c) for acceptance sampling),
N is the number of the items in the lot,
$\bar{p}$ is the process average quality,
p_t is the lot tolerance proportion defective ($P_t = 100p_t$ is the lot tolerance per cent defective, denoted LTPD),
n is the number of items in the sample $(n < N)$,
c is the acceptance number (the lot is rejected when the number of defective items in the sample is greater than c).

The requirement (2) provides a guarantee for the consumer that the lots of the unsatisfactory quality level, with the proportion defective p_t are going to be accepted only with the specified probability β at most (the so-called consumer's risk). The standard value of 0.1 is used for the consumer's risk by dodge and Romig in [1].

3 LTPD single sampling plans for the inspection by variables

The plans for the inspection by variables are introduced in this section. As a more economic alternative to the sampling plans for the inspection by attributes, the LTPD plans for the inspection by variables were designed in [7]. The LTPD plans were designed under the following assumptions: The measurements of a single quality characteristic X are independent and identically distributed normal random variables with unknown parameters μ and σ^2. For the quality characteristic X, either an upper specification limit U (the item is defective if its measurement exceeds U), or a lower specification limit S (the item is defective if its measurement is smaller than S), is given.

Select a random sample of n items in the lot, calculate sample mean $\bar{x}$ and sample standard deviation s and then accept the lot if

$$\frac{U-\bar{x}}{s} \geq k \quad \text{or} \quad \frac{\bar{x}-S}{s} \geq k. \tag{3}$$

The exact operating characteristic for this case is (see the approximative and the exact operating characteristic in [3] and [4])

$$L(p,n,k) = \int_{k\sqrt{n}}^{\infty} g(t,n-1,u_{1-p}\sqrt{n})dt, \tag{4}$$

where $g(t,n-1,u_{1-p}\sqrt{n})$ is probability density of the noncentral t distribution with $(n-1)$ degrees of freedom and noncentrality parameter $u_{1-p}\sqrt{n}$, where u_{1-p} is $(1-p)\cdot 100\%$ quantile of the standard normal distribution.

In situations when the standard deviation of the quality characteristic of interest is known, the procedure for the acceptance sampling is slightly modified and the standard deviation of the quality characteristic σ is used in place of the sample standard deviation s in the procedure described above. The operating characteristic to be used in such case is then [9]

$$L(p,n,k) = \Phi(\sqrt{n}(u_{1-p}-k)), \tag{5}$$

where Φ is the distribution function of the standard normal distribution.

The plan parameters (n,k) are determined so that the plan has optimal economic characteristics and satisfies the requirement (2), when (5) is used as the operating characteristic.

The optimal economic characteristics in this paper mean that the mean inspection cost per lot of the process average quality is minimized

$$I_{ms} = N - (N-n) \cdot L(\bar{p};n;k). \tag{6}$$

The plot of the operating characteristic curve for two example sampling plans is shown in Figure 1. The operating characteristic used in the plot is the function (5) and the plans are to be used in the known standard deviation case. The plot illustrates the effect of increasing the critical value k.

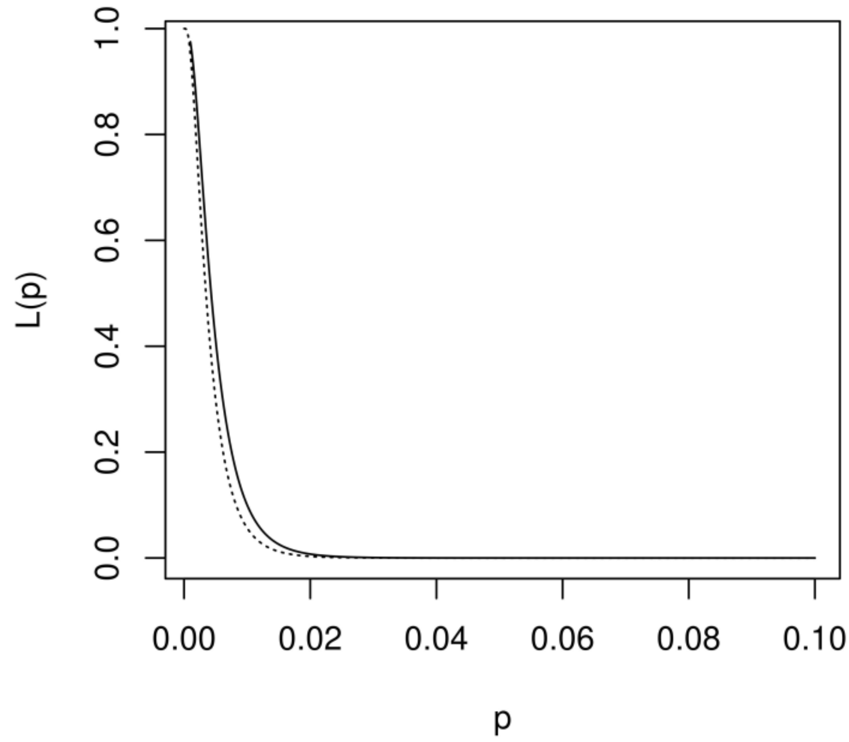


Figure 1 Operating characteristic curves of plans n=26 and k=2.63 or k=2.7 (dotted)

4 LTPD plans for known standard deviation

In the following case study we calculate the LTPD acceptance sampling plan for sampling inspection by variables if the standard deviation of the quality characteristic is known. The economic performance of the plan will be evaluated with the mean inspection cost per lot of the process average quality and compared with the corresponding acceptance sampling plan calculated for the situation of unknown standard deviation of the quality characteristic.

Example 1. A lot with $N = 1000$ items is considered in the acceptance procedure. The lot tolerance proportion defective is given to be $p_t = 0.01$, and the process average quality is $\bar{p} = 0.001$. Find the LTPD acceptance sampling plan for sampling inspection by variables when remainder of rejected lots is inspected, under the assumption of known standard deviation of the quality characteristic.

The plan can be calculated making use of the functions available in the LTPDvar package for the R software [10], see the documentation of the package for a more detailed description.

The solution is n = 26, k = 2.57768. The mean inspection cost per lot of the process average quality for this plan is 30.36. The solution for the case of unknown standard deviation is n = 85, k = 2.627151 with the mean inspection cost per lot of the process average quality equal to 104.67, see [5].

It may be observed that the value of the mean inspection cost per lot of the process average quality of the plan obtained for the known standard deviation case is considerably lower than the mean inspection cost per lot of the process average quality of the plan obtained for the unknown standard deviation case. This reflects lower uncertainty in case of the known standard deviation.

The Figure 2 shows the operating characteristic curve for plan n = 26, k = 2.57768.

The values of the input parameters influence the resulting sampling plan and its economic characteristics. The Table 1 shows the optimal sampling plans calculated for various values of the process average quality $\bar{p}$, keeping

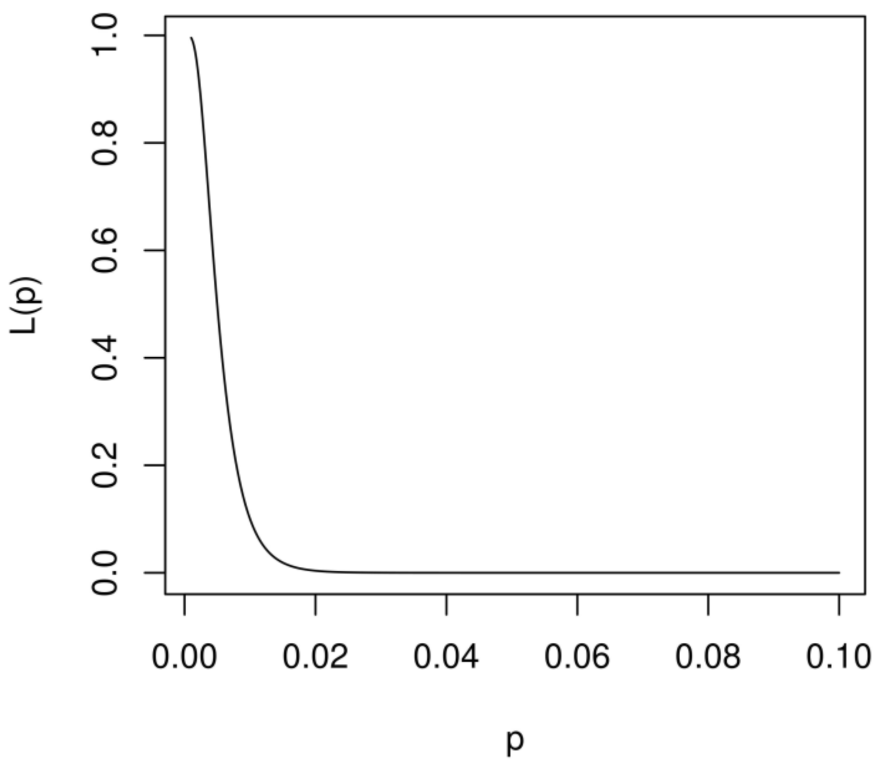


Figure 2 Operating characteristic curve of plan (26, 2.57768)

the other parameters in our example unchanged. It can be observed that the sample size increases if the process average quality becomes closer to the lot tolerance proportion defective. This behaviour is similar to that observed also for attributes sampling plans and plans discussed in [7].

$\bar{p}$	n	k
0.0005	18	2.628412
0.0006	19	2.620356
0.0007	21	2.606005
0.0008	23	2.59357
0.0009	24	2.587943
0.0010	26	2.577681
0.0011	27	2.572983
0.0012	29	2.564326
0.0013	31	2.556521
0.0014	33	2.549437
0.0015	34	2.546132

Table 1 LTPD plans for $p_t = 0.01$, $N = 1000$

Example 2. Let us change some of the input parameter values from Example 1. We will consider the lot tolerance proportion defective p_t = 0.005 first. The optimal plan for this case is $n = 49$, $k = 2.758908$.

The Table 2 and Figure 2 show the situation after parameter update in Example 2. The sample size again increases if the value $\bar{p}$ becomes closer to p_t. From the comparison of Table 2 and Table 1 it can be observed that lower value of p_t (i. e. closer to $\bar{p}$) leads to higher sample size.

Example 3. Let us change the lot size from Example 1. In case that the lot size is N = 2000, we get plan $n = 29$, $k = 2.564326$ for $\bar{p}$=0.001.

$\bar{p}$	n	k
0.0005	29	2.813807
0.0006	32	2.802378
0.0007	36	2.789421
0.0008	40	2.77846
0.0009	44	2.76903
0.0010	49	2.758908
0.0011	53	2.751864
0.0012	58	2.744105
0.0013	63	2.73729
0.0014	68	2.73124
0.0015	73	2.725824

Table 2 LTPD plans for $p_t = 0.005$, $N = 1000$

From the comparison of Table 3 and Table 1 it can be observed that larger lot size has led to an increase in sample size.

$\bar{p}$	n	k
0.0005	20	2.612912
0.0006	22	2.599576
0.0007	23	2.59357
0.0008	25	2.582658
0.0009	27	2.572983
0.0010	29	2.564326
0.0011	31	2.556521
0.0012	33	2.549437
0.0013	35	2.54297
0.0014	37	2.537034
0.0015	39	2.53156

Table 3 LTPD plans for $p_t = 0.01$, $N = 2000$

Remark 1. The assumption of known standard deviation may be strong in many situations in practice. For the cases of unknown standard deviation, the s-plan, which uses the sample standard deviation in the acceptance procedure, may be used, making use of the algorithm already implemented in the CRAN version of [6]. For the case when the procedure for known standard deviation is used, but the real value of σ is different from the assumed standard deviation value used in the acceptance procedure, the resulting value of the operating characteristic will be different from the theoretical value calculated for the assumed standard deviation. Let us consider a simulation study for the plan from Example 1 (n = 26, k = 2.577681). Let us suppose that the upper limit U = 25, the mean μ = 19.5 and true σ = 2. Then in case that the assumed value of the standard deviation $\hat{\sigma} = \sigma = 2$, the probability of acceptance is around 0.8. For lower $\hat{\sigma}$ than the true value of the standard deviation σ, the actual probability of accepting a lot is greater, for example using $\hat{\sigma}$ = 1.9 gives the probability of acceptance 0.94. For greater $\hat{\sigma}$ than the true value of the standard deviation σ, the actual probability of accepting a lot is lower, for example using $\hat{\sigma}$ = 2.1 gives the probability of acceptance 0.59. The plan (n, k) calculated under wrong assumptions may be then suboptimal.

5 Conclusion

This paper for the first time provides the tables of the rectifying LTPD single sampling plans for the inspection by variables under the assumption that the standard deviation of the quality characteristic is known. The mean inspection cost per lot of the process average quality of these plans is lower than mean inspection cost per lot of the process average quality for the unknown standard deviation case. The behaviour of the sample size is similar in both cases, that is the sample size is increasing in the lot size and it increases if the process average quality becomes closer to the lot tolerance proportion defective.

Current implementation of the calculation is based on the R extension package LTPDvar and is reimplemented in julia programming language, which can provide fast computations and clear, easy to maintain code. Such features would be beneficial especially for the calculation of optimal double sampling plans, which is one of the next steps in research.

Acknowledgements

This paper has been produced with contribution of long term institutional support of research activities by Faculty of Informatics and Statistics, Prague University of Business and Economics.

References

[1] Dodge, H., F., & Romig, H. G. (1998). *Sampling Inspection Tables: Single and Double Sampling*. New York: John Wiley.

[2] Hald, A. (1981). *Statistical theory of sampling inspection by attributes*. New York: Academic Press.

[3] Jennett, W. J., & Welch, B. L. (1939). The Control of Proportion Defective as Judged by a Single Quality Characteristic Varying on a Continuous Scale, *Supplement to the Journal of the Royal Statistical Society*, 6, 80–88.

[4] Johnson, N. L., & Welch, B. L. (1940). Applications of the Non-central t distribution, *Biometrika*, 38, 362–389.

[5] Kaspříková, N. (2021). LTPD variables inspection plans and effect of wrong process average estimates. In *International Conference on Mathematical Methods in Economics 2021* (pp. 244–248). Praha : Czech University of Life Sciences Prague.

[6] Kasprikova, N. (2022). *LTPDvar: LTPD and AOQL plans for acceptance sampling inspection by variables.* R package version 1.2.1. http://CRAN.R-project.org/package=LTPDvar.

[7] Klůfa, J. (1994). Acceptance sampling by variables when the remainder of rejected lots is inspected, *Statistical Papers*, 35, 337 – 349.

[8] Klůfa, J. (2010). Exact calculation of the Dodge-Romig LTPD single sampling plans for inspection by variables, *Statistical Papers*, 51(2), 297-305.

[9] Luca, S., Vandercappellen, J. & Claes, J. (2020). A web-based tool to design and analyze single- and double- stage acceptance sampling plans *Quality Engineering*, 32:1, 58–74, DOI: 10.1080/08982112.2019.1641207

[10] R Core Team. (2021). R: *A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org

A Comparison of MCDM and DEA Models

Jana Klicnarová[1], Michaela Brabcová[2]

Abstract. The MCDM and DEA models are widely used in many, not only economical, studies. The main aim of the MCDM models undoubtedly differs from the aim of the DEA ones; on the other hand, somewhere, it could be reasonable to use both methods and compare the results; in some cases, it is not possible. The aim of the paper is to summarize and discuss the differences between these two approaches and show the possible differences between results of these methods.

Keywords: Multiple Criteria Decision Making, Data Envelopment Analysis, criteria preferences

JEL Classification: C830
AMS Classification: 90B50

1 Introduction

Data Envelopment Analysis (DEA) and Multiple-criteria Decision Making (MCDM) are two widely used methods to evaluate some decision units. At first sight, these two types of techniques seem to differ; in the second view, they could seem to coincide from some point of view. The aim of this paper is to discuss the main differences between these methods. Indeed, a huge number of papers on this topic have been written; therefore, the literature review is the main point of this paper. Let us also remark that not all MCDM methods can be compared with DEA models. Especially, Conjunctive and Disjunctive methods are based on an entirely different point of view. Under these methods, the exact values of alternatives under each criterion are important, not their proportion. Therefore, in the following text, we will focus on only such methods of MCDM, which could be comparable with DEA models.

Multiple-criteria Decision Making (MCDM)

First, let us briefly introduce the main assumptions and goals of the MCDM. In the MCDM methods, there are supposed to be some decision units (DMU) and some criteria. Each of the DMU is evaluated under all criteria. There is also a decision-maker (DM) who has some criteria preferences, and she/he is also able to set his alternative preferences under each criterion. The aim of these methods is to order the DMUs from the ”best” to the ”worst” for the DM. Sometimes, the aim of these methods could be only to identify ”good” and ”bad” DMUs. However, the results strongly depend on the DM, more precisely, on her/his preferences.

Data Envelopment Analysis

The Data Envelopment Analysis is based on Farrell’s idea; it was developed in the 80ties. In these methods, some DMUs with known inputs and outputs are supposed. The aim of the method is to measure the efficiency of the DMUs, in the sense of comparing DMU’s inputs needed for produced DMU’s outputs. The basic model for this analysis was given by Charnes, Cooper, and Rhodes, see [5]; after establishing this model, many improvements followed.

2 Comparison between MCDM and DEA methods

As was written above, a huge number of papers have already studied this problem; let us introduce some of them here. However, before we begin with the known results, let us focus on the differences between these approaches.

2.1 The basic differences

We can see some differences between these methods at first glance; let us summarize them. First, the MCDM methods suppose knowledge (or at least existence) of criteria preferences; hence the results depend on these preferences. On the other hand, in DEA, we search for every DMU for weights under which the DMU is viewed in the best possible light. Second, most MCDM methods could order the DMUs from the best to the worst; on the other hand, the basic CCR model identifies the efficient units, and the other ones could be ordered by their efficiency.

[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, brabcm10@ef.jcu.cz

Third, there are some basic assumptions on the data for the CCR model – we suppose all values are positive, only cardinal information is allowed, and the data must be from the convex set. On the other hand, the MCDM methods could also handle ordinal information; convexity is not a necessary condition. Next, the weights in MCDM models are typically supposed to be normalized, i.e., their sum is equal to one. Therefore, the methods usually use some normalization (depending on the method) to get comparable values under different criteria. This step is not used in DEA methods because weights are not standardized the idea is to compare the levels of inputs and outputs. Sure, sometimes, it is helpful to do the DEA evaluation for data that do not satisfy some of the conditions on the CCR model; all values are not positive, ordinal values are included, and so on. Therefore, many authors have studied these problems and tried to avoid such restrictions. Many variations of basic models have been proposed to be able to handle such data.

2.2 Ranking of all DMUs

As we already mentioned, typically, the MCDM methods rank the DMUs from the best one to the worst one (from the point of view of the DM), but the classical DEA methods identify efficient DMUs and rank all others according to their efficiency score. So, there is a natural question if it is also possible to order the efficient DMUs in the DEA model. One of the papers which answer this question was written by Andersen and Petersen (93), see [2]. This paper introduced the method of ranking efficient DMUs in DEA models. The idea of this method is for every efficient DMU to do a new DEA model, which does not take into account the evaluated DMU (the efficient DMU is excluded from the model). And the radial distance between the evaluated DMU and the new production frontier gives us the ranking of efficient units. However, as, for example, Bouyssou, see [3], mentioned, this method suffers from the strong dependence of results on the set of data (for example, the existence of a DMU which is ”close” to the ”best” DMU could strongly affect the result). Another technique that was developed to rank the efficient units is, for example, the Cross-evaluation technique; for more detail, see [7]. The idea is easy – for ”optimal” weights gotten for every DMU to compute the efficient score for all DMUs and then handle with such matrix. However, since the ”optimal” weights are not unique, neither is the matrix unique. It was a reason why Bouyssou, see [3] concluded that if we aim to order the DMUs, it is necessary to have given weights.

2.3 Weights in DEA models

One of the main differences between these two approaches is the existence of criteria preferences in MCDM methods and their no existence in the DEA method. Therefore, it is natural to think about the possibility of including the weight preferences in DEA models. First, it is necessary to remark that no weight restriction is supposed under the original CCR model. Therefore, in several cases, the resulting weights are equal to zero. The possibility of zero weights could be taken as a disadvantage of the method because it causes such criteria to be ignored, i.e., the unrealistic efficiency scores are obtained. On the other hand, as we will discuss in the following, there is no easy way to add weight restrictions into DEA models. Surely, the huge number of authors has discussed this issue and proposed some solutions (see for example [8], [7], [1]). The main problem with weights in DEA models is the idea of the construction of the DEA model, where no weights are supposed. The DEA models compare the proportion of outputs and inputs, not their exact values. Therefore, if we add weight restrictions in some absolute values, then, for example, the solution will not be scale-invariant.

Further, if we set weight restrictions at the same fixed level for all criteria, the strength of such restrictions would strongly depend on the level of values of DMUs under individual criteria. It would be a substantial restriction for some criteria (with a high level of values); on the other hand, it would be a very weak criterion for others, which have low values compared to others. It is also necessary to remark that if we set weight restrictions inappropriately (it is not seen at first glance, what does it mean), there might be no feasible model solution. Moreover, the DEA model uses the given set of DMUs and measures the efficiency of the DMU to the underlying production technology, which is given by observed units; it is a convex hull of observed units. If we add weight restrictions into the model, we, in fact, change the underlying production technology; it is a broader set; for more details, see [11]. Podinski, see [12], proved that for any weight restrictions which satisfied some conditions, the optimal weights for a DMU show the DMU in the best light in comparison to the entire technology expanded by the weight restrictions, whether the weight restrictions are included into the model or not. This fact helps to a meaningful interpretation of such results. Let us also remark that inefficient units without weight restrictions stay inefficient under weight restrictions. On the other hand, some efficient units may be efficient under unrealistic weights – proportion among criteria preferences. The way of setting bounds for criteria weights is discussed, for example, by Dyson and Thanassoulis ([8]), who suggest using regression analysis to establish lower bounds for output weights (they suppose only one input). Later, many authors continue researching this topic; for example, an excellent survey given by Allen et al. [1]. One of the discussed problems is the question of the setting of lower weight bounds to ensure all criteria are taken into account and not destroy the idea of the method. The second widely discussed topic is such weight restrictions ensure the realistic relationship among criteria preferences, see for example [13].

2.4 Data Assumptions and Summarization

As was mentioned above, if we want to apply a basic DEA model, some conditions on data are required. For example, we assume positive values, cardinal values, convex set and so on, see for example [6]. Surely, there exist many modifications which allow to weak or miss out some of them. For example, the special methods that allow handling with ordinal data exist, methods for handling with fractional data were introduced, and the adapted models for fraction data, negative values, categorical variables, etc. However, such discussion in detail, unfortunately, goes beyond the scope of this paper. On the other hand, the data assumptions for MCDM methods depends on chosen method, and they are usually weakly then in DEA cases. This point of view is necessary to involve into the comparison.

In the second half of the eighties and in the nineties, it was very popular to study the comparison between DEA and MCDM methods. Some of the papers studied the comparison of the DEA models results with some chosen MCDM methods; others authors discussed the modifications of the methods to use the advantages of both of them, and many others studied the advantages and disadvantages of such approaches. See, for example [14], [3], [15], and references herein.

3 Using of DEA weights in MCDM models

Now, let us focus on the following question – the comparison of DEA results and results of two chosen MCDM methods. Many authors have already studied a similar question; for example, for comparing DEA and VIKOR results, see [10]. More precisely, let us suppose such data could fit for both types of methods – MCDM and DEA methods. We apply the basic DEA model (the CCR input-oriented model) and, for each DMU, identify the weights under that the DMU achieves the best evaluation. From the DEA idea, it is considered that if we compare the DEA efficiencies of the DMU under different linear optimization (DEA) models (i.e., models for different DMUs, with different weights), they will differ – these efficiencies changes we show in a graph, later.

So, our question is, how does such a change in weights affect the results of MCDM methods. It is well-known that the MCDM results depend on the weight choice. And our question is if there is any connection between DEA and MCDM methods results depending on the weights used. Undoubtedly, both types of results depend on weights, but does it mean that in the same way or not? Hence, we apply MCDM methods with these weights and compare the ordering of the DMUs with the rank according to the achieved efficiency score under the same weights. Therefore, we apply the Spearman Correlation Coefficient to rankings for each DMU's weights and method. We study the relationship between these orderings under obtained weights. Indeed, we keep in mind the resulted weights are not unique, and the results strongly depend on the data set. So, this paper is supposed to be just a case study. Let us remark that the same problem studied Brabcová (see [4]) on different data and got very similar results to ours.

First, let us introduce the data used in our study. These data were inspired by data used in Diploma theses by Brabcová [4]. In the example, we evaluate the districts (NUTS 4) in the Bohemia region in the Czech Republic. Leaving aside the discussion about the appropriation of the choice of variables (it could be further research, and in fact, the districts' evaluation is not the primary goal of this paper), we have data from 50 Czech districts (NUTS 4). These districts (NUTS4) are evaluated under the following criteria: the average family house price (in CzK/m^3), the average unemployment rate, the number of crimes per one thousand inhabitants, hospital bed capacity per one thousand inhabitants, and the average pension.

For this study, we chose the TOPSIS and WSM method (the method based on the assumption of linearity of a utility function) because these two methods are very often used in economics studies, and they are very different in the sense of the standardization of data and measurement of the "optimality". The WSM used a linear standardization; the TOPSIS applies the euclidian metric.

In the DEA, the cost criteria – family house price, the unemployment rate, and the criminality rate – are taken as inputs, the other variables, the benefits criteria – hospital capacity and pension, as outputs. The classical CCR model evaluated each DMU, and we saved the obtained weights. Then we applied two methods of MCDM – TOPSIS and WSM (the technique based on a linear utility function) – to the data with weights obtained in DEA models. We save the ranks of the districts under these weights and the methods. Does the choice of weights affect the results? How do the weights change the efficiencies of the units in the DEA model, and how do the results of MCDM methods depend on the choice of these weights. Is the district's rank in TOPSIS/WSM better under its DEA weights, or if it does not depend? Also, we could ask if the variance of the ranks (for various DEA weights) differs from the variance of the ranks under TOPSIS and WSA (for various weights). Surely, the result strongly depends on the chosen method and the chosen data set; so, if we would make some general conclusions, it would be necessary to study this problem much more deeply. Here is only such a case study on a chosen dataset. Therefore, first, let us show the changes in efficiencies ranks for DMUs under the various optimal weights for individual units. In more detail, we ran 50 DEA models (for each DMU) and got 50 different weights – one for every DMU. We write down the order of every DMU under achieved efficiencies for chosen weights. Therefore, we have a matrix 50 · 50 of efficiencies orders (50 DMUs evaluated under 50 different weights). It is quite difficult to show such a result, so, in the following Figure 1, we show these efficiencies orders for only the top eight DMUs (according to the CCR input model). However, the results for other DMUs seem very similar. Figure shows the top eight DEA evaluated districts, all of which DEA evaluated under all final weights (for each district). Hence, from Figure 1, we can see, as was supposed, that for various weights, the efficiencies of the DMUs differ. As was considered, every district could achieve a high efficiency only under some specific weights; on the other hand, the efficiency of the same district is low for different weights. More precisely, for example, see districts Děčín and Hradec Králové; their efficiencies change from the first position to the 45th one (from 50). We can see that only two districts (from eight displayed in Figure) have position dispersion less than 25 – Jičín and Svitavy. As was already mentioned, if we choose to display some other DMUs, the results seem similar.

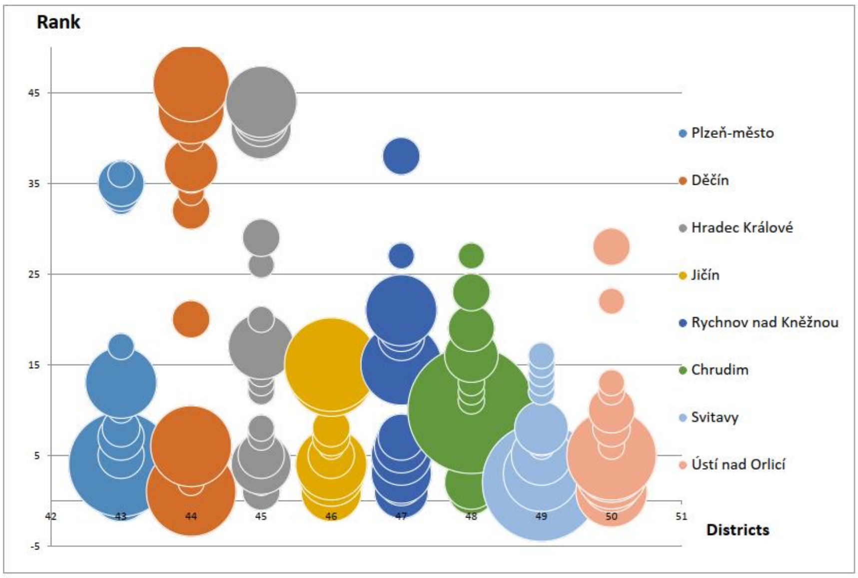


Figure 1 Efficiency Score of the top 8 DMUs

Now, we are interested in the connection with the results of MCDM methods, i.e., how do the rankings under WSM and TOPSIS depend on these weights, and if it is in some relationship with the DEA efficiency results. Therefore, we use the weights from the CCR models and apply WSM and TOPSIS. Our results are again in the forms of 50·50 matrices, so we display the results for only eight chosen districts. Again, the results do not seem to be affected by the original efficiency of the DMUs. Therefore, it is not important which districts we choose. In the following graph, we can see the orderings of the top eight DEA evaluated districts again. More precisely, we ran 50 times TOPSIS, resp. WSM, with all 50 DEA weights and in the graph, we can see the achieved ranks of these districts.

In Figure 2, we can see that there are some districts in which the variance under TOPSIS is higher than under WSM and vice versa. So, from these graphs, we cannot derive any conclusion from this case study. (The situation is the same if we look at the results of all 50 districts.) So, let us look at the correlation among these results. In more detail, we apply Spearman's correlation coefficients between the ranks under DEA and WSM, resp. TOPSIS.

Hence, in the following Figure 3, we can see Spearman's correlation coefficients of DMUs ranking under DEA efficiencies and TOPSIS or WSA ranks. The gray line shows the achieved efficiency in DEA.

From Figure 3, we can see that for the chosen data set, the correlation between DEA efficiencies ranks and ranks under the WSM method is really high - close to one, for all DMUs. More over, it seems that it does not depend on the DMUs efficiency. For the TOPSIS, the situation is different; the correlation coefficient changes from the DMU to the DMU. However, Figure 3 does not show any dependence between the correlation coefficient and the efficiency score of the chosen DMU. Such different results for these two methods are not too much surprising. The WSM uses standardization only by a supposed linear utility function, which is not too far from the DEA model; on the other hand, the standardization and the evaluation technic in TOPSIS are more different from the DEA approach.

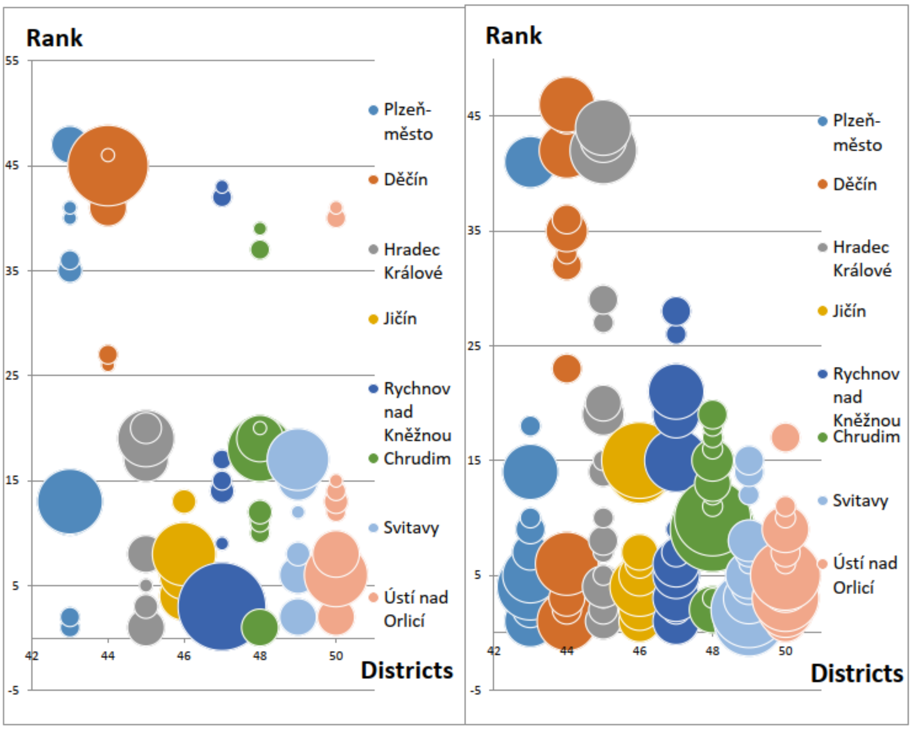


Figure 2 Ranks of the top 8 DMUs under TOPSIS and WSM

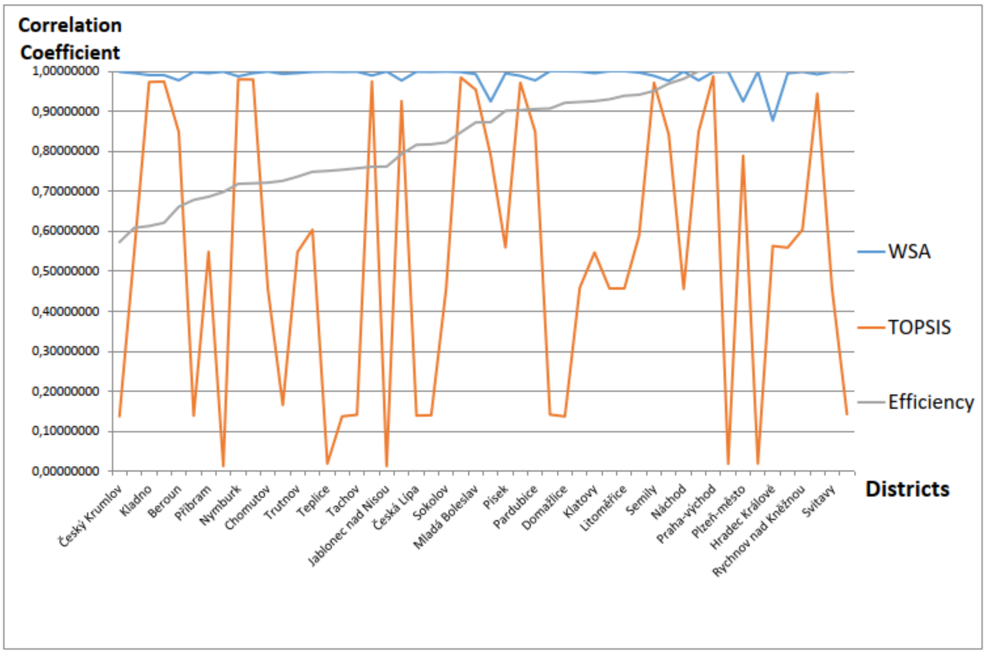


Figure 3 Correlation Coefficients

4 Conclusion

In the paper, we discussed some advantages and disadvantages of DEA models and MCDM methods and their comparison. In the application part, we show the results of some MCDM methods with weights raised from DEA methods, leaving aside the fact if such weights are meaningful and interpretable or not. It is clear that the results of MCDM methods depend on chosen weights and method; however, we are interested in the question if the DEA results (the achieved efficiencies under different weights) are in some connection with MCDM results. We showed the case of an MCDM method (WSM), for which the results are strongly correlated. On the other hand, we showed that it is not a rule, and even though we use the same data, the other method (TOPSIS) could give results that are not in such a strong relationship. Surely, if we would like to do some general conclusion, it is necessary in a future research to use much more different data sets. It would be also necessary to use DEA models with interpretable weights and take into account the weight stability in the MCDM models. More over, for better interpretation of a correlation coefficient it would be better to apply DEA methods which distinguish effective DMUs and explore if the result is affected by chosen method

Acknowledgements

The paper was supported by a grant n. GAJU 129/2022/S.

Many thanks referee for his/her comments and time spent on reviewing this paper.

References

[1] Allen, R., Athanassopoulos, A., Dyson, R. G., & Thanassoulis, E. (1997). Weights restrictions and value judgements in data envelopment analysis: evolution, development and future directions. *Annals of operations research*, 73, 13-34.

[2] Andersen, P., & Petersen, N. C. (1993). A procedure for ranking efficient units in data envelopment analysis. *Management science*, 39(10), 1261-1264.

[3] Bouyssou, D. (1999). Using DEA as a tool for MCDM: some remarks. *Journal of the operational Research Society*, 50(9), 974-978.

[4] Brabcová, M. (2022). *Vícekriteriální rozhodování a analýza obalu dat*. (Diploma Thesis). Jihočeská univerzita v Českých Budějovicích Ekonomická fakulta.

[5] Charnes, A., Cooper, W. W., Rhodes, E. (1978.): Measuring the Efficiency of Decision Making Units. *European Journal of Operational Research*, (2), 6, pp. 429-444.

[6] Cooper, W. W., Seiford, L. M., & Tone, K. (2007). *Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software* (Vol. 2, p. 489). New York: Springer.

[7] Doyle, J., & Green, R. (1994). Efficiency and cross-efficiency in DEA: Derivations, meanings and uses. *Journal of the operational research society*, 45(5), 567-578.

[8] Dyson, R. G., & Thanassoulis, E. (1988). Reducing weight flexibility in data envelopment analysis. *Journal of the operational research society*, 39(6), 563-576.

[9] Kuosmanen, T., & Matin, R. K. (2009). Theory of integer-valued data envelopment analysis. *European Journal of Operational Research*, 192(2), 658-667.

[10] Opricovic, S., & Tzeng, G. H. (2003). Comparing DEA and MCDM method. In *Multi-Objective Programming and Goal Programming* (pp. 227-232). Springer, Berlin, Heidelberg.

[11] Podinovski, V. V. (2004). Production trade-offs and weight restrictions in data envelopment analysis. *Journal of the operational Research Society*, 55(12), 1311-1322.

[12] Podinovski, V. V. (2016). Optimal weights in DEA models with weight restrictions. *European Journal of Operational Research*, 254(3), 916-924.

[13] Roll, Y., & Golany, B. (1993). Alternate methods of treating factor weights in DEA. *Omega*, 21(1), 99-109.

[14] Sarkis, J. (2000). A comparative analysis of DEA as a discrete alternative multiple criteria decision tool. *European journal of operational research*, 123(3), 543-557.

[15] Stewart, T. J. (1996). Relationships between data envelopment analysis and multicriteria decision analysis. *Journal of the operational research society*, 47(5), 654-665.

Facility Layout Problem with Logistic Constraints

František Koblasa[1], Miroslav Vavroušek[2]

Abstract. The great pressure to maximise productivity and minimise logistics costs in manufacturing systems leads to the problem of efficient layout arrangement and logistics path planning. An efficient layout not only minimises product cost but also leads to greater customer satisfaction in the form of earlier delivery dates.
This article aims to develop a constructive algorithm for the Facility Layout Problem of Flexible Manufacturing Systems with material handling. The model of this system consists of cells with unequal sizes oriented in an open field layout and connected with paths. Paths that use facility space create a logistic system between facility entrance and exit and production cells pick-up and drop-off points.
Solution construction mainly consists of selecting a cell to be placed and a suitable free space where the cell can be oriented to fit area limitations and, at the same time to be connected to logistics paths.
There are proposed dispatching rules to make decisions in the before-mentioned steps and tested on developed model instances. The key objective is to minimise material handling costs. However, an additional optimisation objective of potential free space for new cells is also discussed.

Keywords: Facility Layout Problem, Flexible Manufacturing System, Material handling, Constructive Algorithm.

JEL Classification: C60, C63
AMS Classification: 90C27

1 Introduction

Facility Layout Problem (FLP) is one of the oldest [9] Industrial Engineering problems determining the most efficient arrangement of physical departments within a facility space. Efficient facility arrangement decreases manufacturing costs, work in process, production lead times and increases productivity. Material Handling Cost (MHC) is the most mentioned objective to be minimised. The general agreement on MHC participation on the amount of total operating costs of the company is 20-50%, while the total cost of manufacturing a product is 15-70% [4, 5]. The optimisation of FLP can reduce these costs by 10-30% [13], while the inefficient layout cant increase cost [11] and decrease efficiency in general [7] by 35%. That is one of the reasons FLP is one of the most addressed problems in practice as well as in the scientific literature focusing on production management and Industrial Engineering [5, 13].

There are multiple approaches, in industrial engineering, to designing suitable space composition based on manufacturing system type i.e. fixed product layout, process layout, product layout and cellular layout [3]. There is also a long-living concept of the Flexible Manufacturing System layout that is designed to quickly adapt to changes in the type and quantity of the product being manufactured.

The decrease in production resulting in merging the manufacturing space or an expansion of space requirements thanks to new jobs is in the case of real-world problems usually managed manually-intuitively while making proof of concepts with simulation tools [14].

The operation research approach includes various options based on general problem type (static, dynamic), workshop characteristics, problem formulation and solution generation. For deeper characteristics description, see multiple reviews that map not only bibliographical development [4, 12] and classification of problems [5] but also the most advanced approaches to finding a solution using machine learning [1].

The motivation of this article is based on real-world layout requirements of small volume customise production, which has characteristics of job shop material flow. Design FLP of this manufacturing system is very demanding, thanks to its inability to streamline processes into manufacturing lines. It is also hard to merge common operation

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

sequences into manufacturing cells. That leads to intensive material handling demand in the form of path space which is usually neglected in the case of open field FLP by setting its width as 0 or its width is inflated into workshop space.

This article presents a constructive algorithm that provides a feasible solution to FLP with open space MHS with pick-up and drop-off points of rectangular workshops, which have to fit limited facility space with its entrance and exit while paths take additional facility space.

The article is organised as follows. The second chapter defines (based on before mentioned classification reviews) the open field Facility Layout Problem in the scope of the before-mentioned material handling system with space requirements. The third chapter is focused on neighbourhood generation of solution space, including constructive algorithm and dispatching rules used in solution generation. The fourth chapter presents model instances and compares the dispatching rule's ability to meet the goal of minimising material handling costs. The possibility of dealing with the described problem with metaheuristics algorithms is discussed further.

2 Open field FLP with MHS space requirements and logistics access points

This article focuses on the static (greenfield) FLP in which facility and department do not change over time (shape, dimension, number and purpose). The further basic model description is made according to Hosseini-Nasab et al. classification [5] and is based on the model with non-overlapping departments [6, 10].

There are unequal size N departments to be placed in a single floor facility area defined by the height H and width W with the goal to minimise single objective of MHC (1) where f_{ij} represents the intensity of material flow, c_{ij} cost of one unit transportation and d_{ij} distance of a path between i and j logistics access points. These logistic access points include pick-up and drop-off points (P/D) of departments as well as M entrances and K exits to the facility.

$$MHC = min\left\{\sum_{i=1}^{N+M}\sum_{j=1}^{N+K} f_{ij}c_{ij}d_{ij}\right\} \tag{1}$$

Departments have a regular (rectangular) shape with a fixed height h and width w dimensions. Aspect ratio, which would make it possible to change length and width while keeping the same area, is not considered. Dimensions are defined as continual (without defined minimal dimension step as in discrete models). Constraints (2) resp. (3) define the x,y bottom left position boundary of each department and (4)(5) space limits according to the size of the facility [6, 10].

$$x_i' = x_i + (1-u_i)w_i + u_i h_i \quad \forall i \tag{2}$$

$$y_i' = y_i + (1-u_i)h_i + u_i w_i \quad \forall i \tag{3}$$

$$x_i' \le W_i \quad \forall i \tag{4}$$

$$y_i' \le H_i \quad \forall i \tag{5}$$

Position $x^{p(d)}$,$y^{p(d)}$ of pick-up and drop-off stations are defined as (6)and (7)

$$x_i^{p(d)} = x_i + (1-u_i)(1-v_i)P(D)_i^x + P(D)_i^y u_i(1-v_i) + (w_i - P(D)_i^x)(1-u_i) + \left(h_i - P(D)_i^y\right)u_i v_i \quad \forall i \tag{6}$$

$$y_i^{p(d)} = y_i + (1-u_i)(1-v_i)P(D)_i^y + (w_i - P(D)_i^x)u_i(1-v_i) + \left(h_i - P(D)_i^y\right)(1-u_i)v_i + P(D)_i^x)u_i v_i \quad \forall i \tag{7}$$

where variables (u_i, v_i) are desition variables presenting orientation (0,0) – cell i is in original orientation, (1,0) cell i is rotated 90° clockwise, (0,1) – 180° and (1,1) - 270°

During the construction of the facility, the layout department can interfere with the defined path, so a new search for the logistic path has to be performed. However, departments can not overlap (8)(9)(10)

$$l_{ij} + l_{ji} + b_{ij} + b_{ji} \geq 1 \quad \forall i < j \tag{8}$$

$$x_i' \leq l_{ij} x_j + W(1 - l_{ij}) \; \forall i, j \tag{9}$$

$$y_i' \leq b_{ij} y_j + H(1 - b_{ij}) \; \forall i, j \tag{10}$$

where l_{ij} =1 if cell i is placed to the left of cell $j(x_i \leq x_j)$ else l_{ij} =0; b_{ij}=1 if cell i is placed to the below of cell $j(y_i \leq y_j)$ else b_{ij}=0.

Model does not operate with unique path overlapping as paths can overlap each other (merge in one). However, , paths can not overlap with departments. It is impossible to define its position beforehand as a dealt problem is open-field (see Figure 1) without known space constraints of paths other than its width w_p.

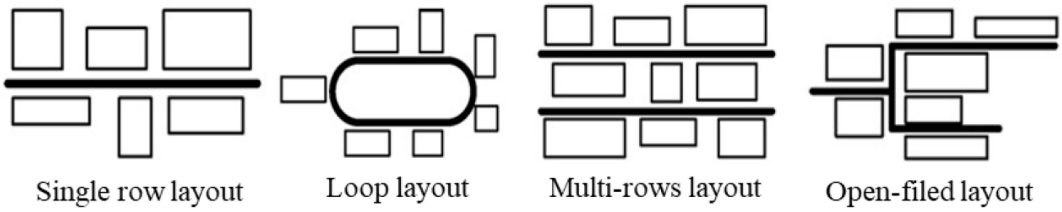


Figure 1 MHS layout design

The following chapter focuses on a constructive algorithm design generating a feasible solution of Regular shape open field FLP of limited space area, with P/D points and single enter and exit (M/K) from the facility, where departments cant overlap. The algorithmic model of MHS consists of w_t >0 logistics paths going from (to) logistics access points (P/D, M/K). Paths are not inflated to departments area and will unite (overlap) in the final MHS.

3 Constructive algorithm for OFLP with MHS space constraints

The proposed constructive algorithm for the Open field FLP with Space-Constrained Material Handling System (OFLP-SCMHS) is inspired by previous space composition problem (bin packing problem) work [8]. During constructing a solution to OFLP-SCMHS (see Figure 2) it is necessary to deal first with the sequencing problem of selecting a suitable department to be placed in one of the defined free spaces (see [8]). Further, it is checked if the department will fit created free space in r=4 basic rotations (0°,90°,180°,270°). Feasible free spaces are then used as areas where possible *rpr* = 36 (4 rotations in 9 placements) space assignment solutions are tested if the department can fit space of *rpr*.

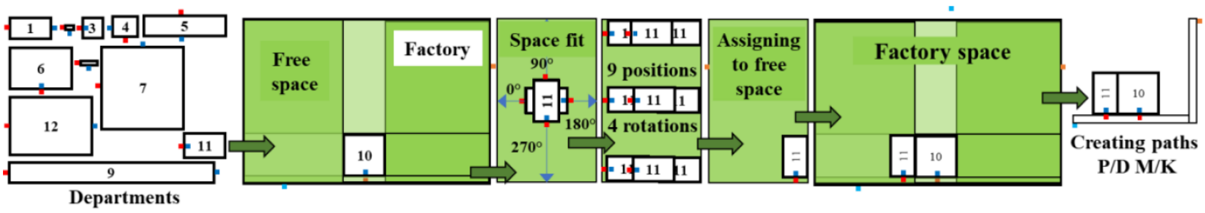


Figure 2 Simplified algorithm steps - department assignment to Factory space

The last step is to find the shortest paths from all P/D and M/K points defined by the transportation intensity matrix. That is the most complex operation of an algorithm which detailed description is out of the scope of this paper. It is based on Dijkstra's algorithm [2]; however, it can't use a classical wave approach with standardised step size, as space is not divided into discrete path subspaces but into rectangular free spaces.

First, it makes a search wave (described by graph representation) from P/D points of the department to be placed, going through free spaces while mapping P/D/M/K of MHS. If it finds free space with possible MHS access points already in the graph, it notes an alternative trajectory while dij deleting the longer alternative. If it finds a way between placed departments and/or departments selected to be placed, it notes position in the graph. After reaching the bottom of the graph, it backtracks to the start of the graph (department to be placed) while calculating the coordinates of the defined free space trajectory.

Detailed constructive algorithm of FLP described in figure 2 follows.

```
Initialise:
01  create list of PB planned and NB un-planed departments N+M+K
02  add enter M and K points as a planned department to PB.
03  create new Bin B in the Bins
04  add new FreeSpace F = (w; h) to B
Place departments
05 for each N placement of department N=(w;h;f;P(x,y);D(x,y))
06 | create conflict set CS from NB, which has flow intensity f>0 with P/D those in PB
07 | select one N of NB from CS to be placed in F
08 |     for each F(w,h) in B
09 |       | for each rotation r (w,h,a) with angle a,
10 |       |     | if F.w>=N.w and F.h>=N.h add r (w,h,a) to R
11 |       | for each rotation r in R,
12 |       |     | for each place rp=9,
13 |       |     |     | for each r=4 rotation in placement rp
14 |       |     |     | create space assigments rpr =( w;h;f;P(x,y);D(x,y))
15 |       |     |     | find shortest path d of every rpr (P,D) and every N in PB
16 |       |     |     | calculate MHC=f*c*d of every P/D and M/K
17 |       |     |     | select rpr with minMHC and add its N in PB
18 |       L     L     L   break
19 |     add new FreeSpaces F = (W; H) to B
20 L     optimizeFreeSpaces in the Bin B (Erase FreeSpace, which is only subspace)
```

It is obvious, from above defined algorithm, that defined problem is computationly demanding. This problem as most of FLP is NP-hard. It consists of n! combinatorial sequencing problem, problem of selecting suitable free space, finding suitable position inside and proper rotation.

Following dispatching rules are designed to reduce full factorial search so reduce the excessive computation time of P/D pathfinding:

1) Department sequencing can be made in step 07 by ordering them by:
   a. total decreasing intensity flow between NB (not planed) and PB (planed) logistics points P/D/M/K
   b. the decreasing total intensity flow P/D/M/K $f^* = \sum_{i=1}^{N+M} \sum_{j=1}^{N+K} f_{ij}$
   c. the decreasing department area A=WH
2) Select free space by the min/max rule of the most fitting department in step 08 – place the department in the smallest feasible area.
3) Selecting 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

There are defined model test instances of this problem in the following chapter.

4 Model instances of OFLP-SCMHS

Following 6-SCMHS and 12-SCMHS Model instances are inspired by previous work of Welgama and Gibson [15]. That includes the size of departments in original rotation (*w,h*), its P/D (*x,y*) coordinates, and the transportation matrix of intensity flow f_{ij} However, as it lacks of logistics and facility area constraints, the model had to be modified to:

- Set maximal *W,H* of the facility.
- Set coordinates of entrance M and exit K to-from facility
- Set flow between departments and M and K (En/Ex).
- Set width of paths w_p=0.5

Figure 3 shows two 6-department instances and Figure 4 two 12-department instances which differ in facility area size (*W,H* = 22,30 and 26,30) as well as the location of both one entrance and exit.

Sized and P/D locations					P/D intensity flow							
N	W	H	P (x,y)	D(x,y)	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	10
2	5	5	0;2.5	5;2.5	2	5	0	1	2	1	2	10
3	20	5	10;0	10;5	3	2	3	0	3	2	1	9
4	8	6	4;0	4;0	4	4	0	0	0	1	2	6
5	12	4	0;2	6;0	5	1	2	0	5	0	0	3
6	9	6	4.5;0	0;3	6	0	2	0	2	10	0	4
En/Ex 6-1	22	30	10;30	22;15	En	0	5	2	2	6	9	0
En/Ex 6-2	26	30	0;16	15;30								

Figure 3 6-1 and 6-2 SCMHS instances

Sized and P/D locations					P/D intensity flow													
N	W	H	P (x,y)	D(x,y)	P/D	1	2	3	4	5	6	7	8	9	10	11	12	Ex
1	10	5	0;2.5	10;2.5	1	0	1	2	3	2	2	3	4	2	3	4	5	0
2	2	1	0;0.5	1;0	2	5	0	1	2	1	1	2	3	3	2	3	4	9
3	5	5	0;2.5	0;.2.5	3	2	3	0	1	2	2	1	2	4	3	2	3	1
4	8	6	4;0	0;3	4	4	0	0	0	3	3	2	1	5	4	3	2	7
5	20	5	10;5	10;0	5	1	2	0	5	0	1	2	3	1	2	3	4	1
6	15	10	7.5;0	7.5;0	6	0	2	0	2	0	0	1	2	2	1	2	3	6
7	20	20	0;10	10;20	7	0	2	0	2	5	5	0	1	3	2	1	2	8
8	4	1	0;0.5	2;0	8	6	0	5	10	1	1	10	0	4	3	2	1	1
9	50	5	0;2.5	50;2.5	9	2	4	5	0	1	1	5	0	0	1	2	3	7
10	10	10	5;0	5;0	10	1	5	2	0	5	5	2	0	0	0	1	2	0
11	10	6	0;3	0;3	11	1	0	2	5	4	4	3	5	10	5	0	1	1
12	20	14	0;7	20;7	12	1	0	2	5	0	0	3	0	10	0	2	0	1
En/Ex 12-1	60	52	25;0	60;30	En	7	2	6	0	0	9	4	6	3	1	0	3	0
En/Ex 12-2	52	52	21;0	21;52														

Figure 4 12-1 and 12-2 SCMHS instances

5 Tests and results

Table 1 describes results obtained by looking for the solution of models 6-1, 6-2, 12-1 and 12-2 using dispatching rules described in the previous section (sequencing-free space search-placement selection e.g. IT-F-ND):

- (IT) stands for option 1a) of selecting maximal total intensity flow first
- (I) is 1b) prioritising department which has maximal flow with already placed
- (Max) is 1c) selecting the biggest area department available to be placed
- (F) represents 2) full search of available free spaces, while (m/M) is preselecting the one which fits placed department best.
- (ND) represent options where all possible 34 positions are tested to find the shortest path while (AD)
- selects placement with the shortest P/D air distance.

The objective function of minimising total MHC (1) by the above-mentioned dispatching rules as well as maximal remaining area are shown in Table 1, where the best results found are marked by green.

	MHC				Max remaining area			
Rule/Model	6-1	6-2	12-1	12-2	6-1	6-2	12-1	12-2
IT-F-ND	6 053	6 565	91 345	90 308	121	90	559	559
IT-m/M-AD	-	10 157	121 873	165 155	-	143	997	484
I-F-ND	-	6 666	77 970	71 851	-	159	795	403
I-m/M-AD	10 675	-	196 948	146 813	138	-	624	572
Max-F-ND	10 031	8 788	80 316	94 707	128	184	741	331
Max-m/M-AD	10 557	14 060	178 875	151 606	85	143	1027	633

Table 1 6 and 12 SCMHS test results

As expected (IT) rule with full (ND) search was the best while minimising MHC in case of small 6-1;2 problems. The biggest current flow of rule (I) was best on average. However, it was not able to find a feasible solution in the case of I-F-ND. Dispatching rules which were not able to find the solution (not all objects can be placed in limited space are marked in Table 1 by (-). That is caused by the very demanding shapes of departments (see Figure 5 model 6-1;2 eg. deparment n.3 or 12-1;2 department n.9). Figure 5 also shoves how the size of free space influences results significantly as 6-1;2 using the same rule IT-F-ND differs only in x facility dimension. The future necessity of multi-criteria objective function can be seen on 12-1 where (Max) is far the best solution in Max remaining space while having the worst result in MHC.

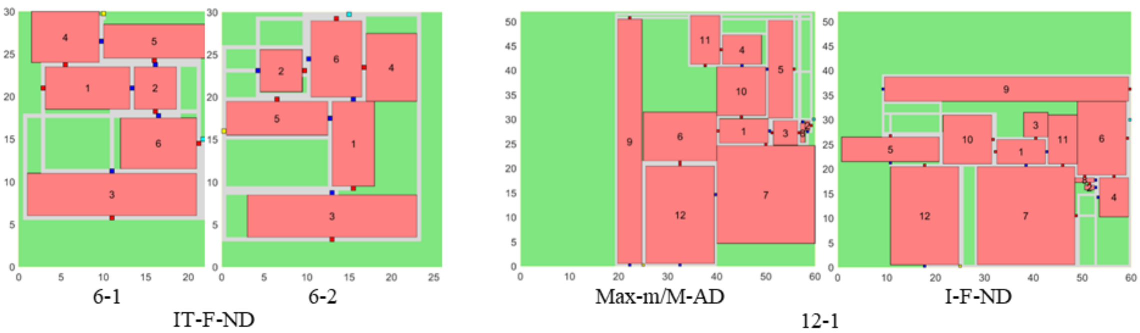


Figure 5 Comparison of Facility size and dispatching rule influence on results

6 Conclusion

Presented research shows that constructing feasible solutions of FLP with MHS space usage while having limited space of facility is very dependent on *W*x*H* aspect of both departments and facilities. The designed algorithm failed to find a feasible solution in 3 out of 24 cases in the last step as 3rd department of both 6-1;2 has the smallest (MH) intensity while having the biggest *w*x*h* ratio, so it was about to be placed last.

Reducing search space by selecting the nearest position (AD) instead of testing all 34 possible combinations (ND) reduced computation time significantly. Computation time was approximately hundreds time shorter in the case of 6-1;2 problems and a thousand times faster in 12-1;2 using AD. ND finds significantly better results in MHC. However, in the case of maximal remaining free space AD performed better on average.

Future research will focus on studying the possibilities of generating feasible solutions of proposed OFLP-SCMHS as well as setting an appropriate size and aspect of Facility space. Multi-objective evaluation and research in the field of time-efficient MHS pathfinding will be key areas to apply metaheuristic optimisation methods in the future.

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 2022. The research reported in this paper was supported by institutional support for nonspecific university research.

References

[1] Burggräf, P., Wagner, J., & Heinbach, B. (2021). Bibliometric Study on the Use of Machine Learning as Resolution Technique for Facility Layout Problems. *IEEE Access*, PP, 1–1. https://doi.org/10.1109/ACCESS.2021.3054563

[2] Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to Algorithms* (3. vyd.). MIT Press.

[3] Dilworth, J. B. (1996). *Operations Management* (Subsequent edition). McGraw-Hill College.

[4] Drira, A., Pierreval, H., & Hajri-Gabouj, S. (2007). Facility layout problems: A survey. *Annual Reviews in Control*, 31(2), 255–267. https://doi.org/10.1016/j.arcontrol.2007.04.001

[5] Hosseini-Nasab, H., Fereidouni, S., Fatemi Ghomi, S. M. T., & Fakhrzad, M. B. (2018). Classification of facility layout problems: A review study. *The International Journal of Advanced Manufacturing Technology*, 94(1), 957–977.

[6] Hu, G. H., Chen, Y. P., Zhou, Z. D., & Fang, H. C. (2007). A genetic algorithm for the inter-cell layout and ma- terial handling system design. *The International Journal of Advanced Manufacturing Technology*, 34(11), 1153– 1163. https://doi.org/10.1007/s00170-006-0694-0

[7] Izadinia, N., & Eshghi, K. (2016). A robust mathematical model and ACO solution for multi-floor discrete layout problem with uncertain locations and demands. *Computers & Industrial Engineering*, 96, 237–248.

[8] 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.

[9] Kusiak, A., & Heragu, S. S. (1987). The facility layout problem. *European Journal of Operational Research*, 29(3), 229–251. https://doi.org/10.1016/0377-2217(87)90238-4

[10] Leno, I. J., Saravanasankar, S., & Ponnambalam, S. G. (2012). Layout Design for Efficient Material Flow Path. *Procedia Engineering*, 38, 872–879. https://doi.org/10.1016/j.proeng.2012.06.110

[11] Ripon, K. S. N., Glette, K., Khan, K. N., Hovin, M., & Torresen, J. (2013). Adaptive variable neighborhood search for solving multi-objective facility layout problems with unequal area facilities. *Swarm and Evolutionary Computation*, 8, 1–12.

[12] Scalia, G., Micale, R., & Enea, M. (2019). Facility layout problem: Bibliometric and benchmarking analysis. *International Journal of Industrial Engineering Computations*, 10(4), 453–472.

[13] Tompkins, J. A., White, J. A., Bozer, Y. A., & Tanchoco, J. M. A. (2010). *Facilities planning*. John Wiley & Sons.

[14] Trebuňa, P., Kliment, M., Edl, M., & Petrik, M. (2014). Creation of Simulation Model of Expansion of Production in Manufacturing Companies. *Procedia Engineering*, 96, 477–482. https://doi.org/10.1016/j.proeng.2014.12.118

[15] 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.

Stochastic Dominance Constrained Portfolio Optimization with Distortion Risk Measures

Miloš Kopa[1]

Abstract. The paper deals with risk-minimizing, stochastic dominance constrained portfolio optimization problems where the risk is modeled by distortion measures. These measures could be seen as a generalization of Value at Risk, Conditional Value at Risk or Expected shortfall. If the associated distortion function is concave the measure is coherent. We analyze several such portfolio selection problems for different choices of a concave distortion function. First, assuming a discrete distribution of returns, we identify in sample optimal portfolios with and without second order stochastic dominance constraints. Then we compute the out–of–sample characteristics. Finally, we compare the in sample and out–of–sample results of all considered models among each other.

Keywords: portfolio optimization, distortion risk measure, stochastic dominance

JEL Classification: D81, G11
AMS Classification: 91B16, 91B30

1 Introduction

Distortion risk measures form an important class of risk measures which emanate from the dual theory of choice under uncertainty proposed by [14]. Later on, [12] developed the axiomatic approach. The idea behind the distortion risk measure is the transformation of the given probability measure of returns or losses in order to quantify the tail risk more accurately and therefore give more weight to higher risk events. Perhaps interesting could be a relation to stochastic dominance which is an attractive tool for random returns comparisons in various applications, see e.g. [8] or [5] for recent applications of stochastic dominance in pension fund management. Therefore, we combine these two tools on the portfolio optimization problems. It means that feasible portfolios are only those which dominate a given benchmark ($1/N$ strategy) with respect to the second–order stochastic dominance. And the goal of the portfolio selection is to identify the risk-minimizing portfolio from these feasible ones. For the sake of comparison, we consider also the risk-minimizing problems without SSD constraints which follows [4] and [7]. Finally, we provide a sensitivity analysis with respect to changes in the underlying distortion function.

The remainder of this paper is structured as follows. Section 2 presents a notation and basic properties of the distortion risk measures. It is followed by a formulation of stochastic dominance constrained portfolio selection problem with distortion measures of risk in Section 3. Empirical study is presented in Section 4 and the paper is concluded in Section 5.

2 Distortion risk measures

We follow [7] in summarizing the most important properties of the distortion risk measures. In the whole text, we assume that $\mathcal{X}$ is a set of random variables on a probability space $(\Omega, \mathcal{F}, P)$. A random variable $X \in \mathcal{X}$ represents a loss random variable (typically, positive values are associated with losses and negative values represent gains) of some financial asset (portfolio).

Definition 1. ([2]) *Suppose that $g : [0, 1] \to [0, 1]$ is a non-decreasing function such that $g(0) = 0$ and $g(1) = 1$ (also known as the **distortion function**) and $X \in \mathcal{X}$ with a distribution function $F_X(x)$. Then, the **distortion risk measure** associated with the distortion function g is defined as*

$$\rho_g(X) = -\int_{-\infty}^{0} [1 - g(1 - F_X(x))]dx + \int_{0}^{\infty} g(1 - F_X(x))dx,$$

provided that at least one of the integrals is finite.

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

When we define the **decumulative distribution function** (also known as the **survival function**) $S_X(x) = 1 - F_X(x) = P(X > x)$ and we use it instead of the distribution function, we obtain

$$\rho_g(X) = -\int_{-\infty}^{0} [1 - g(S_X(x))]dx + \int_{0}^{\infty} g(S_X(x))dx.$$

The interpretation of this definition is that the distortion measure represents the expectation of a new random variable with re–weighted probabilities. In some cases, such as problems related to insurance or capital requirements, it is appropriate to assume that the random variable $X \in \mathcal{X}$ is non–negative. In this case, when $X \in \mathcal{X}$ is a non-negative random variable, then ρ_g reduces to

$$\rho_g(X) = \int_{0}^{\infty} g(S_X(x))dx.$$

The class of distortion risk measures is prospective, because distortion measures, in the general case, fulfill the conditions of monotonicity, positive homogeneity and translation invariance.

Theorem 1. *([9]) (Monotonicity) Suppose that $X, Y \in \mathcal{X}$ and $X \leq Y$. Then $\rho_g(X) \leq \rho_g(Y)$.*
(Positive homogeneity) For a distortion risk measure ρ_g, $X \in \mathcal{X}$ and $\lambda \geq 0 : \rho_g(\lambda X) = \lambda \rho_g(X)$.
(Translation invariance) For a distortion risk measure ρ_g and $X \in \mathcal{X}$ it holds that $\forall c \in \mathbb{R} : \rho_g(X + c) = \rho_g(X) + c$.

Theorem 2. *([13]) The distortion risk measure $\rho_g(X)$ is sub–additive*

$$\rho_g(X + Y) \leq \rho_g(X) + \rho_g(Y),$$

if and only if g is a concave distortion function.

Summarizing, a distortion risk measure $\rho_g(X)$ is coherent iff g is a concave distortion function.

In the following example, we present the representations of risk measures Value–at–Risk (VaR) and Expected Shortfall (ES) as distortion risk measures.

Example 1. Suppose that $X \in \mathcal{X}$, $\alpha \in (0, 1)$. If

$$g(x) = \begin{cases} 0 & \text{if } 0 \leq x < 1 - \alpha \\ 1 & \text{if } 1 - \alpha \leq x \leq 1. \end{cases}$$

then $VaR_\alpha(X) = \rho_g(X)$ and if

$$g(x) = min\left(\frac{x}{1-\alpha}, 1\right), \text{ where } x \in [0, 1].$$

then, $ES_\alpha(X) = \rho_g(X)$.

Another example of distortion risk measure includes the **Proportional Hazard (PH) transform** proposed by [11] as a new risk–adjusted premium for insurance risk pricing. This measure has a distortion function

$$g(x) = x^{1/\gamma}, \quad x \in [0, 1], \gamma \geq 1. \tag{1}$$

Consequently, we define the **PH–transform measure** as:

$$\rho_{PH}(X) = \int_{0}^{\infty} S_X(x)^{1/\gamma} dx, \quad \gamma \geq 1,$$

where $S_X(x) = 1 - F_X(x)$ is defined as previously.

As we can see from the definition of the distortion function g of the PH transform, this function is concave and therefore, the PH–transform measure satisfies the sub–additivity property.

Theorem 3. [4] *Let $g\,(x)$ be a concave distortion function and let y_j be a realization of random loss $X \in \mathcal{X}$ such that $P(X{=}y_j) = \frac{1}{m}$, j=1,. . . ,m. Let $G(i) = g(1 - \frac{i-1}{m}) - g(1 - \frac{i}{m}), i = 1, .., m$. Then*

$$
\begin{aligned}
\rho_g(X) \quad &= \quad \min_{c,d} \sum_{j=1}^{m} c_j + \sum_{i=1}^{m} d_i \\
\text{s.t.} \quad & \quad -G_i y_j + c_j + d_i \geq 0, \quad i,j = 1,..,m.
\end{aligned} \tag{2}
$$

3 Portfolio selection problems

Let us consider a random vector $\boldsymbol{r} = (r_1, r_2, ..., r_N)$ of returns of N assets in m equiprobable scenarios. The returns of the assets for the various scenarios are collected in matrix:

$$
\boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}^1 \\ \boldsymbol{x}^2 \\ \vdots \\ \boldsymbol{x}^m \end{pmatrix}
$$

where $\boldsymbol{x}^t = (x_1^t, x_2^t, \ldots, x_N^t)$ is the t-th row of matrix $\boldsymbol{X}$. A vector of portfolio weights is denoted by $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_N)'$. In this paper, we exclude short sales, that is, the set of all feasible portfolios Λ can be characterized as follows:

$$
\Lambda = \left\{ \boldsymbol{\lambda} \in R^N \;\middle|\; \sum_{n=1}^{N} \lambda_n = 1, \quad \lambda_n \geq 0, \quad n = 1, 2, \ldots, N \right\}.
$$

In this notation, the loss of portfolio $\boldsymbol{\lambda}$ is - $\boldsymbol{r}^T \boldsymbol{\lambda}$ with equiprobable realizations: $-\boldsymbol{x}^t \boldsymbol{\lambda}, \; t = 1, ..,m$ which play the role of vector losses y_j in Section 2. Therefore, employing (2), the distortion risk minimizing problem is formulated as follows:

$$
\begin{aligned}
\min_{c,d,\lambda} \quad & \sum_{j=1}^{m} c_j + \sum_{i=1}^{m} d_i \\
\text{s.t.} \quad & G_i \boldsymbol{x}^j \boldsymbol{\lambda} + c_j + d_i \geq 0, \quad i,j = 1,..,m \\
& \boldsymbol{\lambda} \in \Lambda.
\end{aligned} \tag{3}
$$

Following [8], the problem with SSD constraints could be formulated as follows:

$$
\begin{aligned}
\min_{c,d,\lambda,W} \quad & \sum_{j=1}^{m} c_j + \sum_{i=1}^{m} d_i \\
\text{s.t.} \quad & G_i \boldsymbol{x}^j \boldsymbol{\lambda} + c_j + d_i \geq 0, \quad i,j = 1,..,m \\
& \boldsymbol{X\lambda} \geq W \boldsymbol{X\tau} \\
& 1'W = 1, \quad W'1 = 1, \quad w_{ij} \geq 0 \\
& \boldsymbol{\lambda} \in \Lambda.
\end{aligned} \tag{4}
$$

where $\boldsymbol{\tau}$ is the benchmark portfolio, in this paper: $\tau_k = 0.1, \, k = 1, ..., 10$, and W is a double stochastic matrix with elements w_{ij}.

4 Empirical study

4.1 Data description

We consider ten industry representative portfolios from the Kenneth French library as the base assets in our empirical study. The data of daily returns are divided into two parts: in-sample period (1.1.2019 – 30.6.2020) and out-of-sample period (1.7.2020 – 30.6.2021). Descriptive statistics of both datasets are summarized in Table 1 and Table 2.

Table 1: Basic descriptive statistics of daily returns (in %): in sample period

	mean	st. dev.	min	max	skewness	kurtosis
NoDur	0.043	1.596	-9.870	7.450	-0.746	11.284
Durbl	0.141	2.429	-14.430	15.030	-0.549	9.922
Manuf	0.050	1.964	-11.110	10.830	-0.497	9.475
Enrgy	-0.055	2.871	-19.730	16.000	-0.762	11.521
HiTec	0.152	1.977	-13.180	10.690	-0.530	10.978
Telcm	0.050	1.620	-9.080	9.060	-0.481	9.863
Shops	0.103	1.540	-10.610	7.050	-0.983	12.565
Hlth	0.066	1.579	-9.740	6.980	-0.419	8.387
Utils	0.036	1.964	-11.610	11.760	-0.007	13.133
Other	0.045	2.138	-13.380	12.240	-0.570	11.419

Table 2: Basic descriptive statistics of daily returns (in %): out–of–sample period

	mean	st. dev.	min	max	skewness	kurtosis
NoDur	0.115	0.835	-3.020	2.420	-0.195	0.764
Durbl	0.400	2.862	-12.400	11.510	0.004	2.004
Manuf	0.163	1.073	-2.960	2.740	-0.169	0.294
Enrgy	0.203	2.422	-5.650	14.040	0.902	3.689
HiTec	0.167	1.429	-5.670	4.290	-0.537	1.150
Telcm	0.114	0.960	-3.100	3.450	-0.108	1.115
Shops	0.129	1.066	-3.510	2.890	-0.365	0.894
Hlth	0.100	0.956	-3.010	3.780	-0.043	1.105
Utils	0.073	1.027	-2.950	2.730	-0.143	0.041
Other	0.177	1.163	-3.120	5.660	0.229	1.654

4.2 Results without SSD constraints

As the distortion function we consider only PHT but with several parameters γ. To fulfil the assumptions of Theorem 2, we consider only $\gamma \geq 1$. Note that if $\gamma = 1$ then the distortion measure is expected value and, hence, the optimal portfolio invests everything in the fifth asset which is the most profitable one. The optimal portfolios of (3) for considered parameters γ are summarized in Table 3.

Table 3: Compositions of optimal portfolios: in sample period

γ	NoDur	Durbl	Manuf	Enrgy	HiTec	Telcm	Shops	Hlth	Utils	Other
1					1					
1.5						0.026	0.654	0.320		
2	0.036					0.174	0.398	0.392		
3	0.037					0.465		0.498		
4						0.633		0.367		
5						0.819		0.181		
10						0.910		0.090		

We can see that as the risk aversion expressed by parameter γ increases the portfolio is more concentrated in least risky assets. Moreover, for very large values of parameter γ, (almost) everything is invested in the asset with the highest minimal return. This is due to the fact that investors with extremely large risk aversion want to hedge against the worst realizations – smallest returns.

Finally, we present mean returns and distortion risk measures of the optimal portfolios from Table 3. For the sake of comparison, we express the risk using PHT distortion measure with $\gamma = 2$.

Table 4: Risk – return performance (in %) of optimal portfolios: in sample & out–of–sample period

in sample \ γ	1	1.5	2	3	4	5	10
Mean return	0.1520	0.0899	0.0774	0.0581	0.0563	0.0533	0.0519
Risk	1.6140	1.2717	1.2647	1.2825	1.2910	1.3110	1.3251
Out-of-sample \ γ	1	1.5	2	3	4	5	10
Mean return	0.1674	0.1198	0.1149	0.1074	0.1091	0.1117	0.1130
Risk	0.9583	0.5957	0.5403	0.5230	0.5334	0.5540	0.5693

4.3 Results with SSD constraints

When imposing SSD constraints, the following table summarizes the optimal portfolios of (4) for various values of parameter γ.

Table 5: Compositions of optimal portfolios with SSD constraints: in sample period

γ	NoDur	Durbl	Manuf	Enrgy	HiTec	Telcm	Shops	Hlth	Utils	Other
1					0.363		0.637			
1.5						0.031	0.646	0.323		
2	0.029					0.177	0.403	0.391		
3						0.418	0.098	0.484		
4						0.436	0.107	0.457		
5						0.479	0.125	0.396		
10						0.754	0.246			

Since the set of feasible portfolio is reduced by the SSD constraint, the optimal portfolios are more diversified and their in sample and out-of-sample performace is depicted in Table 6.

Table 6: Risk – return performance (in %) of optimal portfolios with SSD constraints: in sample & out–of–sample period

in sample \ γ	1	1.5	2	3	4	5	10
Mean return	0.1208	0.0895	0.0777	0.0633	0.0634	0.0633	0.0634
Risk	1.3988	1.2712	1.2647	1.2755	1.2756	1.2767	1.3071
Out-of-sample \ γ	1	1.5	2	3	4	5	10
Mean return	0.1432	0.1196	0.1150	0.1090	0.1095	0.1106	0.1180
Risk	0.7553	0.5937	0.5412	0.5240	0.5234	0.5231	0.5472

5 Conclusions

The paper deals with portfolio selection models which minimize risk expressed by a distortion risk measure under SSD constraints. These special constraints express preference between the optimal portfolio and a given bechmark ($1/N$ strategy), such that every non-satiated risk averse investor prefers the optimal portfolio to the benchmark or is indifferent between them.

We first computed the optimal portfolios of the risk minimizing problem without SSD constraints for various choices of distortion function (its parameter). We conclude that the optimal portfolio :

- invests almost everything in the most profitable asset for the small value of risk aversion parameter in the PHT
- distortion function
- is more and more diversified as the risk aversion increases, but is still relatively low
- invest almost everything in the asset with the highest minimal return when the risk aversion is very high.

The in sample and out–of–sample risk–reward performance looks as one would expect – the higher the risk aversion parameter is, the smaller the return and risk of the optimal portfolio is observed. However for the out–of–sample case we can find some exceptions for high values of the parameter.

When imposing the stochastic dominance constraints, this risk–reward performance slightly changes, however, the general message is the same. Comparing to the unconstrained risk minimization, now the optimal portfolios are slightly less profitable for small values of γ but slightly more profitable for large values of the parameter and slightly less risky for all values of γ. The effect of stochastic dominance constraints is the largest for very low risk aversion parameter, negligible for moderate choices and considerable for the highest values of the parameter.

Although the paper presents only static model, the distortion measures could be similarly applied to multistage models with exogenous [15], [16] or endogenous randomness [6].

Acknowledgements

The paper was supported by the grant No. 19-28231X of the Czech Science Foundation.

References

[1] Cherny, A. & Madan, D. (2009). New measures for performance evaluation. *The Review of Financial Studies*, 22, 2571–2606.

[2] Dhaene, J., Kukush, A., Linders, D. &Tang, Q. (2012). Remarks on quantiles and distortion risk measures. *European Actuarial Journal*, 2, 319–328.

[3] Kahneman, D. & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47, 263–292.

[4] Kopa, M. (2021): Risk minimization using distortion risk measures via linear programming, *Proceedings of 13th International conference on Financial Management of Firms and Financial Institutions*, pp. 81–88.

[5] Kopa, M., Kabasinskas, A. & Sutiene, K. (2022). A stochastic dominance approach to pension-fund selection, *IMA Journal of Management Mathematics*, 33 (1) , 139–160.

[6] 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.

[7] Kopa, M. & Zelman, J. (2021): Distortion risk measures in portfolio optimization, *Proceedings of the 39th International Conference on MME2021*, pp. 255-260.

[8] Moriggia, V., Kopa, M. & Vitali, S. (2019). Pension fund management with hedging derivatives, stochastic dominance and nodal contamination. *Omega*, 87, 127–141.

[9] Sereda, E. N., Bronshtein, E. M., Rachev, S. T., Fabozzi, F. J., Sun, W. & Stoyanov, S. V. (2010). Distortion risk measures in portfolio optimization Handbook of portfolio construction, pp. 649–673.

[10] Van der Hoek, J. & Sherris, M. (2001). A class of non-expected utility risk measures and implications for asset allocations. *Insurance: Mathematics and Economics*, 28, 69–82.

[11] Wang, S. (1995). Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance*. *Mathematics and Economics*, 17, 43–54.

[12] Wang, S. (2000). A class of distortion operators for pricing financial and insurance risks. *Journal of risk and insurance*, 67, 15–36.

[13] Wirch, J. & Hardy, M. (1999). A synthesis of risk measures for capital adequacy *Insurance: mathematics and economics*, 25, 337-347.

[14] Yaari, M. E. (1987). The dual theory of choice under risk. Econometrica: *Journal of the Econometric Society*, 55, 95–115.

[15] Vitali, S., Moriggia, V. & Kopa, M. (2017). Optimal pension fund composition for an Italian private pension plan sponsor. *Computational Management Science*, 14, 135–160.

[16] Zapletal, F., Šmíd, M. & Kopa, M. (2020). Multi-stage emissions management of a steel company. *Annals of Operations Research*, 292, 735–751.

The Position of the Czech Republic within the Metallurgical Sector

Michal Krejčí[1], Michaela Staňková[2]

Abstract. This article deals with the evaluation of technical efficiency in the metallurgical sector in selected EU countries. The data envelopment analysis method was chosen to calculate the efficiency. The efficiency evaluation is performed on the basis of data from 2000 to 2015 with an emphasis on the position in the Czech Republic. Radial input-oriented models are constructed separately for individual years. In addition to calculating the efficiency itself, the article also focuses on changes in efficiency and changes in the production possibility frontier through the Malmquist index.
The results of this article show that the Czech Republic lags far behind other countries in terms of efficiency for the entire period under review. In the case of the Czech Republic the largest increase in efficiency was recorded in 2003–2004, when the government decided to transfer several companies into private ownership.

Keywords: data envelopment analysis, efficiency, linear programming, metallurgical sector

JEL Classification: C44, D24
AMS Classification: 90B50, 90C08

1 Introduction

The metallurgical industry, together with the engineering, chemical and food industries, has a privileged position in the Czech Republic. The metallurgical sector is also of great importance in other EU countries. According to a European Commission report [5], this sector is also important in terms of competitiveness and industrial development across the EU. In terms of its share of GDP, the metallurgical industry has 11%, while activities relating to the metallurgical industry represent 46% of total manufacturing value. According to the above-mentioned report, European companies were once among the leaders in this sector, but in the new millennium they are hardly keeping up with the competition from the American and Asian markets.

Many stakeholders (including individual governments and pan-European institutions) emphasize the need to strengthen the metallurgical industry. This is, for example, the pan-European "metallurgical infrastructure" call, in which academic, industrial and governmental organizations participate through a dedicated R&D and Innovation program for metallurgy in Europe. Analysis of technological trends, new applications and innovation issues are a common topic in this sector, see for example [4] and [13]. Instead of the process itself, this article focuses on the evaluation of the technical (production) efficiency of companies in the manufacturing sector at the level of individual countries in the EU.

The evaluation of efficiency in this article is based on a nonparametric method of data envelopment analysis (DEA) derived from Farrell's ideas [6]. Later, several modifications were made to Farrell's original ideas, such as the CCR [8] and BCC models [1]. However, all DEA models, regardless of the different settings (for example, regarding the assumption of returns to scale) have a common idea in the mutual comparison of the number of inputs consumed and the number of outputs produced between the individual decision-making units (so-called DMUs); the DMUs with the highest efficiency form the production possibility frontier.

The DEA method can be considered the most widely used method for evaluating technical efficiency. It has found applications in various sectors. We can name, for example, the areas of education [11], transport [16], pharmacy [7], up to and including the manufacturing sector. Current studies using the DEA method within the manufacturing sector include [10], [14], [17] and [19].

The main aim of this article is to evaluate the technical efficiency of selected EU countries in the metallurgical sector, including a more detailed analysis of the development of efficiency (change in the efficiency of each DMU,

[1] Mendel University in Brno, Department of Statistics and Operation Analysis, Zemědělská 1, 613 00 Brno, Czech Republic, xkrejci9@mendelu.cz.
[2] Mendel University in Brno, Department of Statistics and Operation Analysis, Zemědělská 1, 613 00 Brno, Czech Republic, michaela.stankova@mendelu.cz.

but also a change in the production possibility frontier). This aim will be pursued for the Czech Republic in particular.

2 Material and Methods

Annual aggregated data provided by the EU KLEMS database were used to calculate efficiency. Due to the nature of the dataset (aggregated annual values across countries), the CCR model was used, which uses constant returns to scale. The last choice made when setting up the model was its orientation. With regard to the European Commission strategy, the input orientation of the model was chosen, which assumes a proportional reduction of inputs while maintaining the level of outputs. According to the European Commission [5], there is room in this sector to shorten the development cycle for new product development and also to reduce inputs and thus subsequently reduce a company's costs. The CCR input-oriented model compiled for unit *H*, which is one of the *p* units, is as follows:

$$
\begin{aligned}
\text{max.} \qquad & E_H = \sum_{j=1}^{n} v_{jH} y_{jH}, \\
\text{subject to} \qquad & \sum_{i=1}^{m} u_{iH} x_{iH} = 1, \\
& -\sum_{i=1}^{m} u_{iH} x_{ik} + \sum_{j=1}^{n} v_{jH} y_{jk} \leq 0, \forall k = 1,2,\dots,p, \\
& u_{iH} \geq \varepsilon, \forall i = 1,2,\dots,m, \\
& v_{jH} \geq \varepsilon, \forall j = 1,2,\dots,n,
\end{aligned}
$$

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 $Y = \{y_{ik},\ i = 1, 2, \dots, n, j = 1, 2, \dots, p\}$. The ε is the so-called infinitesimal constant.

A total of four variables were used (two input and two output variables). The input variables representing the labor and capital factors are the number of employees in thousands and nominal gross fixed capital formation in millions of national currency units. Output variables representing the company's product at country level are gross value added and gross output; both at current basic prices in millions of national currency units. In order to be able to correctly compare data from the EU KLEMS database, all the above-mentioned variables (except labor) were converted from the national currency into the euro using annual exchange rates as in [10] or [14]. Due to the availability of data throughout the monitored period in the selected sector, it was possible to include a total of 12 EU countries in these analyses: Austria (AT), the Czech Republic (CZ), Denmark (DK), Finland (FI), France (FR), Germany (DE), Greece (GR), Italy (IT), the Netherlands (NL), Spain (ES), Sweden (SE) and the United Kingdom (UK).

Due to the selected longer period of time, in addition to calculating the efficiency of the DMU individually for each period, attention was also paid to the change in efficiency and also to the change in production possibilities. For this purpose, the Malmquist index was calculated, which hides the influence of both of these components. As a whole, the Malmquist index represents either an improvement in the situation of a DMU (if the index value is greater than 1), a deterioration of the situation (if its value is less than 1) or an invariance of the situation (if it is equal to 1). However, it is possible to decompose this index and monitor separately the change in efficiency (the so-called Catch-up effect) and the change in the production frontier (Frontier shift) using the same logic as in the case of the entire Malmquist index.

To calculate the Malmquist index for two periods, it is necessary to solve four linear programs (i.e. the four CCR input-oriented models in this paper). The Malmquist index (MI) can be defined as the geometric mean of two efficiency ratios (E), where one is the efficiency change measured by the period 1 technology and the other is the efficiency change measured by the period 2 technology:

$$
MI = \left[\frac{E^1((x_H, y_H)^2)}{E^1((x_H, y_H)^1)} * \frac{E^2((x_H, y_H)^2)}{E^2((x_H, y_H)^1)} \right]^{1/2}.
$$

Technical details concerning the CCR models and Malmquist index can be found in [2]. The calculations were performed in DEA SolverPro version 15; figures were performed the MATLAB R2021b.

3 Results and Discussion

The technical efficiency results for selected EU countries are recorded in Figure 1. According to the results of the CCR input-oriented model, only two countries, namely Finland (FI) and Germany (DE), are fully efficient throughout the period under review. Austria (AT), France (FR) and Sweden (SE) can also be included among the countries with the highest efficiency achieved throughout the period under review. On the contrary, the Czech Republic (CZ) took the worst position regardless of the specific year. However, the very low efficiency results are in line with findings [3] that the Czech Republic has problems with labor productivity in this sector. The business environment and its legislation are also responsible for the country's failures in terms of efficiency. According to [12] results, it is possible to detect a strong correlation between GDP and production in the metallurgical sector. Based on the findings in this article, it can be assumed that GDP levels may also have a strong link to efficiency. Because throughout the period under review, the Czech Republic has the lowest GDP per capita compared to other selected countries.

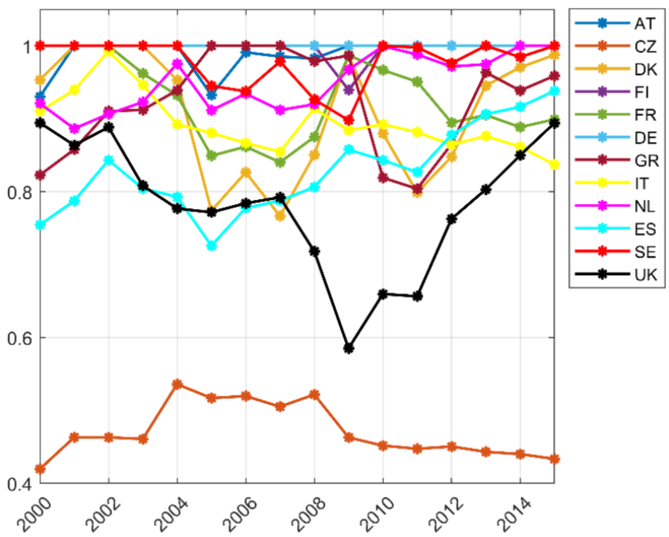


Figure 1 The resulting country efficiency score in individual years based on the CCR input-oriented model

However, if we focus on the values of the Malmquist Index, it can be stated that all selected countries have improved their overall situation from 2000 to 2015, see the values greater than 1 in the Malmquist Index in Table 1.

Country	Catch-up	Frontier shift	Malmquist index
Austria (AT)	1.0748	1.3426	1.4430
Czech Republic (CZ)	1.0329	1.3676	1.4126
Denmark (DK)	1.0357	1.2743	1.3198
Finland (FI)	1.0000	1.3012	1.3012
France (FR)	0.8991	1.3071	1.1752
Germany (DE)	1.0000	1.3480	1.3480
Greece (GR)	1.1651	1.2449	1.4505
Italy (IT)	0.9198	1.2890	1.1856
Netherlands (NL)	1.0858	1.3172	1.4302
Spain (ES)	1.2433	1.3514	1.6801
Sweden (SE)	1.0000	1.2996	1.2996
United Kingdom (UK)	0.9997	1.3124	1.3119

Table 1 Overall change in technical efficiency (Catch-up), change in the production frontier (Frontier shift) and change in the Malmquist index from 2000 to 2015 for individual countries

An increase in the production possibility frontier was also identified in all countries (i.e. values greater than 1 in the Frontier shift area in Table 1). The biggest differences between countries are due to the change in efficiency, see Catch-up effect in Table 1. France, Italy and the United Kingdom performed worse in terms of efficiency in 2015 than in 2000. However, this drop in efficiency has been offset by an increase in the production frontier, and so the resulting Malmquist index is still greater than 1. Three countries (Finland, Germany and Sweden) have the same efficiency in 2015 as in 2000. A positive fact in the case of the Czech Republic is that although it is in the worst position in terms of efficiency values, the efficiency of this country increased by about 3% in 2015 compared to 2000. A detailed view of the changes within the Malmquist index, including its subcomponents, is shown in Figure 2.

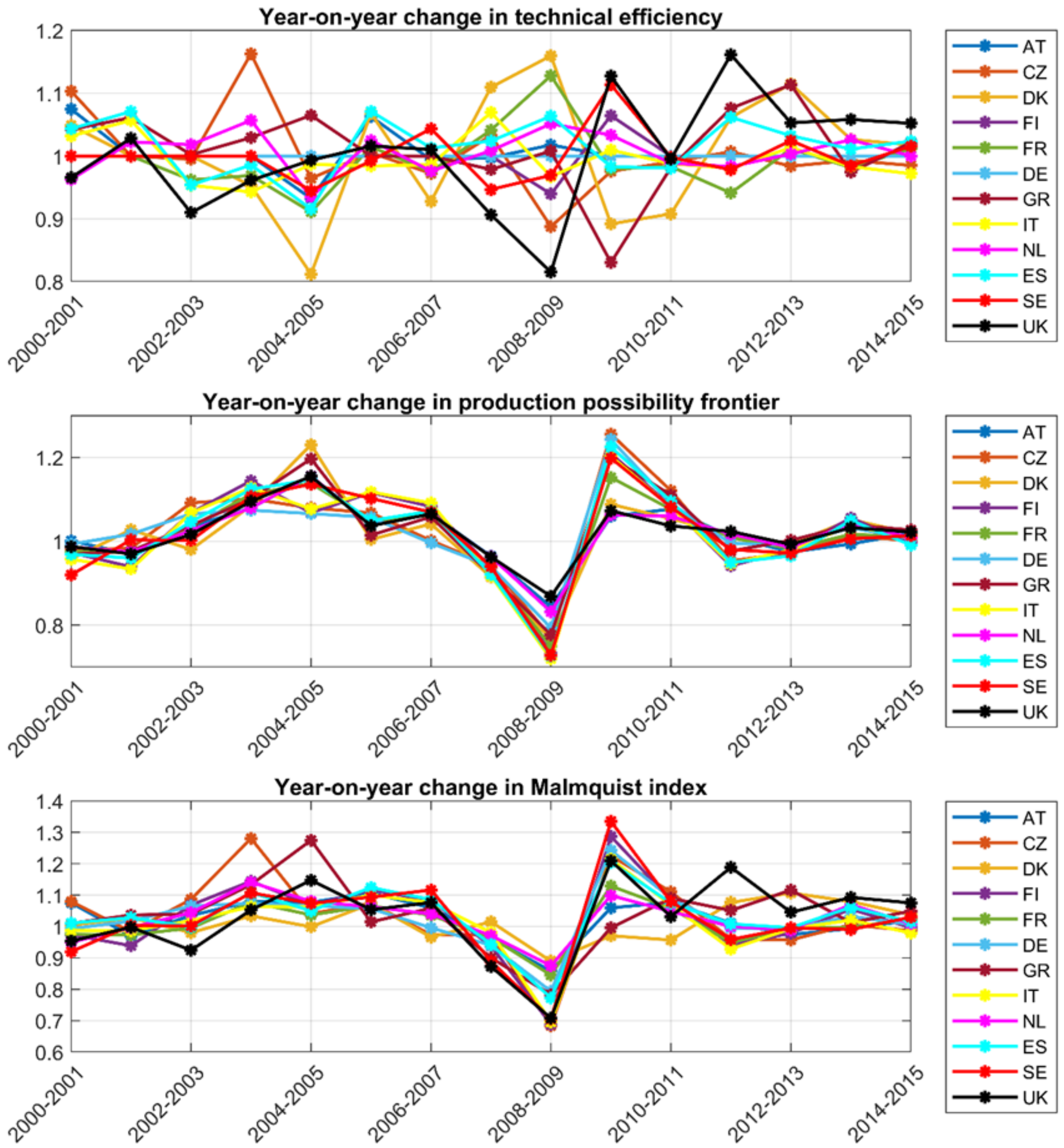


Figure 2 The development of changes in technical efficiency, the production possibility frontier and the Malmquist index itself for selected EU countries

The smallest differences between countries are evident in the changes in the production possibility frontier. At the beginning of the millennium, the frontier was generally growing. According to [12], there has been a general decline in production since 2007, which can be seen in the decline in the frontier in this period. The peak of this decline was observed in 2009 due to the economic crisis. However, after the subsequent recovery and resumption of production, the second decline in the frontier comes in 2012. A parallel with the development of GDP can be seen for this decline. For example, in the Czech Republic, this was the first decline in GDP per capita since the crisis, namely a decline of around 1 percentage point.

As the changes from the frontier shift are similar for all the selected countries, the differences in the Malmquist index of individual countries are mainly due to the change in individual efficiency. The Czech Republic recorded a large increase in efficiency (and thus in its overall situation) in 2004. According to [9], the development of the entire Czech metallurgical industry can be illustrated by the development of the Vítkovice ironworks. And it is this increase in the efficiency of the sector that can be demonstrated in the development of this company. In 2003, the government decided to transfer Vítkovice (as well as other entities) into private ownership. This change of ownership permitted conceptual changes at companies, which increased the productivity and efficiency of the entire industry. Unfortunately, the change of ownership did not have a long-term positive relationship to efficiency. Even according to [9] reports, there was a noticeable improvement in productivity (specifically labor productivity) only in 2003 and 2004, but in the following years there was a systematic decline.

The metallurgy industry is still a relatively little explored area at present and there is only a limited amount of research into the sector's efficiency. However, it can be stated that the results of this article are consistent with other analyzes that have been performed, such as [3], [9] and [12]. To validate the results of the efficiency calculation and their changes in this article, it would be possible to make estimates using the competitive stochastic frontier method. In this case, it would be possible to estimate any of the panel models as in [15]. Unlike the DEA method, however, competitive approaches do not allow the use of multiple output variables.

4 Conclusion

This article dealt with the evaluation of the technical (production) efficiency of the metallurgical industry in selected EU countries. In addition to a calculation of the efficiency score itself, the change in efficiency and the change in the production possibility frontier were also evaluated using the Malmquist index. Special attention was paid to the position of the Czech Republic.

According to the results obtained, the Czech Republic lags far behind other selected EU countries. The low efficiency score is caused not only by the overall economic situation and low labor productivity, but also by legal aspects in the country. On the contrary, Finland and Germany performed best in terms of efficiency scores throughout the period under review.

Acknowledgements

This article was supported by the grant No IGA-PEF-TP-22-003 of the IGA PEF MENDELU Grant Agency.

References

[1] Banker, R. D., Charnes, A. & Cooper, W. W. (1984). The Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management science*, 30(9), 1078–1092.

[2] 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.

[3] Czech Business and Trade (2011). *Metallurgy, Foundry, Industry, and Steel Construction.* Praha: PP Agency.

[4] Dembele, S., Akcil, A. & Panda, A. (2022). Technological trends, emerging applications and metallurgical strategies in antimony recovery from stibnite. *Minerals Engineering*, 175, 107304.

[5] Europeean Commission. (2014). *Metallurgy made in and for Europe*. Luxembourg: Publications Office of the European Union.

[6] Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society*, 120(3), 253–290.

[7] Gaebert, T. & Staňková, M. (2020). Efficiency Development in the German Pharmaceutical Market. *Acta Universitatis agriculturae et silviculturae Mendelianae Brunensis*, 68(5), 877–884.

[8] Charnes, A., Cooper, W. W. & Rhodes, E. (1978). The measurement of productive efficiency. *Journal of the Royal Statistical Society*, 2(6), 429–444.

[9] Knížek, M. (2018). *Příběh Vítkovických železáren odráží historii českého hutního průmyslu. Po rozvoji ve 20. století přišel výrazný útlum na přelomu milénia*. Hospodářské noviny. [Online]. Available at: https://archiv.ihned.cz/c1-66299750-historie-vitkovickych-zelezaren [Accessed: 2022, April 16].

[10] 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.

[11] 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.

[12] Rokicki, T. (2016). Situation of steel industry in European Union. *Metal 2016: 25th Anniversary International Conference on Metallurgy and Materials*. Ostrava: TANGER Ltd., 1981–1986.

[13] Smol, M. & Kulczycka, J. (2019). Towards innovations development in the European raw material sector by evolution of the knowledge triangle. *Resources Policy*, 62, 453–462.

[14] Staňková, M. (2020). Efficiency comparison and efficiency development of the metallurgical industry in the EU: Parametric and non-parametric approaches. *Acta Universitatis agriculturae et silviculturae Mendelianae Brunensis*, 68(4), 765–774.

[15] Staňková, M. & Hampel, D. (2021). On the influence of model setting on stochastic frontier analysis. *Mathematical Methods in the Applied Sciences*, 44(1), 274–283.

[16] 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.

[17] 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.

[18] 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.

[19] 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), in print, https://doi.org/10.5552/crojfe.2022.1541.

Comparison of the Relationship between CPI and PPI in the Czech Republic and Slovakia

Radmila Krkošková[1]

Abstract. The goal of this article is to investigate the existence and character of the relationship between the Consumer Price Index (CPI) and the Producer Price Index (PPI) for the Czech Republic and Slovakia. Is there a relationship between the CPI and PPI? The CPI-PPI indicator also has a strong correlation with the overall growth rate of the economy. The smaller the difference of CPI-PPI, the higher the operating cost of the enterprise, the lower the profit of the enterprise, the slower growth of the profit of the enterprise, and the slower economic growth.

The existence of long-run equilibrium relationship between CPI and PPI has been tested with the Johansen cointegration. The existence of long-run equilibrium relationship between CPI and PPI was confirmed in the case of Slovakia. The short- run dynamics were confirmed through statistical methods such as VEC model, Granger causality, and impulse-response function. Granger causality CPI→PPI was confirmed in Slovakia, the causality PPI→CPI was confirmed in the Czech Republic. The EViews 12 was used for data evaluation. The results are based on the analysis of the data from Eurostat from January 2005 to June 2021 for the Czech Republic and Slovakia.

Keywords: ADF test, CPI, Granger causality, impulse-response function, PPI, VEC model

JEL Classification: C51, F63
AMS Classification: 62P20, 91B02

1 Introduction

The topic of causality between Producer Price Index, (PPI); and Consumer Price Index, (CPI) has been studied by many authors. CPI has been considered as one of the most important economic indicators by economic growth. The structure of the PPI differs country from country, similarly to the structure of industry, and it is the disadvantage of the PPI. There are a lot of results have been found, from the PPI is an anticipated indicator of the CPI, to they are coincident indicators, that the CPI leads the PPI. On this article, there has been studied the relationship between PPI and CPI for the Czech Republic and Slovakia.

These indicators can help determine the status quo and trends of inflation [14]. CPI mainly reflects the price changes of daily life commodities and service items of urban households. It is one of the important indexes for judging the degree and trend of inflation. CPI is divided into food, tobacco, alcohol and supplies, household equipment and maintenance services, health care and personal goods, transportation and communications, entertainment, education, cultural goods and services, and housing. PPI can be regarded as the most important price index of the mid-stream industry, and it plays a role in the product price trend. PPI represents the production materials and consumer goods sold by industrial enterprises to commercial, foreign trade, materials, residents or other industries and sectors. It reflects the impact of ex-factory price changes on industrial output value within a certain period of time.

The aim of this article is to examine existence and character of relationship between CPI and PPI. The existence of long-run equilibrium relationship between CPI and PPI is tested with the Johansen cointegration. The short run dynamics between the variables is examined by Vector Error Correction modelling and the Granger causality test. Moreover, for the validity of the results, diagnostic test such as impulse responses have been checked. The paper is organized in these parts: the introduction, the literature review, the research methods, the results and the 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

The causality between CPI and PPI has been studied by various authors around the world. According to economic theories, supply shocks (changes in raw material prices and factors of production) are immediately reflected in the PPI, while demand shocks (changes in the money supply) are reflected in the CPI.

Intuition tells us that the development of producer prices affects the development of prices for end consumers. Article [4] is one of the first studies to deal with this relationship. It examines the hypothesis that the increased costs at the beginning of the production process can be expected to be reflected in the entire production chain and lead to rising consumer prices. The conclusions of the Clark's study [4] say that the relationship between the PPI and the CPI is weak – changes in the PPI can sometimes help predict changes in the CPI, but not systematically.

The academic literature focuses mainly on finding out what the relationship between the PPI and the CPI is and how they affect each other. Table 1 shows choosing of the literature.

Author/s	**Country**	**Results**
Lima	Brazil	*PPI→CPI*
Gang *et al.*	China	*PPI→CPI*
Caporale *et al.*	USA, Canada, Germany, France, Italy, UK, Japanese	*PPI→CPI*
Woo *et al.*	UK, France, Germany	*PPI→CPI*
Junicke	Poland	*PPI→CPI*
Khan *et al.*	Czech Republic, Slovakia	*PPI→CPI*
Akdi *et al.*	Swedish, UK, Canada	*CPI→PPI*
Losada *et al.*	Brazil, Colombia, Ecuador, Uruguay	*none*
Akcay	Germany	*PPI↔CPI*
Tiwari *et al.*	Mexico	*PPI↔CPI*

Table 1 Studies which analyse the relationship between *CPI* and *PPI*

Table 1 shows diverse results from different countries. The results of this article: CPI→PPI was confirmed in Slovakia, the causality PPI→CPI was confirmed in the Czech Republic.

2.2 Data and Methods

Both CPI and PPI data were obtained from Eurostat. For the study, the Czech Republic and Slovakia have been chosen. The data's periodicity is monthly from January 2005 to June 2021. The program EViews 12 was used to analyze time-series variables. The data from these databases are on percentages. Data are in the form of basic indices (2015 = 100).

The following hypotheses were formulated to fulfil the main aim. The core of this paper is to analyse two hypotheses concerning the relationship between CPI and PPI in the Czech Republic and Slovakia.

H1: There is a long-term relationship between CPI and PPI in the Czech Republic (H1a), Slovakia (H1b).

H2: There is a short-term relationship between CPI and PPI in the Czech Republic (H2a), Slovakia (H2b).

The hypotheses were confirmed or rejected through statistical methods such as VEC model, Granger causality and impulse-response analysis.

The first step is to determine if the variables are stationary. The augmented Dickey-Fuller (ADF) test will be used to decide on the stationarity of variables. In [5] author argues that most time series in macroeconomics are nonstationary or integrated with order I (1).

If all the variables are stationary, then vector autoregressive (VAR) model can be used. If the variables of type I (1) we use the VEC model and test the cointegration and determine the Granger causality connections.

Time series can be analysed based on their short-term and long-term relations. If there is only a short-term relation between the time series, the VAR model is a sufficient tool for analysing this relation. If a long-term relation exists between selected time series, the VEC model can be used for the analysis. The VEC model simultaneously captures and expresses both short-term and long-term relations. The VEC model is based on a cointegration approach that models non-stationary time series the long-term relation of which is expressed through the error correction mechanism.

Model autoregressive distributed lags (1,1) is given by equation: $Y_t = c + \alpha Y_{t-1} + \beta_1 X_t + \beta_2 X_{t-1} + a_t.$ (1)

The general form of the VEC model is: $\Delta Y_t = c + \beta_1 \Delta X_t + \gamma(Y_{t-1} - \beta X_{t-1}) + a_t$ (2)

$$\beta = \frac{(\beta_1 + \beta_2)}{(1-\alpha)}, \quad \gamma = \alpha - 1$$

The correction term $\gamma(Y_{t-1} - \beta X_{t-1})$ corrects deviations from equilibrium. By definition, a negative parameter γ is assumed. This parameter indicates the time it takes for the system to equilibrate. Co-integration parameter β expresses long-term dependence between Y and X.

The cointegration analysis is based on the integrated processes that were first comprehensively addressed by Box and Jenkins. The cointegration analysis examines short-term dynamics and long-term relations between variables. In [7] we can see that the Granger causality test is a statistical hypothesis test for determining whether one time series is useful in forecasting another. Using the term “causality” is a misnomer, as Granger-causality is better described as “precedence” [13], or as Granger later claimed “temporally related” [8]. Rather than testing whether *Y* causes *X*, Granger causality tests whether *Y* forecasts *X* [9]. It is necessary to work with the stationary time series.

Impulse-response analysis allows the use of both the short-term and long-term relations between the analysed variables based on the derived model. We can imagine, that the residuals of the VEC model: ε_{1t} and ε_{2t} are as deviations (impulses, random shocks) of each one of the variables from a theoretical estimation, like Stock and Watson [17] show. The impulse response functions display how the behaviour, of one endogenous variable, changes against a simulated random shock of the other variables. There was used the decomposition method Cholesky (d. f. adjusted), with innovations equal to one standard deviation.

3 Data Analysis

The modelling structure consists of the following steps: testing the presence of unit roots, the VAR model estimation, Granger causality, and response impulse analysis. All values were *seasonally adjusted* and were considered *in logarithmic terms*.

The development of the CPI and the PPI is shown graphically in the following Figure 1. The graphs show that, from a historical perspective, PPI is more volatile than the CPI

Figure 1 CPI and PPI for the Czech Republic and Slovakia in years 2005 - 2021 (Index, 2010 = 100)

The preparatory phase of estimating the VAR model is testing the stationarity of variables included in the model or their first differences. The test results for all variables are provided in Table 2. The Dickey-Fuller test (ADF) was used to test the stationarity. The last column includes the result of testing: N = non-stationary (H0 not rejected), S = stationary (H0 rejected).

Variable	n/c/c+t	T-stat	Signif.	Result	Variable	n/c/c+t	T-stat	Signif.	Result
CPI CZ	c	– 0.22	0.93	N	D(*CPI CZ*)	c	– 12.5	0.00***	S
PPI CZ	n	– 0.91	0.78	N	D(*PPI CZ*)	n	– 9.66	0.00***	S
CPI SK	n	–0.79	0.82	N	D(*CPI SK*)	n	– 8.17	0.00***	S
PPI SK	c	–2.38	0.15	N	D(*PPI SK*)	c	– 9.87	0.00***	S

Statistical significance at the 0.01 level (***), at the 0.05 level (**), at the 0.1 level (*)

Table 2 Testing the unit root of the variables in levels and their first differences

In the second stage of this study, Johansen cointegration test is used to test the long-run relationship of PPI and CPI series, [10]. Since PPI and CPI series are integrated with the same order I (1) for all countries, cointegration test can be conducted. The results are in Table 3. The one cointegration relationship was confirmed for both countries. For both countries was utilized a model assume no deterministic trend in data.

Country	Hypothesis	Eigen Value	Trace Statistic	0.05 Critical	Signif.
Czech Republic	None	0.202	47.42	20.261	0.0000***
	At most 1	0.015	3.017	9.164	0.5775
Slovakia	None	0.179	45.81	20.26	0.0000***
	At most 1	0.034	6.953	9.164	0.1289

Statistical significance at the 0.01 level (***), at the 0.05 level (**), at the 0.1 level (*)

Table 3 Co-integration rank test (Trace)

The results in the Table 3 indicate the presence of a cointegrating vector between the series according to the statistic of the trace and the statistic of the maximum own value to 5 % of significance. Therefore, there can only be a cointegration relation with two variables. It is concluded that there is a long-term stationary relationship between the CPI and the PPI in the both countries.

Formally, VEC (1) models are defined below. The cointegration equations (3) and (4) show that the CPI is positively affected in the long term by PPI. Growth in this variable causes an increase in the CPI

Czech Republic
$$\begin{pmatrix}\Delta CPI\\ \Delta PPI\end{pmatrix}_t = \begin{pmatrix}-0.01 & 0.11^{**}\\ -0.16 & 0.38^{***}\end{pmatrix}\begin{pmatrix}\Delta CPI\\ \Delta PPI\end{pmatrix}_{t-1} + \begin{pmatrix}0.001\\ 0.001\end{pmatrix} EQ_{t-1} + \begin{pmatrix}\mu_{1t}\\ \mu_{2t}\end{pmatrix}$$
$$EQ_{t-1} = 353.3 + CPI_{t-1} - 2.653PPI_{t-1} \tag{3}$$

The positive error correction is not good sign for model because it implies that the process is not converging in the long-run and it could be perhaps due to some instabilities in the model.

Slovakia
$$\begin{pmatrix}\Delta CPI\\ \Delta PPI\end{pmatrix}_t = \begin{pmatrix}0.08 & 0.03\\ 0.74^{***} & 0.19^{***}\end{pmatrix}\begin{pmatrix}\Delta CPI\\ \Delta PPI\end{pmatrix}_{t-1} + \begin{pmatrix}-0.002\\ 0.001\end{pmatrix} EQ_{t-1} + \begin{pmatrix}\mu_{1t}\\ \mu_{2t}\end{pmatrix}$$
$$EQ_{t-1} = 346.5 + CPI_{t-1} - 5.009PPI_{t-1} \tag{4}$$

Causality is captured by the statistically significant value EC1 (−0.002), which indicates that this variable will be modified by 0.2% within 1 month in case of long-term instability of the CPI for Slovakia.

Next part of article deals with the testing of short-term relationships (Granger causality). The Granger causality test is a statistical hypothesis test for determining whether one time series is useful in forecasting another [7]. Rather than testing whether *Y* causes *X*, the Granger causality tests whether *Y* forecasts *X*. It is necessary to work with the stationary time series. The similar procedure is given in [18]. We consider the 5% significance level. The results of the series 1 delay test are shown in Table 4. Table 4 shows, that CPI Granger causes PPI was confirmed in Slovakia, and hypothesis PPI Granger causes CPI was confirmed in the Czech Republic.

Null Hypothesis:	Chi-sq.	Signif.	Rejection of H_0
D(PPI CZ) does not Granger cause	7.08	0.007^{***}	YES
D(CPI CZ) does not Granger cause	1.13	0.288	NO
D(PPI SK) does not Granger cause	1.06	0.302	NO
D(CPI SK) does not Granger cause	25.4	0.000^{***}	YES

Statistical significance at the 0.01 level (***), at the 0.05 level (**), at the 0.1 level (*)

Table 4 VEC Granger Causality/ Block Exogeneity Wald Tests

The impulse-response analysis is related to the question of what reaction in one time series will be caused by an impulse in another time series. Figure 3 shows the response to Cholesky for both of countries. In the Czech Republic, the impact of a shock on the CPI in the PPI is similar like the impact of a shock on the PPI in the CPI. Firstly, it is positive and than the changes in the PPI stabilized around the tenth month by approximately 0.069 %. The changes in the CPI are stabilized around the fourth month by 0.322 %. The impact of a shock on the CPI in the PPI in the case of Slovakia is positive and the changes in the PPI are not stabilized. The function is increasing. The impact of a shock on the PPI in the CPI is firstly positive; for the fourth period it is 0.65 %, and than the changes in the CPI are stabilized around the fifth month by 0.67 %.

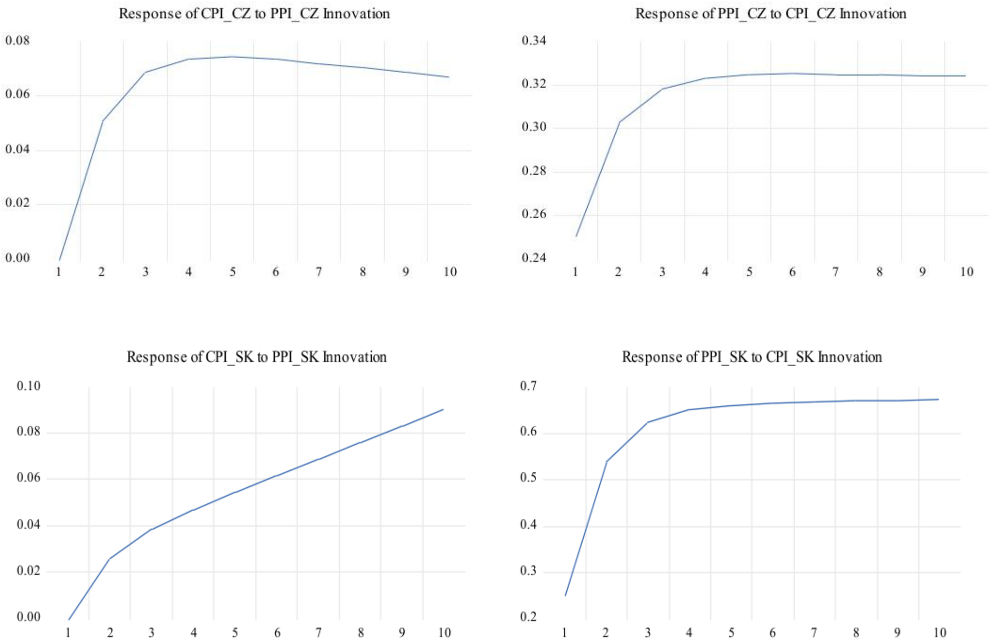


Figure 2 Response to Cholesky One S.D. Innovations

4 Conclusion

The article examines the relationship between the CPI and the PPI, which are most often used to monitor price developments.

Historically, both indices have been moving in a similar direction. A higher volatility of the PPI is mainly due to a faster response to the current state of the economy and a more direct link to the commodities the prices of which react the fastest. However, the available literature does not find a clear relationship between the two indices. The CPI is important in monetary policy; central banks use their instruments to target inflation and thus stabilize consumer price development.

The aim of this article is to find out the relationship between PPI and CPI for two countries: the Czech Republic and Slovakia. The existence of long-run equilibrium relationship between CPI and PPI was tested with the Johansen cointegration, and the short run dynamics between the variables was examined by VEC model. The validity of the results, diagnostic test such as impulse responses were checked. The first hypothesis, analysed if there is a long-term relationship between CPI and PPI, was not confirmed for any of countries. The second hypothesis:

Short-term for relationship: "PPI of the previous period effects current CPI" was confirmed in the Czech Republic. And short-term for relationship: "CPI of the previous period effects current PPI" was confirmed in Slovakia.

What are the economic causes / consequences of different results for the Czech Republic and Slovakia? According to the law on the transfer of prices, the PPI has a certain impact on the CPI, which was confirmed for the Czech Republic. PPIs serve as the main indicator of CPIs, so when manufacturers face input inflation, the increase in their production costs is passed on to retailers and consumers. In the case of Slovakia, the relationship from the CPI to the PPI was confirmed. The same relationship has been confirmed in the UK and Canada.

Further research in this area could be focused on the analysis of the CPI and the PPI during the war in Ukraine, or next research could also be focused on energy prices and their impact on the CPI and the PPI.

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 2022.

References

[1] Akcay, S. (2011). The causal relationship between producer price index and consumer price index: Empirical evidence from selected European countries. *International Journal of Economics and Finance*, 3(6), 227-232.
[2] Akdi, Y., Berument, H., Cilasun, S. M., & Olgun, H. (2006). The relationship between different price indexes: A set of evidence from inflation targeting countries. *Statistical Journal of the United Nations Economic Commission for Europe*, 23(2, 3), 119-125.
[3] Caporale, G. M., Katsimi, M., & Pittis, N. (2002). Causality links between consumer and producer prices: some empirical evidence. *Southern Economic Journal*, 68(3), 703-711.
[4] Clark, T. E. (1995). Do producer prices lead consumer prices? *Economic Revie*, Federal Reserve Bank of Kansas City, 80(QIII), 25-39.
[5] Enders, W. (2010). *Applied econometric time series*. Hoboken: Wiley.
[6] Gang, F. A. N., Liping, H., & Jiani, H. (2009). CPI vs. PPI: Which drives which? *Frontiers of Economics in China*,4(3), 317-334.
[7] Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*. 37(3), 424–438.
[8] Granger, C.W.J. & Newbold, P. (1977). *Forecasting Economic Time Series*. New York: Academic Press.
[9] Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.
[10] Johansen, S. (1995). *Likelihood-based inference in cointegrated vector auto-regressive models*. Oxford: Oxford University Press.
[11] Junicke, M. (2017). Trend inflation and monetary policy in Eastern Europe. *Macroeconomic Dynamics*, 23(4), 1649-1663.
[12] Khan, K., Su, C. W., Tao, R. & Lobont, O. (2018). Producer price index and consumer price index: causality in central and eastern European countries. *Journal of Economics*, 66(4), 367-395.
[13] Leamer, E. E. (1985). Vector Autoregressions for Causal Inference? *Carnegie-Rochester Conference Series on Public Policy*. 22(1), 255–304.
[14] [Li, S., Tang, G., Yang, D. & Du, S. (2019). Research on the Relationship between CPI and PPI Based on VEC Model. *Open Journal of Statistics*, 9, 218-229.
[15] Lima, R. I. D. (2019). Does PPI lead CPI in Brazil? *International Journal of Production Economics*, 214(C), 73-79.
[16] Losada, O. H. C., Arias, J. P. M. & Guzman, J. C. (2018). Relationship between the consumer price index and the producer price index for six south American countries. *APUNTES DEL CENES*, 37(66), 39-74.
[17] Stock, J.H. & Watson, M.W. (2001). Vector Autoregressions. *Journal of Economic Perspectives*, 15(4), 101-115.
[18] Stoklasová, R. (2018). Default rate in the Czech Republic depending on selected macroeconomic indicators. *E&M Economics and Management*, 21(2), 69-82.
[19] Tiwari, A. K., Suresh, K. G., Arouri, M. & Teulon, F. (2014). Causality between consumer price and producer price: Evidence from Mexico. *Economic Modelling*, 36, 432-440.
[20] Woo, K. Y., Lee, S. K. & Ng, C. Y. J. (2019). An investigation into the dynamic relationship between CPI and PPI: Evidence from the UK, France and Ger-many. *Singapore Economic Review*, 64(5), 1081-1100.

Analysis of the Efficiency of the Czech Companies in the NACE Sector "Accommodation and Food Service Activities"

Martina Kuncová[1], Petra Zýková[2], Petra Kozáková[3], Lenka Lízalová[4]

Abstract. The 2019-2020 COVID pandemic has affected many organisations and companies. One of the most affected sectors is accommodation and food service activities. The aim of this paper is to assess technical efficiency of Czech companies in this sector. The research sample included thousands of enterprises (Group I according to the Statistical Classification of Economic Activities in the European Community - NACE). These enterprises were evaluated for the years 2019 and 2020 in order to track changes in pre-pandemic and pandemic years. The data was exported from the Albertina CZ Gold Edition database. DEA models are used for analysis and comparison. Efficiency scores were calculated based on the CCR (Charnes, Cooper and Rhodes) and BCC (Banker, Charnes, Cooper) models with 6 and 4 inputs and 3 and 2 outputs. The results show that there were already many inefficient companies in the given sector in 2019 and this was also observed for 2020, however, there was a significant reduction in the number of subjects in the database, which may also be due to the impact of the COVID-19 pandemic.

Keywords: efficiency, DEA models, accommodation and food service activities

JEL Classification: C44, L83
AMS Classification: 90C15

1 Introduction

The Accommodation and food service activities sector in the Czech Republic is a relatively small industry. In 2017, it accounted for 3.9% of the output of the entire tertiary sector, and 3.5% of gross value added [6]. The sector played a more significant role in employment (6.5%), especially in the case of self-employed persons (one-tenth of all self-employed persons working in services found their main source of economic activity here). The rate of investment in accommodation, catering and hospitality has been at 9% of the EU level over the last five years. This sector is one of the hardest hit by the global COVID-19 pandemic.

This article offers an analysis based on Data Envelopment Analysis (DEA) models. DEA models were first developed by Charnes, Cooper and Rhodes [3] based on the concept introduced by Farrell [9]. These methods and models belong to operational research, especially linear programming or multi-criteria models. They were used many times to evaluate the performances of many different entities (countries, regions, enterprises, schools, hospitals, insurance companies, military units etc.) engaged in many different kinds of activities in many different contexts [5]. The efficiency scores of decision-making units (DMUs) are computed using DEA models. These models classify decision-making units into two subsets – efficient and inefficient. The main idea of DEA models consists of the estimation of an efficient frontier which is created by the best relative ratios of inputs and outputs of the compared DMUs. DMU is taken as efficient when it lies on the efficient frontier. Inefficient units lie under the efficient frontier. Their efficiency can be improved by changing inputs or outputs.

The use of DEA models is very wide. Emrouznejad and Yang [8] report that they found 10300 articles focusing on DEA models in journals published between 1978 and 2016, and especially in recent years, the increase in the use of DEA models in terms of the number of articles is exponential. Analyses in the tourism, accommodation or hospitality industry using DEA models are no exception. Ramírez-Hurtado and Contreras [14] used DEA models to investigate the efficiency of travel agencies in Spain. Hedija, Fiala and Kuncová [10] applied DEA to evaluate the efficiency of Czech travel agencies and tour operators. Reynolds and Thompson [15] evaluated the productivity

[1] College of Polytechnics Jihlava, Department of Economic Studies, Tolstého 16, 58601 Jihlava, kuncova@vspj.cz
[2] Prague University of Economics and Business, Faculty of Informatics and Statistics, Department of Econometrics, nám.W.Churchilla 4, 13067 Prague 3, petra.zykova@vse.cz
[3] Prague University of Economics and Business, Faculty of Management, Department of Management, Jarošovská 1117/II, 37701 Jindřichův Hradec, petra.kozakova@vse.cz
[4] College of Polytechnics Jihlava, Department of Economic Studies, Tolstého 16, 58601 Jihlava, lenka.lizalova@vspj.cz

of restaurants. Min, Min and Joo [12] sanalysed the efficiency of luxury and budget Korean hotels, while Pulina, Detotto and Paba [13] focused on Italian hotels.

2 DEA models

DEA models are a general tool for efficiency and performance evaluation of the set of n homogenous decision-making units (DMUs) that spend multiple (S) inputs and transform them into multiple (R) outputs. A very detailed description of DEA models is in [7]. In general, the DEA models assume that the inputs are minimising, and the outputs are maximising. The measure of the efficiency of this transformation is one of the main results of applying DEA models. Let us denote $\mathbf{Y} = \left(y_{rj},\ r = 1,\dots,R,\ j = 1,\dots n\right)$ a non-negative matrix of outputs and $\mathbf{X} = \left(x_{sj}, s = 1,\dots,S, j = 1,\dots n\right)$ a non-negative matrix of inputs. The efficiency score of the unit under evaluation j_0 is derived as follows:

$$
\begin{aligned}
&\text{Maximise} && U_{j_0} = \frac{\sum_{r=1}^{R} u_r y_{r,j_0}}{\sum_{s=1}^{S} v_s x_{s,j_0}} \\
&\text{subject to} && \frac{\sum_{r=1}^{R} u_r y_{r,j}}{\sum_{s=1}^{S} v_s x_{s,j}} \le 1, \quad j = 1,\ldots,n, \\
& && u_r \ge \varepsilon, \quad r = 1,\ldots,R, \\
& && v_s \ge \varepsilon, \quad s = 1,\ldots,S,
\end{aligned}
\tag{1}
$$

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 one for the efficient units. The inefficient units have an efficiency score lower than 1. Model (1) is not linear in its objective function, but it can be transformed into a linear one. Two ways of linearisation of model (1) depending on the orientation of the linear model are usually used: input-oriented model and output-oriented model. This article is solely focused on the input-oriented models. The linearised version of the input-oriented model, often called the CCR model according to its authors Charnes, Cooper and Rhodes, is as follows:

$$
\begin{aligned}
&\text{Maximise} && U_{j_0} = \sum_{r=1}^{R} u_r y_{r,j_0} \\
&\text{subject to} && \sum_{s=1}^{S} v_s x_{s,j_0} = 1, \\
& && \sum_{r=1}^{R} u_r y_{r,j} - \sum_{s=1}^{S} v_s x_{s,j} \le 0, \quad j = 1,\ldots,n, \\
& && u_r \ge \varepsilon, \quad r = 1,\ldots,R, \\
& && v_s \ge \varepsilon, \quad s = 1,\ldots,S.
\end{aligned}
\tag{2}
$$

The dual formulations of the multiplicative DEA models are called the envelopment DEA models. The CCR input-oriented envelopment model is derived from model (2) and is following:

$$
\begin{aligned}
\text{Minimise} \quad & U_{j_0} = \theta_{j_0} - \varepsilon\left(\sum_{s=1}^{S} s_s^- + \sum_{r=1}^{R} s_r^+\right) \\
\text{subject to} \quad & \sum_{j=1}^{n} x_{s,j}\lambda_j + s_s^- = \theta_{j_0} x_{s,j_0}, \quad s = 1,\dots,S, \\
& \sum_{j=1}^{n} y_{r,j}\lambda_j - s_r^+ = y_{r,j_0}, \quad r = 1,\dots,R, \\
& \lambda_j \ge 0, \quad j = 1,\dots,n, \\
& s_s^- \ge 0, \quad s = 1,\dots,S, \\
& s_r^+ \ge 0, \quad r = 1,\dots,R,
\end{aligned} \tag{3}
$$

where $\boldsymbol{\lambda} = (\lambda_1,\cdots,\lambda_n)$, $\boldsymbol{\lambda} \ge \mathbf{0}$ is a vector of weights assigned to particular decision-making units, $\mathbf{s}^- = (s_1^-,\dots,s_S^-)$ and $\mathbf{s}^+ = (s_1^+,\dots,s_R^+)$ are vectors of slack/surplus variables, U_{j_0} is called the efficiency score for j_0-th unit under evaluation. The efficient units identified by this model have an efficiency score equal to one, and all slack/surplus variables equal to 0. The inefficient units have an efficiency score lower than one.

The CCR input-oriented models assume constant returns to scale (CRS). It means if the decision-making unit with an input/output combination $(\mathbf{x},\mathbf{y})$ is efficient, then the unit with an input/output combination $(\alpha\mathbf{x},\alpha\mathbf{y})$, where $\alpha > 0$ is also efficient. Another possibility of returns to scale are the variable returns to scale (VRS), which were firstly mentioned in [2]. The DEA models with the variable returns to scale are called BCC models, according to Banker, Charnes and Cooper. The multiplicative form of the BCC input-oriented model is following:

$$
\begin{aligned}
\text{Maximise} \quad & U_{j_0}^V = \sum_{r=1}^{R} u_r y_{r,j_0} + \mu \\
\text{subject to} \quad & \sum_{s=1}^{S} v_s x_{s,j_0} = 1, \\
& \sum_{r=1}^{R} u_r y_{r,j} - \sum_{s=1}^{S} v_s x_{s,j} + \mu \le 0, \quad j = 1,\dots,n, \\
& u_r \ge \varepsilon, \quad r = 1,\dots,R, \\
& v_s \ge \varepsilon, \quad s = 1,\dots,S,
\end{aligned} \tag{4}
$$

where μ is a free variable and $U_{j_0}^V$ is 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. The envelopment BCC input-oriented model is the following.

$$
\begin{aligned}
\text{Minimise} \quad & U_{j_0}^V = \theta_{j_0} - \varepsilon\left(\sum_{s=1}^{S} s_s^- + \sum_{r=1}^{R} s_r^+\right) \\
\text{subject to} \quad & \sum_{j=1}^{n} x_{s,j}\lambda_j + s_s^- = \theta_{j_0} x_{s,j_0}, \quad s = 1,\dots,S, \\
& \sum_{j=1}^{n} y_{r,j}\lambda_j - s_r^+ = y_{r,j_0}, \quad r = 1,\dots,R, \\
& \sum_{j=1}^{n} \lambda_j = 1, \\
& \lambda_j \ge 0, \quad j = 1,\dots,n, \\
& s_s^- \ge 0, \quad s = 1,\dots,S, \\
& s_r^+ \ge 0, \quad r = 1,\dots,R.
\end{aligned} \tag{5}
$$

This model and the above models cannot rank the efficient units because of their identical efficiency scores. For this purpose, super-efficiency has been proposed by [1].

3 Data

For the evaluation of the economic efficiency of companies that operated (according to the classification of economic activities NACE) in the I-Accommodation and food service activities, i.e. in the sectors 55 Accommodation and 56 Food and beverage service activities, data from the Albertina CZ Gold Edition database [3] were used. For the year 2019, more than 10,000 subjects were mentioned, but for 2020 it was just under 8,000 (nearly 20% less). It was necessary to select suitable inputs and outputs for the DEA models. Similarly to article [11], we used 6 inputs and 3 outputs for the first analysis – see Table 1.

Inputs	Outputs
Number of employees (I1)	Sales of products, goods and services (O1)
Total assets (I2)	Operating profit/loss (O2)
Equity (I3)	Net profit/loss for the financial period (O3)
Liabilities (I4)	
Cost of sales (I5)	
Staff costs (I6)	

Table 1 List of inputs and outputs for DEA models

As in [11], it was necessary to remove companies with missing data (4,399 for 2019 and 3,518 for 2020), and we also removed companies with zero outputs (around 30 in each year) and with negative inputs or outputs (this group covered 3,459 companies in 2019 and 3,246 companies in 2020) which needs another study especially because of the financial indicators and negative values consequences. So, in the end, the first analysis data for 2,196 companies for 2019 and 1,130 companies for 2020 (which is almost 50% less) was used. We conducted this first analysis to make the results comparable to previous research. But it is not very appropriate to include similar inputs or outputs in DEA models. Based on the high correlation coefficients (higher than 0.9) between the inputs I2 and I3, and I5 and I6 (in both years), respectively, and due to missing data, we decided to exclude inputs I3 (Equity) and I6 (Staff costs) for the second analysis. For the outputs, there was a high correlation (above 0.9) between O2 and O3 outputs, so we discarded O3 (Net profit/loss for the financial period). After this change, it was again necessary to modify the data set. First, the companies with missing data were removed (2,699 companies for 2019 and 2,160 for 2020). Second, the companies with negative inputs or outputs (2,980 for 2019 and 3,179 for 2020) and with zero outputs (868 for 2019, 656 for 2020) were also excluded. Finally, for the second analysis, we used data for 3,538 companies for 2019 and 1,932 companies for 2020.

4 Results

As mentioned above, we conducted two analyses. The first one used 6 inputs and 3 outputs. The following tables show the results for 2019 (Table 2) and 2020 (Table 3). In terms of the absolute values, it is evident that the number of companies analysed in 2020 is half that of 2019. This fact can be influenced by the COVID-19 restrictions and the fact that not all firm data was published and entered into the Albertina database or because firms changed ownership. However, if we take a closer look at the data, we find that of the 1,130 firms observed in 2020, only 669 also appear among the 2,196 firms analysed in 2019. This may again be due to a change in ownership and firm name or missing data in the previous year or the newcomers on the market (this analysis is beyond the scope of the paper and will be addressed in future research). The percentage of efficient companies is slightly higher in 2020 and traditionally in BCC-I models. Of the 158 efficient subjects for 2019, only 43 of them are in the data for 2020, and 23 of them were efficient in both years according to the BCC-I (the analysis was conducted for each year separately; analyses of multiple years together will be part of the following research). It can be assumed that all these drops may be related to the COVID-19 pandemic and the restrictions in place for 2020.

The distribution of efficiency (BCC-I) in terms of firm size (measured by the number of employees) is also interesting. While in 2019, all size categories were represented among efficient firms (Table 2), in 2020, only firms with less than 50 employees were classified as efficient (Table 3). Most efficient companies fall into the category with 0 employees. It might seem that these are primarily entrepreneurs operating in the area - but on closer inspection of the data, we found that this is not always the case. For example, among the companies identified as efficient in both years are Amrest and Aramark, which according to the data in 2019 were large companies (thousands of employees), but for 2020 are in the database as companies with 0 employees. In contrast, for example, the firm Tesař-Mražené registered 0 employees in 2019 and 23 employees in 2020 (and was also marked as efficient in both years). In this respect, some data from the sector under review may be misleading. It is questionable whether companies report the number of full-time employees rather than contract workers or temporary workers.

No. of employees 2019	Number of comp.	%	No. of CCR-I efficient	%	No. of BCC-I efficient	%
>250	7	0.3%	0	0.0%	6	3.8%
50-250	63	2.9%	0	0.0%	7	4.4%
11-49	253	11.5%	5	7.6%	10	6.3%
1-10	508	23.1%	12	18.2%	18	11.4%
0	1365	62.2%	49	74.2%	117	74.1%
SUM	**2196**		**66**	**3.0%**	**158**	**7.2%**

Table 2 Results of DEA models for the year 2019 (6 inputs, 3 outputs)

No. of employees 2020	Number of comp.	%	No. of CCR-I efficient	%	No. of BCC-I efficient	%
>250	0	0.0%	0	0.0%	0	0.0%
50-250	10	0.9%	0	0.0%	0	0.0%
11-49	53	4.7%	0	0.0%	7	5.6%
1-10	200	17.7%	8	16.3%	15	12.1%
0	867	76.7%	41	83.7%	102	82.3%
SUM	**1130**		**49**	**4.3%**	**124**	**11.0%**

Table 3 Results of DEA models for the year 2020 (6 inputs, 3 outputs)

In the second analysis, inputs and outputs with high correlation with another input or output were excluded from the analysis; at the same time, these were parameters with many missing values, due to which several companies were excluded in the previous analysis. For this part, the data set was changed (some companies with missing values in the first analysis can be included). For 2019, we can use data for 3,538 companies; for 2020, only 1,932 companies were included, so again nearly 50% less. Table 4 shows the results for the year 2019, we see that only 21 companies were efficient in the CCR-I model and 77 efficient in the BCC-I model.

No. of employees 2019	Number of comp.	%	No. of CCR-I efficient	%	No. of BCC-I efficient	%
>250	8	0.2%	0	0.0%	7	9.1%
50-250	72	2.0%	0	0.0%	1	1.3%
11-49	337	9.5%	0	0.0%	1	1.3%
1-10	778	22.0%	3	14.3%	5	6.5%
0	2343	66.2%	18	85.7%	63	81.8%
SUM	**3538**		**21**	**0.6%**	**77**	**2.2%**

Table 4 Results of DEA models for the year 2019 (4 inputs, 2 outputs)

No. of employees 2020	Number of comp.	%	No. of CCR-I efficient	%	No. of BCC-I efficient	%
>250	0	0.0%	0	0.0%	0	0.0%
50-250	11	0.6%	0	0.0%	0	0.0%
11-49	72	3.7%	0	0.0%	5	6.4%
1-10	308	15.9%	1	5.6%	4	5.1%
0	1541	79.8%	17	94.4%	69	88.5%
SUM	**1932**		**18**	**0.9%**	**78**	**4.0%**

Table 5 Results of DEA models for the year 2020 (4 inputs, 2 outputs)

From the results for 2020 (Table 5), we see that the numbers of efficient firms are almost identical to 2019, although the total number of firms considered was significantly lower. Only 9 companies were efficient in both years. Similar to the first analysis, efficient subjects in 2019 were identified in all size groups, while in 2020, it was again only subjects with less than 50 employees that were identified as efficient. Figure 1 shows the evolution of the efficiency scores in both years. It can be seen that regardless of the number of companies surveyed and regardless of the year, the efficiency score in the sector Accommodation and food service activities was in most cases very low. Thus, it can be concluded that there are many inefficient entities in the sector, regardless of the impact of the COVID-19 pandemic. The pandemic has most likely had an impact on firm closure or change of ownership. However, in terms of efficiency, as measured by the DEA models, it cannot be conclusively stated that there has been a massive decline in efficiency.

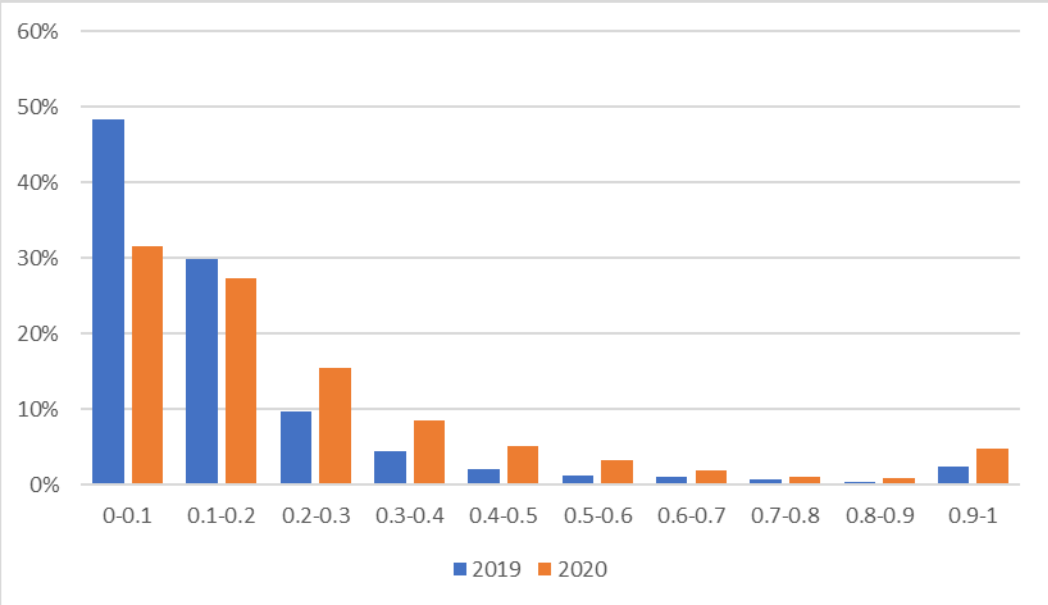


Figure 1 Results – BCC-I efficiency scores – the percentage of companies by efficiency scores

5 Conclusions

The aim of this paper is to assess technical efficiency of Czech companies in the sector Accomodation and food service activities (Group I according to the Statistical Classification of Economic Activities in the European Community - NACE) for the years 2019 and 2020. Efficiency scores were calculated based on the CCR (Charnes, Cooper and Rhodes) and BCC (Banker, Charnes, Cooper) models in two different analysis - with 6 inputs and 3 outputs for the first one and with 4 inputs and 2 outputs in the second one. The results show that the number of subjects with complete data in the database is significantly lower in 2020 than in 2019 (about 50% difference in both analysis). This fact points to the relatively high volatility of the sector, which could also be due to restrictions due to the pandemic COVID-19. In terms of DEA models efficiency, especially BCC-I models, there were already many inefficient companies in the given sector in 2019 and this was also observed for 2020. If we compare the average efficiency in the years for the subjects analysed, the average in both analyses comes out better in 2020 than in 2019. In the first analysis (6 inputs, 3 outputs), the average efficiency for 2020 was 0.46, in 2019 0.36. However, in the second analysis (4 inputs, 2 outputs), where more companies were represented, the average efficiency decreased to 0.16 and 0.24 for 2019 and 2020, respectively. Thus, we see that the inclusion of firms for which some data were missing in the first analysis leads to a decrease in the average efficiency, i.e. these are inefficient subjects. At the same time, we see a slight increase in the average efficiency in the pandemic year 2020, which may be due to the disappearance of subjects operating at the survival frontier that are no longer helped by government subsidies. Thus, the decline in the number of entities and the slight increase in average efficiency signal that the impact of COVID-19 can be perceived as slightly positive. However, the above-mentioned data anomalies (number of employees, missing data) also point to a large degree of informality or “grey economy” in the given sector. This may also be the reason for the very low efficiency in the sector in the Czech Republic.

Comparison of the results of DEA models with specific financial ratio indicators of the monitored sector is the subject of further research. The first results point to the fact that in the given file, the DEA models evaluate a significantly smaller number of enterprises as efficient than those that can be described as financially healthy. At the same time, however, it must be mentioned that not all efficient businesses can be described as financially healthy, especially in 2020.

Acknowledgements

The research was supported by the internal project of the College of polytechnics Jihlava - Creative Support of the Department of Economic Studies - INT/2022/0008 entitled "Atributy řízení výkonnosti podniku a faktory, které ji ovlivňují (Attributes of Business Performance Management and Factors Affecting it)" and also 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.

References

[1] Andersen, P. & Petersen, N.C. (1993). A procedure for ranking efficient units in data envelopment analysis. *Management Science*, 39(10), pp.1261–1264.

[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), pp. 1078–1092.

[3] Bisnode MagnusWeb (2021). Database Albertina-Bisnode Czech Republic [online]. https://www.bisnode.cz/produkty/magnusweb/, [cit. 2022-04-14]

[4] Charnes, A., Cooper, W. & Rhodes, E. (1978). Measuring the efficiency of decision-making units. *European Journal of Operational Research*, 2(6), pp.429-444.

[5] Cooper, W. W., Lawrence, M.S. & Zhu, J. (2004). Handbook on Data Envelopment Analysis. Norwell: Kluwer Academic Publishers

[6] Czech Statistical Office (2017). Analýza ekonomického vývoje v roce 2017 – Ubytování a pohostinství. [online], https://www.czso.cz/documents/10180/97972503/320305-18a15.pdf/32480cbf-d9ac-4da5-bc1e-e64aafac0431?version=1.0 [cit. 2022-04-14]

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

[8] 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

[9] Farrell, M. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 120(3), pp. 253-290.

[10] Hedija, V., Fiala, R., & Kuncová, M. (2017). Is profitability a good proxy for efficiency? Evidence from the Subsector of Tour Operators. *Review of Economic Perspectives*, 17(4), 425–440. https://doi.org/10.1515/revecp-2017-0022

[11] Kuncová, M., Kozáková, P., & Lízalová, L. (2022). Competitiveness in Accommodation and Food Service Activities´ Sector – Efficiency Analysis. In *Proceedings of the 14th Annual International Scientific Conference Competition*. College of Polytechnics Jihlava: College of Polytechnics Jihlava. (in print)

[12] Min, H., Min, H., & Joo, S. J. (2009). A data envelopment analysis on assessing the competitiveness of Korean Hotels. *The Service Industries Journal*, 29(3), 367–385. https://doi.org/10.1080/02642060701849865

[13] Pulina, M., Detotto, C., & Paba, A. (2010). An investigation into the relationship between size and efficiency of the Italian hospitality sector: A window DEA approach. *European Journal of Operational Research*, 204(3), 613–620. https://doi.org/10.1016/j.ejor.2009.11.006

[14] Ramírez-Hurtado, J. M., & Contreras, I. (2016). Efficiency of travel agency franchises: A study in Spain. *Service Business*, 11(4), 717–739. https://doi.org/10.1007/s11628-016-0326-1

[15] Reynolds, D., & Thompson, G. M. (2007). Multiunit restaurant productivity assessment using three-phase data envelopment analysis. *International Journal of Hospitality Management*, 26(1), 20–32. https://doi.org/10.1016/j.ijhm.2005.08.004

Directed Search for Pareto Front Approximation with Path- relinking Method

Marek Kvet[1], Jaroslav Janáček[2]

Abstract. This paper is focused on determination of good approximation of the Pareto front of public service system designs. Generally, an approximation of Pareto front seems to be a very useful tool for a negotiation between the system founder and public representatives, when the system utility is evaluated from two different points of view. The core of presented research consists in generalization of the directed search, which originally inspects solutions in a given direction from a starting solution. The original approach determines the direction in a cone determined by directions derived from two succeeding members of non-dominated solution set. Our generalization replaces the directed search by inspection of the shortest path connecting the two succeeding members. To perform the inspection, the path-relinking method is applied.

Keywords: public service system design problem, bi-criteria optimization, Pareto front, directed search, path-relinking method

JEL Classification: C44, C61
AMS Classification: 90C05, 90C06, 90C10, 90C27

1 Introduction

Emergency Medical Service, more commonly known as EMS, is such a system that provides emergency medical care. Once it is activated by an incident that causes serious illness or injury, the focus of EMS is emergency medical care of the patient(s). EMS is most easily recognized when emergency vehicles are seen responding to emergency incidents. However, EMS is much more than a ride to the hospital. It is a system of coordinated response and emergency medical care, involving multiple people and agencies. A comprehensive EMS system is ready every day for every kind of emergency. The basic role of the system implies the fact that the efficiency of its performance may have a direct impact on the lives and health of the served population [3, 5, 6, 7, 9, 15, 16, 20]. Thus, its mission is irreplaceable and the responsible authorities should design and operate the system in such a way to achieve its maximal possible efficiency.

When designing a new EMS system or optimizing an existing one, many different objectives may be required to follow. Thus, usage of a common mathematical model minimizing usually one given objective function is questionable. If there are two or more contradictory objectives to be taken into account, then a Pareto front of EMS system designs seems to be a sufficient support for making the final decision about the resulting EMS stations deployment. The Pareto front is defined as the set of non-dominated solutions, where each objective is considered equally good. It means that the elements of Pareto front (particular solutions of the associated integer- programming problem) must hold the following characteristic: Neither of them can be improved in one objective unless the other objective is worsened [1, 10, 11, 12]. The Pareto optimal front facilitates decision-makers to select an appropriate operating point as per its suitability. It is obvious that algorithms for computing the Pareto front of a finite set of alternatives have been studied in computer science and power engineering. Since generating the entire Pareto front is often computationally hard [8, 10], the attention of experts in applied informatics and location science is moving towards heuristic or metaheuristic algorithms that produce a good approximation of the Pareto front [11, 12, 18].

The computational process of obtaining a good approximation of the Pareto front may be built on many different strategies of local search within the whole set of feasible solutions. Among all the directions that research in this area takes, we will limit ourselves only to the algorithms based on a simple swap algorithm for neighborhood search [11] or to applying the path-relinking method. The further reported computational study is aimed at comparison of these two approaches to choose better alternative for practical usage.

[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.

2 Emergency Medical Service System Design

The Pareto front of solutions may serve as an excellent output of the computational process, mainly if there are many criteria, the nature of which does not allow their simultaneous minimization. Mentioned criteria are usually in conflict. Within this paper, we restrict ourselves only on such EMS system design problem, where two different objectives functions are to be optimized.

Many algorithms suggested for Pareto front approximation follow the idea of continuous updates of the resulting set of non-dominated system designs further denoted as *NDSS* (non-dominated solutions set). Let us discuss some details. Consider two contradictory objectives f_1 and f_2. The updating process starts with an initial non-empty set of non-dominated solutions ordered according to increasing value of f_2. 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 dominated by predecessor P and is abandoned. Otherwise, if the candidate C is not dominated by the successor S, the candidate becomes the new member of the updated set. 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.

If the quality of the Pareto front approximation is to be evaluated, then the concept of area formed by particular elements of the *NDSS* set may be used [11, 12, 13, 18]. Obviously, the bordering members of the set should be determined in a correct way to make the evaluation valid and meaningful. More details are discussed in [13].

As far as concrete forms of the mentioned objective functions f_1 and f_2 are concerned, they follow the basic standards for EMS designing used also in our previous research activities, the results of which are published in [10, 11, 12, 13, 18, 19] and many others.

To formulate the objectives mathematically, let I denote a finite set of EMS station candidates and consider p to be the number of facilities, which are to be located to provide public with the given kind of urgent pre-hospital healthcare. Then, each subset $P \subseteq I$ containing exactly p elements represents a feasible solution of the EMS system design problem. Each of the solutions can be evaluated according to various criteria, which can be classified either as system or fairness criteria. The average expected response time is often used as a system criterion. This criterion can be defined in the form of the objective function $f_1(P)$ by the formula (1).

$$f_1(P) = \sum_{j \in J} b_j \sum_{k=1}^{r} q_k \min_k \left\{ t_{ij} : i \in P \right\} \tag{1}$$

The objective function (1) assumes that J stands for a finite set of locations, in which the expected patients are concentrated. Furthermore, let symbol b_j denote the expected frequency of emergency calls from the location $j \in J$, which lead to the rescue vehicle departure. Moreover, let t_{ij} correspond to the traversing time from an EMS station located at $i \in \mathrm{I}$ to the patient's location j and finally, let q_k denote the probability value of the case that the k-th nearest EMS station will be the first available one [17]. For completeness, the operator $\min_k\{V\}$ performed on a finite set V of real values gives the k-th smallest value in V. The above mentioned system criterion formulated as the objective function (1) follows the concept of so-called generalized disutility. It allows the operations researchers to model stochastic behavior of real EMS systems by more service centers providing the urgent medical care to the same patients.

The second quality criterion of EMS system design denoted by f_2 takes into account the aspect of fairness [2, 4]. Generally, fairness criteria measure the disutility perceived by the worst situated minority. In this paper, the fair criterion will be computed as the number of emergency calls from all patients' locations, for which the response time from the nearest located EMS station exceeds given threshold T. The associated objective function $f_2(P)$ can be expressed by the following formula (2).

$$f_2(P) = \sum_{j \in J} b_j \max \left\{ 0,\, sign\left(\min \left\{ t_{ij} : i \in P \right\} - T \right) \right\} \tag{2}$$

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 EMS designs with clear consequences of one objective function value improvement on the other characteristic of the EMS stations deployment [12].

3 Search Methods in m-Dimensional Unit Hypercube

The above formulated problem belongs to the family of p-location problems, which consists in deploying p facilities in a set of m possible locations. If each decision on locating or not locating a facility at the location i is modelled by a decision variable $y_i \in \{0, 1\}$, then each feasible solution can be represented by a m-dimensional vector of zeros and ones, where the number of ones must equal to p. Each m-dimensional vector of zeros and ones corresponds with a unique vertex of m-dimensional unit hypercube. If the vector is a feasible solution of the p-location problem, the vertex has also to belong to the simplex in the m-dimensional space. This simplex is described by equality (3).

$$\sum_{i=1}^{m} y_i = p \tag{3}$$

Intersection of the set of unit hypercube vertices and the simplex represent the set $\boldsymbol{Y}$ of feasible solutions of the studied p-location problem. The distance between two elements $\boldsymbol{y}$ and $\boldsymbol{x}$ of $\boldsymbol{Y}$ can be defined by (4) and it is reerred as the Hamming distance. This metric is often used to compare two vectors of the same size containing binary or integer values.

$$H(\mathbf{x}, \mathbf{y}) = \sum_{\substack{i=1 \\ i}}^{m} |x_i - y_i| \tag{4}$$

Obviously, the maximal distance between different elements of $\boldsymbol{Y}$ is $2p$ and the minimal distance is 2. If the Hamming distance of solutions $\boldsymbol{x}$, $\boldsymbol{y}$ equals to 2, then there must exist two subscripts i and j so that $x_i = 0$, $y_i = 1$ and $x_j = 1$, $y_j = 0$ and remainder of components of $\boldsymbol{x}$ and $\boldsymbol{y}$ have equal values. We can define swap operation so that $\boldsymbol{x} = swap(\boldsymbol{y}, i, j)$ or $\boldsymbol{y} = swap(\boldsymbol{x}, j, i)$. Based on the swap operation, two searching algorithms can be suggested to improve a combined objective function $f(\boldsymbol{y}) = w_1 f_1(\boldsymbol{y}) + w_2 f_2(\boldsymbol{y})$. The algorithm *Exchange*($\boldsymbol{y}$) using the best admissible strategy searches the neighborhood of the current solution and if a better solution is met, then it is declared to be a new current solution and the algorithm continues with new neighborhood search. In the opposite case, the algorithm terminates and returns the best found current solution. The algorithm performs according to the steps below.

Exchange($\boldsymbol{y}$)

0. Initialize the current solution $\boldsymbol{y}^c = \boldsymbol{y}$ and $f^c = f(\boldsymbol{y})$. Define sets $P = \{i = 1, \ldots, m, y_i = 1\}$ and $C = \{i = 1, \ldots, m, y_i = 0\}$.
1. Set $f^b = f^c$ and for each pair (i, j), $i \in P$ and $j \in C$ perform the following conditioned commands. If $f^b > f(swap(\boldsymbol{y}^c, i, j))$ then update $f^b = f(swap(\boldsymbol{y}^c, i, j))$, $i^b = i$, $j^b = j$.
2. If $f^b = f^c$, then terminate and return $\boldsymbol{y}^c$, otherwise perform the following commands: $P = P \cup \{j^b\} - \{i^b\}$, $C = C \cup \{i^b\} - \{j^b\}$, $\boldsymbol{y}^c = swap(\boldsymbol{y}^c, i^b, j^b)$, $f^c = f^b$ and go to step 1.

Another algorithm based on the swap operation is a special version of the path-relinking method, which generally searches one of the shortest paths connecting two solutions $\boldsymbol{y}$ and $\boldsymbol{x}$ of $\boldsymbol{Y}$ on the surface of the m- dimensional unit hypercube. The path-relinking method returns the best solution of the inspected path. The algorithm $PR(\boldsymbol{x}, \boldsymbol{y})$ can be described by the following steps.

$PR(\boldsymbol{x}, \boldsymbol{y})$

0. Initialize $\boldsymbol{z} = argmin\{f(\boldsymbol{x}), f(\boldsymbol{y})\}$, $f^c = min\{f(\boldsymbol{x}), f(\boldsymbol{y})\}$. Define sets $P = \{i = 1, \ldots, m, y_i = 1$ and $x_i = 0\}$ and $C = \{j = 1, \ldots, m, y_j = 0$ and $x_j = 1\}$.
1. Set $f^b = +\infty$ and perform the following conditioned commands for each pair (i, j), $i \in P$ and $j \in C$. If $f^b > f(swap(\boldsymbol{x}, i, j))$ then update $f^b = f(swap(\boldsymbol{x}, i, j))$, $i^b = i$, $j^b = j$.
2. Perform $\boldsymbol{x} = swap(\boldsymbol{x}, i^b, j^b)$, $P = P - \{i^b\}$, $C = C - \{j^b\}$.
If $f(\boldsymbol{x}) < f^c$, then update $\boldsymbol{z} = \boldsymbol{x}$, $f^c = f(\boldsymbol{x})$.
If $|C| > 1$, then go to step 1, otherwise terminate and return $\boldsymbol{z}$, and f^c.

4 Incrementing Improvement of Non-Dominated Set of Solutions

The incrementing improvement process starts with a small set of non-dominated solutions, which may consist of only two solutions. The current set *NDSS* of *noNDSS* non-dominated solutions is initialized by the starting set and is continuously updated by a basic sequential process. The basic procedure processes the current *NDSS* from the solution $\boldsymbol{y}^1$ to the solution $\boldsymbol{y}^{noNDSS-1}$ and continuously updates the current set of non-dominated solutions. In the individual step, solution $\boldsymbol{y}^k$ 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 this case, the local search continues with the changed solution, other- wise 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.

The local search can be performed by an arbitrary algorithm, which is able to produce new feasible solutions for *NDSS* updating. In the presented paper, usage of algorithms $Exchange(\boldsymbol{y}^k)$ and $PR(\boldsymbol{y}^k, \boldsymbol{y}^{k+1})$ is studied, when ap- plied to the local search connected with the k-th solution of the current *NDSS*.

5 Computational Study

The computational study reported in this section was aimed at studying the quality of obtained results for both swap and path-relinking methods embedded into the process of Pareto front approximate set creation. Whenever the developers of any solving approach to an integer-programming location problem want to study certain performance characteristics or accuracy of the resulting system design, they need some referential solutions or at least their objective function values, to which the obtained results may be compared. Fortunately, the exact Pareto fronts of non-dominated EMS system designs are available [8, 10].

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 making use of the NetBeans IDE 8.2 environment.

As far as the studied problem instances are concerned, we made use of common benchmarks representing existing EMS system operated in eight self-governing regions of Slovakia [10, 11, 12, 13, 18, 19].

The list of problem instances contains the regions 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 positive coefficients b_j weighting each patients' location $j \in J$ was set to the number of system users sharing the location j and this value was rounded up to hundreds. We assume that the number of emergency calls from the location j is proportional to the number of its inhabitants.

As the system objective function value described by the expression (1) follows the concept of generalized disutility [17], the parameter r was set to 3. The associated coefficients q_k for k=1, …, r were set in percentage in the following way: $q_1 = 77.063$, $q_2 = 16.476$ and $q_3 = 100 - q_1 - q_2$. These values were obtained from a simulation model of existing EMS system in Slovakia [14]. Parameter T used in the fair criterion (2) was set to the value of 10 minutes in accordance with our previous research [10, 11, 12, 13, 18, 19].

Table 1 Benchmarks characteristics and the exact Pareto fronts description

Region	$\|I\|$	p	*NoS*	*Area*	$f_1(MLM)$	$f_2(MLM)$	$f_1(MRM)$	$f_2(MRM)$
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

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 problem instance corresponding to one row of the table, the column denoted by $|I|$ contains the cardinality of the EMS station candidates set. The parameter p expresses the number of stations to be located. In this study, we assume that only one facility (EMS station) can be located at one candidate locations. In the future, the problem could be adjusted in such a way to allow also multiple facility locations. The middle part of Table 1 contains the number of solutions *NoS* forming the Pareto front. In the column denoted by *Area* we provide the readers with the size of area defined by the elements of the Pareto front as suggested in [13, 19]. This value is the most important aspect to evaluate the quality of approximate sets of non-dominated solutions. Finally, the right part of Table 1 summarizes additional information about the bordering points of the Pareto front. Let the symbol *MLM* denote the most left member of the Pareto front and let *MRM* denote the most right member. Then, the objective function value (1) for *MLM* will be denoted as $f_1(MLM)$ and the criterion (2) for *MLM* will be denoted by $f_2(MLM)$. We will use an analogical denotation also for *MRM*.

It must be noted that an individual run of the solving process was limited by 5 minutes of computation. The five minutes limit was chosen deliberately to keep the possibility to compare obtained results to our previously developed approaches. If one wants to compare two or more algorithms, which may be based on different techniques, their computational time restrictions should be the same. Five minutes proved to be sufficient for obtaining enough accurate results. Increasing the time limit does not necessarily mean obtaining better results. On the other hand, if the computational time restriction is too strict, then the algorithm may bring worse results, because better solutions will not be achieved. Of course, it would be helpful to perform an extra computational study, in which the impact of computational time limitation on the result accuracy would be studied. This paper does not contain such a study, but we keep the same settings as in our previous research activities, in which 5 minutes were used.

The obtained results are reported in the following two tables. While Table 2 contains the results for the algorithm making use of a simple swap operation, Table 3 summarizes the obtained results, in which the path-relinking method was applied. Both tables keep the following structure. Each column represents one solved problem instance. The row *CT* contains the computational time in seconds. The objective functions f_1 and f_2 of the most left and the most right member of *NDSS* are not reported here because they are almost the same as in the exact Pareto front. The number of *NDSS* members is reported in the row denoted by *noNDSS*. The size of area formed by the *NDSS* members is given in the row denoted by *FinArea*. The row denoted by *MeanArea* contains the mean value of area as suggested in [19] to evaluate the dynamics of suggested heuristic approach. The last denotation *noOS* expresses the number of outer cycles performed during the solving procedure.

Table 2 Results of numerical experiments for the swap algorithm based directed search

Region	BA	BB	KE	NR	PO	TN	TT	ZA
CT	300.0	436.0	326.2	311.9	587.0	303.4	302.9	309.5
noNDSS	33	210	230	105	264	95	62	90
FinArea	592690	938015	1244591	775288	901065	834807	815095	407905
MeanArea	592806.1	954886.0	1274739.4	781721.7	883822.9	836944.6	816946.4	412037.7
noOS	2098	2	2	11	2	31	82	22

Table 3 Results of numerical experiments for the search using the path-relinking approach

Region	BA	BB	KE	NR	PO	TN	TT	ZA
CT	300.0	300.3	300.0	300.0	300.1	300.0	300.0	300.0
noNDSS	31	152	130	53	81	54	40	43
FinArea	712948	1094802	1817556	946193	1146929	900748	1062544	589698
MeanArea	712952.2	1094936.7	1817814.3	946255.0	1147015.7	900770.7	1062558.7	589711.6
noOS	113007	1048	1486	5123	2337	11147	40129	12616

6 Conclusions

The research reported in this paper was focused on Emergency Medical 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 the directed search based on a swap algorithm to the another approach using the path-relinking method.

Based on performed experimental case study, it has been found that the directed search based on a simple swap algorithm performs much better from the viewpoint of the number of found non-dominated solutions. A secondary consequence of better Pareto front approximation consists in the value of area. Thus, we recommend a swap- based approach for practical usage instead of the path-relinking directed search.

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 Emergency Medical Service System designing.

Acknowledgements

This work was supported by the grants VEGA 1/0216/21 “Design of emergency systems with conflicting criteria using artificial intelligence tools” 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 work was supported also 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 Con- ference 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., Koháni, 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] Doerner, K. F., Gutjahr, W. J., Hartl, R. F., Karall, M. & Reimann, M. (2005). Heuristic Solution of an Extended Double-Coverage Ambulance Location Problem for Austria. *Central European Journal of Operations Research*, 13(4), pp. 325-340.

[7] Drezner, T. & Drezner, Z. (2007). The gravity p-median model. *European Journal of Operational Research* 179, pp. 1239-1251.

[8] 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.

[9] Ingolfsson, A., Budge, S. & Erkut, E. (2008). Optimal ambulance location with random delays and travel times. *Health care management science*, 11(3), pp. 262-274.

[10] 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.

[11] Janáček, J. & Kvet, M. (2021). Swap Heuristics for Emergency System Design with Multiple Facility Location. In: *Proceedings of the 39th International Conference on Mathematical Methods in Economics, 2021*, pp. 226-231.

[12] Janáček, J. & Kvet, M. (2021). Emergency Medical System under Conflicting Criteria. In: *SOR 2021 Proceedings*, pp. 629-635.

[13] Janáček, J. & Kvet, M. (2022). Quality Measure of Pareto Front Approximation. In *Quantitative Methods in Economics 2022*, in print.

[14] 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.

[15] Jánošíková, Ľ. (2007). Emergency Medical Service Planning. In: *Communications* 9(2), pp. 64-68.

[16] Jánošíková, Ľ. & Žarnay, M. (2014). Location of emergency stations as the capacitated p-median problem. In: *Quantitative Methods in Economics (Multiple Criteria Decision Making XVII*). pp. 117-123.

[17] 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.

[18] Kvet, M. & Janáček, J. (2021). Incrementing Heuristic for Non-Dominated Designs of Emergency Medical System. In: *SOR 2021 Proceedings*, pp. 429-474.

[19] Kvet, M & Janáček, J. (2022). Evaluation of Dynamics of Heuristics Used for Pareto Front Approximation. In *Quantitative Methods in Economics 2022*, in print.

[20] Marianov, V. & Serra, D. (2002). Location problems in the public sector, Facility location - *Applications and theory* (Z. Drezner ed.), Berlin, Springer, pp 119-150.

Significance of Web Search Data in Research of Sales in Auto-Motive Industry

Radoslav Lacko[1]

Abstract. Massive development of the internet resulted in many new various types of data. Some of them are indirectly created by users. Almost each of our step on the internet is tracked and written somewhere in the databases. We are able to get the data of users' searches and implement it into econometric research or modelling. Many marketing departments are already doing research of success of their marketing campaigns based on this data. Web search data can relatively fast reflect the changes in customer behavior. We examine how reliable the data are in the long term and evaluate its reliability in this research. Therefore, we have chosen the product that is not fast reversible. Automotive industry, namely passenger cars, would be a great example to test the reliability of web search data in longer period. Industry has gone through de- mand and supply shocks during over past years, such as green economy transformation, pandemic, or global chip shortage. Finally, we discuss the possible utilization of web search data in macroeconomics, or microeconomics.

Keywords: web search data, automotive industry, google trends

JEL Classification: C51
AMS Classification: 91B74

1 Introduction

Many of different new data sources and data types appeared in the past decades. Also, a lot of new tools and practices, to process and visualize data, have been developed in the past years. Statisticians and analysts use these tools to enlarge existing theoretical and practical backgrounds in this field.

Web search data is not the latest source, so we are able to create dataset with longer time periods. The popularity of this data is increasing in past years because it helps to track thoughts and wishes of customers immediately. Therefore, marketing employees are able to evaluate the success of their product campaigns.

This research examines significance of relationship between web search data and real economic indicator of longlasting goods such as passenger cars. So, the main question is whether the web search data are relevant and thus, suitable for further research and processing in the field of macroeconomics, or microeconomics.

2 Literature review

In the previous part we mention that marketing department in companies track outcome and effectiveness of their marketing campaigns by creating various models using Google Trends (GT) data. In other words, marketing employees compare searches on Google, conversion on different level of sales process and real sales in particular time.

Also, the economists and data analysts try to prove a relationship between the GT and the real economy. The first relevant attempt we have found is the article "Predicting the Present with Google Trends" published in 2012. Choi, Hyunyoung, and Hal Varian [3] predict monthly sales of Ford cars based on past 12 months of GT searches. Besides this prediction, they also tried to predict unemployment claims, travel destination planning or consumer confidence.

Boone et al. [1] try to examine the effect of addition of GT data into the models. Authors added a GT data to reduce the forecast errors and provide more robust and relevant outcomes.

Cebrián, Eduardo, and Josep Domenech [2] questions the quality of GT data in the study published in February 2022. The results says that there is problem with the accuracy of the data, but it does not invalidate GT as a data source for other analysis.

[1] Prague University of Economics and Business, Faculty of Informatics and Statistics, nám. W. Churchilla 1938/4 130 67 Praha 3 - Žižkov, lacr02@vse.cz.

3 Data

The dataset consists of two data sources. As the topic says, the first source of the data is used GT [5] – web search data. All the activities of the users are gathered by the Google company and then processed. The results are GT, the statistics of the searches on Google. For each search term or topic is calculated interest over time. In other words, it means "Numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. A score of 0 means there was not enough data for this term", according to the Google[2]. The

The second data source is European Automobile Manufacturers' Association (ACEA). We obtain the numbers of registered passenger cars separately by country and by brand in the monthly period from 2008 to 2020.

We gather the data from GT for each country of EU15[3] with the search term of the particular car brand. The brands Volkswagen and Renault are used in our research. These two brands are the most common according to number of registered passenger cars in the EU. Figure 1 shows the Volkswagen brand and its numbers of registered passenger cars in examined periods. Figure 2 shows the raw Google Trends data of Volkswagen brand in Germany and France. Also, we prepare seasonally adjusted data of all Google Trends data, the Figure 3 shows the raw and seasonally adjusted Google Trends data of Volkswagen in Germany.

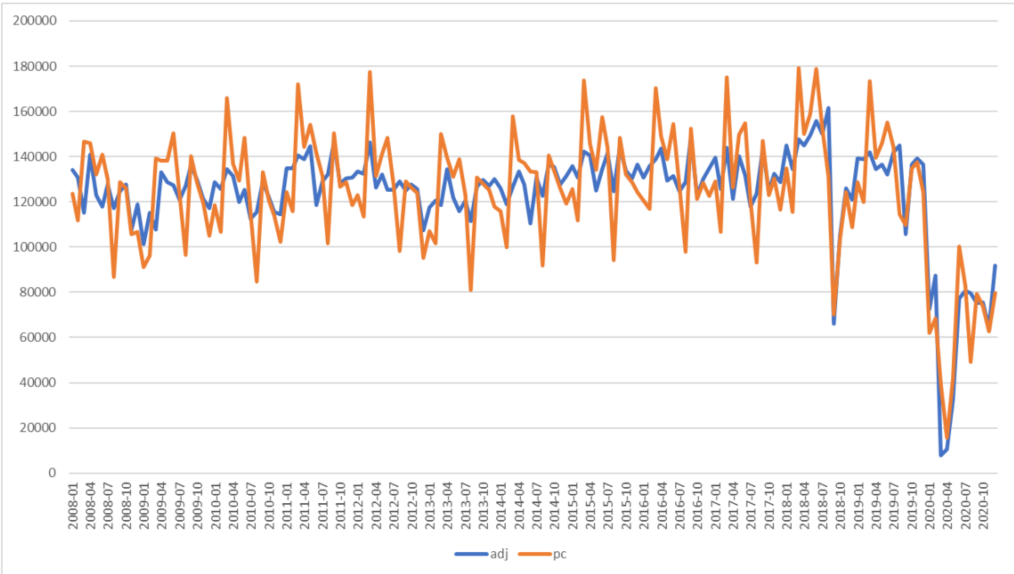


Figure 1 The number of registered passenger cars of Volkswagen brand – real data vs. seasonally adjusted

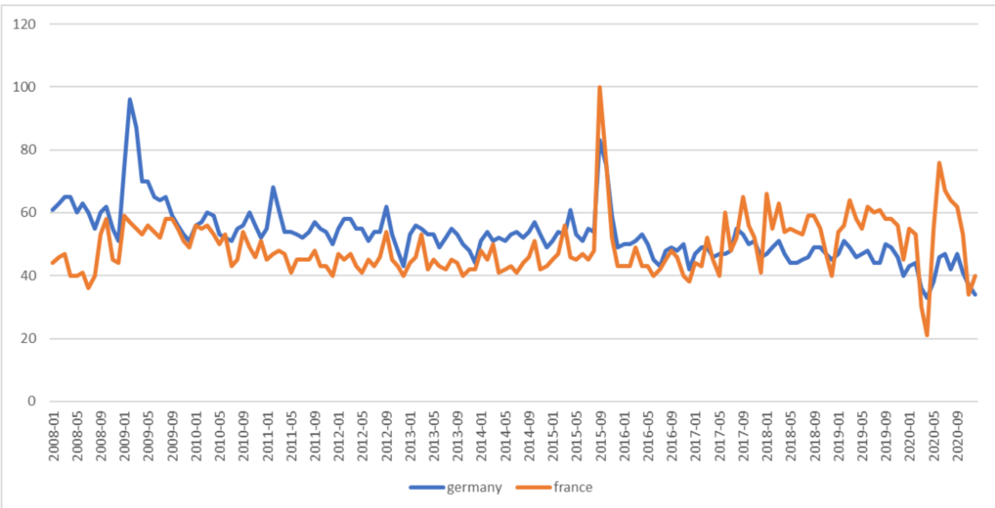


Figure 2 Google Trends data of Volkswagen brand in Germany and France

[2] https://trends.google.com/trends/
[3] European Union members before 2004

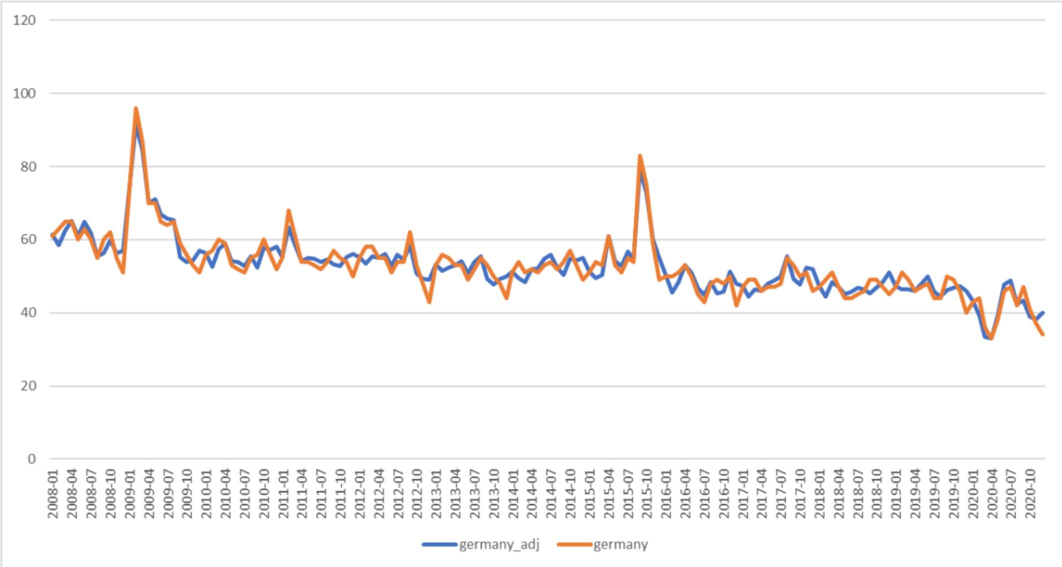


Figure 3 Seasonally adjusted and raw Google Trends data of Volkswagen company in Germany

4 Methodology

The panel data structure defines the models that we use in our estimates. We obtain Least Squares Dummy Variable estimator with a fixed effect of brand represented as *factor(brand).* Paseran [6] shows and defines panel data models and presents the differences in methodology.

Besides the *brand* variable each model has the base independent variables from GT data for particular car brand namely the United Kingdom as *uk*, Netherlands as *netherlands*, Belgium as *belgium*, France as *france*, Germany as *germany*, Italy as *italy*, Spain as *spain*. Each independent variable that represents the country is recalculated as the first difference which helps us with stationarity of each GT time series. We select only countries with share on registered passenger cars higher than 4 %. There are about 8 other countries with shares less than 2 %. Total share of all selected country was about 80 % in 2008 and 78 % in 2020 of EU15 + EFTA.

There are two other dummy variables that represent the covid pandemic period as *cvd* and the diesel gate affair as *dg.*

$$pc_{it} = \beta_0 + \beta_1 uk_{it} + \beta_2 netherlands_{it} + \beta_3 belgium_{it} + \beta_4 france_{it} + \beta_5 germany_{it} + \beta_7 italy_{it} + \beta_7 spain_{it} + \gamma d_{brand} + \epsilon_{it} \tag{1}$$

$$pc_{it} = \beta_0 + \beta_1 uk_{it} + \beta_2 netherlands_{it} + \beta_3 belgium_{it} + \beta_4 france_{it} + \beta_5 germany_{it} + \beta_7 italy_{it} + \beta_7 spain_{it} + \beta_8 cvd + \gamma d_{brand} + \epsilon_{it} \tag{2}$$

$$pc_{it} = \beta_0 + \beta_1 uk_{it} + \beta_2 netherlands_{it} + \beta_3 belgium_{it} + \beta_4 france_{it} + \beta_5 germany_{it} + \beta_7 italy_{it} + \beta_7 spain_{it} + \beta_8 cvd + \beta_8 dg + \gamma d_{brand} + \epsilon_{it} \tag{3}$$

5 Results

As we mentioned in methodology part there are three various estimates, that differ in independent variables used. The results of the estimates are in the Table 1 using seasonally unadjusted data.

Fantazzini [4] estimated models using GT data with obtaining both raw and seasonally adjusted data. The authors examine that estimated models without using GT data performed better compared to seasonally adjusted data in linear models. Our models provided better performance using raw data for all the models. We did all the models using both seasonally adjusted and raw data. In our case there were no significant results using seasonally adjusted data, R-squared value reached top of 0.03. Thus we present the results of estimations using seasonally unadjusted data.

We can conclude that most of the countries independent variables are not significant, but there is observed relationship between *France*, *Italy*, and the number of registered passenger cars, that is significant on 1 % in the first estimation, on 5 % in the second and the third estimation. *Italy* independent variable has negative coefficient which in other words means with more searches the number of registered passenger cars decreases. There is declining trend of the number of registered passenger cars in the period between 2008 and 2020. Thus, increasing searches of Renault and Volkswagen might be connected to changes in the customers' behavior. The overall budget is lower than it was before, and customers are comparing more options and Renault, or Volkswagen might be the second or the third possible choice.

On the other hand, *France* has a positive coefficient, so more searches mean more registered cars. Also, Renault is France-based company and the customers' behavior might be more patriotic. This effect has not been observed in *Germany* and Volkswagen.

The first model provides significant results for Spain, with almost the same but positive effect compared to *Italy*. Moreover, it is not significant on 5 % in the second and the third model. The coefficients of *Spain* and *Belgium* is significant on 10 % in those models.

The variable factor(brand) is significant in all models on less than 1 %. The coefficient represents the difference between Volkswagen car registration compared to Renault.

Only one of our two dummy variables is significant, namely cvd which represents the decreases during the first wave of the covid pandemic from March to May in 2020. Diesel gate affair in 2015 represented by *dg* variable was not significant in any of our model.

The heteroscedasticity was observed in the second and the third model, so we calculate robust standard errors. The results are not so different as it is presented in Table 2. *Belgium* independent variable become significant on 5 % with coefficient of 259.26. The rest of the variables remain unchanged.

Model	**LSDV _brand (1)**	**LSDV_cvd (2)**	**LSDV_dg (3)**
Coefficients:			
constant	**123912**	**125105**	**125043**
	<0.01 ***	<0.01 ***	<0.01 ***
uk	**-235.4**	**-439.8**	**-433.0**
	0.42	0.11	0.12
netherlands	**-425.0**	**-375.6**	**-374.4**
	0.17	0.20	0.20
belgium	**261.2**	**259.3**	**258.1**
	0.089 .	0.072 .	0.073 .
france	**599.0**	**947.2**	**946.0**
	0.034 *	<0.01 ***	<0.01 ***
germany	**-349.6**	**-220.9**	**-218.0**
	0.146	0.329	0.337
italy	**-892.4**	**-1032.7**	**-1038.5**
	<0.01 ***	<0.01 ***	<0.01 ***
spain	**895.2**	**585.7**	**588.3**
	<0.01 ***	0.065 .	0.064 .
factor(brand)	**-42133.9**	**-42142.8**	**-42143.1**
	<0.01 ***	<0.01 ***	<0.01 ***
cvd	-	**-63291.8**	**-63187.2**
		<0.01 ***	<0.01 ***
dg	-	-	**2368.7**
			0.774
Adj. R-squared	0.4483	0.5285	0.5286
p-value (F-test)	<0.01 ***	<0.01 ***	<0.01 ***
Breusch-Pagan test	0.2343	0.0435	0.0447

Table 1 Estimations output

6 Conclusion and discussion

The main point of this research is to consider is the Google Trends data are reliable for the further research in automotive industry. We created several models to prove relationship between web searches and real economy in the field of automotive. Although the results of the estimations are not so robust, but we can consider that there was examined relationship. Therefore, we can continue in the research using Google Trends data and consider it as significant data source. To improve the performance of our estimation, we may include lag variables in further research.

On the other hand, there are some details that would improve the robustness of the estimate. The different granularity of data from ACEA would help to examine relationships between car brand and its web searches in particular country. Also, there are possibilities to provide a better estimate if Google has calculated interest over time for EU as a group not separately for each country. Also, there is the option to created weighted index calculated from the Google Trends index from various search forms including model type such as Volkswagen Golf, Volkswagen Arteon etc.

It would be interesting to see if it will be possible to estimate significant models in periods of microchip shortage, high inflation, and actual war conflict on automotive industry as one of the pillars of European economy.

Acknowledgements

The work on this paper was supported by the grant No. F4/27/2020 of the Internal Grant Agency of University of Economics, Prague.

References

[1] Boone, T. et al. (2018). Can Google Trends Improve Your Sales Forecast?., *Production and Operations Management*, 27(10), (pp. 1770–1774).
[2] Cebrián, E. & Domenech, J. (2022). Is Google Trends a quality data source?. *Applied Economics Letters*, (pp. 1–5).
[3] Choi, H. & Varian, H. (2012). Predicting the Present with Google Trends: PREDICTING THE PRESENT WITH GOOGLE TRENDS. *Economic Record*, 88, (pp. 2–9).
[4] Fantazzini, D. and Toktamysova, Z. (2015). Forecasting German car sales using Google data and multivariate models. *International Journal of Production Economics*, 170, (pp. 97–135).
[5] Google Trends. (2022). Google Trends. Available at: https://www.google.com/trends/.
[6] Pesaran, M.H. (2015). *Time series and panel data econometrics*. First edition. Oxford, United Kingdom:Oxford University Press.

The Impact of COVID19 Lockdown on Imprecision Measure of TrOFNs Portfolio Analysis

Anna Łyczkowska-Hanćkowiak[1], Aleksandra Wójcicka-Wójtowicz[2]

Abstract. Oriented fuzzy numbers (OFNs) can be used in managing an investment portfolio – they include information uncertainty and imprecision related to financial market. A portfolio analysis bases on an expected fuzzy discount factor and an imprecise present value (PV). The main purpose of the paper is to compare a portfolio analysis including stocks identified by PV assessed by trapezoidal oriented fuzzy number (TrOFNs) in times of regular stock exchange session and during unexpected events (COVID19 lockdown). All considerations are illustrated by an empirical case study. The imprecision risk of the investment portfolio is estimated by energy and entropy measures, comparison of which proves that extreme events, had a tremendous impact on both measures. The analysis also shows that the aggregated portfolio measures of energy and entropy in case of regular situation are lower than in case of a crisis and also lower than measures of individual elements of portfolio.

Keywords: OFN; imprecision; PV; discount factor; portfolio analysis

JEL Classification: C44, G11, G24
AMS Classification: 03E72

1 Introduction

Financial markets are characterised by imprecision, hence the use of fuzzy numbers (FNs) in the analysis of financial data is justified. In [7] it is demonstrated that in portfolio analysis, oriented FNs (OFNs) are more useful than FNs. Moreover, in case of financial data imprecision the expected discount factor is a more convenient portfolio tool than the expected return rate [8, 10]. Therefore, the main aim of our paper is to study portfolio discount factor when portfolio assets are evaluated by trapezoidal oriented FNs (TrOFNs). The most general definition of FN was formulated by Dubois and Prade [1]. The notion of ordered FN was introduced by Kosiński et al. [2, 3]. For formal reasons, the Kosiński's theory was revised in [6] where the notion of ordered FN is nar- rowed down to the notion of oriented FN (OFN). The relationships between FNs, ordered FNs, and OFNs are discussed in detail in [8]. A special case of OFNs are TrOFNs [6] defined as follows:

Definition 1. [6] For any monotonic sequence $(a,b,c,d) \subset \mathbb{R}$, TrOFN $\overleftrightarrow{Tr}(a,b,c,d) = \overleftrightarrow{\mathcal{T}}$ is OFN $\overleftrightarrow{\mathcal{T}} \in \mathbb{K}$ determined explicitly by its membership functions $\mu_T \in [0,1]^{\mathbb{R}}$ by

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

If $a < d$ then $\overleftrightarrow{Tr}(a,b,c,d)$ has a positive orientation $\overrightarrow{a,d}$ (an increase in approximated number is expected). If $a > d$ then $\overleftrightarrow{Tr}(a,b,c,d)$ has a negative orientation $\overrightarrow{a,d}$ (a decrease is expected). The space of all negatively and positively oriented TrOFNs is denoted by $\mathbb{K}_{Tr}^-$ and $\mathbb{K}_{Tr}^+$ respectively. Basing on Kosinski's approach, we extend arithmetic operators to the case of $\mathbb{K}_{Tr}$. For any pair $\left(\overleftrightarrow{Tr}(a,b,c,d), \overleftrightarrow{Tr}(p-a,q-b,r-c,s-d)\right) \in \mathbb{K}_{Tr}^2$ and $\beta \in \mathbb{R}$, arithmetic operations of extended sum $\boxplus$ and dot product $\boxdot$ are defined as follows [6, 10]:

$$\overleftrightarrow{Tr}(a,b,c,d) \boxplus \overleftrightarrow{Tr}(p-a,q-b,r-c,s-d) = \begin{cases} \overleftrightarrow{Tr}(\min\{p,q\},q,r,\max\{r,s\}), & (q<r) \vee (q=r \wedge p \le s), \\ \overleftrightarrow{Tr}(\max\{p,q\},q,r,\min\{r,s\}), & (q>r) \vee (q=r \wedge p > s). \end{cases} \tag{2}$$

[1] WSB University in Poznan, ul. Powstańców Wielkopolskich 5, 61-895 Poznań, Poland; anna.lyczkowska-hanckowiak@wsb.poznan.pl
[2] Poznań University of Economics and Business, Al. Niepodległości 10, 61-875 Poznań, aleksandra.wojcicka-wojtowicz@ue.poznan.pl

$$\beta \boxdot \overleftrightarrow{Tr}(a,b,c,d) = \overleftrightarrow{Tr}(\beta \cdot a, \beta \cdot b, \beta \cdot c, \beta \cdot d) \tag{3}$$

Information imprecision is understood as a special case of ambiguity and indistinctness of information [9]. Ambiguity results in an unequivocally recommendation, while indistinctness does not distinguish clearly between recommended and non-recommended alternatives. If OFN ambiguity increases then we are given more recommended alternatives and, in turn, it leads to a higher ambiguity risk of choosing the wrong alternative among given ones. In decision-making analysis, energy measure is used to assess the ambiguity risk while entropy measure – the indistinctness risk. Imprecision risk is a sum of ambiguity and indistinctness risk. For any TrOFN $\overleftrightarrow{Tr}(a,b,c,d)$, its energy and entropy measures are expressed by the following relations [4]:

$$d\left(\overleftrightarrow{Tr}(a,b,c,d)\right) = \frac{1}{2} \cdot |d+c-b-a|, e\left(\overleftrightarrow{Tr}(a,b,c,d)\right) = \frac{1}{4} \cdot |d-c+b-a| \tag{4}$$

$$e\left(\overleftrightarrow{Tr}(a,b,c,d)\right) = \frac{1}{4} \cdot |d-c+b-a| \tag{5}$$

Obtained results may be presented in a form of an asset benefit also regarded as a relative profit of the owned asset can be presented by a return rate or a discount factor, dependent on the profit value and the asset value [5]. In our research benefit indexes are imprecisely estimated by TrOFNs $\overleftrightarrow{\mathcal{K}}, \overleftrightarrow{\mathcal{L}} \in \mathbb{K}_{Tr}$. The portfolio benefit index is determined by a function $\varpi: (\mathbb{K}_{Tr})^2 \times [0,1] \to \mathbb{K}_{Tr}$ given by

$$\left(\overleftrightarrow{\mathcal{K}}, \overleftrightarrow{\mathcal{L}}, \lambda\right) = \left(\lambda \boxdot \overleftrightarrow{\mathcal{K}}\right) \boxplus \left((1-\lambda) \boxdot \overleftrightarrow{\mathcal{L}}\right) \tag{6}$$

Theorem 1 [8]. *For any real number $\lambda \in [0,1]$, we have:*

- *for any pair* $\left(\overleftrightarrow{\mathcal{K}}, \overleftrightarrow{\mathcal{L}}\right) \in (\mathbb{K}_{Tr}^- \times \mathbb{K}_{Tr}^-) \cup \left((\mathbb{K}_{Tr}^+ \cup \mathbb{R}) \times (\mathbb{K}_{Tr}^+ \cup \mathbb{R})\right)$

$$d\left(\varpi(\overleftrightarrow{\mathcal{K}}, \overleftrightarrow{\mathcal{L}}, \lambda)\right) = \lambda \cdot d(\overleftrightarrow{\mathcal{K}}) + (1-\lambda) \cdot d(\overleftrightarrow{\mathcal{L}}), e\left(\varpi(\overleftrightarrow{\mathcal{K}}, \overleftrightarrow{\mathcal{L}}, \lambda)\right) = \lambda \cdot \mathrm{e}(\overleftrightarrow{\mathcal{K}}) + (1-\lambda) \cdot \mathrm{e}(\overleftrightarrow{\mathcal{L}}) \tag{7}$$

- *for any pair* $\left(\overleftrightarrow{\mathcal{K}}, \overleftrightarrow{\mathcal{L}}\right) \in \left((\mathbb{K}_{Tr}^+ \cup \mathbb{R}) \times \mathbb{K}_{Tr}^-\right)$

$$d\left(\varpi(\overleftrightarrow{\mathcal{K}}, \overleftrightarrow{\mathcal{L}}, \lambda)\right) \leq \begin{cases} \lambda \cdot d(\overleftrightarrow{\mathcal{K}}) - (1-\lambda) \cdot d\left(\mathcal{Core}(\overleftrightarrow{\mathcal{L}})\right), & \varpi(\overleftrightarrow{\mathcal{K}}, \overleftrightarrow{\mathcal{L}}, \lambda) \in \mathbb{K}_{\mathrm{Tr}}^+ \cup \mathbb{R}, \\ (1-\lambda) \cdot d(\overleftrightarrow{\mathcal{L}}) - \lambda \cdot d\left(\mathcal{Core}(\overleftrightarrow{\mathcal{K}})\right), & \varpi(\overleftrightarrow{\mathcal{K}}, \overleftrightarrow{\mathcal{L}}, \lambda) \in \mathbb{K}_{\mathrm{Tr}}^- \cup \mathbb{R}, \end{cases} \tag{8}$$

where $\mathcal{Core}(A) = \{x: \mu_A(x) = 1\}$, $A \in \mathcal{F}(\mathbb{R})$

- *for any pair* $\left(\overleftrightarrow{\mathcal{K}}, \overleftrightarrow{\mathcal{L}}\right) \in \left((\mathbb{K}_{Tr}^+ \cup \mathbb{R}) \times \mathbb{K}_{Tr}^-\right) \cup \left(\mathbb{K}_{Tr}^- \times (\mathbb{K}_{Tr}^+ \cup \mathbb{R})\right)$

$$e\left(\varpi(\overleftrightarrow{\mathcal{K}}, \overleftrightarrow{\mathcal{L}}, \lambda)\right) \leq \min\{\lambda e(\overleftrightarrow{\mathcal{K}}), (1-\lambda) e(\overleftrightarrow{\mathcal{L}})\}. \tag{9}$$

The above presented theory on TrOFNs and oriented present value (OPV), discount factor (DF) and portfolio approach, that follows in Section 2 and 3, is the absolute theoretical minimum necessary to explain the most vital part of the paper which is the case study shown in Section 4. Extensions of some equations were removed, only the essential theory is presented. For more information on TrOFNs, OPV and EDF see [4, 5, 7, 9].

2 Oriented present value and discount factor

PV was redefined by Piasecki as a present equivalent of a payment available at a given time in the future. In [5] the estimation of a fuzzy present value was supplemented by a forecast of the closest changes in quoted price. In a portfolio analysis OFNs are more useful than FNs [8]. Therefore, an imprecise PV may be estimated using OFNs [4-7] becoming an oriented present value. Any OPV is described by a monotonic sequence $\left(V_s, V_f, \check{P}, V_l, V_e\right)$ and estimated by trapezoidal oriented fuzzy numbers:

$$\overleftrightarrow{PV} = \overleftrightarrow{Tr}\left(V_s, V_f, V_l, V_e\right) \tag{10}$$

where the monotonic sequence $\left(V_s, V_f, \check{P}, V_l, V_e\right)$ is determined by: $\check{P}$ – quoted price, $[V_s, V_e] \subset \mathbb{R}^+$ – interval of all PV possible values, $\left[V_f, V_l\right] \subset [V_s, V_e]$ – interval of all prices that do not significantly differ from the quoted price $\check{P}$. If an increase/decrease in price is anticipated, the PV is respectively determined by a positively/negatively oriented TrOFN. Japanese candle models are examples of trapezoidal OPVs [7]. Assuming time horizon $t > 0$ of a fixed investment, the analyzed asset is described by an anticipated future value V_t and an assessed present value V_0. The benefits from possessing an asset is a simple return rate r_t given by the identity:

$$r_t = \frac{V_t - V_0}{V_0} = \frac{V_t}{V_0} - 1 \tag{11}$$

FV is a random variable $\tilde{V}_t : \Omega \longrightarrow \mathbb{R}^+$ where Ω is a set of elementary states, ω, of the financial market. PV is identified with the observed quoted price $\check{P}$ and the return rate is a random variable determined by identity:

$$r_t(\omega) = \frac{\tilde{V}_t(\omega) - \check{P}}{\check{P}} \tag{12}$$

In practice of financial markets analysis, the uncertainty risk (lack of knowledge) is usually described by the probability distribution of a return rate (12) given by a cumulative distribution function $F_r(\cdot\,|\bar{r}) : \mathbb{R} \longrightarrow [0,1]$. Assuming that an expected value $\bar{r}$ of the distribution and the expected discount factor (EDF) $\bar{v}$ exist then:

$$\bar{v} = (1 + \bar{r})^{-1} \tag{13}$$

Taking together (11) and (12), we obtain a following formula describing the return rate

$$r_t = r_t(V_0, \omega) = \frac{\check{P} \cdot (1 + r_t(\omega))}{V_0} - 1 \tag{14}$$

Then the imprecise EDF $\mathcal{V} : \mathbb{R}^+ \to \mathbb{R}^+$ as a unary operator transforming PV as follows:

$$\mathcal{V}(V_0) = \left(\frac{\check{P} \cdot (1 + \bar{r})}{V_0}\right)^{-1} = \frac{\bar{v}}{\check{P}} \cdot V_0 \tag{15}$$

PV is imprecisely estimated by TrOFN (10) and the imprecise EDF $\overleftrightarrow{\mathcal{V}}(\overleftrightarrow{PV})$ is also given by TrOFN

$$\overleftrightarrow{\mathcal{V}}(\overleftrightarrow{PV}) = \overleftrightarrow{Tr}\left(\frac{V_s \cdot \bar{v}}{\check{P}}, \frac{V_f \cdot \bar{v}}{\check{P}}, \frac{V_l \cdot \bar{v}}{\check{P}}, \frac{V_e \cdot \bar{v}}{\check{P}}\right) = \overleftrightarrow{\mathcal{V}}\left(\overleftrightarrow{Tr}(V_s, V_f, V_l, V_e)\right) \tag{16}$$

The energy measure and the entropy measure of the OEDF $\overleftrightarrow{\mathcal{V}}(\hat{S})$ are determined by formulas:

$$d\left(\overleftrightarrow{\mathcal{V}}(\hat{S})\right) = \left|\frac{(V_e + V_l - V_f - V_s) \cdot \bar{v}}{2\check{P}}\right|, \tag{17}$$

$$e\left(\overleftrightarrow{\mathcal{V}}(\hat{S})\right) = \left|\frac{(V_e - V_l + V_f - V_s) \cdot \bar{v}}{4\check{P}}\right|. \tag{18}$$

3 Portfolio analysis

For the financial portfolio analysis we use the method proposed in [4, 5]. By a financial portfolio we understand an arbitrary, finite set of assets. Any asset is understood as a fixed security in a long position. If we consider a case of a multi-asset portfolio π^*, built of securities Y_i, we describe the portfolio as a set $\pi^* = \{Y_i : i = 1,2,\ldots,n\}$. Any security Y_i is characterized by its price $\check{P}_i \in \mathbb{R}^+$, by its imprecise PV evaluated by TrOFN

$$\overleftrightarrow{PV}_i = \overleftrightarrow{Tr}\left(V_s^{(i)}, V_f^{(i)}, V_l^{(i)}, V_e^{(i)}\right) \tag{19}$$

and by its EDF $\bar{v}_i$ determined by (13). Any security Y_i is therefore defined by its imprecise EDF

$$\overleftrightarrow{\mathcal{V}}_i = \overleftrightarrow{Tr}\left(D_s^{(i)}, D_f^{(i)}, D_l^{(i)}, D_e^{(i)}\right) = \overleftrightarrow{Tr}\left(V_s^{(i)} \cdot \frac{\bar{v}_i}{\check{P}_i}, V_f^{(i)} \cdot \frac{\bar{v}_i}{\check{P}_i}, V_l^{(i)} \cdot \frac{\bar{v}_i}{\check{P}_i}, V_e^{(i)} \cdot \frac{\bar{v}_i}{\check{P}_i}\right). \tag{20}$$

A portfolio PV is a sum of its components' PVs. In our case the components' PVs are estimated by TrOFNs, we distinguish the portfolio of rising securities $\pi^+ = \{Y_i \in \pi^* : \overleftrightarrow{PV}_i \in \mathbb{K}^+_{Tr}\}$ and the portfolio of falling securities $\pi^- = \pi^* \backslash \pi^+$. Using (2) PV of portfolio π^+ ($\overleftrightarrow{PV}^+$) – PV of portfolio π^- ($\overleftrightarrow{PV}^-$) is calculated analogously:

$$\overleftrightarrow{PV}^+ = \overleftrightarrow{Tr}\left(V_s^{(+)}, V_f^{(+)}, V_l^{(+)}, V_e^{(+)}\right) = \overleftrightarrow{Tr}\left(\sum_{Y_i \in \pi^+} V_s^{(i)}, \sum_{Y_i \in \pi^+} V_f^{(i)}, \sum_{Y_i \in \pi^+} V_l^{(i)}, \sum_{Y_i \in \pi^+} V_e^{(i)}\right). \tag{21}$$

Eventually, the PV of portfolio π^* ($\overleftrightarrow{PV}^*$) is calculated:

$$\overleftrightarrow{PV}^* = \overleftrightarrow{PV}^+ \boxplus \overleftrightarrow{PV}^- = \overleftrightarrow{Tr}\left(V_s^{(+)}, V_f^{(+)}, V_l^{(+)}, V_e^{(+)}\right) \boxplus \overleftrightarrow{Tr}\left(V_s^{(-)}, V_f^{(-)}, V_l^{(-)}, V_e^{(-)}\right). \tag{22}$$

The imprecise EDFs of portfolios π^+, (π^- being analogous to π^+) and π^* are given by the following formulas:

$$\overleftrightarrow{\mathcal{V}}^+ = \overleftrightarrow{Tr}\left(\sum_{Y_i \in \pi^+} \frac{\bar{v}^+ \cdot p_i^{(+)}}{\bar{v}_i} \cdot D_s^{(i)}, \sum_{Y_i \in \pi^+} \frac{\bar{v}^+ \cdot p_i^{(+)}}{\bar{v}_i} \cdot D_f^{(i)}, \sum_{Y_i \in \pi^+} \frac{\bar{v}^+ \cdot p_i^{(+)}}{\bar{v}_i} \cdot D_l^{(i)}, \sum_{Y_i \in \pi^+} \frac{\bar{v}^+ \cdot p_i^{(+)}}{\bar{v}_i} \cdot D_e^{(i)}\right), \tag{23}$$

$$\overleftrightarrow{\mathcal{V}}^{*} = \overleftrightarrow{Tr}\left(D_s^{(*)}, D_f^{(*)}, D_l^{(*)}, D_e^{(*)}\right) = \left(\frac{\bar{v}^{*}\cdot q^{+}}{\bar{v}^{+}} \boxdot \overleftrightarrow{\mathcal{V}}^{+}\right) \boxplus \left(\frac{\bar{v}^{*}\cdot q^{-}}{\bar{v}^{-}} \boxdot \overleftrightarrow{\mathcal{V}}^{-}\right), \tag{24}$$

where

- $M^{+} = \sum_{Y_i \in \pi^{+}} \check{P}_i$, $M^{-} = \sum_{Y_i \in \pi^{-}} \check{P}_i$, $M^{*} = M^{+} + M^{-}$ are the values of portfolios $\pi^{+}, \pi^{-}, \pi^{*}$,
- $p_i^{+/-} = \frac{\check{P}_i}{M^{+/-}}$, are the share $p_i^{+/-}$ of the asset $Y_i \in \pi^{+/-}$ in the portfolio $\pi^{+/-}$ respectively,
- $q^{+/-} = \frac{M^{+/-}}{M^{*}}$ are the share $q^{+/-}$ of portfolio $\pi^{+/-}$ in π^{*}respectively,
- $\bar{\boldsymbol{v}}^{+/-} = \left(\sum_{\boldsymbol{Y}_i \in \boldsymbol{\pi}^{+/-}} \frac{\boldsymbol{p}_i^{+/-}}{\bar{\boldsymbol{v}}_i}\right)^{-1}$, $\bar{\boldsymbol{v}}^{*} = \left(\frac{\boldsymbol{q}^{+}}{\bar{\boldsymbol{v}}^{+}} + \frac{\boldsymbol{q}^{-}}{\bar{\boldsymbol{v}}^{-}}\right)^{-1}$ are the EDF $\bar{\boldsymbol{v}}^{+/-/*}$ of portfolio $\boldsymbol{\pi}^{+/-/*}$ respectively.

The energy measure and the entropy measure of EDF $\overleftrightarrow{\mathcal{V}}^{+/-}$are determined by formulas:

$$d\left(\overleftrightarrow{\mathcal{V}}^{+/-}\right) = \sum_{Y_i \in \pi^{+/-}} \frac{\bar{v}^{+/-} \cdot q_i^{(+/-)}}{\bar{v}_i} \cdot d\left(\overleftrightarrow{\mathcal{V}}(Y_i)\right), e\left(\overleftrightarrow{\mathcal{V}}^{+/-}\right) = \sum_{Y_i \in \pi^{+/-}} \frac{\bar{v}^{+/-} \cdot q_i^{(+/-)}}{\bar{v}_i} \cdot e\left(\overleftrightarrow{\mathcal{V}}(Y_i)\right). \tag{25}$$

The entropy measure of EDF $\overleftrightarrow{\mathcal{V}}^{*}$ of portfolio π^{*} meet the conditions:

$$d\left(\overleftrightarrow{\mathcal{V}}^{*}\right) \le \begin{cases} \frac{\bar{v}^{*}\cdot q^{+}}{\bar{v}^{+}} \cdot d\left(\overleftrightarrow{\mathcal{V}}^{+}\right) - \frac{\bar{v}^{*}\cdot q^{-}}{\bar{v}^{-}} \cdot d\left(\mathcal{Core}\left(\overleftrightarrow{\mathcal{V}}^{-}\right)\right), & \overleftrightarrow{\mathcal{V}}^{*} \in \mathbb{K}_{Tr}^{+} \cup \mathbb{R}, \\ \frac{\bar{v}^{*}\cdot q^{-}}{\bar{v}^{-}} \cdot d\left(\overleftrightarrow{\mathcal{V}}^{-}\right) - \frac{\bar{v}^{*}\cdot q^{+}}{\bar{v}^{+}} \cdot d\left(\mathcal{Core}\left(\overleftrightarrow{\mathcal{V}}^{+}\right)\right), & \overleftrightarrow{\mathcal{V}}^{*} \in \mathbb{K}_{Tr}^{-} \cup \mathbb{R}, \end{cases} \tag{26}$$

$$e\left(\overleftrightarrow{\mathcal{V}}^{*}\right) \le \min\left\{\frac{\bar{v}^{*}\cdot q^{+}}{\bar{v}^{+}} e\left(\overleftrightarrow{\mathcal{V}}^{+}\right), \frac{\bar{v}^{*}\cdot q^{-}}{\bar{v}^{-}} e\left(\overleftrightarrow{\mathcal{V}}^{-}\right)\right\}. \tag{27}$$

4 Case study – empirical analysis

The main objective of the paper is to implement a new method of portfolio analysis using TrOFNs decribed in detail in [4] to compare the portfolio measures in times of regular listing (29.01.2020) and during first COVID19 lockdown in Poland (16.03.2020) to see how the portfolio reacts to extreme situations which has not been researched before. We consider a portfolio π* consisting of 17 companies included in WIG20 quoted on Warsaw Stock Exchange (WSE). The choice of those companies is dictated by the fact that WIG20 index is considered to be the best reflection of economic changes on Polish stock exchange. Standard WIG20 includes 20 companies. However, in the analysed time (regular/lockdown) three of the companies where replaced by other ones, therefore to allow the comparability of the research 17 out of standard number of 20 were analysed. In case of standard time (before lockdown) the portfolio contains the following securities (presented in a following manner {number of stocks/tick/symbol}): {10/CCC/Y_{CCC}}, {17/CDR/Y_{CDR}}, {50/CPS/Y_{CPS}{5/DNP/Y_{DNP}}, {200/JSW/Y_{JSW}}, {8/KGH/Y_{KGH}}, {50/LTS/Y_{LTS}}, {1/LPP/Y_{LPP}}, {100/OPL/Y_{OPL}}, {10/PEO/Y_{PEO}}, {100/PGE/Y_{PGE}}, {1200/PGN/Y_{PGN}}, {50/PKN/Y_{PKN}}, {30/PKO/Y_{PKO}}, {25/PZU/Y_{PZU}}, {10/SPL/Y_{SPL}}, {200/TPE/Y_{TPE}}. For each analysed WSE stock on 28.01.2020 (closing session) we estimate its PV (a single stock price) equal to TrOFN $\overleftrightarrow{PV}_S$ describing its Japanese candle [7]. A quoted price $\check{P}_S$ of each portfolio element becomes an initial price on 29.01.2020. We consider a quarterly horizon of an investment. For an average expected return rate $\bar{r} = 0.0985$ from (13) we get an expected discount factor $\bar{v} = 0.9103$. For each considered stocks Ŝ using (15) we compute its imprecise expected discount factor and, basing on (17) and (18), the energy and entropy measure of OEDF $\overleftrightarrow{V}_S$. PVs and OEDFs are presented in Table 1.

Assets	**Present Value $\overleftrightarrow{PV}_s$**	**Quoted Price $\check{P}_s$**	**OEDF $\overleftrightarrow{V}_s$ – energy measure**	**OEDF $\overleftrightarrow{V}_s$ - entropy measure**
Y_{CCC}	$\overleftrightarrow{Tr}$(83.35; 88.00; 88.00; 89.65)	88.00	0.0326	0.0163
Y_{CDR}	$\overleftrightarrow{Tr}$(271.50; 271.50; 276.30; 276.30)	277.00	0.0158	0.0000
Y_{CPS}	$\overleftrightarrow{Tr}$(26.42; 26.60; 27.04; 27.34)	27.20	0.0228	0.0040
Y_{DNP}	$\overleftrightarrow{Tr}$(155.00; 155.00; 155.10; 157.30)	155.30	0.0070	0.0032
Y_{JSW}	$\overleftrightarrow{Tr}$(18.60; 19.36; 20.14; 20.14)	20.32	0.0520	0.0085
Y_{KGH}	$\overleftrightarrow{Tr}$(91.78; 93.60; 93.70; 94.90)	94.24	0.0156	0.0073
Y_{LTS}	$\overleftrightarrow{Tr}$(83.88; 83.40; 81.16; 80.26)	81.44	0.0328	0.0039
Y_{LPP}	$\overleftrightarrow{Tr}$(8205.00; 8380.00; 8395.00; 8460.00)	8385.00	0.0147	0.0065
Y_{OPL}	$\overleftrightarrow{Tr}$(7.01; 7.05; 7.20; 7.35)	7.17	0.0311	0.0060

Y_{PEO}	$\overleftrightarrow{Tr}(97.22;\ 97.70;\ 98.20;\ 98.66)$	98.20	0.0090	0.0022
Y_{PGE}	$\overleftrightarrow{Tr}(7.08;\ 7.15;\ 7.30;\ 7.40)$	7.30	0.0293	0.0053
Y_{PGN}	$\overleftrightarrow{Tr}(3.91;\ 3.88;\ 3.86;\ 3.82)$	3.87	0.0129	0.0041
Y_{PKN}	$\overleftrightarrow{Tr}(83.22;\ 83.00;\ 81.62;\ 81.18)$	81.90	0.0190	0.0018
Y_{PKO}	$\overleftrightarrow{Tr}(34.59;\ 34.68;\ 34.90;\ 35.26)$	34.93	0.0116	0.0029
Y_{PZU}	$\overleftrightarrow{Tr}(40.72;\ 40.73;\ 40.89;\ 41.11)$	40.88	0.0061	0.0013
Y_{SPL}	$\overleftrightarrow{Tr}(276.20;\ 278.00;\ 281.80;\ 283.80)$	287.00	0.0181	0.0030
Y_{TPE}	$\overleftrightarrow{Tr}(1.51;\ 1.53;\ 1.56;\ 1.56)$	1.56	0.0233	0.0029

Table 1 Recorded values of the portfolio π stocks (pre-lockdown period)

The shares $Y_{CCC}, Y_{CDR}, Y_{CPS}, Y_{DNP}, Y_{JSW}, Y_{KGH}, Y_{LPP}, Y_{OPL}, Y_{PEO}, Y_{PGE}, Y_{PKO}, Y_{PZU}, Y_{SPL}$ and Y_{TPE} belong to portfolio π^+ of rising securities. The shares Y_{LTS}, Y_{PGN}, and Y_{PKN} belong to portfolio π^- of falling securities. Using (21) and (22) we calculate PVs of portfolios π^+, π^-, π^*: $\overleftrightarrow{PV}^+ = \overleftrightarrow{Tr}(27705.14, 28142.95, 28508.45, 28691.45)$, $\overleftrightarrow{PV}^- = \overleftrightarrow{Tr}(13047.00, 12976.00, 12771.00, 12656.00)$, $\overleftrightarrow{PV}^* = \overleftrightarrow{Tr}(40752.14, 41118.95, 41279.45, 41347.45)$. EDFs of portfolios π^+, π^-, π^* are equal $\bar{v}^+ = 0.8886$, $\bar{v}^- = 0.9313$, $\bar{v}^* = 0.9013$. Finally, using (23) and (24) we calculate imprecise EDFs of portfolios π^+, π^-, π^*: $\overleftrightarrow{V}^+ = \overleftrightarrow{Tr}(0.8823, 0.8954, 0.9069, 0.9127)$, $\overleftrightarrow{V}^- = \overleftrightarrow{Tr}(0.9272, 0.9221, 0.9075, 0.8993)$, $\overleftrightarrow{V}^* = \overleftrightarrow{Tr}(0.8957, 0.9034, 0.9071, 0.9087)$. The energy measure of EDF $\overleftrightarrow{V}^*$ of portfolio π^* equals $d(\overleftrightarrow{V}^*) = 0.0083$. Using (25) we get $d(\overleftrightarrow{V}^+) = 0.0210$ and $d(\overleftrightarrow{V}^-) = 0.0213$, which is $d(\overleftrightarrow{V}^*) \le 0.0103$ - the condition (26) is satisfied. We also get $d(\overleftrightarrow{V}^*) = 0.0083 \le 0.0210 \le \min\{d(\overleftrightarrow{V}^+), d(\overleftrightarrow{V}^-)\}$. Similarly, the entropy measure of EDF $\overleftrightarrow{V}^*$ of portfolio π^* is $e(\overleftrightarrow{V}^*) = 0.0023$. Using (25) we get $e(\overleftrightarrow{V}^+) = 0.0047$ and $e(\overleftrightarrow{V}^-) = 0.0033$, meaning that $e(\overleftrightarrow{V}^*) \le \frac{\bar{v}^* \cdot q^+}{\bar{v}^+} e(\overleftrightarrow{V}^+) + \frac{\bar{v}^* \cdot q^-}{\bar{v}^-} e(\overleftrightarrow{V}^-) \le e(\overleftrightarrow{V}^+) + e(\overleftrightarrow{V}^-)$. Moreover, $e(\overleftrightarrow{V}^*) \le \min\{e(\overleftrightarrow{V}^+), e(\overleftrightarrow{V}^-)\}$.

In case of COVID19 lockdown time a portfolio π^* of the same 17 companies is considered to ensure comparability of obtained results. Based on a closing session of March 16, 2020, for each observed share we assess its PV equal to TrOFN $\overleftrightarrow{PV}_S$ describing its Japanese candle [7]. A quoted price $\check{P}_S$ of each portfolio element becomes an initial price on 17.03.2020. For an average expected return rate $\bar{r} = 0.0750$ from (13) we get an expected discount factor $\bar{v} = 0.9302$. For each considered stocks $\hat{S}$ using (15) we get its imprecise expected DF and, basing on (17)-(18), the energy and entropy measure of OEDF $\overleftrightarrow{V}_S$. PVs and OEDFs are presented in Table 2.

Assets	**Present Value $\overleftrightarrow{PV}_s$**	**Quoted Price $\check{P}_s$**	**OEDF $\overleftrightarrow{V}_s$ – energy measure**	**OEDF $\overleftrightarrow{V}_s$ - entropy measure**
Y_{CCC}	$\overleftrightarrow{Tr}(35.00;\ 35.00;\ 29.00;\ 28.36)$	32.98	0.1783	0.0045
Y_{CDR}	$\overleftrightarrow{Tr}(222.00;\ 235.90;\ 252.00;\ 272.00)$	262.00	0.1173	0.0301
Y_{CPS}	$\overleftrightarrow{Tr}(18.71;\ 20.00;\ 21.74;\ 22.40)$	23.50	0.1075	0.0193
Y_{DNP}	$\overleftrightarrow{Tr}(133.20;\ 135.00;\ 136.00;\ 143.20)$	139.80	0.0366	0.0150
Y_{JSW}	$\overleftrightarrow{Tr}(9.51;\ 10.00;\ 11.50;\ 11.64)$	12.00	0.1407	0.0122
Y_{KGH}	$\overleftrightarrow{Tr}(52.78;\ 52.50;\ 51.54;\ 48.01)$	53.00	0.0503	0.0167
Y_{LTS}	$\overleftrightarrow{Tr}(44.90;\ 44.90;\ 44.43;\ 40.62)$	45.46	0.0486	0.0195
Y_{LPP}	$\overleftrightarrow{Tr}(4552.00;\ 4552.00;\ 4450.00;\ 3920.00)$	4450.00	0.0767	0.0277
Y_{OPL}	$\overleftrightarrow{Tr}(4.70;\ 4.92;\ 5.68;\ 5.95)$	5.75	0.1626	0.1986
Y_{PEO}	$\overleftrightarrow{Tr}(61.00;\ 61.00;\ 55.64;\ 54.90)$	57.78	0.0923	0.0030
Y_{PGE}	$\overleftrightarrow{Tr}(2.53;\ 2.81;\ 2.92;\ 3.00)$	3.08	0.0876	0.0272
Y_{PGN}	$\overleftrightarrow{Tr}(2.22;\ 2.31;\ 2.49;\ 2.53)$	2.62	0.0870	0.0115
Y_{PKN}	$\overleftrightarrow{Tr}(42.67;\ 45.69;\ 48.00;\ 49.30)$	49.64	0.0838	0.0202
Y_{PKO}	$\overleftrightarrow{Tr}(23.11;\ 22.94;\ 21.50;\ 20.30)$	22.24	0.0889	0.0143
Y_{PZU}	$\overleftrightarrow{Tr}(23.80;\ 25.91;\ 26.92;\ 26.92)$	27.38	0.0702	0.0179
Y_{SPL}	$\overleftrightarrow{Tr}(180.00;\ 180.00;\ 168.20;\ 157.10)$	170.00	0.0949	0.0152
Y_{TPE}	$\overleftrightarrow{Tr}(0.82;\ 0.89;\ 0.94;\ 0.96)$	0.97	0.0911	0.0216

Table 2 Recorded values of the portfolio π stocks (COVID19 lockdown)

The shares $Y_{CDR}, Y_{CPS}, Y_{DNP}, Y_{JSW}, Y_{OPL}, Y_{PGE}, Y_{PGN}, Y_{PKN}, Y_{PZU}$ and Y_{TPE} belong to portfolio π^+. The shares $Y_{CCC}, Y_{KGH}, Y_{LTS}, Y_{LPP}, Y_{PEO}, Y_{PKO}$, and Y_{SPL} belong to portfolio π^-. The present value $\overleftrightarrow{PV}^+ = \overleftrightarrow{Tr}(13557.00, 14340.55, 15460.00, 16049.00)$, $\overleftrightarrow{PV}^- = \overleftrightarrow{Tr}(10672.54, 10665.20, 10257.22, 9347.68)$, $\overleftrightarrow{PV}^* = \overleftrightarrow{Tr}(24229.54, 25005.75, 25717.22, 25717.22)$. EDFs of portfolios π^+, π^-, π^* are equal $\bar{v}^+ = 0.7766$, $\bar{v}^- = 0.9092$, $\bar{v}^* = 0.8237$. The imprecise EDFs of portfolios are $\overleftrightarrow{V}^+ = \overleftrightarrow{Tr}(0.7846,0.8296,0.8922,0.9250)$, $\overleftrightarrow{V}^- = \overleftrightarrow{Tr}(0.9531,0.9524,0.9158,0.8348)$, $\overleftrightarrow{V}^* = \overleftrightarrow{Tr}(0.8446,0.8733,0.9006,0.9006)$. The energy measure of EDF $\overleftrightarrow{V}^*$ of portfolio π^* equals $d(\overleftrightarrow{V}^*) = 0.0378$. Using (25) we get $d(\overleftrightarrow{V}^+) = 0.1015$ and $d(\overleftrightarrow{V}^-) = 0.0775$ as well as $d\left(\mathcal{C}ore(\overleftrightarrow{V}^-)\right) = 0.0366$, which is $d(\overleftrightarrow{V}^*) \leq 0.05224$ which means that the condition (26) is satisfied. Additionally $d(\overleftrightarrow{V}^*) \leq \min\{d(\overleftrightarrow{V}^+), d(\overleftrightarrow{V}^-)\}$. Similarly, the entropy measure of EDF $\overleftrightarrow{V}^*$ of portfolio π^* is $e(\overleftrightarrow{V}^*) = 0.0053$. On the other hand, using (25) we get $e(\overleftrightarrow{V}^+) = 0.0194$ and $e(\overleftrightarrow{V}^-) = 0.0204$, meaning that $e(\overleftrightarrow{V}^*) \leq \frac{\bar{v}^* \cdot q^+}{\bar{v}^+} e(\overleftrightarrow{V}^+) + \frac{\bar{v}^* \cdot q^-}{\bar{v}^-} e(\overleftrightarrow{V}^-) \leq e(\overleftrightarrow{V}^+) + e(\overleftrightarrow{V}^-)$. Additionally, $\min\{e(\overleftrightarrow{V}^+), e(\overleftrightarrow{V}^-)\}$, which means that the condition (27) is also satisfied.

5 Conclusions

Obtained results allow us to conclude that the portfolio diversification reduces uncertainty and imprecision risk. Moreover, it is worth stressing that different orientations of index profits significantly affect ambiguity and indistinctness of portfolio profit index. If index profits are described by TrOFNs with different orientations then portfolio diversification significantly reduces the ambiguity of portfolio profit index and if the index profits are described by TrOFNs with the same orientations, then portfolio diversification only averages the ambiguity of portfolio profit index [9]. The same is true for indistinctness of portfolio profit index.

Lockdown which was imposed due to COVID19 caused tremendous changes in sessions on worlds' ex- change markets, also in Poland. In most cases the PVs determined by TrOFNs on 16.03.2020 (the beginning of lockdown) are more imprecise. Therefore corresponding energy and entropy measures are higher compared to the quotations of the same companies on 28.01.2020. Similarly, the entropy and energy measures of OEDFs increased. Despite the rise in energy and entropy measures of individual elements, the measures of the portfolio are lower than of the individual increasing and falling elements of given securities. The portfolio diversification reduces uncertainty risk and imprecision risk. Portfolio diversification significantly reduces the ambiguity and indistinctness of portfolio profit index. PV of the portfolio decreased by approximately 40% and as it is estimated by the positively oriented TrOFN, an increase of the portfolio value is expected.

References

[1] Dubois, D. & Prade, H. (1978) Operations on fuzzy numbers. *International Journal of System Science* 9, 613-629.

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

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

[4] Łyczkowska-Hanćkowiak A. (2021) The Use of Trapezoidal Oriented Fuzzy Numbers in Portfolio Analysis. *Symmetry*, 13(9)

[5] Łyczkowska-Hanćkowiak, A. & Piasecki, K. (2018) The Present Value of a Portfolio Of Assets With Present Values Determined by Trapezoidal Ordered Fuzzy Number, *Operations Research and Decisions* 28(2), 41-56.

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

[7] Piasecki, K. & Łyczkowska-Hanćkowiak, A. (2019) Representation of Japanese Candlesticks by Oriented Fuzzy Numbers. *Econometrics* 8(1).

[8] Piasecki, K. & Łyczkowska-Hanćkowiak, A. (2021) Oriented Fuzzy Numbers vs. Fuzzy *Numbers. Mathematics* 9(3), 523.

[9] Piasecki, K. & Łyczkowska-Hanćkowiak, A. (2021) Imprecision Indexes of Oriented Fuzzy Numbers, In K. T. Atanassov et al. (Eds.), International Workshop on Intuitionistic Fuzzy Sets and General Nets - IWIFSGN 2019, *Artificial Intelligence and Soft Computing* 1308, pp. 1–13.

[10] Piasecki, K. & Wójcicka-Wójtowicz, A. (2021). Application of the Oriented Fuzzy Numbers in Credit Risk Assessment. *Mathematics* 9, 5, 535.

Dynamics of the Economic-Ecological System of Aral Sea

Mira Mauleshova[1], Igor Krejčí[2]

Abstract. Paper deals with the implementation of the system dynamics model of the Aral Sea. The proposed model focuses on the impact of the agricultural use of the water and its influence on the environment in the Republic of Uzbekistan. The model explains the causality of agroeconomic activity in the examined area and the enormous decrease of the water in the lake known as the Aral Sea but also the significant deterioration of water's quality in it, namely its salinity.

The paper presents the stock and flow diagram, which describes the basic demographical structure, its connection to the agricultural subsystem and the impact of agriculture on the flow of the river Amu Darya. The construction of the simulation model is described, estimation of the parameters and model validation are discussed. The last part of the paper shows the baseline scenario behaviour of the core variables, such as the volume of water flowing to the Aral Sea, the volume of the lake and salinity. The text is closed with the discussion and model limitations.

Keywords: Aral Sea, Computer simulation, System dynamics

JEL Classification: C44, C63, Q15
AMS Classification: 90B90, 93C15

1 Introduction

The dynamic complexity is a key reason why complex systems surprise us [21], [14]. Investigating the situation of the Aral Sea one should understand that complex systems are characterised by trade-offs. Feedbacks, delays, non-linearity and history dependence often result in a significant difference between short- and long-run results. It is even more complicated by the fact that the short-term benefit is often accompanied by long-term worsening and vice-versa [21]. Besides the confusing character of the complex systems, humans often choose short-term benefits without thinking about the far future.

In the 1930s, cotton production in Middle Asia was supported by the massive construction of irrigation canals. The goal was to double the sown area by the 1960s. Such processes led to the radical decline of the main inflows to the Aral Sea (rivers Amu Darya and Syr Darya). Consequently, the volume in the lake that was called the sea because of its size dried up and the salinity of the remaining water inclined [12]. Nowadays, the Aral Sea is approximately on 10% of its original area and it is split into more separated lakes [5]. The years of drought in the area resulted in the incline in the use of underground water – this effort doesn't represent the long-term solution, it is more the creation of another problem in form of exceeding the limits of sustainability of underground water reservoirs [15].

The paper focuses on the part of the Aral Sea situated in Uzbekistan. The main inflow to this part is the Amu Darya. Despite the Syr Darya flowing through the territory of Uzbekistan, it doesn't inflow to the Uzbek part of the Lake, however, Syr Darya water is also used to satisfy the needs in the Uzbekistan [8], [20]. The agricultural land in Uzbekistan is an area of 5.2 million ha out of which 4.2 million ha are irrigated, 91% of consumed water in Uzbekistan is for agricultural purposes [13], [4]. Uzbekistan is the 4th highest consumer of water per capita in the world with consumption of nearly $2*10^6$ litres per capita [7].

In the paper, we focus on the construction of the system dynamics computer simulation model. The complexity of the described is extremely huge system (with tightly interconnected socio-economic and environmental subsystems) and has consequences in the form of sand storms, health issues of the inhabitants etc. Some of these aspects have been omitted from the actual model and are described in the model boundary table.

2 Material and methods

The modelling approach followed the first four steps of a typical system dynamics process described in [9]. In the first step, the dynamic hypothesis was formulated with the support of the causal loop diagramming [21]. This

[1] CZU Prague, Department of Systems Engineering, Kamycká 129, 165 00 Prague, xmaum002@studenti.czu.cz.
[2] CZU Prague, Department of Systems Engineering, Kamycká 129, 165 00 Prague, krejcii@pef.czu.cz.

diagram identifies the main feedback loops but isn't directly connected with the mathematical model. The second step consists of the transformation of the description into the level and rate equations formalised in the stock and flow diagram. Figure 1 contains the simplified version of the diagram, the overall simulation model consists of 101 symbols (out of which are 5 levels, 45 auxiliaries and constants/parameters 26).

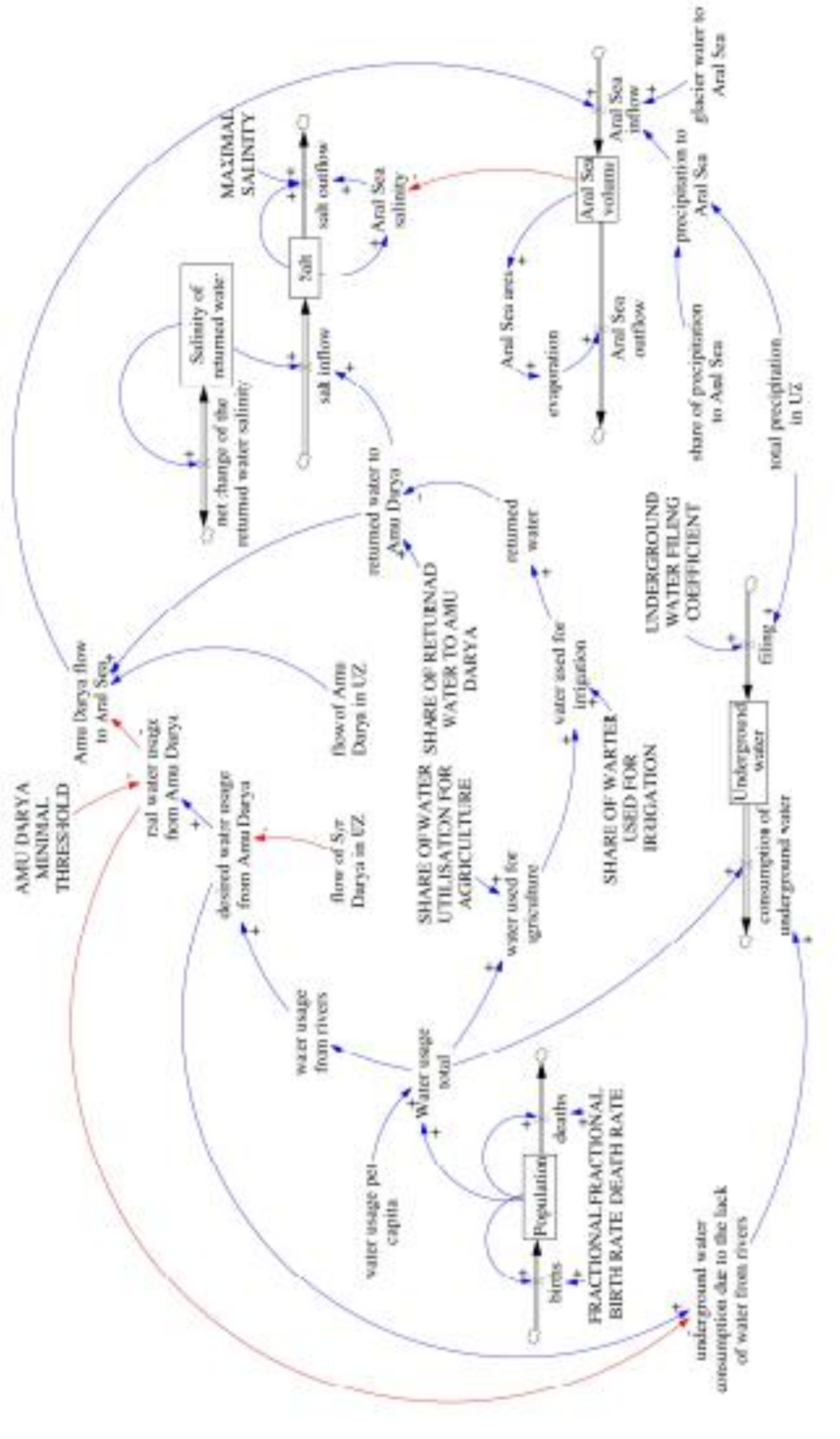


Figure 1 Stock and flow diagram of economic-ecological system of the Aral Sea

Every box variable represents the stock/level, flows/rates are represented by the icon of the pipe with the faucet. From the mathematical perspective, the stock variable is a definite integral [21]:

$$s_T = \int_{T_0}^{T} (i_t - o_t)\mathrm{d}t + s_{T_0}, \tag{1}$$

where s_T is stock, i_t are 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.

A general interpretation of positive causal links' polarity (plus connected to the head of the arrow) is [21]:

$$\frac{\delta y}{\delta x} > 0, \tag{2}$$

and the opposite for the negative link polarity [21]:

$$\frac{\delta y}{\delta x} < 0, \tag{3}$$

where x is the independent (from the perspective of one causal link) variable at the beginning of the arrow and y is the dependent variable at the end of the arrow.

The simulation model was calibrated according to the official statistics. The data on water consumption, including the underground water, are from FAO AQUASTAT core database [8], real water consumption, drainage and returned water is based on data from CAWater [3] supported by [20] to evaluate the water returned to the Amu Darya. Population structure is calibrated to fit the United Nations statistics [22]. Precipitation is again according to the FAO AQUASTAT [8] with the added detail on the precipitation that leads to the Aral Sea [11]. Data on Pamirs glacier melting water flowing to the Aral Sea comes from [2]. For the model calibration, we usually used the time series 1998–2017 if such data existed.

To highlight the most influential endogenous, exogenous and excluded variables [16], table 1 expresses the model boundary. Water usage per capita is used for the analysis of scenarios, therefore, we keep that variable in exogenous form. Despite the weather conditions (especially precipitation and evaporation) represent crucial variables we do not causally close and consider them naturally exogenous as the endogenous form would require a significant model boundary shift and inefficient modelling burden.

Endogenous	**Exogenous**	**Excluded**
Aral Sea volume	precipitation	age cohorts
Aral Sea area	water usage per capita	dynamic life expectancy
salinity of the water	water flow in Amu Darya	salinity of the land
population	water flow in Syr Darya	impact on health
Amu Darya flow to Aral Sea	share of precipitation to Aral Sea	impact on weather
returned water to Amu Darya	iceberg water to Aral Sea	connection of evaporation and precipitation

Table 1 Economic-ecological system of Aral Sea – Model boundary

To calculate the salt in returned water, we used the common structure that generates the S-shape behaviour, see e.g. [21]. To calibrate the missing parameters, we apply the Powell optimisation that minimises the differences between the real and simulated data [6], [17]. Model passes the dimensional analysis test [21].

3 Results and discussion

To evaluate the model, we compared the time series of the core variables with the simulation run for the appropriate years. For the initial validation, we used simple indicators Mean Absolute Percentage Error and R^2 between real and simulated data [21]. For the population, total water consumption, water used for irrigation and water usage from the Amu Darya MAPE is below 2%. Sometimes, higher MAPE is also caused by the high variability of the variable. MAPE for the flow of the Amu Darya is above 20% (mainly due to the years with very small values). However, the R2 between simulated and real data, in this case, is above 0.99.

Figure 2 shows two basic variables flow of Amu Darya to Aral sea and Aral sea volume. The red line of simulated data shows the baseline scenario that will be used (together with the real data) for comparison in future "what if" analysis.

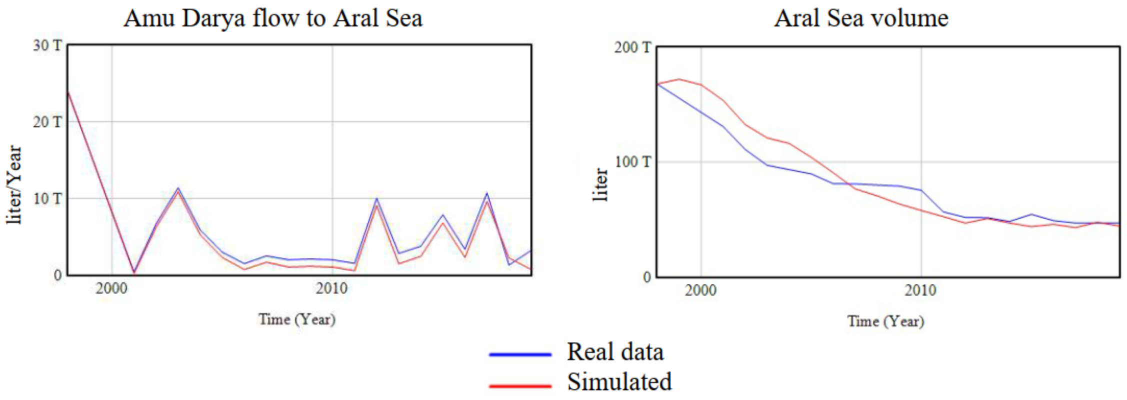


Figure 2 Water in the Amu Darya and in the Aral Sea

Both variables show a significant fall in the past years. Even though the average water usage per capita is decreasing, it can't be considered a positive signal. Most of the water isn't used to satisfy the individual needs of the population but for agricultural purposes. The decline in the usage per capita is caused by the significant growth of the population; the total water consumption was growing in the period used for model validation too.

We have found that the small increase of the Aral Sea volume at the beginning of the simulation could be confusing because of the inflow drop in the corresponding period. The falling inflow doesn't necessarily mean a decrease of the stock variable. Figure 3 compares the evaporation, Amu Darya flow to the Aral Sea and total inflow, which also adds the average inflow from glacier melting and precipitation to the Aral Sea. The stock increases when the total inflow is greater than the total outflow. This explanation could be also applied to the opposite behaviour (increasing inflow and decreasing stock). It is also good to stress that this particular increase in figure 2 is only in the simulated data because of the estimation errors.

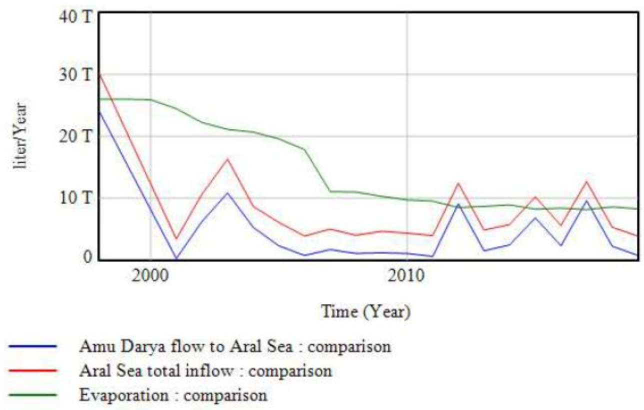


Figure 3 Aral Sea – comparison of flows

To model the accumulation of salt in the Aral Sea, we used a simple balancing loop to model the salt outflow as it is getting harder to increase the salinity – the outflow represents the salt that doesn't remain as a residual in the lake's water. The maximal salinity of the Aral Sea for the model purposes is 236 g/liter, it is not necessarily the real value, the value was estimated by linear regression for year 2020 ($R^2 = 0.9788$, p-value < 0.001). Ratio between the parameter *Maximal Salinity* and variable *Aral Sea salinity* represents the magnitude of the *salt outflow*. For comparison, the salinity of the Dead Sea is above 200 g/liter [1] but the real maximum salinity of water could be much higher depending on the water temperature (more than 350 g/liter for 20 °C).

Figure 4 shows the selected part of sensitivity analysis. The figure shows the sensitivity of Aral Sea salinity subject to change of *Maximal salinity, Share of returned water to Amu Darya*. Both examples show confidence intervals for 2000 simulation runs, ±10% interval, uniform distribution, seed 1234. Maximal salinity showed the highest impact on lake sensitivity. Sensitivity analysis provides various kinds of information [18]. For model testing, it shows whether the small change could result in a significant behaviour change, it helps with the identification of errors. For the policy testing, it helps with the identification of leverage points where the ratio between the effort and the change is efficient.

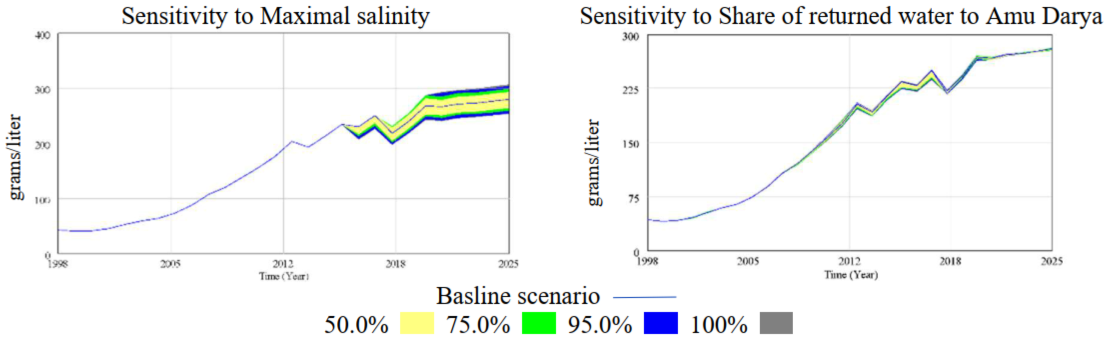


Figure 4 Sensitivity of Aral Sea salinity

The parameter Maximal salinity represents the physical attributes of water. As such, it isn't a matter of the possible policy changes. Despite the high sensitivity, the leverage points must be identified elsewhere.

4 Conclusion

The article describes the basics of the system dynamics model of the economic-ecological system of the Aral Sea. The model describes the current situation and already allows the scenario testing and development estimation, which we will describe in future.

We also plan to focus on the improvement of the model. Currently, the population is modelled by a simple structure. According to the model validation, this is sufficient for the historical data. However, the population in Uzbekistan is rapidly growing. Consequently, the population is very young and the fractional death rate is very small. This will change as the population reaches a stable age structure. Such development could influence the model estimation of future development. We plan to focus on the population and implement the proper ageing chain to improve the estimations.

The Aral Sea problem is well known and reminds the typical failure caused by the lack of system thinking. The growth out of control could result (and often results) in the overshot and collapse scenario [14]. Considering the history of the Aral Sea tragedy, one can identify more than one system archetypes that explain the source of the problematic behaviour – see e.g. Limits to growth, Fixes that fail, Tragedy of commons [19]. The drastic decrease in the volume (together with increasing salinity) caused by human activity significantly affected the biodiversity in the region. Besides the economic implications (such as the end of the commercial fishing industry), the toxic environment causes serious health issues of inhabitants.

The behaviour of the system is given by its structure. Nevertheless, understanding the source of the problem is only the first step. Crucial factor is the response in the proper time – the later is the stage of the system development, the higher are the costs of change [10]. Right now, it is obvious that the response is late and if the correction is possible, it will deserve enormous effort and costs.

Acknowledgements

The research was supported by the project 2021B0003 of the Internal Grant Agency of Faculty of Economics and Management CZU Prague (IGA FEM).

References

[1] Anati, D. A. (1999). The salinity of hypersaline brines: Concepts and misconceptions. *International Journal of Salt Lake Research*, 8, 55–70.

[2] Brun, F., Berthier, E., Wagnon, P., Kääb, A. & Treichler, D. (2017). A spatially resolved estimate of High Mountain Asia glacier mass balances, 2000-2016. *Nature Geoscience*, 10, 668–673.

[3] CAWater-Info, Portal of Knowledge for Water and Environmental Issues in Central Asia (2022). *Database of the Aral Sea*, [online], Available: http://cawater-info.net/aral/data/index_e.htm [6 Feb 2022].

[4] Central Asian Bureau for Analytical Reporting (2020). *Uzbekistan's Water Sector: Environmental and Managerial Issues*, [online], Available: https://cabar.asia/en/uzbekistan-s-water-sector-environmental-and-managerial-issues [26 Feb 2022]

[5] Chen, D.-H. (2018). *The country that brought a sea back to life*, [online], Available: https://www.bbc.com/future/article/20180719-how-kazakhstan-brought-the-aral-sea-back-to-life [4 Feb 2022]

[6] Dangerfield, B. & Roberts, C. (1999). Optimisation as a statistical estimation tool: An example in estimating the AIDS treatment-free incubation period distribution. *System Dynamics Review*, 15, 273–291.

[7] EURASIANET (2022). *Central Asian States Are World's Leading Water Wasters*, [online], Available: https://eurasianet.org/central-asian-states-are-worlds-leading-water-wasters [20 Feb 2022]

[8] Food and Agriculture Organization of the United Nations (2021). *AQUASTAT Core Database. Food and Agriculture Organization of the United Nations*, [online], Available: https://www.fao.org/aquastat/en/databases/maindatabase/ [5 Feb 2022]

[9] Forrester, J. W. (1994). System Dynamics, Systems Thinking, and Soft OR. *System Dynamics Review*, 10, 245–256.

[10] Harrell, C., Ghosh, B. K. & Bowden, R. (2012). *Simulation using ProModel* (3rd edition), New York, NY: McGraw-Hill.

[11] Krapivin, V., Mkrtchyan, F. & Rochon, G. (2019). Hydrological Model for Sustainable Development in the Aral Sea Region. *Hydrology*, 6, 91.

[12] Kulturologia.ru. (2021). *Как в СССР загубили Аральское море и зачем хотели повернуть вспять сибирские реки*, [online], Available: https://kulturologia.ru/blogs/210921/51138/ [6 Feb 2022]

[13] 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.

[14] Meadows, D. H. (2008). *Thinking in systems: a primer*. Wright, D. (Ed.) White River Junction, Vt.: Chelsea Green Pub.

[15] Micklin, P. (2010). The past, present, and future Aral Sea. Lakes & Reservoirs: *Research and Management*. 15, 193–213.

[16] Pierson, K. & Sterman, J. D. (2013). Cyclical dynamics of airline industry earnings. *System Dynamics Review*, 29, 129–156.

[17] 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.

[18] Pruyt, E. (2016). *Systems Modelling and Simulation for Everyone. Migration & Refugeess*. Delft: Delft University of Technology, Balaton group.

[19] Senge, P. M. (2006). *The fifth discipline: The art and practice of the learning organization*, New York: Doubleday/Currency.

[20] Shmakova, U. (2018). *Национальный отчет по управлению возвратными водами в республике Узбекистан*, [online], Available: http://riverbp.net/Отчет+по+управлению+возвратными+водами+в+Республике+Узбек истан.pdf [4 Feb 2022]

[21] Sterman, J. D. (2000). *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Boston: Irwin/McGraw-Hill.

[22] United Nations (2019). *World population prospects*. [online]. Available: https://population.un.org/wpp/ [8 Jan 2022]

On the Ordering of Cycles in the Threshold Digraphs of Concave Monge Matrices in Max-Min Algebra with Respect to Robustness

Monika Molnárová[1]

Abstract. Threshold digraphs with respect to strongly connected components identification and robustness of concave Monge matrices over max-min algebra are studied. Properties of matrices in max-min such as periodicity and robustness are investigated using the corresponding so-called threshold digraphs. Therefore, characterization of their structure is crucial for further research. The nodes of two cycles from different non-trivial strongly connected components satisfy an ordering. Consequently, the nodes of a strongly connected component are numbered by a sequence of consecutive natural numbers and the threshold digraph of a concave Monge max- min matrix has a block form for each threshold. A more efficient algorithm for finding the strongly connected components of the threshold digraphs of a concave max-min Monge matrix is presented. The robustness of considered class of matrices with fixed data is proved.

Keywords: : max-min algebra, concave Monge matrix, robustness, threshold digraph

JEL Classification: C02
AMS Classification: 08A72, 90B35, 90C47

1 Introduction

The extremal algebras are used in many optimization problems regarding discrete dynamic systems (DDS), graph theory, knowledge engineering, cluster analysis, fuzzy logic programs or business, where maximum and minimum operations are involved ([3], [11], [12]). Most frequent are the max-min algebra with operations maximum and minimum and max-plus algebra with operations maximum and plus. For special classes of matrices, such as Toeplitz, Circulant or Monge matrices, more efficient algorithms can be derived using the concept of extremal algebras ([2], [8], [9]). The Monge matrices their structural properties and algorithms solving many problems related to Monge matrices were studied in [1], [4], [5] and [7].

The aim of this paper is to present results related to the structure of the threshold digraphs of concave Monge matrices over max-min algebra including an effective algorithm for finding all strongly connected components throughout all threshold digraphs corresponding to a given matrix and to study stability of discrete event systems represented by concave Monge matrices, i.e. the robustness of the considered type of matrix 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 period and the property of robustness of a max-min matrix are defined and formula for computing the matrix period and necessary and sufficient condition for robustness in general case of a max-min matrix are recalled. Finally, the notion of a concave Monge matrix is introduced. In Section 3, theorems concerning ordering of cycles of two different strongly connected components and implying the block form of a concave Monge matrix are presented. In Section 4, an efficient algorithm for computing the strongly connected components of a threshold digraph is presented and the robustness of concave Monge matrices is proved in Theorem 7.

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., $N = \{1, 2, \ldots, n\}$.

[1] Technical University of Košice, Department of Mathematics and Theoretical Informatics, B. Něemcovej 32, 04200 Košice, Slovakia, Monika.Molnarova@tuke.sk.

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.

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.

Let $A \in B(n,n)$ and $x \in B(n)$. The sequence $O(A,x) = \{x^{(0)}, x^{(1)}, x^{(2)}, \dots, x^{(n)}, \dots\}$ is the orbit of $x = x^{(0)}$ generated by A, where $x^{(r)} = A^r \otimes x^{(0)}$ for each $r \in \mathbb{N}$.
For a given matrix $A \in B(n,n)$, the number $\lambda \in B$ and the n–tuple $x \in B(n)$ are the so-called *eigenvalue* of A and *eigenvector* of A, respectively, if they satisfy the equation $A \otimes x = \lambda \otimes x$. We define the corresponding *eigenspace* $V(A,\lambda)$ as the set $V(A,\lambda) = \{x \in B(n);\ A \otimes x = \lambda \otimes x\}$.
Let $\lambda \in B$. A matrix $A \in B(n,n)$ is *ultimately λ–periodic* if there are natural numbers p and R such that the following holds: $A^{k+p} = \lambda \otimes A^k$ for all $k \geq R$. The smallest natural number p with above property is called the period of A, denoted by $\operatorname{per}(A,\lambda)$. In case $\lambda = I$ we denote $\operatorname{per}(A,I)$ by abbreviation $\operatorname{per} A$.

According to [2] we define

$$\mathrm{SCC}^\star(A) = \cup\{\,\mathrm{SCC}^\star(G(A,h));\ h \in \{a_{ij}; i,j \in N\}\,\} \tag{1}$$

Theorem 1. [2] *Let $A \in B(n,n)$. Then*

$$\operatorname{per} A = \operatorname{lcm}\{\operatorname{per}\mathcal{K};\ \mathcal{K} \in \mathrm{SCC}^\star(A)\}. \tag{2}$$

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*.

We have used the following result (adapted for $\lambda = I$) proved in [10] to study robustness of a concave Monge matrix.

Lemma 1. [10] Let $A = (a_{ij}) \in B(n,n)$. Then A is robust if and only if $\operatorname{per} A = 1$.

Definition 2. We say, that a matrix $A = (a_{ij}) \in B(m,n)$ is a concave Monge matrix if and only if

$$a_{ij} \otimes a_{kl} \geq a_{il} \otimes a_{kj} \qquad \text{for all} \quad i < k,\ j < l\,. \tag{3}$$

Remark 1. Obviously, it is enough to consider thresholds $h \in H = \{a_{ij};\ i,j \in N\}$ to get all threshold digraphs corresponding to the matrix A.

Let us denote by $h^{(1)}, h^{(2)}, \dots, h^{(r)}$ the elements of the set $H = \{a_{ij};\ i,j \in N\}$ ordered into a strictly decreasing sequence, i.e.,

$$h^{(1)} > h^{(2)} > \dots > h^{(r)}. \tag{4}$$

The number r is equal to the number of different inputs of the matrix A.

Lemma 2. [8] Let $A \in B(n,n)$. Then the sequence of threshold digraphs corresponding to the sequence (4) is ordered by inclusion

$$G(A,h^{(1)}) \subseteq G(A,h^{(2)}) \subseteq \ldots \subseteq G(A,h^{(r)}).$$

Remark 2. The threshold digraph $G(A,h)$ for $h = \min H = \min\{a_{ij};\ i,j \in N\}$ equals to $G(A)$, i.e. to the digraph associated with the given matrix A, which is a complete, edge-weighted digraph.

3 Ordering of cycles in threshold digraphs of concave Monge matrices

In this section we present results crucial for study the properties of threshold digraphs of concave Monge matrices. The nodes of two cycles from different non-trivial strongly connected components satisfy an ordering. Namely, the nodes of the cycle of one non-trivial strongly connected component are numbered by a sequence of natural numbers less than the nodes of the cycle of the other non-trivial strongly connected component. The generalization of the theorem for two different strongly connected components, with an useful consequence for deriving a more effective algorithm for finding the strongly connected components of the threshold digraphs of a concave Monge matrix, is presented. Namely, a concave Monge matrix has a block form.

Theorem 2. [7] *Let $A \in B(n,n)$ be a concave Monge matrix. Let $c_1 = (i_0, i_1, \dots, i_k)$ with $i_0 = i_k$ and $c_2 = (j_0, j_1, \dots, j_l)$ with $j_0 = j_l$ be cycles in different non-trivial strongly connected components in $G(A,h)$ for $h \in H$. Let $i_s = \min\{i_0, i_1, \dots, i_{k-1}\}$ and $i_t = \max\{i_0, i_1, \dots, i_{k-1}\}$. Then exactly one of the conditions holds*
(i) $j_m < i_s$ for all $m \in \{0, 1, \dots, l\}$,
(ii) $j_m > i_t$ for all $m \in \{0, 1, \dots, l\}$.

Using the fact that all nodes of a non-trivial strongly connected component lie on a common cycle, we can generalize the theorem.

Theorem 3. [6] *Let $A \in B(n,n)$ be a concave Monge matrix. Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be two different non-trivial strongly connected components of $G(A,h)$ for $h \in H$ generated by the node set $N_{\mathcal{K}_1} = \{i_1, \dots, i_k\}$ and $N_{\mathcal{K}_2} = \{j_1, \dots, j_l\}$, respectively. Let $i_s = \min\{i_1, i_2, \dots, i_k\}$ and $i_t = \max\{i_1, i_2, \dots, i_k\}$. Then exactly one of the conditions holds*
(i) $j_m < i_s$ for all $m \in \{1, 2, \dots, l\}$,
(ii) $j_m > i_t$ for all $m \in \{1, 2, \dots, l\}$.

The above theorem has an useful consequence formulated in the following corollary, which helps to find the nodes of a strongly connected component of a threshold digraph by checking the consecutive nodes of the digraph.

Corollary 1. [6] Let $A \in B(n,n)$ be a concave Monge matrix. Than A has a block form in which the diagonal blocks represent the strongly connected components of $G(A,h)$ for $h \in H$.

4 Structure of threshold digraphs of concave Monge matrices with respect to robustness

In this section we present an overview of results related to the structure of threshold digraphs of concave Monge matrices we have obtained. Namely, there is a cycle of length two for each pair of consecutive nodes in a non-trivial strongly connected component. In addition there is a loop on every node of a non-trivial strongly connected component. These results on one hand allow to find all non-trivial strongly connected components of $G(A,h)$ for all $h \in H$ by a more effective algorithm for concave Monge matrices with computational complexity $O(n^3)$, on other hand are useful to prove the robustness of a concave Monge matrix with fixed data over max-min algebra.

Theorem 4. [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 2. [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 $\mathcal{K}$.

The following Corollary helps to distinguish between a trivial and a non-trivial strongly connected component in case, when the component is generated by one node.

Corollary 3. [6] Let $A \in B(n,n)$ be a concave Monge matrix. Element $a_{kk} < h$ represents a trivial strongly connected component of $G(A,h)$.

Theorem 5. [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 $(i, i+1, i)$.*

We can use the obtained results to derive an algorithm for finding all strongly connected components of the threshold digraph $G(A,h)$ for a threshold $h \in H$ with computational complexity $O(n)$. Since by Theorem 2 the nodes of a cycle of a non-trivial strongly connected component are numbered by a sequence of natural numbers less than the nodes of a cycle of another non-trivial strongly connected component and by the fact that all nodes of a non-trivial strongly connected component lie on a common cycle, the same holds by Theorem 3 for all nodes of two different non-trivial strongly connected components. Namely, all nodes of a non-trivial strongly connected component are numbered by a sequence of natural numbers less than the nodes of the other non-trivial strongly connected component. This property of a concave Monge matrix was applied in the proof of Theorem 5, which together with Corollary 2 and Corollary 3 essentially decreases the computational complexity of the algorithm **Strongly Connected Components** (see bellow). It is namely enough for completing the node set of a strongly connected component to check the consecutive nodes whether there is a common cycle of length two. To distinguish between a trivial and a non-trivial component in a threshold digraph we check whether the corresponding diagonal element a_{ii} is less than the value of the threshold h.

Due to Remark 1 and the fact that the number of inputs of a square matrix is n^2 is the number of threshold digraphs bounded by n^2.

Theorem 6. [6] *There is an algorithm with computational complexity $O(n^3)$ for computing the strongly connected components of $G(A,h)$ for all $h \in H$.*

Now, we present the algorithm for finding all strongly connected components of a threshold digraph $G(A,h)$ for given concave Monge matrix A and threshold $h \in H$.

Algorithm **Strongly Connected Components**

Input. A, h.
Output. '$\mathcal{K}_i$ with the node set $N_{\mathcal{K}_i}$ trivial' if $a_{ii} < h$ or '$\mathcal{K}_i$ with the node set $N_{\mathcal{K}_i}$ non-trivial' if $a_{ii} \geq h$.

begin

Step 1 $i := 0$;

Step 2 $N_{\mathcal{K}_{i+1}} := \{i+1\}$;

Step 3 $i := i + 1$;

Step 4 **If** $i = n$ **then if** $a_{ii} \geq h$ **then** '$\mathcal{K}_i$ with the node set $N_{\mathcal{K}_i}$ non-trivial'; **go to** end; **else** '$\mathcal{K}_i$ with the node set $N_{\mathcal{K}_i}$ trivial'; **go to** end;

Step 5 **If** $a_{ii+1} \geq h$ **and** $a_{i+1i} \geq h$ **then** add $i + 1$ to $N_{\mathcal{K}_i}$; **go to** Step 3;

Step 6 **If** $a_{ii} \geq h$ **then** '$\mathcal{K}_i$ with the node set $N_{\mathcal{K}_i}$ non-trivial'; **go to** Step 2; **else** '$\mathcal{K}_i$ with the node set $N_{\mathcal{K}_i}$ trivial'; **go to** Step 2;

end

Using obtained results we can answer the question whether the discrete event system represented by a square concave Monge matrix $A \in B(n,n)$ reaches always the steady state starting with arbitrary initial n–tuple $x \in B(n)$. In other words, we prove the robustness of a concave Monge matrix over max-min algebra.

Theorem 7. *Let $A \in B(n,n)$ be a concave Monge matrix. Then A is robust.*

Proof. The assertion follows by Theorem 1, by Theorem 4 and by Lemma 1. □

We can illustrate the algorithm **Strongly Connected Components** and the property of robustness in the following example.

Example 1. Let $A \in B(8,8)$ for $B = [0,3]$ be a concave Monge matrix of the form

$$A = \begin{pmatrix} 1 & 3 & 2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 3 & 3 & 0 & 0 & 0 & 0 & 0 \\ 1 & 3 & 3 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 3 & 2 & 2 & 1 & 0 \\ 0 & 1 & 1 & 2 & 2 & 2 & 1 & 0 \\ 0 & 0 & 1 & 2 & 2 & 2 & 1 & 0 \\ 0 & 0 & 0 & 1 & 2 & 2 & 3 & 3 \\ 0 & 0 & 0 & 1 & 2 & 2 & 3 & 3 \end{pmatrix}.$$

Due to Remark 1 all inputs of the given matrix define the set of all thresholds $H = \{a_{ij};\, i, j \in N\} = \{0, 1, 2, 3\}$. Let us check the algorithm for the threshold $h = 3$, i.e. let us find the strongly connected components of the digraph $G(A, 3)$.

We start with the node $i = 1$ and by Step 2 we set the initial node set $N_{\mathcal{K}_1} = \{1\}$ of the first strongly connected component $\mathcal{K}_1$. By Step 5 we verify the existence of the cycle of length two for nodes 1 and 2. Since a_{21} is less than the threshold $h = 3$ is the node set of the first component complete. By Step 6 we verify the existence of the loop on node $i = 1$ and decide whether this component is trivial or non-trivial. Since a_{11} is less than the threshold, is the first component trivial. We proceed to Step 2 by setting the node set $N_{\mathcal{K}_2} = \{2\}$ of the second strongly connected component $\mathcal{K}_2$. Since both values a_{23} and a_{32} in Step 5 are equal to $h = 3$ we extend the node set by adding the node $i + 1 = 3$ to the node set $N_{\mathcal{K}_2}$. Hence, the second component is non-trivial. Thus, the values a_{22} and a_{33} are more or equal to $h = 3$ and there is a loop on both nodes in $G(A, 3)$. By checking the existence of the cycle of length two for next pair of consecutive nodes, namely, nodes 3 and 4 in Step 5, we determine that there is no common cycle. Hence, the node set of the second component $\mathcal{K}_2$ is complete with $N_{\mathcal{K}_2} = \{2, 3\}$. We proceed to Step 2 by setting the node set $N_{\mathcal{K}_4} = \{4\}$ of the next strongly connected component $\mathcal{K}_4$. By Step 5 we verify the existence of the cycle of length two for nodes 4 and 5. Since a_{45} is less than the threshold $h = 3$ is the node set of the component complete. Moreover, in contrast to the first component, the corresponding diagonal element a_{44} is equal to $h = 3$ in Step 6. Consequently there is a loop on the node $i = 4$. Hence, the strongly connected component $\mathcal{K}_4$ is non-trivial. Similarly to the first component finds the algorithm following two trivial components $\mathcal{K}_5$ for node $i = 5$ and $\mathcal{K}_6$ for node $i = 6$. The last component $\mathcal{K}_7$ with node set $N_{\mathcal{K}_7} = \{7, 8\}$ is similarly to the second component non-trivial. Completing the node set of the last component the algorithm ends in Step 4.

To find the strongly connected components for all threshold digraphs (see Figure 1) it is enough due to Remark 2 to apply the algorithm **Strongly Connected Components** gradually for values $h = 1, 2, 3$.

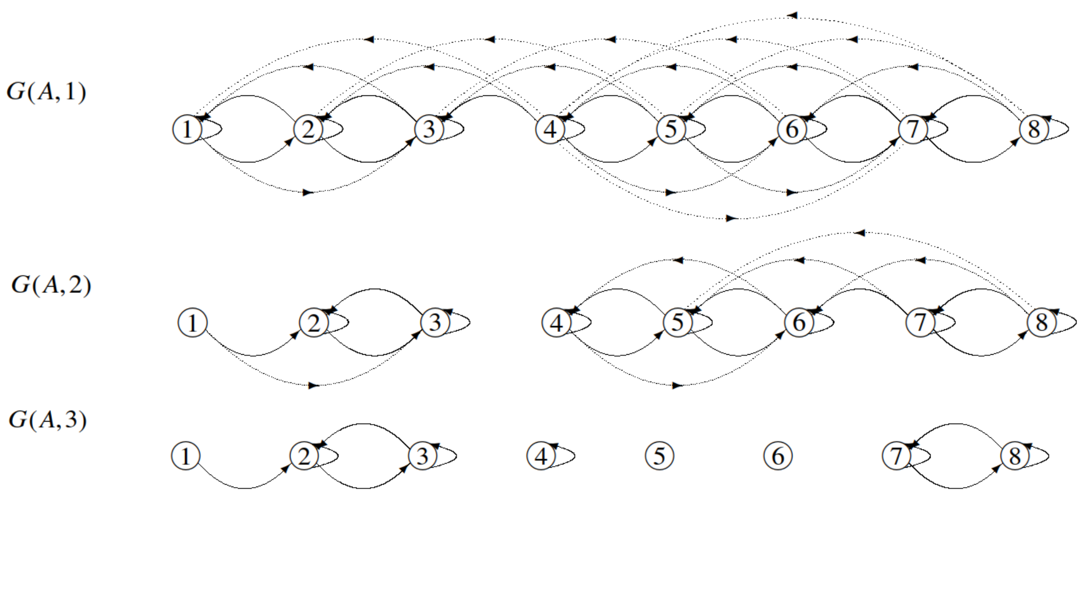


Figure 1 Strongly connected components in threshold digraphs of a concave Monge matrix

Due to Corollary 1 the matrix A has a block form in which the diagonal blocks represent the strongly connected components of $G(A,h)$ for $h \in H = \{a_{ij};\ i,j \in N\} = \{0,1,2,3\}$. Due to Remark 2 is $G(A,0)$ strongly connected. Thus, $G(A,0)$ consists of one non-trivial strongly connected component. The corresponding blocks for $h = 1$ are bounded by single lines in A, for $h = 2$ by single and double lines in A and for $h = 3$ by single, double and triple lines in A

$$A = \left(\begin{array}{ccc|ccccc} 1 & 3 & 2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 3 & 3 & 0 & 0 & 0 & 0 & 0 \\ 1 & 3 & 3 & 0 & 0 & 0 & 0 & 0 \\ \hline 1 & 1 & 1 & 3 & 2 & 2 & 1 & 0 \\ 0 & 1 & 1 & 2 & 2 & 2 & 1 & 0 \\ 0 & 0 & 1 & 2 & 2 & 2 & 1 & 0 \\ 0 & 0 & 0 & 1 & 2 & 2 & 3 & 3 \\ 0 & 0 & 0 & 1 & 2 & 2 & 3 & 3 \end{array}\right) = \left(\begin{array}{ccc|ccc||cc} 1 & 3 & 2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 3 & 3 & 0 & 0 & 0 & 0 & 0 \\ 1 & 3 & 3 & 0 & 0 & 0 & 0 & 0 \\ \hline 1 & 1 & 1 & 3 & 2 & 2 & 1 & 0 \\ 0 & 1 & 1 & 2 & 2 & 2 & 1 & 0 \\ 0 & 0 & 1 & 2 & 2 & 2 & 1 & 0 \\ \hline\hline 0 & 0 & 0 & 1 & 2 & 2 & 3 & 3 \\ 0 & 0 & 0 & 1 & 2 & 2 & 3 & 3 \end{array}\right) = \left(\begin{array}{c|||cc|c||c||c||cc} 1 & 3 & 2 & 0 & 0 & 0 & 0 & 0 \\ \hline\hline\hline 1 & 3 & 3 & 0 & 0 & 0 & 0 & 0 \\ 1 & 3 & 3 & 0 & 0 & 0 & 0 & 0 \\ \hline 1 & 1 & 1 & 3 & 2 & 2 & 1 & 0 \\ \hline\hline\hline 0 & 1 & 1 & 2 & 2 & 2 & 1 & 0 \\ \hline\hline\hline 0 & 0 & 1 & 2 & 2 & 2 & 1 & 0 \\ \hline\hline\hline 0 & 0 & 0 & 1 & 2 & 2 & 3 & 3 \\ 0 & 0 & 0 & 1 & 2 & 2 & 3 & 3 \end{array}\right).$$

Obviously, contains every non-trivial strongly connected component throughout all threshold digraphs loops on the nodes. Hence, by formula (2) for computing the matrix period is the period per A of the considered matrix equal to 1. Consequently by Lemma 1 is the concave Monge matrix robust. In other words the result follows by Theorem 7.

5 Conclusion

We have studied structure of threshold digraphs of concave Monge matrices over max-min algebra in this paper. Obtained results were used to prove robustness of a concave Monge matrix and to derive a more effective algorithm for finding all strongly connected components of the threshold digraphs of the considered class of matrices with fixed data. Moreover, diverse types of robustness of Monge matrices and diverse types of eigenvectors ([3]) with fixed data as well as inexact data can be investigated and more efficient algorithms can be derived using obtained results.

The matrix robustness is connected with the problem of finding the steady state of a discrete event system represented by the given matrix, i.e. finding an eigenvector $x(r)$ of the eigenproblem $A \otimes x(r) = x(r)$. Solving eigenproblem of a max-min matrix can present the following economic application ([3]). A company should evaluate diverse projects characterized by different properties. The value x_i is influenced by all values x_j (the influence is expressed by the inputs a_{ij} of the matrix A) and describes the level of each property i.

References

[1] Burkard, R. E., Klinz, B. & Rudolf, R.: Perspectives of Monge properties in optimization, *DAM*, Volume 70 (1996), 95–161.

[2] Gavalec, M.: *Periodicity in extremal algebras*. GAUDEAMUS, Hradec Králové, 2004.

[3] Gavalec, M., Ramík, J. & Zimmermann, K.: *Decision making and optimization.* Lecture Notes in Economics and Mathematical Systems 677, Springer, 2015.

[4] Hireš, M., Molnárová, M. & Drotár, P.: Robustness of Interval Monge Matrices in Fuzzy Algebra, *Mathematics,* Volume 4 (2020), 1–16.

[5] Molnárová, M.: Periodicity of convex and concave Monge matrices in max-min algebra, In: *Proceedings of 38th Int. Conference Mathematical Methods in Economics 2020*, Brno, 2020, 377–382.

[6] Molnárová, M.: Structure of the threshold digraphs of convex and concave Monge matrices in max-min algebra, In: *Proceedings of 39th Int. Conference Mathematical Methods in Economics 2021*, Praha, 2021, 331–336.

[7] Molnárová, M.: *Convex and concave Monge matrices in fuzzy algebra - comparison with respect to robustness,* (in preparation).

[8] Molnárová, M., Myšková, H. & Plavka, J.: The robustness of interval fuzzy matrices, *DAM*, Volume 438 (2013), 3350–3364.

[9] Myšková, H. & Plavka, J.: X-robustness of interval circulant matrices in fuzzy algebra, *Linear Algebra and its Applications*, Volume 438 (2013), 2757–2769.

[10] Plavka, J. & Szabó, P.: On the λ-robustness of matrices over fuzzy algebra, *DAM*, Volume 159 Issue 5 (2011), 381–388.
[11] De Schutter, B., van den Boom, T., Xu, J. & Farahani, S.S.: Analysis and control of max-plus linear discreteevent systems: *An introduction. Discrete Event Dyn. Syst*., Volume 30 (2020), 25–54.
[12] Zimmermann, H.J.: *Fuzzy Set Theory And Its Applications. Springer Science and Business Media*, Berlin, Germany, 2011.

Weak Solvability of Interval Max-Min Matrix Equations

Helena Myšková[1]

Abstract. Behavior of discrete event systems, in which the individual components move from event to event rather than varying continuously through time, is often described by systems of linear equations or by matrix equations in max-min algebra, 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 be- tween input and output variables. Many practical situations can be described using max-min matrix equations. It often happens that a max-min matrix equation with exact data is unsolvable. If we replace matrix elements with intervals of possible values, we obtain an interval matrix equation. Several types of solvability of interval max-min matrix equation have been studied yet. In this paper, we shall deal with the weak solvability of interval max-min matrix equations. We provide a procedure for checking the weak solvability.

Keywords: max-plus algebra, interval matrix, matrix equation, weak solvability

JEL Classification: C02
AMS Classification: 15A18; 15A80; 65G30

1 Motivation

Fuzzy (max-min) equations have found a broad area of applications in causal models which emphasize relationships between input and output variables. They are used in diagnosis models [8], [9] or models of nondeterministic systems [10]. Diagnostic models are particularly important because they cope with the uncertainty in many real-life situations concerning either medical diagnoses or diagnoses of technical devices. In the simplest formulation we are faced with a space of symptoms and a space of faults. Elements of faults are related with elements of symptoms by means of a fuzzy relation. Usually, the stronger the relationship between the symptom and a fault, the higher is the value of the corresponding argument. The solution of the max-min equation of the form $A \otimes x = b$, where A is a matrix, b and x are vectors of suitable dimensions and classical addition and multiplication operations are replaced by maximum and minimum, provides a maximal set of symptoms that produce the given fault. Publications [2] and [3] can serve as examples of the use of max-min algebra in operational research.

The solvability of the systems of fuzzy linear equations is well reviewed. In this paper, we shall deal with the solvability of fuzzy matrix equations of the form $A \otimes X = B$, where A and B are given matrices of suitable sizes and X is an unknown matrix.

In this paper, we shall deal with interval max-min matrix equations of the form $\boldsymbol{A} \otimes X = \boldsymbol{B}$, where $\boldsymbol{A}$, $\boldsymbol{B}$, are given interval matrices of suitable sizes a X is an unknown matrix. Several solvability concepts have been studied in [5], [7]. The following example is the shortened version of an example given in [5].

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 $D_1, D_2, \ldots, D_r$. Suppose that the air transport to the destination D_k is provided by the airport terminal T_k. Different transportation means provide transporting passengers from places $P_1, P_2, \ldots, P_m$ to airport terminals $T_1, T_2, \ldots, T_r$. We assume that the connection between P_i ($i \in N,\ N = \{1, 2, \ldots, n\}$) and T_k ($k \in R,\ R = \{1, 2, \ldots, r\}$) is possible only via one of the check points $Q_1, Q_2, \ldots, Q_n$.

Denote by a_{ij} the capacity of the connection from P_i to Q_j. If there is no connection from P_i to Q_j we put $a_{ij} = 0$. If the capacity of the connection from Q_j to T_k is x_{jk}, then the capacity of the connection from P_i to T_k is equal to $\min\{a_{ij}, x_{jk}\}$. Suppose that there are b_{ik} passengers traveling from place P_i to destination D_k. Our task is to choose the appropriate capacities $x_{jk},\ j \in N, l \in R$ such that all passengers from place P_i can be transported to the terminal T_k is equal to a given number b_{ik} for all $i \in M$ and for all $k \in R$, i.e.,

$$\max_{j \in N} \min\{a_{ij}, x_{jk}\} = b_{ik}. \tag{1}$$

[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

In the following, we will show how (1) can be solved in max-min algebra.

2 Preliminaries

Max-min algebra is the triple $(\mathcal{I}, \oplus, \otimes)$, where $\mathcal{I}$ is an linearly ordered set with the least element O and the greatest element I and two binary operations defined as follows:

$$a \oplus b = \max\{a, b\} \text{ and } a \otimes b = \min\{a, b\}.$$

The set of all $m \times n$ *matrices* over $\mathcal{I}$ is denoted by $\mathcal{I}(m, n)$ and the set of all column n-*vectors* over $\mathcal{I}$ by $\mathcal{I}(n)$. For given natural numbers m, n, r denote the sets of indices $M = \{1, 2, \ldots, m\}$, $N = \{1, 2, \ldots, n\}$ and $R = \{1, 2, \ldots, r\}$.

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 $A \otimes B \leq C \otimes D$.

For given matrix $A \in \mathcal{I}(m, n)$ and vector $b \in \mathcal{I}(m)$ a system of linear max-min equation can be written in the form

$$A \otimes x = b. \tag{2}$$

The vector $x^*(A, b)$, defined as follows, plays an important role in deciding about solvability of (2):

$$x^*_j(A, b) = \min_{i \in M}\{b_i : a_{ij} > b_i\} \tag{3}$$

for any $j \in N$, whereby $\min \emptyset = I$. The vector $x^*(A, b)$ is called a principal solution of (2).

Theorem 1. [4, 11] *Let $A \in \mathcal{I}(m, n)$ and $b \in \mathcal{I}(m)$ be given. The system $A \otimes x = b$ is solvable if and only if $x^*(A, b)$ is its solution.*

Lemma 2. [6] Let $b \in \mathcal{I}(m)$ and $C, D \in \mathcal{I}(m, n)$ be such that $D \leq C$. Then $x^*(C, b) \leq x^*(D, b)$.

Le us replace a vector b in (2) by a matrix $B \in \mathcal{I}(m, r)$ and an unknown vector x by an unknown matrix $X \in \mathcal{I}(n, r)$. We get the matrix equation of the form

$$A \otimes X = B. \tag{4}$$

It is easy to see that (4) is equivalent to the following r equations:

$$A \otimes X_k = B_k$$

for each $k \in R$, where X_k (B_k) is a k-th column of X (B).

Consequently the principal solution of (4) is the matrix $X^*(A, B) \in \mathcal{I}(n, r)$ with elements

$$x^*_{jk}(A, B) = x^*_j(A, B_k). \tag{5}$$

Theorem 3. [5] *Let $A \in \mathcal{I}(m, n)$ and $B \in \mathcal{I}(m, r)$ be given. The matrix equation $A \otimes X = B$ is solvable if and only if the matrix $X^*(A, B)$ is its solution.*

Lemma 4. [5] Let A, $A^{(1)}$, $A^{(2)} \in \mathcal{I}(m, n)$ and B, $B^{(1)}$, $B^{(2)} \in \mathcal{I}(m, r)$. The following assertions hold:

(i) If $A^{(2)} \leq A^{(1)}$ then $X^*(A^{(1)}, B) \leq X^*(A^{(2)}, B)$.

(ii) If $B^{(1)} \leq B^{(2)}$ then $X^*(A, B^{(1)}) \leq X^*(A, B^{(2)})$.

3 Interval case

A certain disadvantage of the necessary and sufficient condition for the solvability of (4) given in Theorem 3 (iii) stems from the fact that it only indicates the existence or non-existence of the solution but does not indicate any action to be taken to increase the degree of solvability. However, it happens quite often in modelling real situations that the obtained system turns out to be unsolvable.

One of possible methods of restoring the solvability is to replace the exact input values by intervals of possible values.

3.1 Weak solvability of interval systems of linear equations

Similarly to [5, 6] we define an *interval matrix* $\boldsymbol{A}$ and *interval vector* $\boldsymbol{b}$ as follows:

$$\boldsymbol{A} = [\underline{A}, \overline{A}] = \left\{ A \in \mathcal{I}(m,n);\ \underline{A} \leq A \leq \overline{A} \right\}, \quad \boldsymbol{b} = [\underline{b}, \overline{b}] = \left\{ b \in \mathcal{I}(m);\ \underline{b} \leq b \leq \overline{b} \right\},$$

where $\underline{A}, \overline{A} \in \mathcal{I}(m,n)$, $\underline{A} \leq \overline{A}$ and $\underline{b}, \overline{b} \in \mathcal{I}(m)$, $\underline{b} \leq \overline{b}$.

Denote by

$$\boldsymbol{A} \otimes x = \boldsymbol{b} \tag{6}$$

the set of all of max-min systems of linear equations of the form (2) such that $A \in \boldsymbol{A}$, $b \in \boldsymbol{b}$. We shall call (6) a *max-min interval system of linear equations.*

Definition 1. Interval system (6) is *weakly solvable* if there exist $b \in \boldsymbol{b}$ and $A \in \boldsymbol{A}$ such that $A \otimes x = b$ is solvable.

Lemma 5. Interval system (6) is weakly solvable if there exists $A \in \boldsymbol{A}$ such that

$$A \otimes x^*(A, \overline{b}) \geq \underline{b}, \tag{7}$$

Theorem 6. [1] *Interval system (6) is weakly solvable if and only if*

$$\overline{A} \otimes x^*(\underline{A}, \overline{b}) \geq \underline{b}. \tag{8}$$

This theorem gives answer whether a given interval system is weakly solvable or not, but it does not provide an algorithm for finding matrix A and vector b.

For a fixed $b \in \boldsymbol{b}$, let us define the canonical matrix $A(b)$ as follows:

$$a_{ij}(b) = \begin{cases} \overline{a}_{ij} & \text{if } x_j^*(\underline{A}, b) \leq b_i; \\ \min\{b_i, \overline{a}_{ij}\} & \text{if } x_j^*(\underline{A}, b) > b_i. \end{cases} \tag{9}$$

We show that $A(b) \in \boldsymbol{A}$. It is easy to see $A(b) \leq \overline{A}$. To prove the inequality $A(b) \geq \underline{A}$ we have to show that $b_i \geq \underline{a}_{ij}$ in the case $x_j^*(\underline{A}, b) > b_i$. This inequality follows directly from the definition of the principal solution.

Lemma 7. Let $\boldsymbol{A} \otimes x = b$ an interval system with a constant right-hand side. Then

(i) $x^*(A(b), b) = x^*(\underline{A}, b)$;

(ii) $x^*(A, b) = x^*(\underline{A}, b)$ if and only if $A \leq A(b)$;

(iii) if there exists a solvable subsystem then $A(b) \otimes x = b$ is solvable too.

Proof. (i) Let $j \in N$ be arbitrary, but fixed. Let us denote by $M^{(1)}, M^{(2)}, M^{(3)}$ the following sets:

$$M^{(1)} = \{i \in M : \underline{a}_{ij} \leq b_i \text{ and } x_j^*(\underline{A}, b) \leq b_i\}; \quad M^{(2)} = \{i \in M : \underline{a}_{ij} \leq b_i \text{ and } x_j^*(\underline{A}, b) > b_i\};$$

$$M^{(3)} = \{i \in M : \underline{a}_{ij} > b_i\}.$$

We shall prove that $\min\limits_{i \in M^{(k)}} \{b_i : a_{ij}(b) > b_i\} \geq x_j^*(\underline{A}, b)$ for each $k = 1, 2, 3$.

For $k=1$, we have $\min_{i\in M^{(1)}}\{b_i : a_{ij}(b) > b_i\} = \min_{i\in M^{(1)}}\{b_i : \overline{a}_{ij} > b_i\} \geq x_j^*(\underline{A}, b)$ because of $b_i \geq x_j^*(\underline{A}, b)$ for each $i \in M^{(1)}$ or $\{b_i : \overline{a}_{ij} > b_i = \emptyset$.

For $k=2$, we obtain $\min_{i\in M^{(2)}}\{b_i : a_{ij}(b) > b_i\} = \min_{i\in M^{(2)}}\{b_i : \min\{b_i, \overline{a}_{ij}\} > b_i\} = \min\emptyset = I \geq x_j^*(\underline{A}, b)$.

At least, for $i \in M^{(3)}$ the inequality $\underline{a}_{ij} > b_i$ implies $x_j^*(\underline{A}, b) \leq b_i$ which implies

$$\min_{i\in M^{(3)}}\{b_i : a_{ij}(b) > b_i\} = \min_{i\in M^{(3)}}\{b_i : \overline{a}_{ij} > b_i\} = \min_{i\in M^{(3)}}\{b_i : \underline{a}_{ij} > b_i\} \geq x_j^*(\underline{A}, b)$$

We obtain $x_j^*(A(b), b) = \min_{k\in\{1,2,3\}}\min_{i\in M^{(k)}}\{b_i : a_{ij}(b) > b_i\} \geq x_j^*(\underline{A}, b)$. On the other hand, $x_j^*(A(b), b) \leq x_j^*(\underline{A}, b)$ according to Lemma 2, which implies $x_j^*(A(b), b) = x_j^*(\underline{A}, b)$.

(ii) Suppose that $A \leq A(b)$ does not hold, i. e., there exist $i \in M$, $j \in N$ such that $a_{ij} > a_{ij}(b)$. Then $a_{ij}(b) = b_i < \overline{a}_{ij}$ and consequently $x_j^*(\underline{A}, b) > b_i$. The inequality $a_{ij} > b_i$ implies $x_j^*(A, b) \leq b_i < x_j^*(\underline{A}, b)$.

For the converse implication suppose that $A \leq A(b)$. By applying Lemma 2 twice we obtain $x^*(A, b) \geq x^*(A(b), b) = x^*(\underline{A}, b) \geq x^*(A, b)$.

(iii) Supppose that $A(b)\otimes x^*(A(b), b) \neq b$, i. e., there exists $i \in M$ such that $[A(b)\otimes x^*(A(b), b)]_i < b_i$. We prove that $[A \otimes x^*(A, b)]_i < b_i$ for each $A \in \boldsymbol{A}$. Let $A \in \boldsymbol{A}$ be arbitrary.

If $j \in N$ is such that $a_{ij}(b) = \overline{a}_{ij}$ then for each $A \in \boldsymbol{A}$ we have $a_{ij} \otimes x_j^*(A, b) \leq \overline{a}_{ij} \otimes x_j^*(\underline{A}, b) < b_i$.

If $a_{ij}(b) = b_i < \overline{a}_{ij}$ then the inequality $b_i \otimes x_j^*(\underline{A}, b) < b_i$ implies $x_j^*(\underline{A}, b) < b_i$, which is a contradiction with the definition of the matrix $A(b)$.

Since $a_{ij} \otimes x_j^*(A, b) < b_i$ for each $j \in N$, we obtain $[A\otimes x^*(A, b)]_i < b_i$, so $A\otimes x = b$ is not solvable. □

Since $x^*(A, b) \leq x^*(\underline{A}, b)$ for each $A \in \boldsymbol{A}$, Lemma 7 (ii) sayes that the matrix $A(b)$ is the greatest matrix with $x^*(A, b) = x^*(\underline{A}, b)$.

Lemma 8. An interval system (6) is weakly solvable if and only if

$$A(\overline{b}) \otimes x^*(A(\overline{b}), \overline{b}) \geq \underline{b}. \tag{10}$$

Proof. The proof follows directly from Lemma 5. □

3.2 Weak solvability of interval matrix equations

We define

$$\boldsymbol{A} = [\underline{A}, \overline{A}] = \left\{ A \in \mathcal{I}(m, n);\ \underline{A} \leq A \leq \overline{A} \right\} \text{ and } \boldsymbol{B} = [\underline{B}, \overline{B}] = \left\{ B \in \mathcal{I}(m, r);\ \underline{B} \leq B \leq \overline{B} \right\}.$$

Denote by

$$\boldsymbol{A} \otimes X = \boldsymbol{B} \tag{11}$$

the set of all matrix equations of the form (4) such that $A \in \boldsymbol{A}$ and $B \in \boldsymbol{B}$. We call equation (11) an *interval max-min matrix equation.*

Definition 2. Interval system(6) is *weakly solvable* if there exist $B \in \boldsymbol{B}$ and $A \in \boldsymbol{A}$ such that the equation $A \otimes X = B$ is solvable.

We shall deal with interval matrix equations with the constant right-hand side $\underline{B} = \overline{B} = B$, i. e.

$$\boldsymbol{A} \otimes X = B. \tag{12}$$

If $\underline{A} \otimes X^*(\underline{A}, B) = B$ or $\overline{A} \otimes X^*(\overline{A}, B) = B$ then the weak solvability of (11) trivially follows. In the following we shall assume that $\underline{A} \otimes X^*(\underline{A}, B) \neq B$ and $\overline{A} \otimes X^*(\overline{A}, B) \neq B$.

We shall create the sequence of matrices $\{A^{(l)}\}_{l=1}^{T}$ such that $A^{(T)}$ is such that the matrix equation $A^{(T)} \otimes X = B$ is solvable in the case that (11) is weakly solvable. In which follows we show the procedure of creating the sequence $\{A^{(l)}\}_{l=1}^{(T)}$ and an auxiliary sequence $\{C^{(l)}\}_{l=0}^{(T)}$.

Let us denote $C^{(0)} = \underline{A}$. If $C^{(l)}$ is known, define the matrices $A^{(l+1)}(B_k)$ for each $k \in R$ as follows:

$$a_{ij}^{(l+1)}(B_k) = \begin{cases} \overline{a}_{ij} & \text{if } x_j^*(C^{(l)}, B_k) \leq b_{ik}; \\ \min\{b_{ik}, \overline{a}_{ij}\} & \text{if } x_j^*(C^{(l)}, B_k) > b_{ik}. \end{cases} \tag{13}$$

Lemma 9. Let $\boldsymbol{A} \otimes X = B$ be an interval system with a constant right-hand side.

(i) $x^*(A^{(l+1)}, B_k) = x^*(C^{(l)}, B_k)$

(ii) If $x^*(A, B_k) = x^*(C^{(l)}, B_k)$ for $A \in \boldsymbol{A}$, $A \geq C^{(l)}$, then $A \leq A^{(l+1)}(B_k)$.

Proof. The assertions is similar to the proof of Lemma 7. □

In general, for $k \neq r$ we have $A^{(l)}(B_k) \neq A^{(l)}(B_t)$. To find the greatest matrix $A \in \boldsymbol{A}$ such that $X^*(A, B) = X^*(C^{(l)}, B)$, we define the matrix $A^{(l)}$ as follows:

$$A^{(l)} = \min_{k \in R} A^{(l)}(B_k) \tag{14}$$

Lemma 10. If $X^*(A, B) = X^*(C^{(l)}, B)$ for some $A \in \boldsymbol{A}$, $A \geq C^{(l)}$ then $A \leq A^{(l+1)}$.

Proof. The assumption is equivalent to $X^*(A, B_k) = X^*(C^{(l)}, B_k)$ for each $k \in R$. According to Lemma 9 we obtain $A \leq A^{(l+1)}(B_k)$ for each $k \in R$. Then $A \leq \min_{k \in R} A^{(l+1)}(B_k) = A^{(l+1)}$. □

Suppose that for a fixed $l \in \mathbb{N}$ the equality $A^{(l)} \otimes X^*(A^{(l)}, B) = B$ is not satisfied. Denote

$$U^{(l)} = \{(i, k) \in M \times R : [A^{(l)} \otimes X^*(A^{(l)}, B)]_{ik} < b_{ik}\}.$$

Let $(r, t) \in U^{(l)}$. We shall give the procedure of creating the matrix $C^{(l)}$ such that $[C^{(l)} \otimes X^*(C^{(l)}, B)]_{rt} = b_{rt}$ and $[A^{(l)} \otimes X^*(A^{(l)}, B)]_{ik} = b_{ik}$ implies $[C^{(l)} \otimes X^*(C^{(l)}, B)]_{ik} = B_{ik}$ if such matrix exists.

Let $u \in N$ be arbitrary but fixed. Define the matrix $C^{(l)}$ as follows:

$$c_{ij}^{(l)} = \begin{cases} b_{it} & \text{for } i = r,\ j = u \\ a_{ij}^{(l)} & \text{otherwise.} \end{cases} \tag{15}$$

We can see that $x_j^*(A^{(l)}, B_k) = x_j^*(C^{(l)}, B_k)$ for each $k \in R$, $j \in N - \{u\}$, but $x_u^*(A^{(l)}, B_k) \neq x_u^*(C^{(l)}, B_k)$ in general. Denote

$$R_u^{(l)} = \{k \in R : x_u^*(A^{(l)}, B_k) = x_u^*(C^{(l)}, B_k)\}$$

.

Lemma 11. $t \in R_u^{(l)}$ if and only if $b_{rt} \leq a_{ru}^{(l)}(B_t)$.

Proof. Since $x^*(A^{(l)}, B_t) = x^*(C^{(l-1)}, B_t)$, the equality $x_u^*(A^{(l)}, B_t) = x_u^*(C^{(l)}, B_t)$ is equivalent to $x^*(C^{(l-1)}, B_t) = x^*(C^{(l)}, B_t)$. According to Lemma 9, this is equivalent to $C^{(l)} \leq A^{(l)}(B_t)$. For $(i, j) \neq (r, u)$ the inequality $c_{ij}^{(l)} = a_{ij}^{(l)} \leq a_{ij}^{(l)}(B_t)$ trivially holds and therefore it is sufficient for the inequality $b_{rt} \leq a_{ru}^{(l)}(B_t)$ to apply. □

Denote $N_i^{(l)}(B_k) = \{j \in N : a_{ij}^{(l)} \otimes x_j^*(A^{(l)}, B_k) = b_{ik}\}$. It is easy to see that $[A^{(l)} \otimes X^*(A^{(l)}, B)]_{ik} = b_{ik}$ if and only if $N_i^{(l)}(B_k) \neq \emptyset$.

Lemma 12. Suppose that $(r, t) \in U^{(l)}$ and there exists $u \in N$ which satisfies the following conditions:

i) $b_{rt} \leq a_{ru}^{(l)}(B_t)$ and $x_u^*(A^{(l)}, B_t) \geq b_{rt}$;

ii) if $k \in R$ is such that $k \notin R_u^{(l)}$ then for each $i \in M, i \neq r$ the following holds¿ inequality

either $[N_i^{(l)}(B_k) \neq \emptyset \Rightarrow N_i^{(l)}(B_k) - \{u\} \neq \emptyset]$ or $[(u \in N_i(B_k) \wedge b_{rt} > b_{rk}) \Rightarrow b_{rk} \geq b_{ik}]$.

Then the following is true:

(i) $[C^{(l)} \otimes X^*(C^{(l)}, B)]_{rk} = b_{rk}$ for each $k \in R$;

(ii) If $i \neq r$ and $k \in R$ are such that $[A^{(l)} \otimes X^*(A^{(l)}, B)]_{ik} = b_{ik}$ then $[C^{(l)} \otimes X^*(C^{(l)}, B)]_{ik} = b_{ik}$.

Proof. (i) For $k = t$ we have

$$[C^{(l)} \otimes X^*(C^{(l)}, B)]_{rt} = \max_{j \in N}\{c_{rj}^{(l)} \otimes x_{jt}^*(C^{(l)}, B)\} \geq c_{ru}^{(l)} \otimes x_u^*(C^{(l)}, B_t) = b_{rt} \otimes x_u^*(A^{(l)}, B_t) = b_{rt}.$$

Together with the inequality $[C^{(l)} \otimes X^*(C^{(l)}, B)]_{rt} \leq b_{rt}$ we obtain $[C^{(l)} \otimes X^*(C^{(l)}, B)]_{rt} = b_{rt}$.
For $k \neq t$ we obtain $x_u^*(C^{(l)}, B_k) = b_{rk}$, $a_{ru}^{(l)} > b_{rk}$ which implies

$$[C^{(l)} \otimes X^*(C^{(l)}, B)]_{rk} = \max_{j \in N}\{c_{rj}^{(l)} \otimes x_j^*(C^{(l)}, B_k)\} \geq c_{ru}^{(l)} \otimes x_u^*(C^{(l)}, B_k) = a_{ru}^{(l)} \otimes b_{rk} = b_{rk}.$$

Similarly as in previous case we obtain $[C^{(l)} \otimes X^*(C^{(l)}, B)]_{rk} = b_{rk}$.

(ii) If $[A^{(l)} \otimes X^*(A^{(l)}, B)]_{ik} = b_{ik}$ and $X^*(A^{(l)}, B_k) = X^*(C^{(l)}, B_k)$, then $c_{ij}^{(l)} \otimes x_j^*(C^{(l)}, B_k) = a_{ij}^{(l)} \otimes x_j^*(A^{(l)}, B_k)$ holds for each $j \in N$ and the assertion trivially follows.

If $x_u^*(C^{(l)}, B_k) < x_u^*(A^{(l)}, B_k)$ and then the assumption $N_i^{(l)}(B_k) \neq \emptyset \Rightarrow N_i^{(l)}(B_k) - \{u\} \neq \emptyset$ holds then there exists $j \in N$, $j \neq u$ such that $a_{ij}^{(l)} \otimes x_j^*(A^{(l)}, B_k) = b_{ik}$ which implies

$$c_{ij}^{(l)} \otimes x_j^*(C^{(l)}, B_k) = a_{ij}^{(l)} \otimes x_j^*(A^{(l)}, B_k) = b_{ik}$$

because of $x_j^*(C^{(l)}, B_k) = x_j^*(A^{(l)}, B_k)$ and $a_{ij}^{(l)} = c_{Ij}^{(l)}$.

If $x_u^*(C^{(l)}, B_k) < x_u^*(A^{(l)}, B_k)$ and then the assumption $b_{rt} > b_{rk} \Rightarrow b_{rk} \geq b_{ik}$ holds then $x_u^*(C^{(l)}, B_k) = b_{rk}$ and $c_{iu}^{(l)} \otimes x_u^*(C^{(l)}, B_k) = a_{iu}^{(l)} \otimes b_{rk} \geq b_{ik}$. Together with $C^{(l)} \otimes X^*(C^{(l)}, B) \leq B$ the assertion follows. □

Algorithm: Weak solvability

Input: $\boldsymbol{A}$, $\boldsymbol{B}$
Output: aswer 'yes' in ws if $\boldsymbol{A} \otimes X = \boldsymbol{B}$ is weakly solvable and 'no' otherwise
begin

$\boxed{1}$ $l := 0$, $C^{(0)} := \underline{A}$; $N_i := N$ for each $i \in M$;

$\boxed{2}$ Compute $A^{(l+1)}$ by (13), (14) and find $U^{(l+1)}$. If $U^{(l+1)} = \emptyset$ then $ws :=' yes'$, go to **end**;

$\boxed{3}$ Choose $(r, t) \in U^{(l+1)}$. If $N_r = \emptyset$ then $ws :=' no'$, go to **end**;

$\boxed{4}$ Find $u \in N_r$ which satisfies conditions i) and ii) of Lemma 12. If such $u \in N_r$ does not exist and $l > 0$ then $l := l - 1$ go to $\boxed{2}$, else $ws :=' no'$ and go to **end**;

$\boxed{5}$ $N_r := N_r - \{u\}$;

$\boxed{6}$ Compute $C^{(l+1)}$ by formula (15); $l := l + 1$, go to $\boxed{2}$.
end

In general, Algorithm Weak solvability has exponential complexity, but in practice algorithm gives answer in most cases in polynomial time.

Example 2. Let us have $\mathcal{I} = [0, 10]$. Decide whether $\boldsymbol{A} \otimes X = B$ is weakly solvable, if

$$\boldsymbol{A} = \begin{pmatrix} [1,5] & [3,5] & [1,2] \\ [3,6] & [3,7] & [5,8] \\ [1,3] & [5,7] & [3,6] \end{pmatrix}, \qquad B = \begin{pmatrix} 3 & 4 & 3 \\ 6 & 6 & 5 \\ 4 & 5 & 4 \end{pmatrix}.$$

We can easily show that $\underline{A} \otimes X^*(\underline{A}, B) \neq B$ and $\overline{A} \otimes X^*(\overline{A}, B) \neq B$. We are looking for the matrix $A \in \boldsymbol{A}$ such that $A \otimes X^*(A, B) = B$. For $C^{(0)} = \underline{A}$ we have

$$X^*(C^{(0)}, B) = \begin{pmatrix} 10 & 10 & 10 \\ 4 & 10 & 4 \\ 10 & 10 & 10 \end{pmatrix}.$$

We compute the matrices $A^{(1)}(B_k)$ for $k = 1, 2, 3$ by formula (13) and the matrix $A^{(1)}$ by (14). We obtain

$$A^{(1)}(B_1) = \begin{pmatrix} 3 & 3 & 2 \\ 6 & 7 & 6 \\ 3 & 7 & 4 \end{pmatrix}, \quad A^{(1)}(B_2) = \begin{pmatrix} 4 & 4 & 2 \\ 6 & 6 & 6 \\ 3 & 5 & 5 \end{pmatrix}, \quad A^{(1)}(B_3) = \begin{pmatrix} 3 & 3 & 2 \\ 5 & 7 & 5 \\ 3 & 7 & 4 \end{pmatrix}, \quad A^{(1)} = \begin{pmatrix} 3 & 3 & 2 \\ 5 & 6 & 5 \\ 3 & 5 & 4 \end{pmatrix}.$$

The matrix $A = A^{(1)}$ is the greatest matrix such that $X^*(A, B) = X^*(C^{(0)}, B)$. We compute

$$A^{(1)} \otimes X^*(A^{(1)}, B) = \begin{pmatrix} 3 & \mathbf{3} & 3 \\ \mathbf{5} & 6 & 5 \\ 4 & 5 & 4 \end{pmatrix} \neq B,$$

so matrix $A^{(1)}$ is not a searched matrix. We have $U^{(1)}) = \{(1, 2), (2, 1)\}$.

Let us take $r = 1, t = 2$. We have to find $u \in N$ for which the assumptions of Lemma 12 are satisfied.

For $u = 1$, we have

- $b_{12} \leq a_{11}^{(1)}(B_2)$ and $x_1^*(A^{(1)}, B_2) \geq b_{12}$
- For $k = 1$ we have $N_2^{(1)}(B_1) = \emptyset$, $N_3^{(1)}(B_1) = \{2, 3\}$, so $N_3^{(1)}(B_1) - \{1\} \neq \emptyset$.

  For $k = 3$ we have $N_2^{(1)}(B_3) = \{1, 3\}$, so $N_2^{(1)}(B_3) = \{3\} \neq \emptyset$, $N_3^{(1)}(B_3) = \{2, 3\}$, so $N_3^{(1)}(B_3) - \{1\} \neq \emptyset$.

Since the assumptions of Lemma 12 are satisfied, we can compute the matrix $C^{(1)}$ by (15) and check whether $C^{(1)} \otimes X^*(C^{(1)}, B) = B$. We obtain

$$C^{(1)} \otimes X^*(C^{(1)}, B) = \begin{pmatrix} \mathbf{4} & 3 & 2 \\ 5 & 6 & 5 \\ 3 & 5 & 4 \end{pmatrix} \otimes \begin{pmatrix} \mathbf{3} & 10 & \mathbf{3} \\ 4 & 10 & 4 \\ 10 & 10 & 10 \end{pmatrix} = \begin{pmatrix} 3 & 4 & 3 \\ \mathbf{5} & 6 & 5 \\ 4 & 5 & 4 \end{pmatrix}.$$

Using formulas (13) and (14) we obtain the matrix $A^{(2)}$, the greatest matrix such that $X^*(A^{(2)}, B) = X^*(C^{(1)}, B)$.

$$A^{(2)}(B_1) = \begin{pmatrix} 5 & 3 & 2 \\ 6 & 7 & 6 \\ 3 & 7 & 4 \end{pmatrix}, \quad A^{(2)}(B_2) = \begin{pmatrix} 4 & 4 & 2 \\ 6 & 6 & 6 \\ 3 & 5 & 5 \end{pmatrix}, \quad A^{(1)}(B_3) = \begin{pmatrix} 5 & 3 & 2 \\ 6 & 7 & 5 \\ 3 & 7 & 4 \end{pmatrix}, \quad A^{(2)} = \begin{pmatrix} 4 & 3 & 2 \\ 6 & 6 & 5 \\ 3 & 5 & 4 \end{pmatrix}.$$

and check whether the equality $A^{(2)} \otimes X^*(A^{(2)}, B) = B$ is satisfied. We obtain

$$A^{(2)} \otimes X^*(A^{(2)}, B) = \begin{pmatrix} 4 & 3 & 2 \\ 6 & 6 & 5 \\ 3 & 5 & 4 \end{pmatrix} \otimes \begin{pmatrix} 3 & 10 & 3 \\ 4 & 10 & 4 \\ 10 & 10 & 10 \end{pmatrix} = \begin{pmatrix} 3 & 4 & 3 \\ \mathbf{5} & 6 & 5 \\ 4 & 5 & 4 \end{pmatrix}.$$

Since $U^{(2)} = \{(2, 1)\}$ we have $r = 2$, $t = 1$. For $u = 3$ we have $R_3^{(2)} = \{2, 3\}$ and the following assertions hold:

- $b_{21} \leq a_{23}^{(2)}(B_1)$ and $x_3^*(A^{(2)}, B_1) \geq b_{21}$;

- For $k=1$ we have $N_1^{(2)}(B_1)=\{1,2\}$, so $N_1^{(2)}(B_3)-\{3\}\neq\emptyset$ and $N_3^{(2)}(B_1)=\{2,3\}$, so $N_2^{(2)}(B_3)-\{3\}\neq\emptyset$.

Since the assumptions of Lemma 12 are satisfied, we can compute the matrix $C^{(2)}$ and verify whether $C^{(2)}\otimes X^*(C^{(2)},B)=B$. We obtain

$$C^{(2)}\otimes X^*(C^{(2)},B)=\begin{pmatrix}4&3&2\\6&6&5\\3&5&4\end{pmatrix}\otimes\begin{pmatrix}3&10&3\\4&10&4\\5&10&10\end{pmatrix}=\begin{pmatrix}3&4&3\\6&6&5\\4&5&4\end{pmatrix}.$$

Since $C^{(2)}\otimes X^*(C^{(2)},B)=B$, it is not necessary to compute the matrix $A^{(3)}$. There exists a matrix A, namely $A=C^{(2)}$ such that $A\otimes X^*(A,B)=B$, so the given interval matrix equation is weakly solvable.

3.3 Conclusion

In this paper, we dealt with the solvability of matrix equations in max-min algebra. Max-min algebra is a useful tool for describing real situation in economics and industry. In Example 1, the values a_{ij}, x_{jl} 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) b_{ik} are also from certain intervals rather than fixed numbers. Therefore, it would be appropriate if the matrix B were also interval. The study of weak solvability for interval right side is our main goal for the future. Another challenge is to modify the algorithm so that it is polynomial.

References

[1] Cechlárová, K. (2001). Solutions of interval systems in max-plus algebra, In: V. Rupnik, L.Zadnik-Stirn & S. Drobne (Eds.), *Proc.of SOR 2001* (321-326), Preddvor, Slovenia.

[2] Cuninghame-Green, R.A. (1960) Process synchronisation in a steelworks –a problem of feasibility. In: Banbury, M. (eds.) *Proceedings of 2nd international conference on operational research*, 323–328.

[3] Cuninghame-Green, R.A. (1962) Describing Industrial Processes with Interference and Approximating Their Steady-State Behaviour. *J Oper Res Soc* 13, 95–100.

[4] Cuninghame-Green, R. A. (1979). *Minimax Algebra. Lecture notes in Economics and Mathematical ystems*. Berlin: Springer.

[5] E. Draženská & H. Myšková, H. (2017). Interval fuzzy matrix equations. *Kybernetika*, 53, 99–112.

[6] Myšková, H. (2005). Interval systems of max-separable linear equations. *Lin. Algebra Appl.*, 403, 263–272.

[7] Myšková, H. & Plavka, J. (2020). On the solvability of interval max-min matrix equations. *Lin. Algebra and its Appl.*, 590, 85–96.

[8] Di Nola, A., Salvatore, S., Pedrycz, W. & Sanchez, E. (1989*). Fuzzy Relation Equations and Their Applications to Knowledge Engineering*. Dordrecht:Kluwer Academic Publishers.

[9] Sanchez, E. (1979). Medical diagnosis and composite relations. In: M. M. Gupta, R. K. Ragade, R. R. Yager,(Eds), *Advances in Max-plus Set Theory and Applications* (437-444). North-Holland, Amsterdam.

[10] Zadeh, L. A. (1971) Toward a theory of max-plus systems, In: R. E. Kalman & N. De Claris (Eds), *Aspects of Network and Systems Theory* (209-245). New York: Hold, Rinehart and Winston.

[11] Zimmermann, K. (1976). *Extremální algebra*. Prague: Ekonomicko-matematická laboratoř Ekonomického ústavu ČSAV.

Application of ARMA–GARCH Returns Generation in Portfolio Selection Process

David Neděla[1]

Abstract. In this paper, we examine the application of a return scenario generation procedure in the portfolio optimization strategy based on an ARMA–GARCH model. We assume that residuals follow a stable distribution and the dependency structure of residuals is determined by Student t and skewed t copula. Moreover, we analyse the modification in dependency estimation of residuals obtained from the ARMA–GARCH model in order to find an appropriate setting. Then, we compare the effect of selected return approximation methods on the ex-post portfolio wealth and statistics determined using a portfolio model maximizing selected widely used reward-risk measures. Additionally, the strategy consists of monthly re-optimization and transaction costs expressed proportionally. The following empirical analysis on the U.S. market data allows us to evaluate the impact of return approximation in the portfolio optimization process. Results obtained using analysed approaches generate lower risk with affecting portfolio performance for certain models compared to the benchmark.

Keywords: ARMA–GARCH model, Copula, Performance measures, Portfolio selection

JEL Classification: C53, C61
AMS Classification: 62M10, 90C05, 90C06

1 Introduction

The construction of ideal portfolio is still an actual problem for both investors and researchers, respectively. Since the work of Markowitz [8], the mean-variance model has become the most well-known portfolio model recently. The concept of monitoring the expected return of the portfolio with respect to the risk expressed as the variance is easily feasible and satisfactory for small investors, whose minimize the risk incurred. Nevertheless, the model has supporters but also opponents, see [9]. Main reasons are following: set of assumptions that need to be met in the model is not consistent with possible investor preferences, and inappropriate portfolio risk measure in the model, respectively. For this reason, alternative approaches and models containing advanced measures for optimization have been provided, see, among others, [5], [13], [14].

A significant task in the portfolio optimization process is to precisely estimate the future behaviour of assets. Portfolio selection models are sensitive to estimated statistical parameters describing the behaviour and interdependency of asset returns. For this reason, it is necessary to generate accurately these return series. The problems of return approximation by parametric or non-parametric models and scenario generation are broadly discussed in the literature, see [4], [7], [9], or [10]. In [9], authors examined the use non-parametric regression estimator based on locally weighted least squares. However, the application of the parametric ARMA–GARCH model with modified residuals is applied in [4]. This approach allow us to appropriate capture the behaviour of return series. Therefore, it is used in our analysis.

Overall, the major contribution of this analysis is to investigate and compare the effect of ARMA(1,1)–GARCH(1,1) model variations to generate return scenarios. In particular, we assume that residual series follow stable distribution and their dependency is described by two types of t copula function. Finally, we compare this method with application of the simple ARMA(1,1)–GARCH(1,1) model considering the reward-risk portfolio selection strategy.

The rest of this paper is divided as follows. In Section 2, the methodology of return approximation and the reward-risk portfolio optimization task is characterized. In Section 3 we provide the empirical analysis of the selected dataset. The results of the analysis are summarized in Section 4.

[1] VSB–Technical university of Ostrava, Department of Finance, Sokolská třída 33, Ostrava , david.nedela@vsb.cz

2 Returns Approximation and Portfolio Optimization Problem

This section explains the main theoretical background of the return approximation by the ARMA–GARCH model with the stable estimation and Student and skewed *t* copula dependency between residuals.

We suppose that a portfolio contains z assets with the random daily return vector denoted as $r = [r_i, \ldots, r_z]$ and the vector of weights of a particular asset $r = [x_i, \ldots, x_z]$ for $i = 1 \ldots z$. Then, we can express the random portfolio return as $x'r$.

2.1 Generation of Scenarios

Firstly, let us briefly describe the algorithm for scenarios generating of asset return proposed by [1] and [2].

We generally assume that the individual return of asset $r_{i,t}$ follows the ARMA(1,1)–GARCH(1,1) process in (1) while residuals are stable distributed with the dependency structure expressed by asymmetric Student t copula valued on those residual series, see among others [2], [4], or [12]. The model is formulated as follows:

$$\begin{aligned} r_{i,t} &= a_{i,0} + a_{i,1} r_{i,t-1} + b_{i,1}\varepsilon_{i,t-1} + \varepsilon_{i,t} \\ \varepsilon_{i,t} &= \sigma_{i,t} u_{i,t} \\ \sigma^2_{i,t} &= c_{i,0} + c_{i,1}\sigma^2_{i,t-1} + d_{i,1}\varepsilon^2_{i,t-1} \end{aligned} \tag{1}$$

where $i = 1, \ldots, z$; $t = 1, \ldots, T$. We use the maximum likelihood method to estimate parameters of model (1).

From the estimated standardized residuals, we approximate the standardized residuals $\hat{u}_{i,t} = \hat{\varepsilon}_{i,t}/\sigma_{i,t}$ using α_i-stable distribution. We simulate S stable distributed scenarios for each of the future standardized residuals to capture the marginal distribution of each residual. The sample distribution functions for these simulated series $\hat{u}_{i,t}$ are then computed as follows:

$$F_{\hat{u}_{i,T+1}}(x) = \frac{1}{S}\sum_{s=1}^{S} I_{\left\{\hat{u}^{(s)}_{i,T+1} \le x\right\}}, x \in \mathbb{R}, i = 1, \ldots, z \tag{2}$$

where $\hat{u}^{(s)}_{i,T+1} (1 \le s \le S)$ is the s-th value simulated with the fitted stable distribution for future standardized residuals (valued in $T+1$) of the i-th return.

Then we capture the dependency structure by fitting the vector of standardised residuals $\hat{u} = [\hat{u}_1, \ldots, \hat{u}_z]$ with asymmetric t-distribution $v = [v_1, \ldots, v_z]$. Considering $U^{(s)}_i = F_{Vi}(V^{(s)}_i)$; $1 \le i \le z$; $1 \le s \le S$, we can generate S scenarios $(U^{(s)}_1, \ldots, U^{(s)}_z)$; $s = 1, \ldots, S$ of the uniform random vector $(U_1, \ldots, U_z)$ and with the distribution following the Student t or skewed t copula as follows:

$$C(t_1, \ldots, t_z) = F_V\left(F^{-1}_{V_1}(t_1), \ldots, F^{-1}_{V_z}(t_z)\right); 0 \le t_i \le 1; 1 \le i \le z \tag{3}$$

Student t and skewed t distributions are deeply described in [4] and [7], respectively. Furthermore, we are able to generate S scenarios of the stable distributed standardized residual vector considering dependency modelling as $u^{(s)}_{T+1} = \left(u^{(1,s)}_{T+1}, \ldots, u^{(z,s)}_{T+1}\right)$, $s = 1, \ldots, S$. In doing so, we can generate the vector of the model's standardized residuals at time $T+1$ from the model (1) as: $\varepsilon^i_{T+1} = \left[\varepsilon^{1,i}_{T+1}, \ldots, \varepsilon^{z,i}_{T+1}\right] = \left[\sigma_{1,T+1}u^{1,i}_{T+1}, \ldots, \sigma_{z,T+1}u^{z,i}_{T+1}\right]$.

Finally, to generate the vector of future return behaviour $r^i_{T+1} = \left[r^{1,i}_{T+1}, \ldots, r^{z,i}_{T+1}\right]$ at time $T+1$, we use the first equation in model (1). The whole process with the derivation of particular equations is broadly described in [2].

2.2 Performance Measures and Optimization Model

Until now, many performance measures of the portfolio have been developed and provided in the literature, see [2], [13]. Thus, we briefly characterize only three frequently used performance ratios, which are applicable in this analysis, see, among others, [6].

Sharpe ratio (SR). The most known reward-risk measure is the Sharpe ratio. According to the original Markowitz work [8], Sharpe [15] introduced the ratio involving the portfolio excess return and the standard deviation defined as follows:

$$SR = \frac{E(x'r - r_f)}{(x'Qx)^{\frac{1}{2}}}, \tag{4}$$

where r_f is a risk-free return (or a benchmark return) and Q represents the covariance matrix. The SR value expresses the return for the unit of risk.

Rachev ratio (RR). The Rachev ratio measures the ratio between the Conditional Value-at-Risk (CVaR) of earnings and the mean of losses beyond Value-at-Risk (VaR), see [13]. The equation is as follows:

$$RR = \frac{CVaR_\beta(r_f - x'r)}{CVaR_\alpha(x'r - r_f)}. \tag{5}$$

where $CVaR_\alpha(x'r) = \frac{1}{\alpha}\int_0^\alpha VaR_y(x'r)dy$, α and β are significance values.

Farinelli-Tibiletti ratio (FTR). The final performance ratio including the partial moments of a return series is the Farinelli-Tibiletti ratio calculated as follows:

$$FTR\,(x'r) = \frac{\left\{E\left[(x'r - r_f)_+^\gamma\right]\right\}^{1/\gamma}}{\left\{E\left[(x'r - r_f)_-^\delta\right]\right\}^{1/\delta}}. \tag{6}$$

where γ and δ are the orders of the partial moments, see [3]. The usually used setting is $(\gamma = 0.5, \delta = 2)$, which is applied in this analysis. If $(\gamma = \delta = 1)$, *FTR* transforms into a ratio called *Omega ratio (OR)*.

Assuming optimization of the portfolio based on maximization of mentioned performance measures above, the optimal portfolio is obtained by solving the following optimization framework:

$$\begin{aligned} \max\ \Gamma(x'r) \\ x'\mathrm{e} = 1 \\ x_i \geq 0;\ i = 1, \ldots, z, \end{aligned} \tag{7}$$

where $\Gamma(x'r)$ is one of the performance measures formulated in equations (4), (5), and (6), see [13] or [11].

3 Empirical Evidence

The content of this section is to empirically investigate the validity and performance of the ARMA–GARCH scenario generation process of returns in the portfolio selection process.

For empirical evidence, we use active stocks (on date 4 November 2021) in the SP 100 index from 1 January 2010 to 30 June 2021, meaning in total 2893 daily observations. The data is gathered from the Bloomberg Database. As a risk-free rate, the 3-month U.S. Treasury bill return is used for the need of performance measures calculation. We recalibrate the portfolio monthly (20 trading days) using a rolling window of one year (250 trading days). Furthermore, short sales of assets are not allowed, and an upper limit of weight is not set. The initial wealth is set as $W_0 = 1$. Some of the stock series are excluded from the analysis due to the unavailability of data. The whole empirical procedure is divided into several steps:

Step 1. Based on 250 historical observations, we approximate the future return scenarios by either the simple ARMA(1,1)–GARCH(1,1) model considering only equation (1) marked as (ARMA-GARCH), the ARMA(1,1)–GARCH(1,1) with stable distributed residuals and skewed tc opula dependency structure between them described in subsection 2.1 (stab-skew-t-ARMA-GARCH), or ARMA(1,1)–GARCH(1,1) with stable distributed residuals and Student t copula dependency structure (stab-Student-t-ARMA-GARCH). We generated the matrix G with 2000 future scenarios of each asset using Monte Carlo simulation.

Step 2. Determine the optimal vector of asset weights x applying the portfolio problem (7) to historical returns with rolling window of 250 days and the approximated returns from the previous steps.

Step 3. After recalibration of the asset weight structure in the portfolio every 20 days, we calculate the ex-post final wealth as follows:

$$W_{t+1} \begin{cases} W_t(1 + r_{t+1} - tc) & \text{if recalibrated} \\ W_t(1 + r_{t+1}) & \text{otherwise} \end{cases} \quad (8)$$

where $tc = 0.002 \sum_{i=1}^{z} \left| x_{i,t} - \frac{x_{i,t-1}(1+r_{i,t})}{\sum_{i=1}^{z} x_{i,t-1}(1+r_{i,t})} \right|$ represents the transaction costs as the proportion of 20 basis points. In Table 1, we present the ex-post results of the constructed portfolio assuming: the approximated returns by ARMA(1,1)–GARCH(1,1) model with different residual estimation techniques, historical returns, and a benchmark defined as an index S&P 100. Maan is the mean return, Std is the standard deviation of portfolio returns, VaR and CVaR are Value-at-Risk and Conditional-Value-at-Risk with 5% significance, respectively.

Model	Mean (%)	Wealth	Std(%)	VaR(%)	CVaR(%)	SR(%)	RR
	stab-skew-t-ARMA-GARCH						
maxSR	0.0623	3.5422	1.1139	1.6764	2.6796	5.5741	0.9096
maxRR	0.0422	1.7760	1.3586	2.1510	3.3504	3.0953	0.9078
maxFTR	0.0379	1.8465	0.9277	1.3582	2.2014	4.0694	0.8715
maxOR	0.0702	4.4101	1.1373	1.7273	2.7260	6.1566	0.9037
	stab-t-ARMA-GARCH						
maxSR	0.0629	3.5110	1.1209	1.7007	2.6666	5.5992	0.9344
maxRR	0.0435	1.5507	1.5419	2.2188	3.6275	2.8113	0.9513
maxFTR	0.0380	1.8443	0.9183	1.3567	2.1732	4.1142	0.8936
maxOR	0.0686	4.1049	1.1441	1.7565	2.7188	5.9813	0.9407
	simple ARMA-GARCH						
maxSR	0.0441	2.0616	1.0687	1.5937	2.6709	4.1133	0.8523
maxRR	0.0371	1.5576	1.2674	1.6857	3.1265	2.9105	0.8811
maxFTR	0.0425	3.2888	1.0197	1.5165	2.5273	4.1474	0.8607
maxOR	0.0450	2.1135	1.0626	1.5805	2.6563	4.2172	0.8525
	Without returns approximation						
maxSR	0.0579	3.3382	1.2623	1.8795	3.0836	4.5763	0.9216
maxRR	0.0756	4.9617	1.3328	1.9859	3.2083	5.6569	0.9671
maxFTR	0.0330	1.9470	0.8809	1.2409	2.0870	3.7245	0.8734
maxOR	0.0688	4.4448	1.2993	1.8700	3.1148	5.2854	0.9382
S&P100	0.0470	2.9659	1.0823	1.6258	2.7012	4.3303	0.8652

Table 1: Daily statistics of ex-ante returns obtained by maximizing different portfolio measures

The results obtained in Table 1 can be viewed from two different perspectives of comparison: return approximation approaches and reward-risk maximization models. Starting with a different model perspective, we observe the differences in the results when incorporating historical returns and approximated returns. Looking at the use of historical returns, the highest wealth and performance (expressed by SR or RR) are visible while the maxRR and maxOR models are applied. On the contrary, the worst result generates the maxFTR model, where the final wealth is even lower than the selected benchmark. If the approximated return scenarios are included in the optimization model, the most profitable models are those where we maximize the Omega and Sharpe ratios, especially when applying more sophisticated approaches for residual approximation. In those cases, we outperform the profitability of both the benchmark and the optimization strategy assuming historical data. Simultaneously, we achieve the risk reduction (exxpresed by Std or VaR), which is required mainly by risk-averse investors. However, maximizing RR of a portfolio based on approximated returns surprisingly does not reach the efficiency as for historical returns. Analysing the maxFTR model favorability, its combinations with the approximated model outperform the strategy without return approximation, however, it is still under the benchmark level.

If we focus on the comparison of return approximation techniques, the results correspond to the expectations, meaning that the simple ARMA–GARCH approach generates consistent results, even if we change the portfolio model, except for the maxFTR model. Furthermore, in the majority, the profitability of the simple model is lower than if using only historical observations, nor the benchmark is not exceeded. For this reason, the simple model without relations between assets does not meet the need for the most accurate prediction, due to the shortcomings caused by random estimation of residuals.

Otherwise, when we capture the dependency relationships in the approximation procedure, and the proportion of assets in the portfolio is based on future estimation, we see more different results with higher final wealth and profitability of the maxSR portfolio and similar for the maxOR portfolio. The most important conclusion is that we can reduce the risk even if we examine Std or VaR. Looking at the differences between the skewed t and Student t copula dependency of residuals, we can observe that the skewed t copula fits precisely to the approximation model and the following portfolio reports slightly higher final wealth, mean return, and performance ratios while the riskiness is lower.

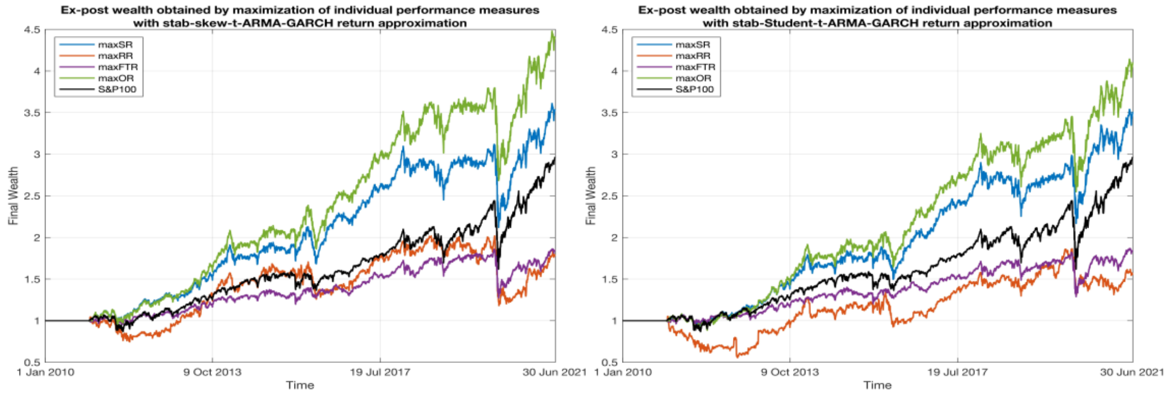


Figure 1: Ex-post wealth with using stab-skew-t-ARMA-GARCH model for return approximation (left) and stab-Studen-t-ARMA-GARCH model (right) for each portfolio model

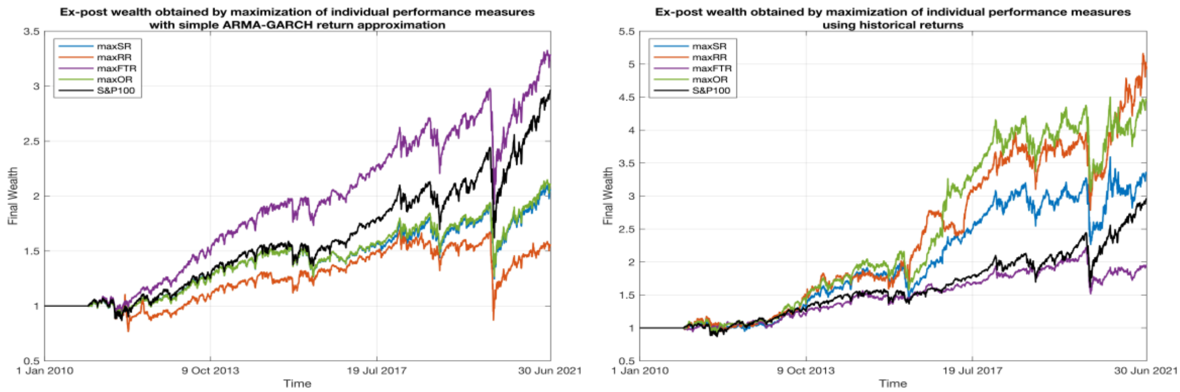


Figure 2: Ex-post wealth while using simple ARMA–GARCH model for return approximation (left) and using historical returns (right) for each portfolio model

For illustration, we also depict in Figures 1 and 2 the wealth evolution of individual portfolio models applied to different data. The black curve in all figures represents the benchmark. We can observe that during the first few years, the wealth of portfolios has a slightly increasing trend until the middle of 2015 when the trend started to grow faster with higher volatility. At the beginning of 2020, due to the COVID-19 pandemic, the market fell rapidly for a few days, but the subsequent trend is a highly increasing slope compared to the time before the slump.

If we look in detail at the evolution of portfolios using ARMA–GARCH with estimated dependent residuals in Figure 1, the maxOR and maxSR models mutually copy the trend even in volatile periods in the market. Otherwise, at the beginning of the investment, the maxRR model drops significantly, and then the performance is not as sufficient as the historical data are used for optimization in Figure 2. On the left side of this figure, we can determine the main weakness of a simple model, which is a misidentification of the impending decline. Initially, created portfolios have copied and in some cases even exceeded the benchmark until the beginning of the year 2020 affected by the COVID-19 pandemic. At this point, the portfolios fell below the benchmark, which subsequently grew at a higher rate to the surprise of investors and analysts.

Overall, these results confirm that the ARMA–GARCH model with stable estimated residuals and dependence structure determined by t copula function provides a sufficiently accurate approximation of future returns. Depending on the data used, maximizing reward-risk portfolio model with Sharpe and Omega ratios generates the highest profitability with a lower risk rate and always performs better than the S&P 100 index. This strategy is chiefly suitable for risk-averse investors.

4 Conclusion

In this paper, we compare the effect of different return approximation approaches based on the ARMA–GARCH model, where the ARMA model captures the autocorrelation and the GARCH model indicates the volatility. To capture the relationship structure between assets, the estimated residuals of the ARMA-GARCH model follow the Student t or skewed t copula function. Then, we included the approximation procedure in the portfolio optimization strategy comprising the reward-risk portfolio model for four selected ratios and analysed the main characteristics of the portfolio.

The results obtained showed that the ARMA(1,1)–GARCH(1,1) model with estimated residuals following a stable distribution and dependence structure determined by one of t copula functions led to a sufficiently accurate approximation of the future return scenario. This analysis was applied to the active stocks included in the index S&P 100. When maximizing reward-risk portfolio model with Sharpe and Omega ratios, it generated the higher profitability with a lower rate of risk compared to the simple application of this model to the historical data. All portfolio strategies were compared with the benchmark defined as the index S&P 100. Maximization of Sharpe and Omega ratios in the optimization framework provided higher profitability with a lower risk rate and performs better than the benchmark. Therefore, we can conclude that this strategy is recommended to risk-averse investors.

Concurrently, we are considering an application of multifactor model employing the principal component analysis in order to reduce the dimensionality of the large-scale portfolio selection problem. Given that the components are obtained from the dependency matrix, we could compare the impact of the linear and trend-dependent correlations. In addition, we should analyze behaviour of residuals of the multifactor model and precisely predict its evolution in the future.

Acknowledgements

The author greatly acknowledged support through the Czech Science Foundation (GACR) under the project GA20-16764S, SGS research project SP2022/4 of VSB-TU Ostrava, and the Moravian-Silesian region by the RRC/02/2020 project.

References

[1] Biglova, A., Ortobelli, S. & Fabozzi, F.J. (2014). Portfolio selection in the presence of systemic risk. *Journal of Asset Management*, 15(5), 285–299.

[2] Biglova, A., Ortobelli, S., Rachev, S. & Fabozzi, F.J.. (2010). Modeling, estimation, and optimization of equity portfolios with heavy-tailed distributions. In: *Satchell, S.: Optimizing Optimization: The next Generation of Optimization Applications and Theory.* Amsterdam: Academic Press, 117–141.

[3] Farinelli, S., Ferreira, M., Rossello, D., Thoeny, M. & Tibiletti, L. (2008). Beyond Sharpe Ratio: Optimal Asset Allocation Using Different Performance Ratios. *Journal of Banking & Finance*, 32(10), 2057—2063.

[4] Georgiev, K., Kim, Y. S. & Stoyanov, S. V. (2015). Periodic portfolio revision with transaction costs. *Mathematical Methods of Operations Research*, 81(3), 337–359.

[5] Konno, H.,&Yamazaki,H.(1991).Mean-Absolutedeviationportfoliooptimizationmodelanditsapplications to Tokyo stock-market. *Management Science*, 37(5), 591–531.

[6] Kouaissah, N. & Hocine, A. (2020). Forecasting systemic risk in portfolio selection: The role of technical trading rules. *Journal of Forecasting*, 1–22.

[7] Kouaissah, N., Ortobelli, S. & Jebabli, I. (2021). Portfolio Selection Using Multivariate Semiparametric Estimators and a Copula PCA-Based Approach. *Computational Economics*, doi:10.1007/s10614-021-10167- w.

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

[9] Ortobelli, S., Kouaissah, N., & Tichy, T. (2017). On the impact of conditional expectation estimators in portfolio theory. *Computational Management Science*, 14(4), 535–557.

[10] Ortobelli, S., Kouaissah, N., & Tichy, T. (2019). On the use of conditional expectation in portfolio selection problems. *Annals of Operations Research*, 274(1–2), 501–530.

[11] Ortobelli, S. & Tichý, T. (2015). On the impact of semidefinite positive correlation measures in portfolio theory. *Annals of Operations Research*, 235(1), 625–652.
[12] Rachev, S. & Mittnik, S. (2000). *Stable Paretian Models in Finance*. New York, USA: John Wiley & Sons.
[13] Rachev,S.T., Stoyanov,S.V.&Fabozzi,F.J.(2008).*Advanced stochastic models, risk assessment and portfolio optimization: The ideal risk, uncertainty and performance measures*. New York: Wiley Finance.
[14] Rockafellar, R.T. & Uryasev, S.P. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26(7), 1443–1471.
[15] Sharpe, W.F. (1994). The Sharpe ratio. *Journal of Portfolio Management*, 21(1), 49–58.

Impact of Financial Development and Trade Openness on the GDP of CR

Václava Pánková[1]

Abstract. Economic development, always conditioned by economic growth, is one of the most important indicators of a healthy economy. The main factors of economic growth could be supported directly or indirectly by financial development and international trade. The financial development and the trade openness are defined, measured, and related to the GDP *per capita* of the Czech Republic.
Econometric techniques are applied comprising data generating process allowing for break points. The break indicated in 2008 is treated as a difference between financial crisis 2008 – 09 and other years using multiplicative dummies. Relevant econometric model is formulated and estimated.
From the results a conclusion follows that both financial development and trade openness show their positive role just during the period of a crisis.

Keywords: economic growth, structural break, model ECM

JEL Classification: C40, F13, G01
AMS Classification: 60H30, 62G05

1 Introduction

Economic development is one of the central issues explored by economic theory and is also one of the most important indicators of a healthy economy. It is a broader concept than economic growth, however, economic development is always conditioned by economic growth.

The main factors of economic growth can be split up into consumer spending factors and productive capacity factors. All of them could be supported directly or indirectly by financial development and international trade.

Economic growth is defined as an increase in the amount of goods and services produced per head of the population over a period of time. Quantitatively it is given as GDP *per capita*.

1.1 Importance of Financial Development

One of the main drivers of economic growth is capital, its formation and accumulation. Countries with better-developed financial systems tend to grow faster over long periods of time; an evidence suggests that financial development contributes to this growth.

In practice, it is difficult to measure financial development as it is a vast concept and has several dimensions. Empirical work done so far is usually based on standard quantitative indicators available for a long time series for a broad range of countries. To understand the impact of financial development on economic growth, a ratio of a chosen financial indicator to GDP usually is applied. A wide overview of literature discussing different convenient concepts is given e. g. in [8]. The impact of the financial crisis on the relationship between finance and growth is another area of recent exploration in the financial literature.

In this article, financial development variable is defined as the ratio of financial system deposits to GDP (source [4]).

1.2 Importance of Trade Openness

A brief characterisation of trade openness and its relation to other important macroeconomic and financial variables is given e. g. in a handbook [1]. The authors state that bilateral equity investment is strongly correlated with underlying patterns of trade. Investors are better able to attain accounting and regulatory information on foreign markets through trade and thereby invest in foreign assets. Further, trade transactions may directly generate cross-border financial flows including trade credits, export insurance, payment facilitation.

[1] University of Economics and Business, Department of Econometrics, 130 67 Praha 3, nám. Winstona Churchilla 4, pankova@vse.cz

Trade openness also is one of the determinants which play a role in the determination of currency crises. Several studies find that greater trade integration reduces a country's financial fragility; hence, higher trade integration tends to reduce the frequency of external financial crises.

The nature of the relationship between trade openness and economic growth is described e. g. in [11] exploring existing economic theories and relevant conclusions. Besides other aspects, trade openness promotes the efficient allocation of resources, factor accumulation, technology diffusion, and knowledge spillovers. They conclude that export growth is considered to be the main driver of economic growth.

In this article, a common definition of trade openness as a ratio of sum of exports and imports to GDP is used (source [3] – all variables).

2 Data Generating Process

Data cover the period from 1995 to 2020, T=26. In 2008 and 2009, the financial crisis struck the main macroeconomic variables; an influence of covid pandemic started during 2020. That is why the DGP (Data Generating Process) should pay attention not only to unit roots but also to possible break points. In [9], Perron shows that failure to allow for an existing break makes the ability to reject a false unit root null hypothesis reduced. Such a time-series is then incorrectly treated as non-stationary.

If the structural breaks are known or exogenous they can be modelled using appropriate dummy variables. Extension of ADF unit root test is in [9] allowing for a break in the intercept or in the slope. The former means a change in a level, the latter represents a changing growth. Both effects can also be tested simultaneously. In all cases, null hypothesis is “unit root with a break”, the alternative is “broken trend stationary process”. Later, in [12]. and [10] the authors proposed determining the break point endogenously from the data. It means a unit root hypothesis in the presence of structural change at the unknown time of the break. Endogenous structural break is performed as a sequential test using the full sample and different dummy variables modelling each possible break date. As a resulting break point the date is selected where the t-statistic of unit root coefficient in ADF test is minimal. So, the break point is chosen where the evidence is most near to reject the unit root null. This test is implemented in Eviews and will be used in the application part. The choice of the break point itself is a technicality; being selected an analysis can be made whether it corresponds to certain known event as economic crisis, important changes in government policy, etc.

An argumentation exists that one structural break is insufficient what opens a topic of multiple structural breaks. A short survey as well as the one break point history is given e.g. in [5]. The problems arising with a multiple break points specification are formulated and solved in [2] and comprehensibly described also in [7] or [6]. They propose a Fourier approximation of a time-series in which the essential characteristics of a series can be captured using a small number of its low frequency components. Nor the number nor the break dates need to be known *a priori*.

Starting by Dickey – Fuller test in the form

$$\Delta y_t = \rho y_{t-1} + \gamma_1 + \gamma_2 t + \sum_{k=1}^{n} \gamma_{3k} \sin\left(\frac{2\pi k t}{T}\right) + \sum_{k=1}^{n} \gamma_{4k} \cos\left(\frac{2\pi k t}{T}\right) + \varepsilon_t\,, \quad n \le T/2 \tag{1}$$

with the trigonometric part representing a nonlinearity.

If $\gamma_{31} = \gamma_{41} = \cdots = \gamma_{3n} = \gamma_{4n} = 0$, the process is linear and the traditional unit root test should be used. On the opposite, if there is a break or nonlinear trend, at least one Fourier frequency must be present in the DGP. In practice, a large value of n cannot be applied because of degrees of freedom. That is why the choice of one proper frequency is recommended (see [2]). (1) then changes into

$$\Delta y_t = \rho y_{t-1} + \gamma_1 + \gamma_2 t + \gamma_3 \sin\left(\frac{2\pi k t}{T}\right) + \gamma_4 \cos\left(\frac{2\pi k t}{T}\right) + \varepsilon_t \tag{2}$$

what is regressed for all integer values $1 \le k \le 5$. The regression with smallest RSS gives k to be used. F-test with $H_0: \gamma_3 = \gamma_4 = 0$ not to be reject excludes nonlinearity. In [6], the authors point out the non-standard distribution of relevant F-statistic and publish critical values found by a simulation. Having $T \approx 25$ the critical value is 4.77 (90 %), 5.56 (95 %), 7.61 (99 %).

Data involved in this article are GDPPC (= GDP *per capita*) in millions of Euro. FD (= financial development) measured as financial system deposit divided by GDP, TO (= trade openness) defined as (exports + imports)/GDP, source [3]. Figure 1 shows the graphics.

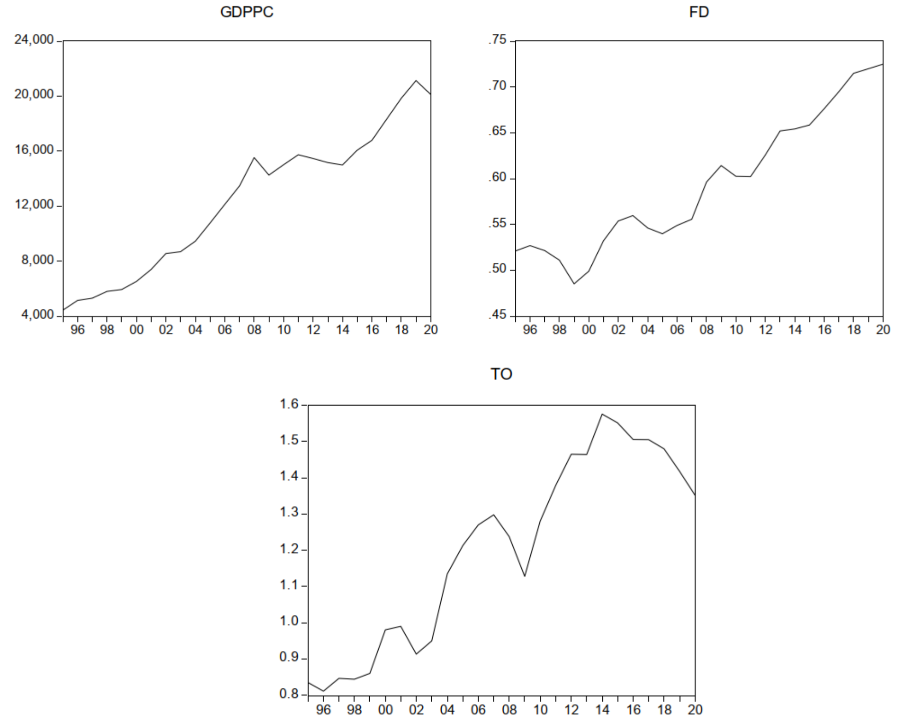


Figure 1 GDPPC, FD, TO, respectively.

For our data, the results of testing according to (2) are in Table 1.

variable	k	F-statistic
GDPPC	2	4.91
FD	5	7.30
TO	1	3.59

Table 1 Fourier test results

We do not reject null at 5 % level for GDPPC and TO and at 1 % level for FD, too. Hence, we exclude a nonlinearity of each time series in question.

ADF Breakpoint unit root test follows with H_0: variable has a unit root. The summarisation is in Table 2.

variable	t-statistic	prob	Break date
GDPPC	-3.9341	0.4134	2006
DGDPPC	-5.8361	0.0100	2008
FD	-5.3686	0.0100	2003
TO	-3.7654	0.5221	2018
DTO	-4.6094	0.0320	2014

Table 2 ADF Breakpoint unit root test results

Analysing the break dates we conclude that 2008 is the first year of the financial crisis 2008 – 09; other dates do not coincide with a known important economic event.

As for the ADF tests (without a break assumed) the results are as follows in Table 3. H_0: variable has a unit root is tested.

variable	t-statistic	prob
GDPPC	-1.9920	0.5772
DGDPPC	-3.8989	0.0282
FD	-4.1236	0.0177
TO	-0.9836	0.9285
DTO	-3.8371	0.0320

Table 3 ADF test results

The ADF test was performed including intercept and trend. The results show that GDPPC and TO are I(1), FD is I(0).

Possible cointegration of the variables was tested by using the Engle – Granger procedure. It is based on the following idea: if all variables in a model are cointegrated then the residuals form a stationary time series. ADF test should be applied without an intercept term because of the zero expected value of disturbances.

(i) $gdpppc = \theta_0 + \theta_1 to + \epsilon$ $\quad\hat{\epsilon}$: *t-stat* = -2.6103 *Prob* = 0.0114

(ii) $gdpppc = \theta_0 + \theta_1 to + \theta_2 fd + \epsilon$ $\quad\hat{\epsilon}$: *t-stat* = -2.4696 *Prob* = 0.0159

In both cases, H_0: variable $\hat{\epsilon}$ has a unit root is rejected (in case (i) the variables included are integrated at the same degree).

The above foundations lead to a proposition to model the process by using the ECM formulation with a dummy variable accenting 2008 - 09 financial crisis.

3 Model and Computations

The ECM part of the model is based on the relation

$$gdppc_t = \propto_0 + \propto_1 fd_t + \propto_2 D * fd_t + \propto_3 to_t + \propto_4 D * to_t + \varepsilon_t$$

in which D equals 1 for 2008, 2009 and 0 in other years and

$$ecm_t = \widehat{gdppc_t} - \widehat{\propto_0} - \widehat{\propto_1} fd_t - \widehat{\propto_2} D * fd_t - \widehat{\propto_3} to_t - \widehat{\propto_4} D * to_t.$$

Parameters $\widehat{\propto_2}, \widehat{\propto_4}$ represent the difference in *fd*, *to*, respectively between the years of financial crisis and the others.

Finally, the model

$$\Delta gdppc_t = \beta_0 + \beta_1 \Delta fd_t + \beta_2 D * \Delta fd_t + \beta_3 \Delta to_t + \beta_4 D * \Delta to_t + \beta_5 ecm_{t-1} + u_t$$

is estimated; parameter β_5 is expected to be negative. Table 4 shows the results (output is harmonized with the variables introduced in equations).

Dependent Variable: ΔGDPPC
Method: Least Squares
Sample (adjusted): 1996 2020
Included observations: 25 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
ΔFD	60.32966	111.6938	0.540134	0.5954
D*ΔFD	866.7797	287.2992	3.016994	0.0071
ΔTO	5.264033	23.28122	0.226106	0.8235
D*ΔTO	317.1772	97.01383	3.269402	0.0040
ECM(-1)	-0.150670	0.125565	-1.199937	0.2449
C	571.1102	189.9598	3.006479	0.0073
R-squared	0.4165	Prob(F-stat)	0.0501	

Table 4 Regression output - Eviews

Parameter β_5 should be negative to functionate as a corrective; the condition is fulfilled but the parameter is statistically non-significant. So are also the parameters associated with increments ΔFD and ΔTO. Evidently, non-zero coefficients occur in crisis years and there is also a non-zero increment.

A possible interpretation is that the process is relatively stable (described by a constant) nevertheless, potential crises show a good characteristic of the system: financial development as well as trade openness are supporting factors of the GDP.

4 Conclusions

The importance of financial development as well as of trade openness on the economic growth is indisputable. Both the variables are increasing during the studied period, though the trade openness is a little bit more turbulent while financial development is trend stationary. According to the definitions, the variables are represented by significantly smaller values than GDP *per capita* and its main components, especially a consumption which is a driver of the GDP growth. It is probably an explanation why an impact of both the variables on GDP *per capita* is not quantitatively apparent by using a standard econometric model. The significant constant could indicate that the process is relatively stable in the long-run.

Nevertheless, the results concerning the years of crisis show that the positive role of growing financial development as well as of trade openness is important, is appreciated and becomes one of supporting factors of an economic growth during turbulent events.

Acknowledgements

The research is 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.

References

[1] Arouri, M., Bourbaker, S. & Nguyen, D. (2014). *Emerging Markets and the Global Economy. A Handbook. Academic* Press, ISBN 978-0-12-411549-1
[2] Enders, W., Lee, J. (2012). The Flexible Fourier Form and the Dickey-Fuller Type Unit Root Tests. Economics Letters, Elsevier, Vol. 117 No. 1 (pp 196-199).
[3] Eurostat, ec.europa.eu/eurostat/data/database
[4] GlobalEconomy.com, Business and economic data, https://www.theglobaleconomy.com/rankings/financial_system_deposits_GDP/
[5] Glynn, J., Perera, N. &Verma, R. (2007). Unit root tests and structural breaks" a survey with applications. https://ro.uow.edu.au/commpapers/455
[6] CHi-Wei-Su. (2012). Flexible Fourier Stationary Test in Purchasing Power Parity for Central and Eastern European Countries, *Ekonomický časopis*, Vol. 60 No. 1 (pp 19 – 31).
[7] Jones, P. M., Enders, W. (not dated, read 2022). On the Use of the Flexible Fourier Form in Unit Roots Tests, Endogenous Breaks, and Parameter Instability, University of Alabama, http://www.time-series.net/yahoo_site_admin/assets/docs/ Flexible_Fourier_Form06172012.3985105.pdf
[8] Paun , C.V., Musetescu, R.C., Topan, V.M. & Danuletiu, D.C. (2019). The Impact of Financial Sector Development and Sophistication on Sustainable Economic Growth, *Sustainability* 2019, 11 https://www.mdpi.com/2071/10/50/11/6/ 1713.pdf
[9] Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis, *Econometrica*, Vol. 57 No. 6 (pp 1361 – 1401).
[10] Perron, P. (1997). Further Evidence of Breaking Trend Functions in Macroeconomic Variables, *Journal of Econometrics*, Vol. 80 No. 2 (pp 355 – 385).
[11] Wang Wei (2016). *Achieving Inclusive Growth in China through Vertical Specialization*. Chandos Publishing, ISBN 978-0-08-100627-6
[12] Zivot, E. & Andrews, K. (1992). Further Evidence On The Great Crash, The Oil Price Shock, and The Unit Root Hypothesis, *Journal of Business and Economic Statistics*, Vol. 10 No. 10 (pp 251 – 270).

Identification of Investment Strategies for Portfolio Selection Utilizing the Markov Switching Model and Optimization Model of Portfolio Selection with Conditional Value-at-Risk

Juraj Pekár [1], Ivan Brezina[2], Marian Reiff [3]

Abstract. The investor decides which products or securities to invest in and chooses the time of investment. The investor is faced with the question of the possible return investment, taking into account the investment risk. The possible return and risk are different in bull and bear markets. The paper presents a method to determine the bear and bull markets using the Markov switching model and then determine investment strategies for these markets using a portfolio selection model using the CVaR risk measure. The proposed method of investment strategy selection was realized based on the historical data of the Dow Jones Industrial Average (DJIA) stock index components from January 1, 2007, to February 22, 2022.

Keywords: Markov switching model, portfolio selection, CVaR, bear market, bull market

JEL Classification: G11, C60
AMS Classification: 91B30, 90C90

1 Introduction

Nowadays, we observe a more frequent alternation of crisis and non-crisis periods. As markets behave differently during these periods, different investment strategies need to be implemented. The Markov switching model can be used to determine hidden states, revealing the state of the market. The result of this approach is the identification and subsequent division of the period analyzed into a bear and bull market, whose revenues are used to analyze the selection of efficient portfolios. One approach to creating an effective equity portfolio is to use a Conditional Value at Risk (CVAR) optimization model. The decision-making assumptions may then include the assumption of the emergence of a crisis and thus the use of the investor's strategy based on the situation in which the market finds itself. The paper aims to present the possible use of the Markov switching model for the division of periods and then determine the investor's strategies. To obtain effective portfolios, we will use the optimization portfolio selection model based on CVaR risk. The analysis was performed on historical data of the components of the Dow Jones Industrial Average (DJIA).

2 Use of financial assets data to create a portfolio

To compare the impact of different types of markets, the Dow Jones Industrial Average (DJIA) components, which is one of the world's best-known stock indices, were selected. The DJIA is a stock index of thirty US companies with the largest and most comprehensive publicly traded stocks in the United States.

In our analysis, we selected the following stocks included in the Dow Jones Industrial Average (DJIA) Index, namely: Apple Inc. (AAPL), Amgen, Inc. (AMGN), American Express Company (AXP), Boeing (BA), Caterpillar Inc. (CAT), Salesforce.com, Inc. (CRM), Cisco Systems (CSCO), Chevron Corporation (CVX), The Walt Disney Company (DIS), Goldman Sachs (GS), The Home Depot (HD), Honeywell International Inc. (HON), IBM (IBM), Intel (INTC), Johnson & Johnson (JNJ), JPMorgan Chase (JPM), The Coca-Cola Company (KO), McDonald's (MCD), 3M (MMM), Merck & Co. (MRK), Microsoft (MSFT), Nike, Inc. (NKE), Procter & Gamble (PG), The

[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.

Travelers Companies (TRV), UnitedHealth Group (UNH), Visa Inc. (V), Verizon Communications (VZ), WBA Walmart (WMT).

Different levels of risk can be used to construct a standard investor task (portfolio selection task) that maximizes the expected return on the portfolio. One such measure is the contingent value at risk (CVaR) rate. Therefore, the aim is to formulate such an investment strategy that will bring the investor, at the chosen expected level of appreciation at the end of the investment horizon, a minimum risk in the form of a CVaR risk measure.

By solving this task, we determine the optimal weights of assets in the portfolio at a set minimum expected return value while minimizing the CvaR risk function. Based on historical prices (weekly data) of selected shares contained in the DJIA stock index, comparisons and analyses of their behavior in the bull and bear markets were carried out from 1.1.2007 to 22.2.2022.

3 Markov switching model

Let's have a Markov chain $\{X_t\} = \{x_0, x_1, x_2, \ldots, x_m\}$, where t acquires non-negative integer values. Imagine that $\{X_t\}$ is "wrapped" in noise-damaged signals. Markov chain $\{X_t\}$ it is hidden due to noise and is not observable in practice. What is observed in the real world is the process $\{Y_t\} = \{y_0, y_1, y_2, \ldots, y_n\}$, which is a function of $\{X_t\}$. Time series $\{Y_t\}$ is a distorted version of the time series $\{X_t\}$ due to the noise we assume, it can be described, for example, by a normal probability distribution. The process $\{Y_t\}$ is a series of signals containing information of the "true" state or mode of the modeled system.

By definition [2], the hidden Markov model is a bivariate discrete process $\{X_t, Y_t\}$, where $\{X_t\}$ is a Markov chain and $\{Y_t\}$ is a sequence of random values of a variable y, which are conditional $\{X_t\}$.

By MacDonald and Zucchini definition [9], hidden Markov chain is a stochastic process consisting of two parts, where the first basic part is an unobservable process $\{X_t\}$, which satisfies the Markov property $P\{x_t | x_{t-1}, x_{t-2} \ldots, x_{t_0}\} = P\{x_t | x_{t-1}\}$ and the second part is an observable process $\{Y_t\}$ satisfying $P\{y_t | y_{t-1}, y_{t-2} \ldots, y_{t_0}, x_{t-1}, x_{t-2} \ldots, x_{t_0}\} = P\{y_t | x_t\}$, which is referred to as a conditionally independent property. Independent because it does not depend on historical values $y_{t-1}, y_{t-2} \ldots, y_{t_0}$ and is conditional only x_t. Then a pair of stochastic processes $\{X_t, Y_t\}$ we call the hidden Markov model with m states. The conditional independent property can be explained as follows. In case we know the condition x_t, value y_t depends only on x_t and is not dependent on previous historical states of the system and previous observations of the variable y, thus, it is conditionally independent. A natural extension of the hidden Markov chain is a variant where the observable process depends not only on the current hidden state but also on previous emission observations. This variant is shown in Figure 1.

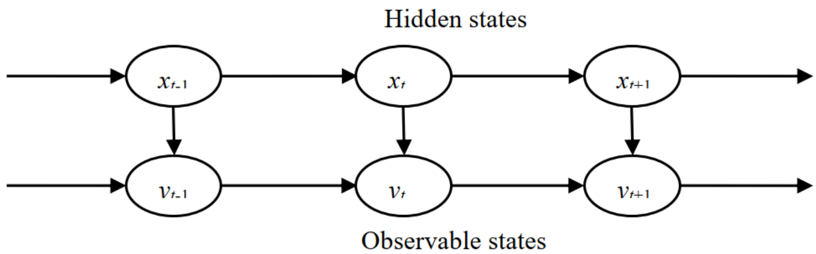


Figure 1 Description of the structure of the Markov switching model, where the variable x_k shows the hidden Markov string and y_k shows the observed process. Source: Own creation

For this variant of the hidden Markov model, the name Markov switching model has become established or Markov regime switching model [4] and [5].

4 Market identification using Markov switching model

In this paper, we analyze the dynamic behavior of the DJIA stock index using the indicators "adjusted close"[4] and its delayed value by one period and information on inflation "Producer Price Index by Commodity: All Commodities (PPIACO)." Time series of weekly data are processed from 1.1.2007 to 22.2.2022. The data source is Yahoo.Finance database [12], where the stock index is kept under the name DJIA, and the Federal Reserve Economic Data database [3], for information on the development of PPIACO inflation information.

All estimates of hidden states were performed using EViews 9 software [1]. We chose the resulting model with two states (modes) for further analysis by comparing different estimated models with different explanatory variables by comparing the statistical significance of the estimated parameters and selecting a model with the minimum value of the Akaike information criterion.

We present the following results for the resulting model:

- matrix of transition probabilities between hidden states of the system and the average length of stay in the respective state: bear and bull market (Figure 2),

- sequence of probabilities of termination in hidden states of the system (Figure 3).

Transition summary: Constant Markov transition probabilities and expected durations
Sample (adjusted): 1 791
Included observations: 791 after adjustments

Constant transition probabilities:
P(i, k) = P(s(t) = k | s(t-1) = i)
(row = i / column = j)

	1	2
1	0.907185	0.092815
2	0.022559	0.977441

Constant expected durations:

	1	2
	10.77409	44.32780

Figure 2 Probabilities of transition between the two hidden states (bear and bull market) and the average length of stay in the respective state. Source: Own calculations

In the next part, we visualize the sequence of estimated probabilities for a given order of observations of DJIA closing course emissions. In Figure 3, these probabilities are shown graphically. In this case, we interpret hidden states as a regime (market mood), a bull, or a bear market. The mean values and the variance of the closing price difference compared to the previous day are estimated for each regime.

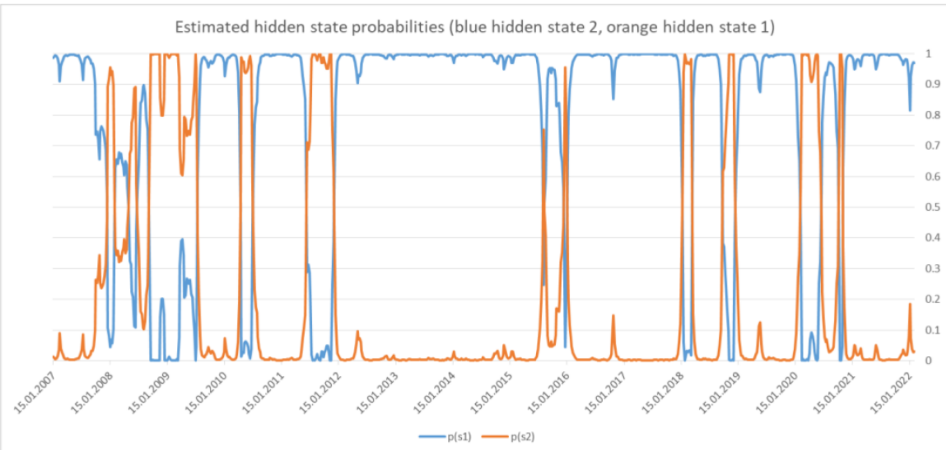


Figure 3 The estimated sequence of termination probabilities in one of the two hidden states of the system. Source: Own calculations

Figure 4 shows the development of the closing price of the DJIA index for the period from 1.1.2007 to 22.2.2022, representing the system's observable state. The bottom line on the graph shows the respective estimated hidden states in which the system is modeled using the Markov switching model with two hidden states. For example, the

[4] https://finance.yahoo.com

interpretation of the orange line is as follows: the modeled system starts on 17.2.2020 in state 2 and 29.6.2020 switches to state 1, and so on.

Figure 4 Markov switching model with two hidden states of the system and the value of the closing price of the DJIA index from January 1, 2007, to February 22, 2022. Source: Own calculations

5 Recommendations for investing based on a portfolio selection model for different market regimes

To compile the portfolio based on the considered historical data (historical weekly returns in the period from 1.1.2007 to 22.2.2022) for 29 DJIA companies (n = 29), we use the mathematical programming model of portfolio selection with Conditional Value-at-Risk [6], [7], [8], [10] and [11]. By solving the problem, we obtain efficient portfolios at the set of different values of expected weekly returns listed in Tables 1 and 2.

Risk CVaR (%)	Expected re-turn (%)	AAPL (%)	AXP (%)	BA (%)	CAT (%)	CRM (%)	DIS (%)	HD (%)	JNJ (%)	JPM (%)	MCD (%)	MMM (%)	MRK (%)	MSFT (%)	PG (%)	TRV (%)	UNH (%)	V (%)	WMT (%)
2.1	0.435	6.0	0.0	2.6	0.0	0.0	5.2	2.3	5.3	0.0	16.6	8.2	3.6	0.0	18.9	5.1	6.2	8.7	11.4
2.2	0.475	5.2	0.0	6.4	0.0	0.0	9.0	7.0	0.0	0.0	14.6	7.7	0.0	2.6	21.2	5.8	8.2	6.4	6.1
2.4	0.516	8.1	2.3	2.0	1.0	0.0	11.6	12.7	2.3	0.0	14.8	2.7	0.0	1.2	12.9	6.1	11.7	5.4	5.4
2.5	0.556	12.4	5.9	1.7	0.0	0.0	12.9	16.5	0.0	0.0	17.8	0.0	0.0	0.0	9.5	6.0	14.9	1.7	0.8
2.8	0.597	11.9	2.3	1.6	0.0	0.0	11.4	21.9	0.0	0.0	13.2	0.0	0.0	6.2	0.0	11.6	20.0	0.0	0.0
3.1	0.637	13.3	0.0	8.8	0.0	0.9	8.3	25.3	0.0	1.0	3.7	0.0	0.0	6.1	0.0	7.1	25.6	0.0	0.0
3.5	0.678	17.2	0.0	9.3	0.0	1.1	0.5	30.3	0.0	1.0	0.0	0.0	0.0	11.4	0.0	0.6	28.1	0.5	0.0
4.1	0.718	36.2	0.0	0.0	0.0	8.7	0.0	36.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	18.4	0.0	0.0

Table 1 Distribution of investments in efficient portfolios during the period of bull markets (unlisted shares have weights equal to 0) Source: Own calculations

Tables 1 (bull markets) and 2 (bear markets) show the calculated solutions. The value of the objective function, representing the minimum CVaR value, is given in the column labeled CVaR Risk. In the following columns, the proportions invested in the individual shares are listed at different expected return values.

Risk CVaR (%)	Expected return (%)	AAPL (%)	AMGN (%)	MRK (%)	V (%)	VZ (%)	WMT (%)
7.1	-0.102	10.6	5.7	0.0	5.4	61.0	17.3
7.1	-0.085	11.4	7.8	0.0	2.0	72.9	5.9
7.2	-0.067	11.5	13.5	2.2	0.0	72.8	0.0
7.5	-0.050	13.7	13.8	16.9	0.0	55.6	0.0
7.9	-0.033	2.7	24.7	21.7	0.0	50.9	0.0
8.4	-0.016	0.2	28.8	32.9	0.0	38.2	0.0
8.9	0.001	0.0	30.9	45.8	0.0	23.3	0.0
9.5	0.018	0.0	39.4	51.8	0.0	8.9	0.0

Table 2 Distribution of investments in efficient portfolios in the period of bear markets (unlisted shares have weights equal to 0) Source: Own calculations

It is clear from Table 1 that the recommendation based on the portfolio selection model using bull market input data is to invest in AAPL, AXP, BA, CAT, CRM, DIS, HD, JNJ, JPM, MCD, MMM, MRK, MSFT, PG, TRV, UNH, V, WMT, which will form the investment portfolio.

When applying the model for input data for bear markets, the investment portfolio consists of shares AAPL, AMGN, MRK, V, VZ, WMT (Table 2). From a comparison of the values in Table 1 and Table 2 obtained from applying the mathematical programming model of portfolio selection with Conditional Value-at-Risk, it is clear that the recommended investments are the same in both periods only for AAPL, MRK, V, WMT shares, which occur in both investment portfolios.

Another result is greater diversification among assets in bull markets, which causes an increase in the return on several types of assets at lower risk values in the given period. From the portfolio structure in the bear market, we observe a recommendation to invest in assets from segments that are not sensitive to consumer sentiment, i.e., IT, healthcare, and consumer goods sales.

6 Conclusion

The alternation of different types of markets affects stock markets. The effects of the change in market sentiment are causing a significant drop in the value of investments in the financial markets, while a temporary fall in stock prices can be observed in the stock markets.

Based on the assumption that worldwide or suppose a regional crisis can arise over some time for any reason. In that case, investors should have the tools to decide which assets under consideration will be possible and necessary to invest with the greatest profit at the lowest risk under different market moods. Therefore, every investor should be interested in alternative investing ways, respectively, different investment models. However, every crisis is temporary, and therefore every investor must decide which assets to continue investing in. Therefore, the investor should know how the situation in the financial markets could develop. Therefore, the investor's attention should focus on alternative ways of investing, respectively, investment models. Investment firms offer various investment forms, for example, a stock market index or a portfolio of shares created by the investment firm. The possible approach and tool for such a decision were the subject of the presented paper.

The Markov switching model can be used to identify the type of market when determining the investment strategy. The authors tried to compare the impact of the identified market type on the return and investment portfolio indicators on the Dow Jones Industrial Average (DJIA) stock index.

A price-weighted DJIA stock index was selected to compare the impact of the change in the type of market. Based on the historical stock prices contained in the DJIA stock index, analyses of this impact were performed in the bull and bear market periods.

For the analysis, the average weekly returns of DJIA shares from 1 January 2007 to 22 February 2022 were used. After dividing the yields into individual groups based on the Markov switching model, the optimization model was used to compile the portfolio based on the considered historical data. The optimization problem solution provides efficient portfolios at set different values of expected weekly returns (Tables 1 and 2). They list the proportions to be invested in individual shares.

The paper analyzes the impact of market type (market mood) on the choice of investor's strategy. The main goal is to analyze the impact of market change reflected in demand for individual stocks. Based on the obtained solutions, a significant influence of the market type can be stated, as the investor's strategy in individual markets is diametrically different. The market restructuring was reflected in the fact that companies that increased sales in the bear market and belonged to the IT, healthcare, and consumer goods sales segment came to the forefront of investment.

Acknowledgements

This work was supported by the Grant Agency of Slovak Republic – VEGA grant no. 1/0339/20 „ Hidden Markov Model Utilization in Financial Modeling “.

References

[1] Aljandali, A. & Tatahi, M. (2018). Economic and financial modelling with eviews. *A Guide for Students and Professionals*. Switzerland: Springer International Publishing.
[2] Cappé, O., Moulines, E. & Rydén, T. (2009). Inference in hidden markov models. In P*roceedings of EUSFLAT conference* (pp. 14-16).
[3] Federal Reserve Economic Data. (2022). Economic data. [Online]. Available at: https://fred.stlouisfed.org/series/PPIACO [Accessed: 14 February 2022].
[4] Hamilton, J. D. (1988). Rational-expectations econometric analysis of changes in regime: An investigation of the term structure of interest rates. *Journal of Economic Dynamics and Control*, 12(2-3), 385-423.
[5] Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, 357-384.
[6] Kang, Z., Li, X., & Li, Z. (2019). Mean-CVaR portfolio selection model with ambiguity in distribution and attitude. *Journal of Industrial & Management Optimization*, 13(5), 1.
[7] Krokhmal, P., Uryasev, T., & Palmquist, J. (2001). Portfolio optimization with conditional value-at-risk objective and constraints. *Journal of Risk*, 4, 43-68.
[8] Liu, N., Chen, Y., & Liu, Y. (2018). Optimizing portfolio selection problems under credibilistic CVaR criterion. *Journal of Intelligent & Fuzzy Systems*, 34(1), 335-347.
[9] MacDonald, I. L. & Zucchini, W. (1997). *Hidden Markov and other models for discrete-valued time series.* Chapman and Hall/CRC.
[10] Pekár, J. (2015). *Modely matematického programovania na výber portfólia*. Bratislava: Vydavateľstvo EKONÓM.
[11] Rockafellar, R. T., & Uryasev, S. (2002). Conditional Value-at-Risk for General Loss Distributions. *Journal of Banking and Finance.* 26(7), 1443−1471.
[12] Yahoo.Finance. (2022). Data on individual stock indices. [Online]. Available at: https://finance.yahoo.com [Accessed: 14 February 2022].

Measuring the Performance of Czech Cluster Organizations: Malmquist Index Approach

Natalie Pelloneová[1]

Abstract. The main aim of this research is to use the Malmquist Index to evaluate the influence of Czech cluster organizations on the financial performance of member companies. Due to the considerable diversity of individual clusters in terms of their date of establishment and due to the availability of financial statements, the research was focused on the period 2012–2017. All cluster organizations operating in the Czech Republic were selected for this research. There were a total of 30 cluster organizations. The input variables selected for the model are total assets and long-term invested capital. The output was the economic value added. The research found that in the Czech Republic, out of the total number of 30 examined clusters, only 12 clusters increased their financial performance in the period 2012–2017. On the other hand, the remaining 18 clusters saw a decline in financial performance during the period under review. The possible causes of this finding are discussed at the end of the paper.

Keywords: Malmquist index, economic value added, cluster organization, performance

JEL Classification: C10, C67, L25
AMS Classification: 90B90, 90C90

1 Introduction

It has been observed in the past that competing firms have a very strong tendency to agglomerate or cluster. These were mainly firms located in a certain narrowly defined geographical area [1]. The agglomeration of related economic activity is a major element of economic geography, which has been studied by a large number of scholars, such as Marshall [12]; Porter [16]; Krugman [11]. Cooperation between several completely independent enterprises can lead to the formation of broad business networks, so-called clusters [4]. Maskell, Bathelt and Malmberg [14] consider clusters to be geographically located agglomerations of firms with similar or highly complementary capabilities. In recent decades, clusters have been widely used as a tool to increase competitive- ness at regional and national level [6]. Kaźmierski [9] perceives clusters as so-called accelerators of progress and competitiveness. This means that stimulating their development has become an important element of regional policy. The cluster concept has become widely used and recognized as an essential part of regional development and policy-making strategies in many countries [10]. The fundamental question is what are the positive effects of cluster policy and what are the implications for entrepreneurs and the enterprises themselves [6].

According to Kincaid [10], clusters offer a range of benefits to all stakeholders. Clusters create an environment conducive to innovation and knowledge creation. For this reason, regions with strong clusters are considered to be leaders in innovation. Clusters also have a positive effect on the economic performance of firms. In particular because they are driven by advantages such as: higher efficiency (lower cost), flexibility (labour mobility), but also innovation (cooperation). Research by many authors has led to the finding that the financial performance of the units involved can be enhanced by geographic (territorial) concentration of economic activity [7]. Many researchers in agglomeration economics believe that the benefits of cluster membership lead to higher financial performance of firms through improved production or increased demand (e.g., Krugman [11]; Marshall [13]). The authors Jirčíková, Remeš and Pavelková [8] also state that it is generally accepted that the geographical colocation of firms has a positive effect on the financial performance of firms in the cluster. Overall, these streams of cluster literature suggest that compared to isolated firms, cluster firms benefit from geographic proximity. If this is the case, it is reasonable to expect that the benefits of clusters should lead to better financial per- formance for cluster firms. However, there is only a limited number of empirical studies in the literature that address this relationship. At the same time, even this limited number of studies that investigate the impact of spatial proximity on financial performance yield ambiguous results [19]. The main objective of the paper is to measure changes in efficiency over the reference period 2012–2017 on the basis of the Malmquist index and to analyze the level of productivity

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

in individual clusters. If the productivity of the whole cluster increases, there is an assumption that the productivity of its members also increases.

2 Malmquist index

In recent years, the Malmquist Index has become the standard approach to measuring productivity. The Malmquis index measures the change in efficiency over time. The Malmquist index is based on Data Envelopment Analysis models and is one of the important indicators for measuring the change in relative efficiency of DMUs over different time periods. The Malmquist index allows the evaluation of multiple inputs and outputs. In assessing changes in efficiency over time, it allows its decomposition into two components, i.e., the change in the relative efficiency of a unit relative to the set of remaining units and the change in the production possibilities frontier induced by technology. Unlike the econometric stochastic frontier approach, it offers a different rate of technological change for each individual. Moreover, since it is estimated with a non-parametric methodology, it needs neither to impose any functional form on the data nor to make any distributional assumptions for the inefficiency term.

The Malmquist index was only processed theoretically until its improvement by Fare et al. [5]. Fare et al. [5] defined an input-oriented productivity index as the geometric mean of the two Malmquist indices developed by Caves et al. [2]. The input-oriented Malmquist index, which measures the change in the efficiency of a production unit q between successive periods t and t+1, can be expressed by equation (1). In this mathematical expression, $EFFCH_q$ is a change in relative technical efficiency and $TECH_q$ expresses a change in the production pos- sibilities frontier due to technological progress. The individual components are defined by relations (2) and (3).

Where $D_q^t(x^t, y^t)$ is the efficiency of production unit q relative to the existing technology in period t with inputs and outputs from period t, $D_q^t(x^{t+1}, y^{t+1})$ is the efficiency of production unit q relative to the existing technology in period t with inputs and outputs from period $t + 1$, $D_q^{t+1}(x^t, y^t)$ is the efficiency of production unit q relative to existing technology in period $t + 1$ with inputs and outputs from period t, $D_q^{t+1}(x^{t+1}, y^{t+1})$ is the efficiency of production unit q relative to existing technology in period $t + 1$. The values of $D_q^t(x^t, y^t)$, $D_q^t(x^{t+1}, y^{t+1})$, $D_q^{t+1}(x^t, y^t)$, $D_q^{t+1}(x^{t+1}, y^{t+1})$ can be obtained by solving DEA models, assuming either CRS or VRS. The Malmquist index can also be expressed directly as the product of the two components $EFFCH_q$ and $TECH_q$ see relation (4).

$$MI_q(x^{t+1}, y^{t+1}, x^t, y^t) = EFFCH_q TECH_q \tag{1}$$

$$EFFCH_q = \frac{D_q^{t+1}(x^{t+1}, y^{t+1})}{D_q^t(x^t, y^t)} \tag{2}$$

$$TECH_q = \left[\frac{D_q^t(x^{t+1}, y^{t+1}) D_q^t(x^t, y^t)}{D_q^{t+1}(x^{t+1}, y^{t+1}) D_q^{t+1}(x^t, y^t)}\right]^{\frac{1}{2}} \tag{3}$$

$$MI_q(x^{t+1}, y^{t+1}, x^t, y^t) = EFFCH_q TECHP_q = \frac{D_q^{t+1}(x^{t+1}, y^{t+1})}{D_q^t(x^t, y^t)} \left[\frac{D_q^t(x^{t+1}, y^{t+1}) D_q^t(x^t, y^t)}{D_q^{t+1}(x^{t+1}, y^{t+1}) D_q^{t+1}(x^t, y^t)}\right]^{\frac{1}{2}} \tag{4}$$

When Farrell's measurement of technical efficiency is used is used in the construction of the Malmquist index, we obtain productivity growth if $MI_q > 1$ and productivity regression if $MI_q < 1$. If $MI_q = 1$, productivity remains unchanged. If $EFFCH_q > 1$, it indicates an improvement in technical efficiency. If $EFFCH_q = 1$, technical efficiency remains unchanged. A value of $EFFCH_q < 1$ implies a regression of technical efficiency. If the value of $TECH_q > 1$, there is a technological progress. If $TECH_q = 1$, there has been no change in technology. A value of $TECH_q < 1$ indicates technological decline.

3 Data and methodology

The main objective of the research was to evaluate the potential impact of cluster organization (CO) on the financial performance of member businesses. Due to the considerable diversity of individual COs in terms of their date of origin and due to the availability of financial statements, the research focused on the period 2012–2017. For this research, all COs for which there is available data were selected. A total of 30 COs were included.

1. **Creation of a database of cluster organisations and member entities**. In the first step, a database of cluster organisations existing in the Czech Republic was created. Then a database of member entities was created for each cluster organisation containing basic information about each member entity.

2. **Collection of financial statements and obtaining data from financial statements**. For the member entities, it was necessary to obtain the necessary data from the financial statements, in particular the balance sheet and the profit and loss account for the years 2012–2017. The obstacle to this step is that not all companies comply with the obligation to publish the selected data in the collection of documents. The commercial database MagnusWeb was used as the main source of accounting data. The second source used is the public register and the collection of documents in the Commercial Register. For the companies, it was necessary to use data from a total of 3,240 financial statements. Of this number, approximately 87% of the financial statements were obtained from the MagnusWeb database. Approximately 5%, i.e. about 170 financial statements had to be manually transcribed and added from the scanned financial statements published in the collection of documents into MS Excel spread- sheets. The remaining approximately 8% of the accounts were not published at all or only the balance sheet was published, which was not usable for further research. Due to the unavailability of financial statements in some years, the number of companies had to be reduced. Due to the relatively short time series, firms with missing financial statements for one year or more had to be eliminated.

3. **Calculation of economic value added**. The economic value added (EVA) indicator was calculated for all business entities. The EVA indicator was calculated using the EVA equity method (see relation 5). Where ROE is the return on equity and E is the book value of equity. The CAPM model was used to estimate the cost of equity (r_e) (see relation 6).

$$EVA = (ROE - r_e)E \tag{5}$$

$$r_e = r_f + \beta(r_m - r_f) \tag{6}$$

Where r_f is the risk-free interest rate and can be defined as the rate of return on government bonds (these rates are shown in Table 1 for the period 2012–2017), $(r_m - r_f)$ is the systematic market risk premium, which is defined as the difference between the expected total return of the capital market (r_m) and the risk-free interest rate (r_f). The market risk premium is based on the rating of the Czech Republic. The market risk premium based on country risk was 6.94%. And β is a coefficient expressing the degree of specific market risk by measuring the sensitivity of the stock to changes in the market portfolio. The β coefficient is calculated according to equation (7). Where d is the income tax rate, D is debt and $\beta_{unlevered}$ is beta of the company financed only by equity (these values were obtained from the website of Damodaran). It should also be noted that the EVA indicator according to the chosen methodology can only be determined for firms with a positive equity value. Therefore, firms that had zero or negative equity value in at least one year were excluded from the research.

Year	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**
Risk-free interest rate (r_f)	2.31 %	2.26 %	1.58 %	0.58 %	0.43 %	0.98%

Table 1 Risk-free rate in 2012–2017 in the Czech Republic [15]

$$\beta_{levered} = \beta_{unlevered}\left[1 + (1-t)\left(\frac{D}{E}\right)\right] \tag{7}$$

4. **Input and output specifications and Malmquist index calculation**. The Malmquist index was then applied to the data. The input variables selected for the model were total assets and long-term invested capital. The out- put was the economic value added (EVA). For each company and period, the values of the distance functions and the individual components of the Malmquist index were determined in the MaxDEA 7 Ultra software environment according to relationships (2) and (3). Finally, the value of the Malmquist index was calculated using relation (4).

4 Research results

Table 2 provides an overview of the average values of the Malmquist index and its components - change in technical efficiency (EFFCH) and technological change (TECH) for COs over the period 2012–2017. The geometric mean was used to calculate these average values.

Cluster organisation	Malmquist index	EFFCH	TECH	Industrial production index
Regional Food Cluster	1.3491	1.3200	1.0221	1.0186
National Wood Cluster	1.1880	1.3367	0.8888	1.0136
Safety and Security Technology Cluster	1.1244	1.0577	1.0631	0.9958
Moravian Aviation Cluster	1.0779	1.4625	0.7371	1.0265
Mechatronics Cluster	1.0713	1.1247	0.9525	1.0347
CGMC	1.0605	0.9905	1.0706	1.0589
Czech Furniture Cluster	1.0445	1.0303	1.0138	1.0473
National Energy Cluster	1.0315	0.9125	1.1304	0.9964
Energy Cluster	1.0238	1.0473	0.9775	0.9964
Packaging Manufacturers Cluster	1.0180	0.9556	1.0653	1.0448
Moravian Forestry Cluster	1.0061	1.0295	0.9772	1.0136
Czech-Slovak Industrial Cluster	1.0027	1.1648	0.8608	1.0366
South Moravian Construction Cluster	0.9988	0.8791	1.1361	1.0099
Atomex Group	0.9981	0.9820	1.0164	0.9964
Hi-Tech Innovation Cluster	0.9900	1.0435	0.9488	0.9958
Biocluster	0.9893	1.0819	0.9144	1.1457
Czech Pellet	0.9765	0.9714	1.0053	1.0136
Clutex	0.9718	0.9633	1.0089	1.0322
Moravian-Silesian Automotive Cluster	0.9657	1.1818	0.8171	1.0857
Precision Engineering Cluster	0.9605	1.1646	0.8247	1.0589
Plastics Cluster	0.9432	0.9850	0.9576	1.0517
Moravian-Silesian Dynamic Drives Cluster	0.9425	0.9542	0.9878	1.0407
National Engineering Cluster	0.9307	0.9641	0.9654	1.0407
CREA Hydro&Energy	0.8730	0.9410	0.9278	0.9964
Network security monitoring cluster	0.8673	0.9446	0.9182	1.0982
Nanoprogress	0.8503	0.9671	0.8792	1.0232
Czech IT Cluster	0.8431	0.9194	0.9170	1.0982
MedChemBio	0.8419	0.8661	0.9721	1.0232
IT Cluster	0.8330	0.8582	0.9706	1.0982
CzechBio	0.8221	0.9961	0.8253	1.0232

Table 2 Malmquist index for each CO in the period 2012–2017

The analysis of the cluster organisations leads to the following conclusions. In the Czech Republic, a total of 12 COs out of the 30 examined increased their financial performance in the period 2012–2017. Of these 12 organisations, 7 were established in a top-down manner (National Energy Cluster, National Wood Cluster, Packaging Manufacturers Cluster, Czech Furniture Cluster, Energy Cluster, CGMC and Safety and Security Technology Cluster) and 5 in a bottom-up manner (Mechatronics Cluster, Moravian Forestry Cluster, Moravian Aviation Cluster, Regional Food Cluster and Czech-Slovak Industrial Cluster).

It can be concluded that out of the 30 COs surveyed, 12 organisations (40% of the total) improved their financial performance. The largest overall change in financial performance was in companies in the Regional Food Cluster, whose performance increased by an average of 35% per year. The second significant change in performance was recorded in the National Wood Cluster, whose members were able to increase financial performance by an average of 19% per year as a result of improved internal efficiency. The third significant change in performance was recorded by the Safety and Security Technology Cluster, whose members were able to increase financial performance by an average of 12% per year as a result of improved internal efficiency and innovation activities. For the remaining 9 organisations, financial performance growth was only moderate or almost stagnant.

According to the decomposition of the Malmquist index into individual components - change in technical efficiency (EFFCH) and technological change (TECH), the growth in performance of these 12 organisations was caused by both components only in the Safety and Security Technology Cluster, the Czech Furniture Cluster and the Regional Food Cluster. These three COs were able to increase the efficiency of the firm's processes (increase in the value of EFFCH) and at the same time to develop higher innovation activity (increase in the value of TECH) during the period studied. For the remaining 9 COs, performance growth was always induced by only one component of the Malmquist index. In the vast majority of cases it was an improvement in the internal technical efficiency of the member companies (Energy Cluster, National Wood Cluster, Mechatronics Cluster, Moravian Forestry Cluster, Moravian Aviation Cluster and Czech-Slovak Industrial Cluster). Such a trend of improvement in internal technical efficiency points to managerial measures aimed at improving the internal efficiency of firms or to economies of scale for firms. For the remaining COs, on the other hand, the increase in financial performance was caused only by positive technological change (CGMC, Packaging Manufacturers Cluster and Energy Cluster). For these COs, managers seem to have focused their attention on innovation activities without major changes in the internal structure of the production organism.

The remaining 18 COs, on the other hand, experienced a decline in financial performance over the period under review. For some organisations, the decline in financial performance was only slight (COs were almost at the level of stagnation) by an average of 0.12% to 1.07% per year (Atomex Group, Biocluster, Hi-Tech Innovation Cluster and South Moravian Construction Cluster). A slight decline in performance by an average of 2.4% to 4% per year was recorded by the Czech Pellet, Clutex, Moravian-Silesian Automotive Cluster and Precision Engineering Cluster. Decreases in performance by an average of 6% to 10% per year were recorded in the Moravian- Silesian Dynamic Drives Cluster, the Plastics Cluster and the National Engineering Cluster. For 7 COs a relatively significant decrease in performance of more than 10% per year was recorded (CREA Hydro&Energy, CzechBio, Czech IT Cluster, IT Cluster, MedChemBio, Nanoprogress and Network security monitoring cluster).

While some of these COs experienced a decline in overall performance, at least one component of the Malmquist index experienced an increase. The Biocluster, the Precision Engineering Cluster, the Moravian-Silesian Automotive Cluster, and the Hi-Tech Innovation Cluster all recorded increases in internal technical efficiency. Positive technological change was recorded by Atomex Group, Clutex, Czech Pellet and the South Moravian Construction Cluster. In terms of financial performance, the worst performers are the firms in the 10 COs, which recorded a decline in both components of the Malmquist index (CREA Hydro&Energy, CzechBio, Czech IT Cluster, IT Cluster, Med-ChemBio, Moravian-Silesian Dynamic Drives Cluster, Nanoprogress, National Engineering Cluster, Plastics Cluster and Network Security Monitoring Cluster). In these firms, internal technical efficiency has deteriorated and at the same time there has been a technological decline.

5 Discussion

The development of the financial performance of firms in Czech COs can be explained by several causes and factors. On the one hand, the causes of this development can be seen in the overall economic development. After the financial crisis, the Czech Republic was in a recessionary gap. The economy experienced an economic downturn in 2012 and 2013. In 2012, GDP fell by 0.8% year-on-year and in 2013 it fell by 0.5% [3]. In contrast, since 2014 the Czech economy has shown growth. Therefore, it can be concluded that the development of the Malmquist index can to some extent follow the development of the Czech economy, especially in the years 2012–2014, and that the overall economic situation in the Czech Republic has thus influenced the development of the performance of cluster companies. However, the dependence of the Malmquist index on economic growth cannot be verified in such a short time series. The Spearman correlation coefficient for all COs reached an indicative value of 0.42 on average.

Another possible factor that can be used to explain the development of the performance of individual clustered enterprises is the development of individual economic sectors. For the 13 COs, a similar performance trends (growth and decline) can be observed (measured by the Malmquist index) as for the industrial production index. For the remaining COs, the development of performance was opposite to that of the industrial production index in the economic sector concerned. However, the dependence of the Malmquist index on the development of the industrial production index cannot be verified over such a short time series. The Spearman correlation coefficient for all KOs was on average 0.45.

Other external factors, such as political decisions or unemployment rates in a particular industry, may also have partly influenced the performance of cluster organisations. Another possible factor may be changes in the production characteristics of individual companies in the CO and other internal factors.

6 Conclusion

The present research was focused on verifying the assumption that membership in CO brings benefits to member firms in the form of increased financial performance over time. For only 12 COs were members able to increase financial performance. In contrast, members of the remaining 18 COs experienced a decline in financial performance over the period studied. However, the results of the research conducted did not confirm a significant effect of CO membership on financial performance growth, as also reported by Kukalis [12] and Ruland [17]. According to the findings, the research is inclined to the opinion of Skokan and Zotykova [18] who believe that the impact of the cluster on the performance of the member firm and on its business results is highly individual and depends on a number of factors. It is also important to recall that during the period under review, the financial crisis in 2012–2013 may have negatively affected the economic performance of some companies and, consequently, the CO as a whole. It is also important to mention that the economic crisis in 2012–2013 affected each industrial area with a different force. It is also interesting to compare the change in the financial performance of COs with the amount of public support provided. For example, the CGMC cluster has drawn the largest amount of all the COs assessed and at the same time its members have increased their financial performance in the peri- od under review. On the other hand, e.g. the Czech IT Cluster also drew significant amounts of public funding, but the financial performance of its members decreased. It is therefore necessary to evaluate each cluster organisation separately with regard to individual influences.

Acknowledgements

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

References

[1] Alcácer, J. & Chung, W. (2014). Location Strategies for Agglomeration Economies. *Strategic Management Journal*, 35(12), 1749–1761.

[2] Caves, D. W., L. R. Christensen & Diewert W. E. (1982). The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica*, 50(6), 1393–1414.

[3] *CZSO. (2020). HDP, národní účty* [online]. Praha: Český statistický úřad [cit. 2020-02-03]. Available at: https://www.czso.cz/csu/czso/hdp_narodni_ucty.

[4] Dvořáček, J. & Tyll, L. (2010). Outsourcing a offshoring podnikatelských činností. Praha: C.H. Beck.

[5] Fare, R. et al. (1994). Productivity developments in swedish hospitals: a malmquist output index approach. In: A. Charnes, W. W. Cooper, A. Y. Lewin & L. M. Seiford (Eds.), *Data envelopment analysis: theory, methodology and applications* (pp. 253–272). Boston: Kluwer Academic Publishers.

[6] Hernández-Rodríguez, C. & Montalvo-Corzo, R. F. (2012). Entrepreneurial Clusters in China and Mexico – implications for Competitiveness. *Revista de Globalización, Competitividad y Gobernabilidad*, 6(1), 55–90.

[7] Irawati, D. (2012). *Knowledge transfer in the automobile industry: global-local production networks*. New York: Routledge.

[8] Jirčíková, E., D. Remeš & Pavelková, D. (2006). Zvyšování výkonnosti podniků prostřednictvím zapojení do průmyslových klastrů. In L. Gurčík et al. (Eds.), *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.

[9] Kazmierski, J. (2013). Redefinicja roli administracji samorządowej w polityce wspierania rozwoju klastrów. *Zarzadzanie Publiczne*, 2(22), 149–159.

[10] Kincaid, B. L. (2005). Competitive advantage of clusters within lesser developed countries of the South Pacific: An empirical case study extending the Porter diamond model. *Ann Arbor*, 246 s. Dissertation (Ph.D.). Capella University. Adviser Anthony Pizur, Ph.D.

[11] Krugman, P. (1991). Increasing Returns and Economic Geography. *The Journal of Political Economy*, 99(3), 483–499.

[12] Kukalis, S. (2010). Agglomeration Economies and Firm Performance: The Case of Industry Clusters. *Journal of Management*, 36(2), 453–481.

[13] Marshall, A. (1920). *Principles of Economics*. London: Macmillan.

[14] Maskell, P., H. Bathelt & Malmberg, A. (2004). *Temporary Clusters and Knowledge Creation: The Effects of International Trade Fairs, Conventions and Other Professional Gatherings*. SPACES Working Paper. Marburg: Fachbereich Geographie, Philipps-University of Marburg.

[15] MPO. 2017. *Finanční analýza podnikové sféry za rok 2017* [online]. Praha: Ministerstvo průmyslu a obchodu [cit. 2018-11-28]. Dostupné z: https://www.mpo.cz/cz/rozcestnik/analyticke-materialy-a- statistiky/analyticke-materialy/financni-analyza-podnikove-sfery-za-rok-2017--237570/.

[16] Porter, M. E. (1990). *The Competitive Advantage of Nations*. New York: Free Press.

[17] Ruland, W. (2013). Does Cluster Membership Enhance Financial Performance? *iBusiness*, 5(1), 1–11.

[18] Skokan, K. & Zotyková, L. (2014). Evaluation of Business Cluster Performance During Its Lifecycle. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 62(6), 1395–1405.

[19] Stuart, T. & Sorenson, O. (2003). The geography of opportunity: Spatial heterogeneity in founding rates and the performance of biotechnology firms. *Research Policy*, 32(2), 229–253.

Measuring the Efficiency of Football Players by DEA Model

Natalie Pelloneová[1], Michal Tomíček[2]

Abstract. Football is a very popular subdivision of sports not only in our country, but also all around the world. This article expands the ideas from the economic literature on efficiency to develop method for evaluating the performance of football players that take into account many dimensions of football performance. The evaluation of football players has always been an important input for strategic decisions in the football industry. The aim of this paper is to apply an input-oriented Data Envelopment Analysis (DEA) model in order to measure football players' efficiency and to identify the technically efficient players with regard to their position on the field. Efficient players will be further ranked using the Andersen-Petersen super efficiency model. The model is empirically applied to players of the first and second Slovak football leagues in the 2020/21 season. The model proposed in this paper seeks to incorporate more objectivity into decision-making and can thus be an important step in developing a systematic methodology for evaluating football players.

Keywords: football player, DEA, efficiency, super-efficiency, sport

JEL Classification: C10, L83, C67, C44
AMS Classification: 90B90, 90C90

1 Introduction

Economics has long been successfully extending its sphere of influence into diverse areas. Football has not been left out of its interest. Many authors are concerned with football, whether at the level of players and clubs or at the level of national teams and international tournaments. Identifying talented players and then forming a team is a very difficult task for football club managers. Evaluating player performance is currently a key issue in the sports industry. Player performance evaluation itself is quite a challenging problem. All over the world, football fans try to determine the ranking of players based on their subjective views or on the basis of various key parameters. Player performance varies from position to position, but it is also based on competition, time played and team style of play. Recently, there has been an emphasis on evaluating player performance in football using statistical methods, aided by the availability of a wide range of data.

In football, financial aspects, such as the transfer market and club investment, influence player performance. The aim of every club is to achieve the best possible performance. It is therefore important to identify players who meet the requirements, with the best potential for return on investment. In general, the market price of football players is also linked to their performance. However, the game performance of football players is related to specific technical skills in offensive or defensive activities [2]. In this paper, we propose an approach to measure the performance of football players based solely on the use of Data Envelopment Analysis (further DEA). This re- search aims to measure the impact of the market value of players in the 1st and 2nd Slovak football leagues on their efficiency level using data from the 2020/21 season with respect to playing position and technical skills. The CCR-I model was used to analyse the relative efficiency of football players. Using Andersen-Petersen (further AP) model, a super-efficiency analysis was further performed to differentiate the efficiency scores between the efficient units and to determine the ranking of the players. Determining the ranking of the best player among goalkeepers, defenders, midfielders and forwards is a difficult task. The research will evaluate each group of players separately and different output variables will be selected for each group of players in the CCR-I model. Among goalkeepers, for example, their successful saves will be evaluated, as well as the interventions of defenders, the assists of midfielders and the goals of attackers.

In the past, the DEA approach has been used to measure efficiency in other contexts. In the last 10 to 15 years, DEA has also been recognised as a common methodology for measuring efficiency in sports such as football, handball, volleyball, etc. For example, Cooper, Ruiz and Sirvent [3] apply the DEA method to the evaluation of basketball players in the Spanish Basketball League. Cooper, Ramón, Ruiz and Sirvent [4] further applied DEA to evaluate basketball players using cross-efficiency evaluation. Suk [13] evaluated the relative efficiency of

[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, michal.tomicek@tul.cz

players in a Korean baseball organization using DEA. His research investigates whether a national baseball team composed of the best players can achieve outstanding results in international competitions. Ramón, Ruiz and Sirvent [11] ranked tennis players using the CCR-O model. Santín [12] uses an output-oriented nonincreasing returns to scale super-efficiency Data Envelopment Analysis model in order to measure the performance of Real Madrid football players. The aim of his research is to identify the historically best players of this famous big club. Papahristodoulou [10] measures the efficiency of several scorers for different football teams.

The present research focuses on Slovak football. In global terms, the popularity of football in Slovakia is quite high. Slovakia has a high number of clubs and players per capita and a relatively long history of football. The number of registered football players in 2008 was up to 428 968, which is about 8% of the population of Slovakia. The highest football competition in Slovakia is called Fortuna Liga and 12 teams participate in it. The second highest Slovak football league is called the 2nd Slovak Football League. A total of 16 teams have participated in the Second Slovak Football League since the 2017/18 season. In the 2020/21 season, 375 players started at least one match in the Fortuna Liga. There were 441 players in the 2nd Slovak League. Measuring the efficiency of these players will be the subject of the research below.

2 Methodology

Data Envelopment Analysis was used to analyse the relative efficiency of Slovak football players. Data Envelopment Analysis can be classified as a modern approach to performance evaluation of production units. The first Data Envelopment Analysis models were proposed in the 1970s. The first DEA models were used to evaluate the efficiency of homogeneous production units. The units evaluated can be, for example, institutions, territorial units, enterprises or even athletes as in the case of this paper. Another assumption is that the evaluated units are comparable to each other [5]. Units are evaluated on the basis of inputs used and outputs produced. The ad- vantage of DEA models is the ability to compare multiple inputs and outputs, with no a priori knowledge or weighting of input data required. Due to this assumption, variables with different units of measurement can be included in the model. DEA models are based on the Farrell model, which dealt with the technical efficiency of production units in the 1950s. He was followed up in the 1970s and 1980s by other experts who built the first DEA models. The two basic DEA models are the CCR model by Charnes, Cooper and Rhodes and the BCC model by Banker, Charnes and Cooper [6;7].

The CCR model with constant returns to scale (further CRS) is based on the assumption that the same change in inputs causes the same change in outputs. In contrast, the BCC model is considered under the assumption of variable returns to scale (further VRS). The models can be further divided into input-oriented and output- oriented models. Using an input-oriented model, it is possible to determine what the quantity of inputs should be to make an inefficient unit efficient. A unit with a technical efficiency coefficient equal to 1 is efficient, a coefficient less than 1 indicates an inefficient unit and determines the degree of input reduction needed to make the unit efficient. Using the output-oriented model, it is possible to determine the quantity of outputs that should be produced to make an inefficient unit efficient. A unit with a technical efficiency coefficient equal to 1 is efficient, a coefficient greater than 1 indicates an inefficient unit and determines the degree of increase in outputs needed to make the unit efficient [7;9].

The CCR-I model is the first model that will be used to analyse the efficiency of Slovak football players. This model assumes constant returns to scale [9] and can be written mathematically by relations (1) and (2). 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 DMU_q.

$$\text{Minimize} \qquad \theta_q \tag{1}$$

$$\begin{aligned} \text{S.t.} \qquad & \sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- = \theta_q x_{iq}, \qquad i = 1,2,\dots,m, \\ & \sum_{j=1}^{n} y_{kj}\lambda_j - s_k^+ = y_{kq}, \qquad k = 1,2,\dots,r, \\ & \lambda_j \geq 0, \quad j = 1,2,\dots,n, \end{aligned} \tag{2}$$

Standard DEA models have a large number of applications and modifications. One of the most important extensions of the DEA model is the formulation of super-efficiency models, which are used to determine the ranking of DMUs with a single efficiency score. The best known model is that of Andersen and Petersen (1993). This model allows an efficient unit to achieve an efficiency greater than 1 for input-oriented models or less than 1 for output-oriented models. The whole concept of super-efficiency is based on extracting a specific efficient production unit from the set under consideration, thus shifting the original efficient frontier. Their main advantage is the possibility to further classify the efficient production units [1]. This is because the arrangement of efficient units is not offered by conventional DEA models. Jablonský and Dlouhý [9] formulate an input-oriented AP model under the CRS assumption by using relations (3) and (4). If the unit under consideration is identified as efficient, then $\theta_q^{AP} > 1$.

$$\text{Minimize} \qquad \theta_q^{AP} \tag{3}$$

$$\begin{aligned} S.t. \qquad & \sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- = \theta_q^{AP} x_{iq}, \quad i = 1, \dots, m, \\ & \sum_{j=1}^{n} y_{kj}\lambda_j - s_k^+ = y_{kq}, \quad k = 1, \dots, r, \\ & \lambda_j \geq 0, \quad j = 1, \dots, n, j \neq q, \\ & \lambda_q = 0. \end{aligned} \tag{4}$$

3 Data and variables

Transfermarkt.com and the commercial database InStat were used as data sources. These sources were supplemented with data from Livesport.cz. Only football players who played at least one match in the 2020/21 season were included in the analysis. The database consists of 375 footballers, including 47 forwards, 179 midfielders, 117 defenders and 32 goalkeepers, who played in the Fortuna Liga in the 2020/21 season. Furthermore, the database consists of 441 football players, including 76 forwards, 166 midfielders, 152 defenders and 47 goalkeepers, who played in the 2nd Slovak Football League in the 2020/21 season.

The only input is the market value of each player (further MV). On the other hand, the research will consider several outputs depending on the playing position of the players. The first common output is the number of minutes played. The importance of a player in a team increases every time he is selected to play. This variable takes into account whether the player played the whole game or came on as a substitute. For the group of goal- keepers, the following output variables were included in the research: close range shots saved, mid range shots saved, long range shots saved, stopped shots, supersaves and accurate passes. For the group of defenders, the following output variables were included in the research: goals, assists, chances created, fouls suffered, successful actions, blocked shots, accurate passes, crosses, defensive challenges won, dribbles successful, ball interceptions, free ball pick ups, ball recoveries. For the group of midfielders, the following output variables were included in the research: goals, assists, chances created, fouls suffered, successful actions, shots on target, accurate passes, crosses, challenges won, dribbles, ball interceptions, free ball pick ups and ball recoveries. For the group of forwards, the following output variables were included in the research: goals, assists, chances created, successful actions, shots on target, accurate passes, crosses and attacking challenges won.

4 Research results

The results of the research are divided into four groups of players according to their position on the field: forwards, midfielders, defenders and goalkeepers. Two variants of the DEA model, the CCR-I model and the AP super-efficiency model, were applied to each group of players. Thanks to the super-efficiency model, the ranking of players can be determined.

The CCR-I model identified a total of 4 forwards as efficient out of a total of 47 forwards in the 1st league (see Table 1). None of the efficient forwards reached the highest market value. These were forwards of average or below average market value. In terms of goals scored, with the exception of Erik Jendrisek, they scored an above average number of goals. Tomas Malec scored the most goals (8 total) of the efficient forwards. The CCR-I model identified a total of 5 forwards out of a total of 76 forwards in the 2nd league as efficient (see Table 1). On average, forwards in the 2nd league needed less minutes played (1658) to be efficient than forwards in the 1st league (1950). On the other hand, efficient forwards in the 2nd league scored more goals on average than efficient forwards in the 1st league. Lukas Gasparovic scored the most goals (11 total) of the efficient forwards. Efficient forwards in the 2nd league also had on average more assists, successful actions, accurate passes and crosses than forwards in the 1st league. The market value of the efficient forwards of the 2nd league was at a lower financial level.

Name	League	Club	MV (th. EUR)	AP	Ranking
Tomas Malec	1^{st}	Senica	250	1.8524	1.
Matej Trusa	1^{st}	Michalovce	200	1.4896	2.
Milos Lacny	1^{st}	Sered	150	1.4583	3.
Erik Jendrisek	1^{st}	Nitra	200	1.1436	4.
Lukas Gasparovic	2^{nd}	Petrzalka	100	1.4805	1.
Gabor Toth	2^{nd}	Komarno	50	1.4676	2.
Patrik Rumansky	2^{nd}	Poprad	25	1.3197	3.
Lutfi Biljali	2^{nd}	Partizan Bardejov	50	1.2341	4.
Marek Kuzma	2^{nd}	Dubnica nad Vahom	100	1.0995	5.

Table 1 Efficient forwards in the 1st and 2nd Slovak football league

The second group consisted of midfielders. From the perspective of the CCR-I model, a total of 4 midfielders were identified as efficient out of a total of 179 midfielders (see Table 2). None of the efficient midfielders reached the highest market value. These were midfielders of below average market value. In the case of the 2nd league, a set of 166 midfielders was analysed. In terms of the CCR-I model, a total of 8 players were identified as efficient (see Table 2). On average, midfielders in the 2nd league needed smaller values of the output variables to be efficient than midfielders in the 1st league. The market value of efficient midfielders in the 2nd league was also in the lower financial level.

Name	League	Club	MV (th. EUR)	AP	Ranking
Igor Zofcak	1^{st}	Michalovce	100	3.0000	1.
Juraj Piroska	1^{st}	Senica	100	1.9041	2.
Alieu Fadera	1^{st}	Pohronie	300	1.2314	3.
Dimitris Popovits	1^{st}	Michalovce	200	1.0918	4.
Matej Rosenberger	2^{nd}	Slovan Bratislava	25	2.1398	1.
Guytho Mijland	2^{nd}	Bardejov	25	1.6951	2.
Sebastian Gembicky	2^{nd}	Petrzalka	25	1.6250	3.
Dominik Malinak	2^{nd}	Poprad	25	1.5860	4.
Vladimir Bajtos	2^{nd}	Trebisov	25	1.3390	5.
Filip Szetei	2^{nd}	Komarno	25	1.0832	6.
Michal Petras	2^{nd}	Dubnica nad Vahom	50	1.0658	7.
Wisdom Uda Kanu	2^{nd}	Trebisov	125	1.0500	8.

Table 2 Efficient midfielders in the 1st and 2nd Slovak football league

In the next part of the research the CCR-I and AP model was applied to a set of 117 defenders of the 1st league. Approximately 8% of the defenders were considered efficient in terms of the CCR-I model. These were defenders who played on average around 2000 minutes. In the 2nd league, out of a total of 152 defenders, a total of 13 defenders were marked as efficient. On average, defenders in the 2nd league needed less minutes played to be efficient than defenders in the 1st league. On the other hand, efficient defenders in the 2nd league scored more goals on average. For the other output variables, the 2nd league defenders achieved similar or lower values. The market value of efficient defenders in the 2nd league was also in a lower financial level than in the 1st league (see Table 3).

Name	League	Club	MV (th. EUR)	AP	Ranking
Peter Mazan	1^{st}	Pohronie	100	2.9980	1.
Lubomir Michalik	1^{st}	Sered	100	1.5408	2.
Martin Chren	1^{st}	Zlate Moravce	75	1.3853	3.

Tomas Hucko	1st	Sered	100	1.3549	4.
Bernard Petrak	1st	Pohronie	150	1.3360	5.
C. Blackman	1st	Dunajská Streda	500	1.1579	6.
Martin Toth	1st	Zlate Moravce	100	1.1044	7.
Jan Maslo	1st	Ruzomberok	200	1.0947	8.
J. Mendez	1st	Dunajská Streda	100	1.0702	9.
Vladimir Kukol	2nd	Podbrezova	50	2.0241	1.
Jakub Balaz	2nd	Slovan Bratislava	25	2.0000	2.
Patrik Leitner	2nd	Zilina II	25	1.5271	3.
Michal Ranko	2nd	Skalica	175	1.4286	4.
Juraj Lacko	2nd	Puchov, Slovan Bratislava	25	1.2933	5.
Peter Vojtovic	2nd	Trebisov	25	1.2916	6.
Juraj Martincek	2nd	Puchov	75	1.1525	7.
Denis Knizka	2nd	Dubnica nad Vahom	50	1.1460	8.
Lukas Simko	2nd	Trebisov	50	1.1327	9.
Jakub Parkan	2nd	Dubnica nad Vahom	50	1.1071	10.
Peter Vosko	2nd	Liptovsky Mikulas	125	1.0936	11.
David Kocik	2nd	Poprad	50	1.0697	12.
Marek Frimmel	2nd	Banska Bystrica	125	1.0587	13.

Table 3 Efficient defenders in the 1st and 2nd Slovak football league

Finally, the CCR-I model and the AP model were applied to the set of 32 goalkeepers of the 1st league. The goalkeepers who emerged from the analysis as efficient are listed in Table 4. These are 5 goalkeepers. It is typical for all goalkeepers to have above-average values in the variables related to saves. For example, Adrian Cho- van was the best goalie in terms of the variables mid-range shots saved and super-saves. The highest market value was attributed to Dominik Greif (1 million euros). However, Dominik Greif has a very low OTE score (0.197) and can be considered inefficient according to the CCR-I model. Table 4 also shows the effective goal- keepers in the 2nd league. From the set of 47 goalkeepers, the goalkeepers Matej Luksch and Milan Vincler were identified as effi- cient from the perspective of the CCR-I model. These goalkeepers were above average on all outcome variables. Matej Luksch recorded maximum shots saved, mid and long range shots saved, stopped shots and accurate passes. Thanks to these great stats, he ultimately achieved an efficient score. It was a goalkeeper with a relatively low market value (75 thousand euros). On average, goalkeepers in the 2nd league had to play more minutes to be efficient than goalkeepers in the 1st league. Efficient goalkeepers in the 2nd league also had more saves on average, except for supersaves, and also more accurate passes. The market value of efficient goalkeepers in the 2nd league was also in a lower financial level than in the 1st league.

Name	League	Club	MV (th. EUR)	AP	Ranking
Tomas Frystak	1st	Senica	150	1.6087	1.
Adrian Chovan	1st	Zlate Moravce	300	1.1917	2.
Tomas Jenco	1st	Pohronie	100	1.1828	3.
Matej Markovic	1st	Michalovce	150	1.1781	4.
Igor Semrinec	1st	Trencin	150	1.0093	5.
Matej Luksch	2nd	Liptovsky Mikulas	75	1.4427	1.
Milan Vincler	2nd	Trebisov	50	1.4403	2.

Table 4 Efficient goalkeepers in the 1st and 2nd Slovak football league

5 Conclusion

The main aim of the paper was to propose an approach for the analysis and evaluation of the best players in two Slovak football competitions. The CCR-I model was used to select the best players. Then the AP super- efficiency model was used to determine the ranking of the best players. The method of Data Envelopment Analysis focused on the analysis of players of two selected football competitions - 1st and 2nd Slovak football leagues. Players were divided according to playing positions on the field and the most appropriate game factors were selected for each group.

The highest number of efficient players in both competitions was found for the defenders. This result is consistent with those published by Tiedemann et al. [14] and Fernández et al. [8]. These authors believe that the largest number of efficient players should be among the defenders. Fernandez et al. [8] believes that the attack actions are initiated in to defense, continue with the midfield and just them in the opposite field. In neither case were the most valuable players of the given positions assigned to the set of efficient players. This fact can be interpreted in such a way that the most expensive players in both competitions should, according to the CCR-I model, achieve higher performance as measured by the aforementioned game statistics. The research also showed that the market value of the efficient players of the 1st league was in a higher financial level than that of the 2nd league players.

Given the current economic and financial situation of football clubs, there is an increased need to know how efficiently a club is using its resources. Efficiency analysis is used to calculate the performance scores of the players and also to determine the lack of aspects and the amount of lack of the inefficient players. The DEA methodology has an advantage to set benchmarks for inefficient players and identifies sources of inefficiency. Along with their general observations and experience, sport managers can take into account the DEA efficiency analysis, when creating teams.

Acknowledgements

The paper was supported by the Technical University of Liberec grant no. 21456 "Factors influencing the economic performance of European sports enterprises" (SGS-2022-1017).

References

[1] Andersen, P. & Petersen, N. C. (1993). A Procedure for Ranking Efficient Units in Data Envelopment Analysis. *Management Science*, 39(10), 1261–1264.
[2] Aydemir, A. E., et al. (2021). A Dimension Reduction Approach to Player Rankings in European Football. *IEEE Access*, 9, 1–1.
[3] Cooper, W. W., Ruiz, J. L. & Sirvent, I. (2009). Selecting non-zero weights to evaluate effectiveness of basketball players with DEA. *European Journal of Operational Research*, 195(2), 563–574.
[4] Cooper, W. W., Ramón, N., Ruiz, J. L. & Sirvent, I. (2011). Avoiding large differences in weights in cross-efficiency evaluations: application to the ranking of basketball player. *Journal of centrum cathedra*: *The Business and Economics Research Journal*, 4(2), 197–215.
[5] Cooper, W. W. (2011). *Handbook on Data Envelopment Analysis*. New York: Springer.
[6] Cooper, W. W., Seiford, L. M. & Zhu, J. (2004). *Handbook on Data Envelopment Analysis*. Boston: Kluwer Academic.
[7] 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.
[8] Fernández, R. C., Núñez, T. G., & Sala-Garrido, R. (2020). Analysis of the Efficiency of Spanish Soccer League Players (2009/10). *Using the Metafrontier Approach. Studies of Applied Economics*, 30(2), 565.
[9] Jablonský, J. & Dlouhý, M. (2015). *Modely hodnocení efektivnosti a alokace zdrojů*. Praha: Professional Publishing.
[10] Papahristodoulou, C. (2007). *The relative efficiency of UEFA Champions League scorers*. Mälardalen University: MPRA paper no. 4943.
[11] Ramón, N., Ruiz, J. L. & Sirvent, I. (2012). Common sets of weights as summaries of DEA profiles of weights: With an application to the ranking of professional tennis players. *Expert Systems with Applications*, 39(5), 4882–4889.
[12] Santín, D. (2014). Measuring the technical efficiency of football legends: who were Real Madrid's all-time most efficient players? *International transactions in operational research*, 21(3), 439–452.
[13] Suk, Y. K. (2014). Selecting the Batters of National Baseball Squad using Data Envelopment Analysis. *Journal of the Korea Academia-Industrial cooperation Society*, 15(1), 165–172.
[14] Tiedemann, T., Francksen, T., & Latacz-Lohmann, U. (2011) Assessing the performance of German Bundesliga football players: a non-parametric metafrontier approach. *Central European Journal of Operations Research*, 19(4), 571–587.

Transport Infrastructure Investment Project Portfolio Optimization Using a Cascade Approach to Solving the Min-Max Problem

Daniel Pilát[1], Karel Ječmen[2], Dušan Teichmann[3], Olga Mertlová[4]

Abstract. Transport infrastructure investment projects are usually very costly and time consuming. The limited capacity of resources does not allow the implementation of all prepared projects at the same time; therefore it is necessary to select a portfolio of projects for implementation. The selection must be made in such a way that not only the requirements of transport policy are met, but also the rules of the investment funds from which the projects are financed. The quality of the selected project portfolio is crucial due to their significant socio-economic impacts represented by indicators representing the usefulness of the buildings in the effective use of available funds. The aim of optimization is to minimize the maximal negative deviations from the minimum values of indicators representing the degree of non-fulfilment.
The article presents one of the possible approaches to solving this problem based on linear programming with the addition of a cascade approach providing improved results.

Keywords: linear programming, min-max problem, cascade approach

JEL Classification: C610
AMS Classification: 90C47

1 Introduction

The scope of the transport network and the quality of its infrastructure have a major impact on the comprehensive development of the economy and the well-being of society. Its growth has a positive impact not only on macroeconomic development, but can also have a major impact on other related areas, such as the energy and exhalation impact of transport, as mentioned by [5].

As part of the implementation of the national and transnational transport strategy, infrastructure managers [5] are preparing investment projects ensuring the development and modernization of the transport network. However, the number of projects proposed for implementation usually exceeds the amount of available financial resources, and therefore it is necessary to select a limited portfolio of projects for implementation by the decision-making bodies. In addition to the criterion of the volume of financial resources, this portfolio must also meet the conditions of investment programs (in the Czech Republic, for example, the Operational Program Transport), which are mainly the fulfilment of indicators considering the social benefits of individual projects in accordance with [6].

In order to make the decision on the selection of the project portfolio as efficient as possible, it can be supported by a systematic optimization approach. An approach based on mathematical programming seems appropriate. E.g. paper [4] compares the use of two approaches to mathematical programming based on linear and goal programming. The output of the LP-based method ensured maximization of the fulfilment of the cumulative value of the indicators, with significant deviations between the individual achieved values of the indicators. In practice, however, it is usually required that all indicators be met as evenly as possible. In the case of LP, it is possible to use a min- max type task to ensure the above requirement when selecting projects. The use of min-max mathematical programming methods is cross-sectional and can be used in many cases of optimization approaches. A typical min- max optimization task mentioned e.g. in [1] is the search for the p-centre. Its principle is to optimize the location of the service centre, often the intervention team, as shown, for example [2], in an effort to find a solution

[1] Czech Technical University in Prague, Faculty of Transport Sciences, Department of Logistics and Management of Transport, pilatdan@fd.cvut.cz.
[2] Czech Technical University in Prague, Faculty of Transport Sciences, Department of Logistics and Management of Transport, jecmekar@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, pokorol1@fd.cvut.cz.

that minimizes the maximum deviation from the required minimum values. When applied to the solved task, it is then a matter of minimizing the maximum negative deviation of the fulfilment of indicators.

In order to ensure an even distribution of the fulfilment of indicators, the classic min-max problem is extended in the article by a cascade approach. An example of the application of the cascade approach is [3] solving the task of optimizing the even distribution of connections on the section of the transport network. The use of cascade approach in project portfolio selection is not well explored in literature yet. In our case, this approach draws inspiration from [3] and ensures that the group of constraints is updated at each step so that those constraints that do not allow further improvement of the result are fixed and only those that allow further improvement are used.

2 Problem formulation

Consider the set I of isolated investment projects, intended for implementation in a predefined period and the set of indicators of a given period K. The indicators can be interpreted as a quantified form of social benefit. Each of the $i \in I$ projects has defined investment costs n_i, which need to be spent on its implementation, and its contribution to the fulfilment of the $k \in K$ indicator called a_{ik}. In addition, for each indicator $k \in K$, assume the quantity w_k, which quantitatively expresses its weight (determined, for example, on the basis of the penalty rate per unit of negative deviation). The task is to create a portfolio of projects for implementation from the set of projects I so that the maximum negative deviation of the fulfilment of $k \in K$ indicators is minimized.

Assume predetermined values of the requirements of the fulfilment of the indicator b_k for $k \epsilon K$, the negative deviation from this maximum value will then be expressed for $k \epsilon K$ as $y_k = b_k - \sum_{i \in I} a_{ik} x_i$. Let $y = max\{y_k; k \in K\}$. The aim of the calculation is to create such a portfolio of selected investment projects that will create the most even distribution of the fulfilment of the value of indicators while minimizing the maximum possible value of the negative deviation y_k for $k \epsilon K$.

Recapitulation of sets and input quantities used in the model:

I	set of projects to be implemented
K	set of indicators
n_i	total financial costs of the projects $i \in I$
N	available funds
b_k	minimum required value of indicator filling $k \in K$
a_{ik}	contribution of the project $i \in I$ to the fulfilment of the indicator $k \in K$
w_k	weight of indicator $k \in K$
M	Prohibitive constant

Variables used in the model:

x_i	binary variable representing project selection (financing) $i \in I$ for implementation; if $x_i = 1$applies after the optimization calculation, the project $i \in I$ will be financed, if $x_i = 0$ applies, the project will not be financed
y	a non-negative variable representing the value of the maximum negative deviation
u_k	perturbation non-negative variable, which identifies the indicators $k \in K$, whose negative deviations from the minimum required values can be reduced in further phases of the calculation; if $u_k > 0$, the indicator remains flexible in the next phase, if $u_k = 0$, the indicator will be fixed with a constant value for all subsequent phases

The mathematical model of the optimization task has the form:

$$f(x, y, u) = M * y - \sum_{k \in K} u_k w_k \rightarrow min \tag{1}$$

subject to

$$\sum_{i \in I} x_i n_i \leq N \tag{2}$$

$$b_k - \sum_{i \in I} a_{ik} x_i \leq y - u_k \qquad for\ k \in K \qquad (3)$$

$$x_i \in \{0,1\} \qquad for\ i \in I \qquad (4)$$

$$y \geq 0 \qquad (5)$$

$$u_k \leq \varepsilon \qquad for\ k \in K \qquad (6)$$

$$u_k \geq 0 \qquad for\ k \in K \qquad (7)$$

The objective function (1) ensures the minimization of the optimization criterion - maximum negative deviations and the cumulation of weighted perturbation variables u_k, where $k \in K$. The prohibitive constant M serves to prioritize the minimization of the maximum negative deviation. Restrictive condition (2) ensures that the available budget is not exceeded. The constraint group (3) creates links between the model variables and defines the upper bound for the variable y. The groups of constraints (4), (5) and (7) define the domains of the variables used in the model. The group of constraint (6) creates an upper bound for the perturbation variable u_k, where $k \in K$ so that its activation cannot affect the gradual fixing of indicators, therefore $\varepsilon \ll min\{a_{ik}\}$. If the variable u_k was not limited from above, it would be set to the highest possible value during the optimization calculation to minimize the objective function. In this case, however, this would not be a cascade approach, as there would be no desirable gradual reduction of the maximum deviation in phases (only the minimization of the negative deviation in the case of one indicator would be guaranteed).

The principle of cascade approach and its application in the model (1) – (7) is based on gradually decreasing or increasing (depending on the type of optimization task) the value of the objective function during the calculation. The calculation is therefore divided into sub-phases - sub-optimization calculations. The number of sub-phases may vary according to the specific task. In the case of the problem addressed in this article, it is a gradual reduction of the maximum negative deviation of the indicators. The negative deviation of the indicators $k \in K$ can be expressed as $y_k = b_k - \sum_{i \in I} a_{ik} x_i$. During the calculation using the model (1) - (7) we then obtain the value y, which is identical with at least one value y_k, where $k \in K$. For one or more indicators for which $y_k = y$ applies, it is necessary to fix the limiting conditions for the next phases, because after the end of the partial optimization calculation, the maximum value has already been reached, which cannot be further reduced. Fixation is performed by replacing the variable y with a constant value for the respective value of $k \in K$ in the constraint condition from the group of constraint conditions (3), which corresponds to the achieved result of the maximum negative deviation y_k in the last completed partial calculation. The calculation is repeated until all indicators are fixed and the target value of the minimum of the maximum negative deviations is reached. The value of the maximum number of partial optimization calculations in the solved problem corresponds to the value from the interval $\langle 1; |K| \rangle$ depending on the number of fixed indicators in individual phases. To fix more indicators in the same partial optimization calculation, if for more than one perturbation variable $u_k = 0$, where $k \in K$. The example of cascade approach and its concrete phases can be seen in chapter 4

3 Computational experiments

Computational experiments with model (1) - (7) were performed on two data sets. The first set of data contains indicators for which no context can be sought and the fulfilment of one does not affect the fulfilment of another indicator. The second set of data contains indicators that are closely related. For these, it can be assumed that if one of the indicators is met, the other indicators will be met at the same time (but this may not always be the case). An example is given in Table 1.

Set 1 - Unrelated	**Set 2 - Related**
Length of reconstructed or modernized railways - TEN-T	Length of reconstructed or modernized railways - TEN-T
Length of reconstructed or modernized roads – TEN-T	New or modernized railway stations and stops
Infrastructure for alternative fuels (filling / charging station)	Length of railway lines in operation equipped with the European Rail Traffic Management System - TEN-T

Table 1 Examples of indicators for data sets

For the first set of data, it is logical that the indicators are not related - if, for example, it is a railway construction project, no roads and infrastructure for alternative fuels will be built in the same project.

However, for the second set of data, a situation often arises in which one project contributes to more than one indicator at a time. For example, if a railway line is to be modernized, it is very likely that railway stations will be modernized in the relevant section, or that railway traffic management systems will be built / modernized.

For a computational experiment with the model, assume in both sets a set of 10 projects and 3 indicators and the same values of the parameter n_i for $i \in I$ for both sets. The input data for the computational experiment are shown in Table 2. The values of parameter w_k has been set to 1 for all $k \in K$ for simplification purposes. The value of prohibitive constant M has been set to 100.

Project	Costs n_i	set 1 - a_{ik}			set 2 - a_{ik}		
		$k = 1$	$k = 2$	$k = 3$	$k = 1$	$k = 2$	$k = 3$
$i = 1$	15	13	0	0	14	7	13
$i = 2$	4	0	2	0	5	2	2
$i = 3$	5	3	0	0	4	1	3
$i = 4$	6	0	0	5	5	2	5
$i = 5$	1	0	0	3	0	0	3
$i = 6$	7	6	0	0	6	2	2
$i = 7$	3	0	1	0	4	1	4
$i = 8$	5	0	3	0	5	3	5
$i = 9$	5	0	1	3	3	1	3
$i = 10$	5	4	0	1	4	1	1
Constraint	N	b_1	b_2	b_3	b_1	b_2	b_3
Con. Value	16	15	10	10	20	15	15

Table 2 input parameters for computational experiment with the model

Note: The meaning of all parameters used in this table is explained in chapter “Problem formulation”.

4 Evaluation of the computational experiment with the model

In the computational experiment with the first set of input data, the result was identical in all phases of the calculation. This is due to the input data, namely the contributions of the projects to the fulfilment of the indicators. As the indicators are not related, the first solution can no longer be changed, because the projects are not able to substitute each other for individual indicators. Changing the project portfolio to reduce the maximum deviation of the indicator cannot then occur because the restrictive conditions for fixed indicators could not be met. An overview of the outputs in all phases is given in Table 3.

Set 1	Project realization										Total costs	Deviation		
	1	2	3	4	5	6	7	8	9	10		1	2	3
Phase 1			X		X			X		X	16	8	7	6
Phase 2			X		X			X		X	16	8	7	6
Phase 3			X		X			X		X	16	8	7	6

Table 3 outputs of a computational experiment on the first data set

Note: fixed values of negative deviations are marked in orange.

In the first phase, the maximum value of the negative deviation was $y = 8$ and corresponded to the negative deviation of indicator 1, ie $y_1 = 8$. This corresponds to the selection of projects for implementation 3, 5, 8 and 10. At the end of the first phase, the value of the negative deviation for the first indicator was fixed at a constant value of 8, followed by the second phase of the calculation. In the second and third phases, the achieved values of negative deviations were gradually fixed to constant values, but the selection of projects did not change, as another selection of projects for implementation could not provide already fixed values of indicators - specifically the value of maximum negative deviation 8 for indicator 1 cannot be achieved by selection other projects, therefore the investment portfolio remains the same at all stages.

In the calculation experiment on the second set of data, the project portfolio gradually changed during the calculation, as the fixed values of the indicator fulfilment could be achieved by another combination of projects. The phases thus ensured a gradual reduction of the maximum negative deviation for the individual indicators.

Set 2	Project realization										Total costs	Deviation		
	1	2	3	4	5	6	7	8	9	10		1	2	3
Phase 1	X										15	6	8	2
Phase 2		X		X				X			15	5	8	3
Phase 3		X		X	X			X			16	5	8	0

Table 4 outputs of a computational experiment on the second data set

Note: fixed values of negative deviations are marked in orange.

In the first phase of the calculation, the maximum negative deviation $y = 8$ was reached, which corresponds to the negative deviation of the filling from the second indicator $y_2 = 8$. Project 1 was selected for implementation and the value of the negative deviation for indicator 2 was fixed for further calculation by replacing the variable y with a constant value of 8. In the second phase the maximum value of negative deviations not yet fixed was $y = 5$. These values correspond to the investment portfolio from projects 2, 4 and 8. For further calculation, the value of the negative deviation from the first indicator was fixed by replacing the variable y with a constant value of 5. In the third phase, the maximum flexible negative deviation $y = 0$ was calculated. of the third indicator $y_3 = 0$. These resulting values correspond to the selection of projects 2, 4, 5 and 8. We replace the flexible variable y for indicator 3 with a constant value of 0, which fixes all indicators and therefore the calculation ends.

5 Conclusion

In terms of the suitability of the article presented in investment decision-making in transport infrastructure, the cascade approach can be assessed on the basis of a computational experiment as unsuitable for a type 1 data set, which is characterized by unrelated projects. For a type 2 data set, a cascading approach can already be used, but it carries risks that can affect finding the optimal solution. For example, it may happen that the maximum deviations in a certain phase can be achieved by different combinations of projects for different indicators and the specific selection of projects and fixation of indicators in a given phase will affect the calculation in subsequent phases and the final solution. In order to guarantee a global optimal solution, the method would need to be significantly modified and interphases should be introduced, which would examine other permissible solutions with the same value of the maximum negative deviation y after each phase, thus dividing the calculation into alternatives. An example is given for a better understanding.

Consider the same type of task as in the chapter Problem formulation and the following input data:

Project ID	Costs n_i	Contribution to indicators a_{ik}		
		$k = 1$	$k = 2$	$k = 3$
$i = 1$	4	1	1	0
$i = 2$	6	2	3	1
$i = 3$	6	3	2	2
Constraint	N	b_1	b_2	b_3
Con. value	10	5	5	2

Table 5 input data for illustrative example

The outputs of the computational experiment are presented in the following 2 tables, which present 2 alternative solutions to the problem.

Alternative 1	Project realization			Total costs	Deviation		
	1	2	3		1	2	3
Phase 1	X	X		10	2	1	1
Phase 2	X	X		10	2	1	1
Phase 3	X	X		10	2	1	1

Table 6 outputs of alternative 1 of the illustrative example

Alternative 2	Project realization			Total costs	Deviation		
	1	2	3		1	2	3
Phase 1	X		X	10	1	2	0
Phase 2	X		X	10	1	2	0
Phase 3	X		X	10	1	2	0

Table 7 outputs of alternative 2 of the illustrative example

In the first phase, there are 2 project selection alternatives that have ensured the value of the maximum negative deviation $y = 2$:

- alternative 1 projects 1 a 2, the selection of which will ensure $y_1 = 2$,
- alternative 2 projescts1 a 3, the selection of which will ensure $y_2 = 2$.

The computer software selected only 1 alternative and continued the calculation. However, as Tables 5 and 6 show, option 2 will provide a better solution to the problem, as the overall negative deviation is 1 less than alternative 1. To guarantee the global optimum of the problem, it is therefore necessary to introduce interphases. The authors of the article will deal with the issue of creating interphases in future research.

The aim of the cascade approach in the solved problem is to minimize the maximum negative deviation from the minimum requirement to meet the indicators. This could be used in practice if the goal was to minimize penalties for non-compliance with the minimum required value of indicators.

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] Hassin, R., Levin, A. & Morad, D. (2003). Lexicographic local search and the p-center problem. *European Journal of Operational Research*, 151(2), pp. 265-279, doi: 10.1016/S0377-2217(02)00825-1.

[2] Janáček, J. & Kvet, M. (2018). Min-Max Fair Emergency System with Randomly Occupied Centers. *Operations Research Proceedings 2016*. pp. 307-312, doi: 10.1007/978-3-319-55702-1_41.

[3] Janáčková, M. (2010). Kaskádový přístup k řešení minimaxových úloh. *Úlohy diskrétní optimalizace v dopravní praxi*, pp. 94-102.

[4] Ječmen, K., Pilát, D., Teichmann, D., & Mertlová, O. (2022). Comparison of Two Selected Approaches to the Transport Infrastructure Investment Projects Portfolio Selection for Implementation. *Proceedings of the International Scientific Conference Quantitative Methods in Economics*, In print.

[5] Kalantzis, F., Arnoldus, P. & Brons, M. (2015). Infrastructure in the EU: developments and impact on growth. *Publications Office*. doi:10.2765/85301.

[6] Regulation (EU) No 1303/2013 of the European Parliament and of the Council of 17 December 2013 laying down common provisions on the European Regional Development Fund, the European Social Fund, the Cohesion Fund, the European Agricultural Fund for Rural Development and the European Maritime and Fisheries Fund and laying down general provisions on the European Regional Development Fund, the European Social Fund, the Cohesion Fund and the European Maritime and Fisheries Fund and repealing Council Regulation (EC) No 1083/2006 (2013) *Official Journal of the European Union* L 347, 20 December, pp. 320-469.

Interval Versions of Eigenspaces in Idempotent Semirings

Ján Plavka[1]

Abstract. Idempotent semiring is a bounded linearly ordered set S equipped with two binary operations addition and multiplication, where addition is idempotent, S with addition is a commutative monoid, S with multiplication is a monoid and multiplication left and right distributes over addition. A vector x is said to be an eigenvector of a square matrix A if Ax = λx for some λ ∈ S. This paper investigates the properties of eigenspace for matrices and vectors with interval coefficients in idempotent semirings. An interval vector X is said to be a strong eigenvector of a square matrix A if Ax = λx holds for each x in X and for some λ ∈ S. We suppose that an interval vector X and an interval matrix A can be split into two subsets according to forall–exists quantification of its interval entries. The properties of various versions of eigenspaces in idempotent semirings are studied and characterizations of equivalent conditions are presented.

Keywords: eigenspace, interval, eigenvector

JEL Classification: C60
AMS Classification: 08A72, 90B35, 90C47

1 Motivation

This paper deals with a problem of additively idempotent (a ⊕ a = a) semirings, which are one of the sub-areas of tropical mathematics. Note that the operation of taking maximum of two numbers is the simplest and the most useful example of an idempotent addition.

Idempotent semirings can be used in a range of practical problems related to scheduling and optimization, statistics and game theory, information and data fusion, decision making support, risk management and probability theory. Let us also mention some connections between idempotent algebra and fuzzy sets theory [2], [3].

The idempotent semiring is defined over a bounded linearly ordered set and uses the binary operation of maximum and one of the triangular norms, T, instead of the conventional operations of addition and multiplication. As usual, two arithmetical operations are naturally extended to matrices and vectors.

Application model: The development of linear algebra over idempotent semirings was motivated by multi-entity interaction processes. In these processes we have n entities which work in stages, and in the algebraic model of their interactive work, entry $x_i^{(k)}$ of a vector $x^{(k)} \in \mathbb{B}(n)$ where $i \in \{1,\dots,n\}$ and $\mathbb{B}$ is an idempotent semiring, represents the state of entity i after some stage k, and the entry a_{ij} of a matrix $A \in \mathbb{B}(n,n)$, where $i, j \in \{1,\dots,n\}$, encodes the influence of the work of entity j in the previous stage on the work of entity i in the current stage. Summing up all the influence effects multiplied by the results of previous stages, we have $x_i^{(k+1)} = \bigoplus_j a_{ij} \otimes x_j^{(k)}$. In the case of $\oplus = \max$ this "summation" is often interpreted as waiting till all the processes are finished and all the necessary influence constraints are satisfied. When a multi-entity interaction processes reaches a steady state, after some time of operation, then the state vectors of the steady states are eigenvectors of A ($A \otimes x = x$). The orbit $x,\ A \otimes x, \dots A^k \otimes x$, where $A^k = A \otimes \dots \otimes A$, represents the evolution of such process.

The other application can be seen in the paper [7], where Interactive Cash-Flow System is considered with Łukasiewicz T-norm.

The eigenproblem in idempotent semirings has been described in many monographs and papers, see [1], [5], [17], [22]-[26]. Interesting results describing the structure of the eigenspace and several algorithms for computing the largest eigenvector of a given matrix have been published, for example, in [8]. The eigenvectors in idempotent semirings are useful in fuzzy set theory. Such eigenvectors have been studied in [6], [22].

This paper generalizes the results presented in [5], [17] for max-plus/min algebras and investigates the properties of eigenspace for matrices and vectors with interval coefficients in idempotent semirings. An interval vector X is said to be a strong eigenvector of a square matrix A if Ax = λx holds for each x in X and for some λ ∈ S. We

[1] Technical University, Department of Mathematics and Theoretical Informatics, Nemcovej 32, 04200 Košice, Slovakia, Jan.Plavka@tuke.sk

suppose that an interval vector X and an interval matrix A can be split into two subsets according to forall–exists quantification of its interval entries. The properties of various versions of eigenvectors, namely strong, EA/AE strong eigenvectors and EA strong EA/AE-eigenvectors in idempotent semirings are studied and characterizations of equivalent conditions are presented.

2 Preliminaries and basic definitions

Idempotent semiring is a triple (S, ⊕, ⊗), where S is a bounded linearly ordered set S equipped with two binary operations addition ⊕ and multiplication ⊗, where ⊕ is idempotent, (S, ⊕) is a commutative monoid, (S, ⊗) is a monoid and ⊗ left and right distributes over ⊕.

In the paper we will suppose that an idempotent semiring contains values in the unit interval $S = \langle 0, 1\rangle$ and uses the binary operation of maximum and one of the triangular norms, T (Gödel norm, Łukasiewicz norm and product norm), instead of the conventional operations of addition and multiplication.

The symbol $S(m, n)$, respectively, $S(n)$, denotes the set of all matrices (respectively, vectors) of the given dimensions over S.

The triangular norms (T-norms, in short) were introduced in the context of probabilistic metric spaces. The T-norms are interpreted as the conjunction in multi-valued fuzzy logics, or as the intersection of fuzzy sets ([9]).

The simplest semiring is the Gödel semiring ($x \oplus y = \max(x, y)$, $x \otimes y = \min(x, y)$), and the conjunction is defined as the minimum of the entries: the truth degrees of the constituents. Gödel logic is considered as a logic of relative comparison.

The Łukasiewicz semiring ($x \oplus y = \max(x, y)$, $x \otimes y = \max\{x + y - 1, 0\}$) is often considered as a logic of absolute (or metric) comparison.

The product semiring ($x \oplus y = \max(x, y)$, $x \otimes y = x \cdot y$) is used in applications of discrete event dynamic systems.

The operations $\oplus, \otimes$ are extended to the matrix-vector algebra over S by the direct analogy to the conventional linear algebra. If each entry of a matrix $A \in S(n, n)$ (a vector $x \in S(n)$) is equal to 0 we shall denote it as $A = 0$ ($x = 0$).

Let $x = (x_1, \ldots, x_n) \in S(n)$ and $y = (y_1, \ldots, y_n) \in S(n)$ be vectors. We write $x \leq y$ if $x_i \leq y_i$ holds for each $i \in N$.

For a given matrix $A \in S(n, n)$, the number $\lambda \in S$ and the n-tuple $x \in S(n)$ are the so-called *eigenvalue* and *eigenvector* of A, respectively, if

$$A \otimes x = \lambda \otimes x.$$

An *eigenspace* $V(A, \lambda)$ is defined as the set of all eigenvectors of A with associated eigenvalue λ, i.e.,

$$V(A, \lambda) = \{x \in S(n);\ A \otimes x = \lambda \otimes x\}.$$

3 Interval versions of eigenvectors

Analogously to [4]–[6], [10]-[17], [19]-[21] consider interval matrices $\boldsymbol{A}$ with bounds $\underline{A}, \overline{A} \in S(m, n)$ and an interval vector $\boldsymbol{X}$ with bounds $\underline{x}, \overline{x} \in S(n)$ which are defined as follows:

Definition 3.1. Let $\underline{A} = (\underline{a}_{ij})$, $\overline{A} = (\overline{a}_{ij}) \in S(n, n)$, $\underline{A} \leq \overline{A}$ and $\underline{x}, \overline{x} \in S(n)$, $\underline{x} \leq \overline{x}$. An interval matrix $\boldsymbol{A}$ with bounds $\underline{A}, \overline{A}$ and interval vector $\boldsymbol{X}$ are defined as follows

$$\boldsymbol{A} = [\underline{A}, \overline{A}] = \left\{ A \in S(n, n);\ \underline{A} \leq A \leq \overline{A} \right\}, \quad \boldsymbol{X} = [\underline{x}, \overline{x}] = \left\{ x \in S(n);\ \underline{x} \leq x \leq \overline{x} \right\}.$$

If each element of $\boldsymbol{X}$ is associated either with the universal, or with the existential quantifier, then we can split the interval vector as $\boldsymbol{X} = \boldsymbol{X}^{\forall} \oplus \boldsymbol{X}^{\exists}$, where $\boldsymbol{X}^{\forall}$ is the interval vector comprising universally quantified coefficients and $\boldsymbol{X}^{\exists}$ concerns existentially quantified coefficients. More precisely:

Definition 3.2. Let $\boldsymbol{X} \subseteq S(n)$ be given and $N = N^{\exists} \cup N^{\forall}$, $N^{\exists} \cap N^{\forall} = \emptyset$. Interval vector $\boldsymbol{X}^{\exists}$ is called

- *EA–vector* if $\underline{x}_i^{\forall} = \overline{x}_i^{\forall} = 0$ for each $i \in N^{\forall}$ and $[\underline{x}_i^{\exists}, \overline{x}_i^{\exists}] = [\underline{x}_i, \overline{x}_i]$ for each $i \in N^{\exists}$,
  and interval vector $\boldsymbol{X}^{\forall}$ is called
- *AE–vector* if $\underline{x}_i^{\exists} = \overline{x}_i^{\exists} = 0$ for each $i \in N^{\exists}$ and $[\underline{x}_i^{\forall}, \overline{x}_i^{\forall}] = [\underline{x}_i, \overline{x}_i]$ for each $i \in N^{\forall}$,

Example 3.1. Suppose that $S = [0, 1]$. Consider interval vector $\boldsymbol{X}$. Then $\boldsymbol{X}^{\exists}$, $\boldsymbol{X}^{\forall}$ have the forms

$$\boldsymbol{X} = \begin{pmatrix} [0.1, 0.2] \\ [0.1, 0.3] \\ [0.1, 0.2] \\ [\,0, 0.1] \\ [0.3, 0.4] \end{pmatrix} \quad \boldsymbol{X}^{\exists} = \begin{pmatrix} [0.1, 0.2] \\ [0.1, 0.3] \\ [\,0,\ 0] \\ [\,0,\ 0] \\ [0.3, 0.4] \end{pmatrix} \text{ and } \boldsymbol{X}^{\forall} = \begin{pmatrix} [\,0,\ 0] \\ [\,0,\ 0] \\ [0.1, 0.2] \\ [\,0, 0.1] \\ [\,0,\ 0] \end{pmatrix},$$

where $N^{\exists} = \{1, 2, 5\}$ and $N^{\forall} = \{3, 4\}$.

Similarly, 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 $(i, j) \in \tilde{N}^{\exists}$.

In the rest of this paper, we assume that semiring $(S, \oplus, \otimes)$, $\boldsymbol{A} = [\underline{A}, \overline{A}]$, $\boldsymbol{X} = [\underline{x}, \overline{x}]$ are given and that $\lambda = 1$.

Definition 3.3. Let A, $\boldsymbol{X} = \boldsymbol{X}^{\forall} \oplus \boldsymbol{X}^{\exists}$ be given. Interval vector $\boldsymbol{X}$ is called

- *EA–eigenvector* of A if
$$(\exists x^{\exists} \in \boldsymbol{X}^{\exists})(\forall x^{\forall} \in \boldsymbol{X}^{\forall})\, A \otimes (x^{\exists} \oplus x^{\forall}) = (x^{\exists} \oplus x^{\forall}),$$
- *AE–eigenvector* of A if
$$(\forall x^{\forall} \in \boldsymbol{X}^{\forall})(\exists x^{\exists} \in \boldsymbol{X}^{\exists})\, A \otimes (x^{\exists} \oplus x^{\forall}) = (x^{\exists} \oplus x^{\forall}).$$

For each pair $i, j \in N$, we define $A^{(ij)} \in S(n, n)$ and $x^{(i)} \in S(n)$ by putting for every $k, l \in N$,

$$a_{kl}^{(ij)} = \begin{cases} \overline{a}_{ij}, & \text{for } k = i,\ l = j \\ \underline{a}_{kl}, & \text{otherwise,} \end{cases} \quad , \quad x_k^{(i)} = \begin{cases} \overline{x}_i, & \text{for } k = i \\ \underline{x}_k, & \text{otherwise.} \end{cases}$$

It is easy to show that every $A \in \boldsymbol{A}$ can be written as a max-min linear combination of *generators* $A^{(ij)}$ with $i,\ j \in N$. Similarly, every $x \in \boldsymbol{X}$ is equal to a max-min linear combination of generators $x^{(i)}$ with $i \in N$.

Lemma 3.1. [5, 6, 17, 18] Let $x \in S(n)$ and $A \in S(n, n)$. Then,

(i) $x \in \boldsymbol{X}$ if and only if $x = \bigoplus_{i \in N} \beta_i \otimes x^{(i)}$ for some $\beta_i \in S \cap Y_i$

(ii) $A \in \boldsymbol{A}$ if and only if $A = \bigoplus_{i,j \in N} \alpha_{ij} \otimes A^{(ij)}$ for some $\alpha_{ij} \in S \cap Y_{ij}$,

where

$$Y_i = [\underline{x}_i, \overline{x}_i],\ Y_{ij} = [\underline{a}_{ij}, \overline{a}_{ij}] \text{ for max-min semiring,}$$

$$Y_i = [\frac{\underline{x}_i}{\overline{x}_i}, 1],\ Y_{ij} = [\frac{\underline{a}_{ij}}{\overline{a}_{ij}}, 1] \text{ for max-prod semiring,}$$

$$Y_i = [\underline{x}_i - \overline{x}_i + 1, 1],\ Y_{ij} = [\underline{a}_{ij} - \overline{a}_{ij} + 1, 1] \text{ for max-Łukasiewicz semiring.}$$

Theorem 3.1. [5, 17] Suppose that $(S, \oplus, \otimes)$, $A \in S(n, n)$, $\boldsymbol{X} = \boldsymbol{X}^{\forall} \oplus \boldsymbol{X}^{\exists}$ are given and $x^{(n+1)} := \underline{x}^{\forall}$. Then $\boldsymbol{X}$ is EA–eigenvector if and only if

$$(\exists x^{\exists} \in \boldsymbol{X}^{\exists})(\forall i \in N^{\forall} \cup \{n + 1\})\, A \otimes (x^{\exists} \oplus x^{(i)}) = x^{\exists} \oplus x^{(i)}.$$

4 Strong eigenvectors

Definition 4.1. Let $\boldsymbol{A}$, $\boldsymbol{X}$ be given. Then interval vector $\boldsymbol{X}$ is called *strong eigenvector* of $\boldsymbol{A}$ if

$$(\forall A \in \boldsymbol{A})(\forall x \in \boldsymbol{X})\, A \otimes x = x.$$

Theorem 4.1. Suppose that $(S, \oplus, \otimes)$, $\boldsymbol{A} = [\underline{A}, \overline{A}]$ and $\boldsymbol{X} = [\underline{x}, \overline{x}]$ are given. Then $\boldsymbol{X}$ is a strong eigenvector of $\boldsymbol{A}$ if and only if $\underline{A} \otimes x^{(k)} = x^{(k)}$ and $\overline{A} \otimes x^{(k)} = x^{(k)}$ for all $k \in N$.

Proof. Let us assume that $x \in \mathbf{X}$, $\underline{A} \otimes x^{(k)} = x^{(k)}$ and $\overline{A} \otimes x^{(k)} = x^{(k)}$ for all $k \in N$. Then for arbitrary $x \in \boldsymbol{X}$, $x = \bigoplus_{i \in N} \beta_i \otimes x^{(i)}$ for some $\beta_i \in S \cap Y_i$ we get

$$\underline{A} \otimes x = \underline{A} \otimes \bigoplus_{i=1}^{n} \beta_i \otimes x^{(i)} = \bigoplus_{i=1}^{n} \beta_i \otimes (\underline{A} \otimes x^{(i)}) = \bigoplus_{i=1}^{n} \beta_i \otimes x^{(i)} = x$$

and

$$\overline{A} \otimes x = \overline{A} \otimes \bigoplus_{i=1}^{n} \beta_i \otimes x^{(i)} = \bigoplus_{i=1}^{n} \beta_i \otimes (\overline{A} \otimes x^{(i)}) = \bigoplus_{i=1}^{n} \beta_i \otimes x^{(i)} = x.$$

The assertion follows from the monotonicity of operations, i.e., $x = \underline{A} \otimes x \leq A \otimes x \leq \overline{A} \otimes x = x$ for each $A \in \boldsymbol{A}$. The converse implication is trivial. □

Definition 4.2. Let $\boldsymbol{A}$, $\boldsymbol{X}$ be given. Then interval vector $\boldsymbol{X}$ is called

- *strong EA-eigenvector* of $\boldsymbol{A}$ if

$$(\forall A \in \boldsymbol{A})(\exists x^{\exists} \in \boldsymbol{X}^{\exists})(\forall x^{\forall} \in \boldsymbol{X}^{\forall})\, A \otimes (x^{\exists} \oplus x^{\forall}) = x^{\exists} \oplus x^{\forall},$$

- *strong AE-eigenvector* of $\boldsymbol{A}$ if

$$(\forall A \in \boldsymbol{A})(\forall x^{\forall} \in \boldsymbol{X}^{\forall})(\exists x^{\exists} \in \boldsymbol{X}^{\exists})\, A \otimes (x^{\exists} \oplus x^{\forall}) = x^{\exists} \oplus x^{\forall}.$$

Theorem 4.2. Suppose that $(S, \oplus, \otimes)$, $\boldsymbol{A} = [\underline{A}, \overline{A}]$ and $\boldsymbol{X} = \boldsymbol{X}^{\forall} \oplus \boldsymbol{X}^{\exists}$ are given. An interval vector $\boldsymbol{X}$ is strong EA-eigenvector of $\boldsymbol{A}$ if and only if

$$(\exists x^{\exists} \in \boldsymbol{X}^{\exists})(\forall x^{\forall} \in \boldsymbol{X}^{\forall})\, \underline{A} \otimes (x^{\exists} \oplus x^{\forall}) = x^{\exists} \oplus x^{\forall} \ \wedge\ \overline{A} \otimes (x^{\exists} \oplus x^{\forall}) = x^{\exists} \oplus x^{\forall}.$$

Proof. The assertion follows from Theorem 4.1. □

Theorem 4.3. Suppose that $(S, \oplus, \otimes)$, $\boldsymbol{A} = [\underline{A}, \overline{A}]$ and $\boldsymbol{X} = [\underline{x}, \overline{x}]$ are given. An interval vector $\boldsymbol{X}$ is strong EA-eigenvector of $\boldsymbol{A}$ if and only if

$$(\exists x^{\exists} \in \boldsymbol{X}^{\exists})(\forall k \in N^{\forall} \cup \{n+1\})\underline{A} \otimes (x^{\exists} \oplus x^{(k)}) = x^{\exists} \oplus x^{(k)} \ \wedge\ \overline{A} \otimes (x^{\exists} \oplus x^{(k)}) = x^{\exists} \oplus x^{(k)}.$$

Proof. The assertion follows from Theorem 3.1 and Theorem 4.2. □

Theorem 4.4. Suppose that $(S, \oplus, \otimes)$, $\boldsymbol{A} = [\underline{A}, \overline{A}]$ and $\boldsymbol{X} = \boldsymbol{X}^{\forall} \oplus \boldsymbol{X}^{\exists}$ are given. An interval vector $\boldsymbol{X}$ is strong AE-eigenvector of $\boldsymbol{A}$ if and only if

$$(\forall x^{\forall} \in \boldsymbol{X}^{\forall})(\exists x^{\exists} \in \boldsymbol{X}^{\exists})\, \underline{A} \otimes (x^{\exists} \oplus x^{\forall}) = x^{\exists} \oplus x^{\forall} \ \wedge\ \overline{A} \otimes (x^{\exists} \oplus x^{\forall}) = x^{\exists} \oplus x^{\forall}.$$

Proof. The proof is similar to the proof of Theorem 4.1. □

4.1 EA/AE strong eigenvectors

Denote the set of indices of $\boldsymbol{X}^{\forall}$ corresponding with universal quantifier by $N^{\forall}$ and the set of indices of $\boldsymbol{X}^{\exists}$ corresponding with existential quantifier by $N^{\exists}$.

Definition 4.3. Let $\boldsymbol{A}$, $\boldsymbol{X}$ be given. Interval vector $\boldsymbol{X}$ is called

- *EA–strong eigenvector* of $\boldsymbol{A}$ if there is $A^{\exists} \in \boldsymbol{A}^{\exists}$ such that for any $A^{\forall} \in \boldsymbol{A}^{\forall}$ the vector $\boldsymbol{X}$ is strong eigenvector of $A^{\exists} \oplus A^{\forall}$,
- *AE–strong eigenvector* of $\boldsymbol{A}$ if for any $A^{\forall} \in \boldsymbol{A}^{\forall}$ there is $A^{\exists} \in \boldsymbol{A}^{\exists}$ such that $\boldsymbol{X}$ is strong eigenvector of $A^{\exists} \oplus A^{\forall}$.

Theorem 4.5. Suppose that $(S, \oplus, \otimes)$, $\boldsymbol{A} = \boldsymbol{A}^{\forall} \oplus \boldsymbol{A}^{\exists}$ and $\boldsymbol{X} = [\underline{x}, \overline{x}]$ are given. Then $\boldsymbol{X}$ is EA–strong eigenvector of $\boldsymbol{A}$ if and only if

$$(\exists A^{\exists} \in \boldsymbol{A}^{\exists})(\forall x \in \boldsymbol{X})\,(A^{\exists} \oplus \underline{A}^{\forall}) \otimes x = x \ \wedge\ (A^{\exists} \oplus \overline{A}^{\forall}) \otimes x = x.$$

Proof. Suppose that there is $A^{\exists} \in \boldsymbol{A}^{\exists}$ such that $(A^{\exists} \oplus \underline{A}^{\forall}) \otimes x = x$ and $(A^{\exists} \oplus \overline{A}^{\forall}) \otimes x = x$ hold for each $x \in \boldsymbol{X}$. By monotonicity of the operations $\oplus$ and $\otimes$ for an arbitrary matrix $A^{\forall} \in \boldsymbol{A}^{\forall}$ we get

$$x = (A^{\exists} \oplus \underline{A}^{\forall}) \otimes x \leq (A^{\exists} \oplus A^{\forall}) \otimes x \leq (A^{\exists} \oplus \overline{A}^{\forall}) \otimes x = x.$$

The reverse implication trivially holds. □

Theorem 4.6. Suppose that $(S, \oplus, \otimes)$, $\boldsymbol{A} = \boldsymbol{A}^{\forall} \oplus \boldsymbol{A}^{\exists}$ and $\boldsymbol{X} = [\underline{x}, \overline{x}]$ are given. Then $\boldsymbol{X}$ is EA–strong eigenvector of $\boldsymbol{A}$ if and only if

$$(\exists A^{\exists} \in \boldsymbol{A}^{\exists})(\forall i \in N\,[(A^{\exists} \oplus \underline{A}^{\forall}) \otimes x^{(i)} = x^{(i)} \ \wedge\ (A^{\exists} \oplus \overline{A}^{\forall}) \otimes x^{(i)} = x^{(i)}].$$

Proof. By Lemma 3.1, if $\beta_j \in S \cap Y_j$ for $j \in N$ then $x = \bigoplus_{j=1}^{n} \beta_j \otimes x^{(j)}$ belongs to $\boldsymbol{X}$, and if $x \in \boldsymbol{X}$ then we can find $\beta_j \in S \cap Y_j$ for $j \in N$ such that $x = \bigoplus_{j=1}^{n} \beta_j \otimes x^{(j)}$. Then we have

$$(A^{\exists} \oplus \underline{A}^{\forall}) \otimes x = (A^{\exists} \oplus \underline{A}^{\forall}) \otimes \bigoplus_{i \in N} \beta_i \otimes x^{(i)} \bigoplus_{i \in N} (A^{\exists} \oplus \underline{A}^{\forall}) \otimes x^{(i)} \otimes \beta_i = \bigoplus_{i \in N} \beta_i \otimes x^{(i)} = x.$$

Similarly we can prove the second equality and by Theorem 4.5 the assertion follows. The reverse implication trivially holds. □

4.2 EA strong EA/AE-eigenvectors

Definition 4.4. Let $\boldsymbol{A} = \boldsymbol{A}^{\forall} \oplus \boldsymbol{A}^{\exists}$, $\boldsymbol{X} = \boldsymbol{X}^{\forall} \oplus \boldsymbol{X}^{\exists}$ be given. Then $\boldsymbol{X}$ is called

- *EA strong EA-eigenvector* of $\boldsymbol{A}$ if there is $A^{\exists} \in \boldsymbol{A}^{\exists}$ such that for each $A^{\forall} \in \boldsymbol{A}^{\forall}$ interval vector $\boldsymbol{X}$ is EA-eigenvector of $A^{\exists} \oplus A^{\forall}$,
- *EA strong AE-eigenvector* of $\boldsymbol{A}$ if there is $A^{\exists} \in \boldsymbol{A}^{\exists}$ such that for each $A^{\forall} \in \boldsymbol{A}^{\forall}$ interval vector $\boldsymbol{X}$ is AE-eigenvector of $A^{\exists} \oplus A^{\forall}$.

Theorem 4.7. Suppose that $(S, \oplus, \otimes)$, $\boldsymbol{A} = \boldsymbol{A}^{\forall} \oplus \boldsymbol{A}^{\exists}$, $\boldsymbol{X} = \boldsymbol{X}^{\forall} \oplus \boldsymbol{X}^{\exists}$ are given. Then interval vector $\boldsymbol{X}$ is EA strong EA-eigenvector of $\boldsymbol{A}$ if and only if

$$(\exists A^{\exists} \in \boldsymbol{A}^{\exists})(\exists x^{\exists} \in \boldsymbol{X}^{\exists})(\forall x^{\forall} \in \boldsymbol{X}^{\forall})\,[(A^{\exists} \oplus \underline{A}) \otimes (x^{\exists} \oplus x^{\forall}) = (x^{\exists} \oplus x^{\forall}) \ \wedge (A^{\exists} \oplus \overline{A}) \otimes (x^{\exists} \oplus x^{\forall}) = (x^{\exists} \oplus x^{\forall})].$$

Proof. $(\Longleftarrow)$ The assertion follows from the monotonicity of the operations, i.e.,

$$x^{\exists} \oplus x^{\forall} = (A^{\exists} \oplus \underline{A}) \otimes (x^{\exists} \oplus x^{\forall}) \leq (A^{\exists} \oplus A^{\forall}) \otimes (x^{\exists} \oplus x^{\forall}) \leq (A^{\exists} \oplus \overline{A}) \otimes (x^{\exists} \oplus x^{\forall}) = x^{\exists} \oplus x^{\forall}.$$

The converse implication is trivial. □

Theorem 4.8. Suppose that semiring $(S, \oplus, \otimes)$, $\boldsymbol{A} = \boldsymbol{A}^{\forall} \oplus \boldsymbol{A}^{\exists}$, $\boldsymbol{X} = \boldsymbol{X}^{\forall} \oplus \boldsymbol{X}^{\exists}$ are given. Then interval vector $\boldsymbol{X}$ is EA strong EA-eigenvector of $\boldsymbol{A}$ if and only if

$$(\exists A^{\exists} \in \boldsymbol{A}^{\exists})(\exists x^{\exists} \in \boldsymbol{X}^{\exists})(\forall k \in N^{\forall} \cup \{n+1\})\,[(A^{\exists} \oplus \underline{A}) \otimes (x^{\exists} \oplus x^{(k)}) =$$
$$(x^{\exists} \oplus x^{(k)}) \ \wedge (A^{\exists} \oplus \overline{A}) \otimes (x^{\exists} \oplus x^{(k)}) = (x^{\exists} \oplus x^{(k)})].$$

The assertion follows from Theorem 3.1 and Theorem 4.7. □

Theorem 4.9. Suppose that $(S, \oplus, \otimes)$, $\boldsymbol{A} = \boldsymbol{A}^{\forall} \oplus \boldsymbol{A}^{\exists}$, $\boldsymbol{X} = \boldsymbol{X}^{\forall} \oplus \boldsymbol{X}^{\exists}$ are given. Then interval vector $\boldsymbol{X}$ is EA strong AE-eigenvector of $\boldsymbol{A}$ if and only if

$$(\exists A^{\exists} \in \boldsymbol{A}^{\exists})(\forall x^{\forall} \in \boldsymbol{X}^{\forall})(\exists x^{\exists} \in \boldsymbol{X}^{\exists})\,[(A^{\exists} \oplus \underline{A}) \otimes (x^{\exists} \oplus x^{\forall}) = (x^{\exists} \oplus x^{\forall}) \,\wedge (A^{\exists} \oplus \overline{A}) \otimes (x^{\exists} \oplus x^{\forall}) = (x^{\exists} \oplus x^{\forall})].$$

Proof. The proof is similar to the proof of Theorem 4.7. □

References

[1] Butkovič, P. (2010). *Max-linear Systems: Theory and Algorithms*, Springer Monographs in Mathematics, Springer-Verlag.

[2] Di Nola, A. & Gerla, B. (2005). Algebras of Łukasiewicz's logic and their semiring reducts. In G. L. Litvinov and V. P. Maslov, editors, *Idempotent mathematics and mathematical physics*, 131–144.

[3] Di Nola, A. & Russo, C. (2007). Łukasiewicz transform and its application to compression and reconstruction of digital images. *Information Sci*., 177, 1481–1498.

[4] Gavalec M. & Plavka, J. (2010). Monotone interval eigenproblem in max-min algebra, *Kybernetika* 46(3), 387–396.

[5] Gavalec, M., Plavka, J.& Ponce, D.(2020).EA/AE-Eigen vectors of Interval Max-Min Matrices. *Mathematics*, 8(6), 882; https://doi.org/10.3390/math8060882.

[6] Gavalec, M., Plavka, J. & Němcová, Z. (2020). Strong Tolerance and Strong Universality of Interval Eigenvectors in a Max-Łukasiewicz Algebra. *Mathematics* , 8, 1504; doi:10.3390/math8091504

[7] Gavalec, M. & Němcová, Z. (2020). Parametric System in Max-Łukasiewicz Algebra. *Mathematics*, 8, 1026; doi:10.3390/math8061026

[8] [8] Gavalec, M . & Plavka, J. (2010). Monotone interval eigenproblem in max-min algebra. Kybernetika 46 (3), 387–396.

[9] Gottwald, S. (2001). *Treatise on Many-Valued Logics; Studies in Logic and Computation*; Research Studies Press:

[10] Molnárová, M. (2018). Possible and universal robustness of special classes of matrices with inexact data. In: *Mathematical methods in economics – Proceedings of the 36th international conference* – Praha, 348–353.

[11] Molnárová, M. (2019). Fuzzy interval Monge matrices with respect to robustness. *Mathematical methods in economics – Proceedings of the 37th international conference České Budějovice*, 409-414.

[12] Myšková, H. (2012). On an algorithm for testing T4 solvability of fuzzy interval systems *Kybernetika* 48(5), 924–938.

[13] Myšková, H. (2012). An iterative algorithm for testing solvability of max-min interval systems. *Kybernetika* 48(5), 879–889.

[14] Myšková, H. & Plavka, J. (2019). XAE and XE A robustness of max–min matrices, *Discrete Applied Mathematics* 267, 142–150.

[15] Myšková, H. & Plavka, J. (2013). X-robustness of interval circulant matrices in fuzzy algebra. *Linear Algebra and its Applications* 438, 2757–2769.

[16] Myšková, H. & Plavka, J. (2014). The robustness of interval matrices in max-plus algebra. *Linear Algebra and its Applications* 445, 85–102.

[17] 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

[18] Myšková, H. & Plavka, J. (2020). AE and EA robustness of interval circulant matrices in max-min algebra, *Fuzzy Sets and Systems* 384, 91–104.

[19] Plavka, J. (2013). On the weak robustness of fuzzy matrices. *Kybernetika* 49, 128–140.

[20] Plavka, J. (2012). On the O(n3) algorithm for checking the strong robustness of interval fuzzy matrices, *Discrete Applied Math*. 160, 640-647.

[21] Plavka, J. & Sergeev, S. (2016). Characterizing matrices with X-simple image eigenspace in max-min semiring. *Kybernetika* 52, 497–513.

[22] Rashid, I., Gavalec, M. & Cimler, R. (2016). Eigenspace structure of a max-prod fuzzy matrix. *Fuzzy Sets and Systems* Volume 303, 136–148.

[23] Sanchez, E. (1978). Resolution of eigen fuzzy sets equations, *Fuzzy Sets and Systems* 1, 69–74.

[24] Terano, T. & Tsukamoto, Y. (1977): Failure diagnosis by using fuzzy logic. In: *Proc. IEEE Conference on Decision Control* (New Orleans, LA), 1390–1395.

[25] 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, New York, 1971), 209–245.

[26] Zimmermann, K. (1976). *Extremální algebra*. Ekonomicko-matematická laboratoř EÚ ČSAV, Praha.

Reverse Channel Competition in a Dual Sustainable Closed-Loop Supply Chain

Petr Pokorny[1]

Abstract. In this paper we present a closed-loop supply chain (CLSC) model with a retailer selling products produced by a manufacturer. Products can be returned like a returnable packaging used in beverage industry. The products are returned through a 3rd party collector who makes effort to acquire them and sells them to a manufacturer who can recycle them and reuse the recycled raw material to produce a new product at a lower marginal cost. The customer is stimulated to buy sustainable products at a guaranteed buy-back price received at the point of collection. First, we model the case where only the 3rd party collects and we compare this model with the case where the 3rd party collector competes with the retailer who also enters the market to collect. We derive the CLSC's stability conditions of the solution with respect to the intensity level of the competition between the retailer and the 3rd party collector. We show that the customer benefits from the competition in the reverse chain boosting the whole market size and increasing the profit of the total supply chain. We also discuss the competition impact on the 3rd party collector who is impacted adversely by the market power of the retailer and the manufacturer.

Keywords: lorem, Closed-Loop Supply Chain, Reverse Chain, Game Theory, Nash Equilibrium, Reverse Channel Competition

JEL Classification: C72
AMS Classification: 91A06, 91A10

1 Introduction

Environmental sustainability and social responsibility have been gaining attention in supply chain management leading to the concept of Closed-Loop Supply Chains (CLSC). Besides the traditional forward flows of goods, we also consider reverse goods flows in a CLSC that represent the products at their end of use/life. They are returned either to be recycled, refurbished, or remanufactured. In this paper we analyze a CLSC consisting of a manufacturer (*M*) who produces new products and sells them through a retailer (*R*). The manufacturer can recycle the end of use products and generates a unit saving on the production marginal cost if he uses a recycled material instead of a new one. Products are collected by a 3rd party collector (*C*). We have modified and extended the approach presented in [9] to capture the effect of a dual reverse chain where *R* and *C* compete to collect. The customer is stimulated to buy sustainable products with a guaranteed buy-back price received at the point of collection. If *R* and *C* compete, the customer can benefit from a lower new product price which drives up the demand. Hence, we demonstrate how the competition in the reverse chain can boost the market with sustainable products and how the CLSC can improve its total performance. We drop the quality impact because we are dealing with recycling instead of remanufacturing and the manufacturer's wholesale buy-back price, at which he purchases the returns from *C* and *R*, is a variable instead of a parameter set. In our analysis we also explore different market power structures such a centralized chain case, a Vertical Nash model, and the case where *M* is a Stackelberg leader. We show their impact on the individual as well as on the total chain (*SC*) profits.

1.1 Competition in CLSC

Given a little space we only present the most relevant papers that deal with competition in CLSC, namely in the reverse parts of the chain. A very useful survey was compiled by De Giovanni and Zacour in [2], where the authors specifically focus on CLSCs with the presence of return functions. Papers that deal with SC partners competing in the market for product returns focus usually on finding the optimal channel to collect the returns as in [7], where the situation is analyzed whether a manufacturer should collect alone or if the collection should be subcontracted to competing retailers. A similar model where a retailer and a 3PL compete in a dual recycling channel is studied in [4] with a specific return function. Different market power structures are studied in [1] as to who should be the

[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.

Stackelberg leader in a CLSC. In [3], the authors show that the retailer collecting is the best choice. An OEM and a retailer competing in a dual reverse chain is studied in [5] with the conclusion that both channels should collect. Two manufacturers competing for the returns are analyzed in [10]. A systematic study of an OEM bargaining power is studied in [6] assuming that the EOM can be the leader in both the forward and reverse chains or take the second position. In [8], the role of a government subsidy and a corporate social responsibility (CSR) investment is analyzed and it is concluded that the government subsidy makes sense, if the manufacturer makes a CSR investment by donning a part of their profit.

2 Model

2.1 Notation and Assumptions

CLSC parameters and decision variables are partially taken from [9] with a changed notation and amended to formulate the competition model.

c_n/c_r	Unit production cost of products produced from new /recycled raw materials
c_a/c_s	Acquisition/Sales effort cost
$r_C/r_R/r_{SC}$	Quantity of returned product acquired by *C* / *R* / *SC*, such that $r_{SC} = r_C + r_R$
q	Demand quantity for new products
w	Wholesale price charged for the new products by *M* to *R*
p	Retail price charged for the new products by *R* to customers
m_R	*R*'s margin on new products, i.e. $m_R = p - w$
b	Buy-back price offered for the returns by *M* to *C*/*R*
$\varepsilon_C/\varepsilon_R$	Acquisition price offered for the returns by *C*/*R* to customers
mr_C	*C*'s margin on returns sold to *M*, i.e. $mr_C = b - \varepsilon_C$
mr_R	*R*'s margin on returns sold to *M*, i.e. $mr_R = b - \varepsilon_R$
a	Acquisition (Collection) efforts of *C*/*R*
s	Sales effort of *R* to sell new products
r_e	Reference market price point
N_p	Minimum net price for a customer as $N_p = p - max\{\varepsilon_C, \varepsilon_R\}$

π_j^i is a profit function of agent $j, j \in \{M, C, R, SC\}$ in model, $i \in \{CC, MS, VN, CCD, MSD, VND\}$, $\pi_{SC}^i = \pi_M^i + \pi_R^i + \pi_C^i$. We will also refer to *M* as him, to *R* as her, and to *C* as it. First three models refer to the non-competition model as presented in [9] without the quality consideration and with b as a variable. CC stands for the centralized control model, MS for the Stackelberg model with *M* as the leader and VN is the Vertical Nash model. The other threesome refers to the competition models with dual reverse channels. CCD shows a centralized model with two parallel collecting channels. MSD models a sequential Stackelberg game with *M* as the leader that purchases returned products from the competing *R* and *C*. VND is the Vertical Nash model where both *R* and *C* collect to sell to *M* and all SC partners take decisions simultaneously.

2.2 Forward and Reverse Channels

CLSC's deal with backward and forward product flows. In our case the returns are collected, recycled and the recycled material is used to produce new products. These properties have to be reflected in the structure of the demand for both products. First, we set the demand functions of new products and the return functions of the reverse channels.

Demand Function in Forward Supply Chain

The retail demand quantity for new products is set in accordance with [9], while omitting the quality impact, as

$$q^i = \alpha - \beta p^i + \gamma \varepsilon_C^i + \delta s^i - \theta(p^i - r_e). \tag{1}$$

Parameters $\alpha, \beta, \gamma, \delta, \theta \geq 0$ are all positive, α is the total market size for products made of new raw materials, β is the price elasticity, γ is the customer's sensitivity to the acquisition price of the collector, δ captures the sensitivity to the sales effort (e.g. advertising), and θ represents the so-called reference price effect. In case both R and C compete for the returns, the demand function takes the form of (2).

$$q^i = \alpha - \beta p^i + \gamma(\varepsilon_C^i + \varepsilon_R^i) + \delta s^i - \theta(p^i - r_e) \tag{2}$$

Reverse Supply Chain

Total number of returns collected by the 3rd party is calculated in (3) as a function of price incentives, ε, and collection efforts, a, for the case where 3rd party collector is the only collecting reverse channel, based on [9].

$$r_C^i = r_0 + r_1 \varepsilon_C^i + r_2 a_C^i \tag{3}$$

We will now consider the situation when both the retailer and the 3rd party collector compete to acquire the recyclable packaging. In [2], the authors show a sensitivity model where the total number of returns of 3rd party, r_C, (retailer, r_R,) depend positively on their own price incentives, ε, and their collection efforts, a ,and negatively on the competitors' price incentives, ε, and collection efforts, a.

$$r_C^i = \omega r_0 + r_1 \varepsilon_C^i + r_2 a_C^i - (r_3 \varepsilon_R^i + r_4 a_R^i) \tag{4}$$

$$r_R^i = (1 - \omega) r_0 + r_1 \varepsilon_R^i + r_2 a_C^i - (r_3 \varepsilon_C^i + r_4 a_C^i), \tag{5}$$

Parameters $r_k \geq 0$, $k = 1,2, \ldots, 4$ represent sensitivities of the return function to own acquisition price, own collection effort, competitor's acquisition price, and competitor's collection effort, respectively, ω is the share of the neutral market with recyclables.

2.3 Profit functions

Decentralized CLSC

Modified profit functions of *M, R,* and *C* for $i \in \{CC, MS, VN\}$ without the quality impact are as follows:

$$\pi_M^i(w^i, b^i) = (w^i - c_n) q^i + (c_n - c_r - b^i) r_C^i \tag{6}$$

$$\pi_R^i(s_R^i, p^i) = (p^i - w^i) q^i - 0.5 c_s (s_R^i)^2 \tag{7}$$

$$\pi_C^i(\varepsilon_C^i, a_C^i) = (b^i - \varepsilon_C^i) r_C^i - 0.5 c_a (a_C^i)^2 \tag{8}$$

Profit functions of *M, R,* and *C* for $i \in \{CCD, MSD, VND\}$ under the assumption that *R* and *C* compete for the returns are as follows:

$$\pi_M^i(w^i, b^i) = (w^i - c_n) q^i + (c_n - c_r - b^i)(r_C^i + r_R^i) \tag{9}$$

$$\pi_R^i(s_R^i, p^i, \varepsilon_C^i, a_R^i) = (p^i - w^i) q^i - 0.5 c_s (s_R^i)^2 + (b^i - \varepsilon_R^i) r_R^i - 0.5 c_a (a_R^i)^2 \tag{10}$$

$$\pi_C^i(\varepsilon_C^i, a_C^i) = (b^i - \varepsilon_C^i) r_C^i - 0.5 c_a (a_C^i)^2 \tag{11}$$

When all the agents (*M, R, C*) make their decisions simultaneously, there is no leader with a dominating bargaining power to set the price. The price and margin settings have to be an outcome of a joint decision making, where all the parties enjoy full information available to them, This is called the (Vertical) Nash model denoted as *VN(D)* in our cases. Stackelberg leader-follower games $MS(D)$ are models where the leader sets the price based on the observed best response reaction of the followers in order to maximize his own profit. We have explored the situation where *M* is the leader (controls the margins) and *R* and *C* are followers.

Centralized CLSC

If a CLST is one legal entity or controlled by one decision maker, we consider it to be integrated or centralized. Eqs. (12) and (13) show the total profit functions, $\pi_{SC}^i = \pi_M^i + \pi_R^i + \pi_C^i$, of the respective non-competitive, CC and the competitive dual, CCD, chains, respectively:

$$\pi_{SC}^i = (p^i - c_n) q^i + (c_n - c_r) r_C^i - 0.5 c_s (s_R^i)^2 - \varepsilon_C^i r_C^i - 0.5 c_a (a_C^i)^2, \tag{12}$$

$$\pi^i_{SC} = (p^i - c_n)q^i + (c_n - c_r)(r^i_C + r^i_R) - 0.5c_s(s^i_R)^2 - \varepsilon^i_C r^i_C - \varepsilon^i_R r^i_R - 0.5c_a(a^i_R a^i_C)^2. \quad (13)$$

3 Numerical Example

We have taken over the values of the parameters from [9], used them in our model and amended the cross-sensitivities r_3 and r_4 as shown in Table 1. We have explored that the collection effort parameter r_4 has a low impact on the profits, so we focused on analyzing the impact of r_3 on the stability of the equilibrium.

α	β	γ	δ	θ	r_e	ω	r_1	r_2	r_3	r_4	c_A	c_E	c_r	c_n
1100	10	5	6	1	90	0.5	8	7	*	4	500	300	50	80

Table 1 Model example parameters

3.1 Equilibrium Results

Since we omitted the quality of the returns as it does not make an economic sense in our setting, the model results will differ from the ones published in [9]. The other reason is that we have introduced the competition between the R and C to acquire the returns and sell them to M for price b. Contrary to [9], we also consider b a variable, which significantly impacts the profit functions of both players. Following model conditions (14)-(16) must be fulfilled;

$$c_n - c_r > b^i \quad (14)$$

states that the savings on the marginal cost of producing from recycled materials is greater than M's purchase price of the to-be-recycled item.

$$q^i \geq r^i_C + r^i_R \quad (15)$$

stipulates the closed-loop behavior of the model. No additional recycled items can be acquired outside the CLSC, hence, there is always enough products in the market to acquire the ones for recycling.

$$r_4 < r_2, \text{ and } r_3 < r_1 \quad (16)$$

If violated, the competitor's power would drive the opponent out of the market or deny them an entry to it. We will show the competition impact using the r_3 price sensitivity parameter.

Proposition 1. *The price sensitivity parameter r_3 can be set for each of the models, $r^i_3 = \langle lb^i, ub^i \rangle, \forall\, i \in \{CCD, MSD, VND\}$ such that the above conditions (14-16) are met.*
Proposition 1 is verified and validated via equilibrium sensitivity analysis. Further, we present all equilibrium values for the subject competitive dual models in the form of stability intervals. In economic terms, the lower and upper bounds represent the respective powers of a competitor over their opponent. The lower the boundary, the lower the strength of the competitor. The higher the boundary, the fiercer the competition becomes from which the customer can benefit.
We show the optimal values of the variables in Table 2 for all considered models and put "-", where non-applicable. The results are obtained using the backwards induction process. For the non-competitive models (CC, MS, VN) the results are straightforward as each channel is dedicated to one operation only.

Variable	CC	VN	MS	CCD		VND		MSD	
$r^*_3 = \langle lb^*, ub^* \rangle$	-	-	-	3.5	4.2	5.1	5.6	5.6	6.7
w^*	-	90.6	93.6	-	-	95.9	96.8	99.4	104.4
p^*	97.4	101.2	102.5	106.5	107.8	111.8	113.7	110.9	118.5
q^*	199	116	82	291	305	175	185	126	154
$r^* = r^*_C + r^*_R$	191	114	82	290	278	174	169	126	120
ε^*_C	17.4	7.9	3.89	27.1	29.9	19.5	22.5	14.2	25.9
ε^*_R	-	-	-	27.1	29.9	23.3	26.4	16.9	28.2
b^*	-	22.3	14.3	-	-	27.3	29.8	19.8	29.8
s^*	0.22	0.13	0.09	0.32	0.33	0.19	0.2	0.14	0.17
a^*_C	0.25	0.29	0.21	0.02	0	0.16	0.15	0.11	0.09

Variable	CC	VN	MS	CCD		VND		MSD	
a_R^*	-	-	-	0.02	0	0.08	0.07	0.06	0.03
m_R^*	-	10.61	8.7	-	-	15.9	16.8	11.5	14.0
mr_R^*	-	-	-	-	-	4.1	3.3	2.9	1.6
mr_C^*	-	14.3	10.4	-	-	7.8	7.3	5.6	4.7
π_M^*	-	2,128	2,417	-	-	3,250	3,155	3,734	3,795
π_R^*	-	1,235	730	-	-	3,226	3,484	1,693	2,283
π_C^*	-	1,638	853	-	-	488	422	252	176
π_{SC}^*	5.987	5,002	4,000	8.541	8.484	6,965	7,063	5,679	6,256
N_p	79.9	93.3	98.6	79.4	77.8	88.5	87.2	93,9	90,2
$(p - N_p)/p$, [%]	18	8	8.9	25.5	27.9	20.9	23.3	15.4	13.6

Table 2 Equilibrium results

3.2 Results Discussion and Sensitivity Analysis

Conditions (14)-(16) are all met as we can read in Table 2. We begin with the non-competitive models results. In the MS model M pays a lower b when having a leader position, which leaves him with a higher unit saving profit, $c_n - c_r - b$, than in the VN model. At the same time the 3rd party collector's margins is lower as its market power is weaker. Evidently, the highest profit as well as the lowest N_p of the CLSC are achieved under the CC setting providing the customer with the highest utility. When the manufacturer sets the prices b and w in MS, the customer's utility is the lowest among the non-competitive models due to N_p =98.6 and C's profit is halved compared to VN result, which is considered to describe the situation where all the agents decide simultaneously. Under the MS settings, we see that M pushes down both the R's margin m_R on the new products as well as the C's margin on the returned products mr_C. This forces C to offer a low retail return price ε_c to customers, thus limiting the total size of the recyclable product market r by more than 50 % vs. the CC case.

Introducing the competition into the reverse chain drives up the retail buy-back prices, $\varepsilon_c, \varepsilon_R$ as both R and C attempt to attract the customers which increases the customer's deposit paid at the purchase calculated as a ratio $(p - N_p)/p$. Customers are positively motivated to purchase returnable products despite the total higher price of the new product since they have a 100% buy-back guarantee and because N_p is lower in a pair-wise model comparison CC vs. CCD, VN vs. VND, and MS vs. MSD. Hence, the competitive environment motivates the customer to purchase a returnable product because $N_p^{MS*} > N_p^{MSD*} > N_p^{VN*} > N_p^{VND*} > N_p^{CC*} > N_p^{CCD*}$. In the dual chain M has to offer higher buy-back price b as the competion intensity, measured by r_3, increases, $b^{MSD*} > b^{MS*}, b^{VND*} > b^{VN*}$. He will always be motivated to leverage on the unit cost savings up until the condition (14) is violated. We can see that the CCD model uses up the entire space provided by (14) to attract as many returns as possible. The centralized chain maximizes the total market size at the expense of decreasing the marginal recycling unit cost saving, $c_n - c_r$, to its minimum.

Looking closely at the decentralized competitive models, we observe the collector spends more effort and money to attract the customer since $a_C^* > a_R^*$. The competitive advantage of R is that her profit is generated by the sale of new products and we can see that R's margin on new products m_R increases as the competition in the reverse chain increases. The opposite effect takes place on the returned product margins mr_R^*, mr_C^*, evidently. C can realize a higher unit margin than R, $mr_R^* < mr_C^*$, at the expense of a lower offered retail return price, $\varepsilon_C^* < \varepsilon_R^*$, which in turn means that the total C's share on the product returns market is lower than the one of R, $r_C^* < r_R^*$.

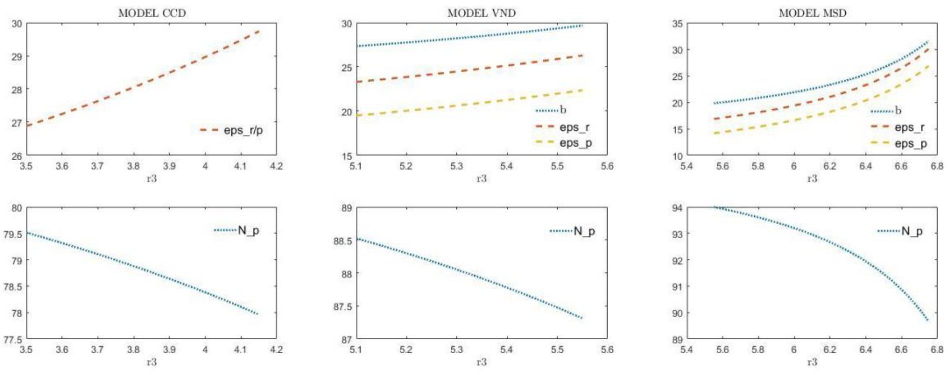


Figure 1 M's buy-back price, retail return and net consumer prices

Figure 1 shows the behavior of return and net consumer prices in the domain stability region per respective competitive dual model. Analyzing the r_3 stability parameter domain we read that $r_3^{CCD*} < r_3^{VND*} \leq r_3^{MSD*} < r_1$, which meets (16), and also implies that a strong competition is not desired under the CCD model where both channels belong to the same company. On the other hand, when M is the price setter, he benefits a great deal if both R and C compete, which favors his profit and adversely affects the profits of R and C. The fiercer the competition, the lower the profit of C. There is a potential role of a government subsidy to balance off the market power of R and C in order to motivate the specialized collecting companies to enter the market.

4 Conclusion

We have focused on analyzing the effect of supply agents competing to acquire the product returns in order to sell them to a manufacturer at a profit to be further recycled for a new product manufacturing. We have assumed a fully closed-loop supply chain structure of forward and reverse flows, which we have explicitly conditioned in the model. Contrary to most of the traditional papers we studied the customer's motivation to start purchasing a recyclable product despite its higher initial price. We have shown through the net consumer price measure that the competition increases the value of a recyclable product for a customer via a reduced net price under a guaranteed buy-back purchase. We concluded that the manufacturer is pushed towards higher effectiveness in the recycling process as the level of competition increases in the reverse chain because the retail return prices increase. We can also see that the 3rd party's market condition and profit worsened as oppose to the situation when it was the exclusive collection channel. The reason being is both the power of the manufacturer and the retailer competing with it on the return market. In order for the 3rd party collector to stay in the market a governmental subsidy or a coalition with the manufacturer could be considered for future research to coordinate this CLSC.

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] Choi, T.-M., Li, Y., & Xu, L. (2013). Channel Leadership, Performance and Coordination in Closed Loop Supply Chains. *International Journal of Production Economics*, 146(1), 371–380.
[2] 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), 1–44.
[3] Hong, X., Xu, L., Du, P., & Wang, W. (2015). Joint Advertising, Pricing and Collection Decisions in a Closed-Loop Supply Chain. *International Journal of Production Economics*, 167, 12–22.
[4] Huang, M., Song, M., Lee, L. H., & Ching, W. K. (2013). Analysis for Strategy of Closed-Loop Supply Chain with Dual Recycling Channel. *International Journal of Production Economics*, 144(2), 510–520.
[5] Liu, L., Wang, Z., Xu, L., Hong, X., & Govindan, K. (2017). Collection effort and reverse channel choices in a closed-loop supply chain. *Journal of Cleaner Production*, 144, 492–500. https://doi.org/10.1016/j.jclepro.2016.12.126

[6] Liu, Y., & Zhang, Y. (2018). Closed Loop Supply Chain under Power Configurations and Dual Competitions. *Sustainability*, 10(5). http://dx.doi.org/10.3390/su10051617

[7] Savaskan, R. C., & Van Wassenhove, L. N. (2006). Reverse Channel Design: The Case of Competing Retailers. *Management Science*, 52(1), 1–14. https://doi.org/10.1287/mnsc.1050.0454

[8] Song, L., Yan, Y., & Yao, F. (2020). Closed-Loop Supply Chain Models Considering Government Subsidy and Corporate Social Responsibility Investment. *Sustainability*, 12(5), 2045. https://doi.org/10.3390/su12052045

[9] Taleizadeh, A., Moshtagh, M. S., & Moon, I. (2017). Optimal decisions of price, quality, effort level and return policy in a three-level closed-loop supply chain based on different game theory approaches. *European J. of Industrial Engineering*, 11, 486.

[10] Wang, W., Fan, L., Ma, P., Zhang, P., & Lu, Z. (2017). Reward-penalty mechanism in a closed-loop supply chain with sequential manufacturers' price competition. *Journal of Cleaner Production*, 168, 118–130. https://doi.org/10.1016/j.jclepro.2017.08.104

Shooting Method for Boundary Value Problems of Ordinary Differential Equations in Economics

Pavel Pražák[1]

Abstract. Optimal control problems over the finite time horizon often lead to the solution of nonlinear ordinary differential equations with given boundary conditions. These problems are known as boundary value problems. There are several numerical methods for solving such problems. This paper mainly deals with the shooting method. Then, a variant of bisection method for solving nonlinear equations is introduced to find a suitable initial condition. The given procedure is also applied to a particular nonlinear dynamics economic model dealing with optimal consumption. All calculations are realized using MATLAB.

Keywords: bisection method, boundary value problems, optimal control, shooting method, life-cycle model

JEL Classification: C44, C61
AMS Classification: 34H05, 49K15, 34B05

1 Introduction

Optimal control theory was developed to find optimal processes to control dynamic systems. Many applications of this theory can be also found in the field of economics and management, see [12], [5], These application include economic growth, extraction of natural resources, optimal investment or optimal taxation among others, see [10], [11], [13] or [7]. The necessary conditions for solution to optimal control problems are known as Pontryagin maximum principle, [6], [3], [8]. If this principle is applied a system of ordinary differential equation together with initial or terminal conditions are gained. Ordinary differential equations are usually formulated as initial value problems or boundary value problems. An initial value problem is given if all necessary conditions are known to find a solution to a given differential equation for a specific value of the independent variable. It means that an initial value problem depends on local conditions. A boundary value problem is a problem of determining a solution to a given differential equation subject to conditions on the unknown function specified at two or more different independent variable values. There exist several types of numerical methods for two-point boundary value problems: e.g. shooting method, finite-difference method, method of collocation, or Galerkin's method, see [3]. We concentrate on the shooting method here. In general, the shooting method transforms the boundary value problem into a system of first-order ordinary differential equations, which can be solved by the initial value methods - in this paper Runge-Kutta methods are considered, [2], and the MATLAB function `ode45` is exploited. More particularly, the initial conditions on one side of the boundary condition (usually on the left side of the interval for the independent variable) are varied until the boundary conditions on the second side are satisfied. Alternatively, the missing initial conditions can be determined by a nonlinear equation solver - in this paper bisection method is considered and the MATLAB function `fzero` is exploited. Number of MATLAB licence that was used for preparation of this paper is Site ID 708501.

2 Methods

This section shortly describes a particular boundary value problem that can appear in economic applications of optimal control. The following commentary uses [3], [11] and [4].

2.1 Boundary Value Problems and Shooting Method

For ordinary differential equations, boundary value conditions are typically specified at two endpoints of an interval, so the two-point boundary value problem is defined. For given T, $T \in \mathbb{R}$ and $T > 0$, $f : R^{n+1} \to R^n$, $g : R^{2n} \to R^n$ and $x(0), x(T) \in \mathbb{R}^n$, a general first-order two-point boundary value problem has form

$$\dot{x} = f(t, x), \quad x \in \mathbb{R}^n,\ t \in [0, T] \tag{1}$$

[1] University of Hradec Králové, Faculty of Informatics and Management, Department of Informatics and Quantitative Methods, Rokitanského 62, Hradec Králové, Czech Republic, pavel.prazak@uhk.cz

with boundary conditions

$$g(x(0), x(T)) = 0. \tag{2}$$

When solving boundary value problem (1), (2), we choose an arbitrary vector $\nu = (\nu_1, \nu_2, \ldots, \nu_n) \in M$, where $M \subset \mathbb{R}^n$ is a region, i.e. a nonempty open and connected subset of $\mathbb{R}^n$, such that functions f and g are defined on $[0, T] \times M$ and $M \times M$ respectively. Then the initial value problem with the equation (1) and the initial condition

$$x(0) = \nu \tag{3}$$

is solved. Under known conditions, the initial value problem (1), (3) has exactly one solution, cf. [3], which is a function of parameter ν. This solution can be denoted $\varphi(t, \nu)$ so $\varphi(0, \nu) = \nu$ and for $t = T$ we have $x(T) = \varphi(T, \nu)$. The principle of the shooting method is that we are looking for such $\nu = \widehat{\nu}$ that the function $\varphi(t, \widehat{\nu})$ satisfies the boundary conditions, i.e.

$$g(\widehat{\nu}, \varphi(T, \widehat{\nu})) = 0.$$

Thus, we actually look for a solution $\nu_1, \nu_2, \ldots, \nu_n$ of a system of n (generally nonlinear) equations for n unknowns

$$G(\nu) = 0, \tag{4}$$

where $G(\cdot) = g(\cdot, \varphi(T, \cdot))$ and $G : \mathbb{R}^n \to \mathbb{R}^n$. To solve this system of equations that is nonlinear in general, a numerical method has to be used. Moreover, the underlying reason that shooting method works for a boundary value problem that has an unique solution is the continuous dependence of solutions to initial value problems on the initial conditions, [3]. The general shooting method is summarized in Algorithm 1, where $||\cdot||_\infty$ is the max norm and $||\cdot||$ is the Euclidean norm.

Algorithm 1: Shooting method

Input : Functions f, g, the time horizon T, the boundary condition $x(T)$ defining problem (1), (2), an initial guess ν_{init} of $x(0)$ and a stopping tolerances $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$

Output: Approximation $\widehat{\nu}$ of $x(0)$

1 Set $k \leftarrow 0$ and $\nu_k \leftarrow \nu_{init}$

2 Solve the initial value problem (given by the equation (1) and the initial value $x(0) = \nu_k$) for $\varphi(T, \nu_k)$

3 **if** $||\varphi(T, \nu_k) - x(T)||_\infty < \varepsilon_1$ *and* $||G(\nu_k)|| < \varepsilon_2$, *where G is given as in (4)* **then**

4     $\widehat{\nu} \leftarrow \nu_k$

5   Stop

6 **else**

7   Based on previous values ν_k choose a new ν_{init} (such that $||G(\nu_{init})|| < ||G(\nu_k)||$)

8   Set $k \leftarrow k + 1$ and $\nu_k \leftarrow \nu_{init}$

9   Go to the line 2

10 **end**

2.2 Bisection Method in Shooting Method

The weak point of the Algorithm 1 is in line 7. How to find a better approximation of the initial value? There exist several possible ways to do it. One of them is introduced in this subsection. Before we do it we will describe more characteristics of boundary values. It is said that boundary conditions are separated if any given component of the function g in (2) involves solution values only at the point 0 or at the point T, but not both. Moreover, boundary conditions are linear if they are of the form

$$B_0 x(0) + B_T x(T) = c \tag{5}$$

where $B_0, B_T \in \mathbb{R}^{n \times n}$ are matrices and $c \in \mathbb{R}^n$ is a vector. The following special two-point boundary problem formulated as a system of ordinary differential equations is considered here

$$\begin{aligned} \dot{x}_1 &= f_1(t, x_1, x_2), \\ \dot{x}_2 &= f_2(t, x_1, x_2) \end{aligned} \tag{6}$$

together with the boundary conditions

$$x_1(0) = x_0, \quad x_2(T) = x_T \tag{7}$$

where $x_1 \in \mathbb{R}^r$, $x_2 \in \mathbb{R}^s$ and $f_1 : \mathbb{R}^{r+s+1} \to \mathbb{R}^r$, $f_2 : \mathbb{R}^{r+s+1} \to \mathbb{R}^s$. The given boundary conditions (7) can be written as separated linear conditions of the form

$$\begin{pmatrix} I_r & 0_{r\times s} \\ 0_{s\times r} & 0_{s\times s} \end{pmatrix} \cdot \begin{pmatrix} x_1(0) \\ x_2(0) \end{pmatrix} + \begin{pmatrix} 0_{r\times r} & 0_{r\times s} \\ 0_{s\times r} & I_s \end{pmatrix} \cdot \begin{pmatrix} x_1(T) \\ x_2(T) \end{pmatrix} = \begin{pmatrix} x_0 \\ x_T \end{pmatrix},$$

where I_r or I_s are identity matrices of the type $r \times r$ or $s \times s$ and $0_{r\times r}$, $0_{r\times s}$, $0_{s\times r}$ or $0_{s\times s}$ are zero matrices of the types $r \times r$, $r \times s$, $s \times r$ or $s \times s$.

If $r = s = 1$ we can enrich the Algorithm 1 in the line 7 using bisection method, see [2]. We consider boundary value problem (6), (7) with real functions. Let

$$(\varphi_1(t, x_1(0), x_2(0)),\ \varphi_2(t, x_1(0), x_2(0)))$$

be solution to the system (6) with initial condition $(x_1(0), x_2(0)) \in \mathbb{R}^2$. Assume that it is possible to find two initial conditions $\nu_a \in \mathbb{R}$ and $\nu_b \in \mathbb{R}$ such that $\varphi_2(T, x_1(0), \nu_a) > x_2(T)$ and $\varphi_2(T, x_1(0), \nu_b) < x_2(T)$. Because the solution $(\varphi_1(T, x_1(0), x_2(0)), \varphi_2(T, x_1(0), x_2(0))$ to the equation (6) continuously depends on initial values there is $\widehat{\nu} \in (\nu_a, \nu_b)$ (or (ν_b, ν_a) provided that $\nu_a > \nu_b$) such that $\varphi_2(T, x(0), \widehat{\nu}) = x_2(T)$ - it is sufficient to apply the intermediate value theorem for continuous function $\varphi_2(T, x_1(0), \cdot))$ on the interval $[\nu_a, \nu_b]$ (or $[\nu_b, \nu_a]$ provided that $\nu_a > \nu_b$). Now it is possible to introduce shooting algorithm with bisection for the problem (6), (7).

Algorithm 2: Shooting method with bisection.

Input : Functions f_1, $f_2 : \mathbb{R} \to \mathbb{R}$, the time horizon $T > 0$, the boundary conditions $x_1(0)$ and $x_2(T)$ defining problem (6), (7), an initial guess ν_a and ν_b of $x_1(0)$ such that for the solution to (6) is $(\varphi_2(T, x_1(0), \nu_a) - x_2(T)) \cdot (\varphi_2(T, x_1(0), \nu_b) - x_2(T)) < 0$ and a stopping tolerance $\varepsilon > 0$

Output: Approximation $\widehat{\nu}$ of $x_2(0)$

Set $k \leftarrow 0$ and $dist \leftarrow 2\varepsilon$
while $dist > \varepsilon$ **do**
  $\widehat{\nu} \leftarrow (\nu_a + \nu_b)/2$
  Solve the initial value problem (given by the system (6) and the initial values $x_1(0)$ and $x_2(0) = \widehat{\nu}$) for $\varphi_2(T, x(0), \widehat{\nu})$
  if $(\varphi_2(T, x_1(0), \nu_a) - x_2(T)) \cdot (\varphi_2(T, x_1(0), \widehat{\nu}) - x_2(T)) < 0$ **then**
    $\nu_b \leftarrow \widehat{\nu}$
  else
    $\nu_a \leftarrow \widehat{\nu}$
  end
  $dist \leftarrow |\nu_a - \nu_b|$
  $k \leftarrow k + 1$
end

2.3 The Standard Optimal Control Problem

The reason why we deal with boundary value problem here is the fact that this problem could be an important part of solving optimal control problems. The standard end-constrained problem of optimal control theory is described here according to [11] and [3]. This problem is based on choosing the input control vector so that the objective function of the system attains its maximum value on the given interval with fixed time horizon T, $T > 0$, and with respect to a criterion given by a system of the first-order ordinary differential equations. Suppose that the state of an economic system at time t, $t \in [0, T]$, is described by a vector $x(t) \in \mathbb{R}^r$. Further, there is a decision-maker who is able to affect the state $x(t)$ by choosing a value $u(t)$ of control function and $u(t) \in U$. This control function is considered to be a piecewise continuous function on control region $U \subset \mathbb{R}^m$. The standard end-constrained problem with r state variables and m control variables is to find a pair of vector functions $(x(t), u(t))$ defined on the interval $[0, T]$ in order to maximize functional

$$\int_0^T g(t, x(t), u(t))dt,$$

where $g : \mathbb{R} \times \mathbb{R}^r \times \mathbb{R}^m \to \mathbb{R}$. It is also assumed that the evolution of the state is given by ordinary differential equation

$$\dot{x}(t) = f(t, x(t), u(t)),\ t \in [0, T],$$

where $f : \mathbb{R} \times \mathbb{R}^r \times \mathbb{R}^m \to \mathbb{R}^r$ with initial state

$$x(0) = x_0$$

and terminal condition

$$x(T) \text{ free.}$$

In general, the terminal conditions can be specified in more details as relations describing each coordinates of the vector $x(T)$ and it is possible to use equalities or inequalities, see e.g. [11], but the latter formulation of the terminal condition is sufficient for our purposes. Pontryagin maximum principle gives necessary condition for optimal solution to the given optimal control problems, [11], [3], [6]. Its simplest form is as follows: Let $(\widehat{x}(t), \widehat{u}(t))$ be an optimal solution to the standard end-constrained control problem. Then there exists a continuous and piecewise differentiable adjoint function $\lambda(t) = (\lambda_1(t), \lambda_2(t), \ldots, \lambda_r(t))$ defined on $[0,T]$ such that:

1. The control function $\widehat{u}(t)$ maximizes the Hamiltonian $H(t, \widehat{x}(t), u, \lambda(t))$ for $u \in U$, i.e.

$$H(t, \widehat{x}(t), u, \lambda(t)) \leq H(t, \widehat{x}(t), \widehat{u}(t), \lambda(t)),\ u \in U,$$

   where $H(t, x(t), u(t), \lambda(t)) = g(t, x(t), u(t)) + (\lambda(t) | f(t, x(t), u(t)))$ and $(.|.)$ denotes the Euclidean inner product in $\mathbb{R}^r$.
2. If $\widehat{u}(t)$ is continuous the adjoint function satisfy the ordinary differential equation

$$\dot{\lambda}(t) = -H_x(t, \widehat{x}(t), \widehat{u}(t), \lambda(t)).$$

3. Furthermore, the transversality condition

$$\lambda(T) = 0.$$

   holds.

To summarize, the important part of necessary optimality conditions for the optimal process $(\widehat{x}(t), \widehat{u}(t))$ is given as boundary value problem of the form

$$\begin{aligned} \dot{x}(t) &= H_\lambda(t, \widehat{x}(t), \widehat{u}(t), \lambda(t)) = f(t, x(t), \widehat{u}(t)) \\ \dot{\lambda}(t) &= -H_x(t, \widehat{x}(t), \widehat{u}(t), \lambda(t)) \end{aligned} \tag{8}$$

subject to

$$x(0) = x_0,\ \lambda(T) = 0. \tag{9}$$

It can be directly seen that the problem (8), (9) is a special form of the problem (6), (7). Finally it is necessary to mention that if the state equation is given with both the initial condition $x(0) = x_0$ and with the terminal condition $x(T) = x_T$ then there is no transversality condition $\lambda(T)$ for adjoint variable λ. It means that instead of (9) the boundary conditions are

$$x(0) = x_0,\ x(T) = x_T. \tag{10}$$

3 Applications

Some of the two-point boundary value problems that can be solved using Algorithm 2 are introduced in this section.

3.1 Example of Standard Optimal Control Problem with Free Endpoint

Consider, see [12], the following nonlinear optimal control problem

$$\max \int_0^1 -\frac{1}{2}\left(x(t)^2 + u(t)^2\right) dt$$

subject to

$$\dot{x}(t) = -x(t)^3 + u(t), \quad x(0) = 5.$$

The Hamiltonian of the problem is

$$H(t, x(t), u(t), \lambda(t)) = -\frac{1}{2}(x(t)^2 + u(t)^2) + \lambda(t)(-x(t)^3 + u(t))$$

and the adjoint function λ is the solution to the equation

$$\dot{\lambda}(t) = x(t) + 3x(t)^2\lambda(t), \quad \lambda(1) = 0.$$

Because $u \in \mathbb{R}$ is not restricted the maximum principle yields $H_u(t, \widehat{x}, \widehat{u}, \lambda) = -\widehat{u} + \lambda = 0$ and thus $\widehat{u} = \lambda$. To abbreviate the notation we omit the symbol $\widehat{}$ above the variables x and u. Now the state equation and the adjoint equation can be written as boundary value problem

$$\begin{aligned} \dot{x}(t) &= -x(t)^3 + u(t), & x(0) &= 5, \\ \dot{u}(t) &= x(t) + 3x(t)^2 u(t), & u(1) &= 0. \end{aligned} \tag{11}$$

The MATLAB implementation of this system that returns the boundary condition $u(1)$ and that allows to seek the appropriate initial condition $u(0)$ can be found in the Table 1 on the left side. Having this function it is possible to use an implementation of bisection method, see [2], or it can be directly used the MATLAB function `fzero`. In particular the command `[u0,fval] = fzero(@fce, [-1 1])` finds the right initial value $u(0)$ in the interval $[-1,1]$. The result that was found is $u(0) \doteq -0.50000$ with the terminal value $u(T) \doteq 1.13039 \cdot 10^{-17}$.

```
function [uT, t, y] = fce(u0)
t0=0; T=1;
time=[t0, T];
x0=5;
init=[x0, u0];

[t, y]=ode45(@fce, time, init);
[m, n]=size(y);
uT=y(m, 2);

function dydt=fce(t, y)
x=y(1);
u=y(2);
dydt=[-x^3+u;x+3*x^2*u];
end

end
```

```
function [wT, t, x] = wealth(c0)
t0=0; T=80;
time=[t0, T];
w0=0;
init=[w0, c0];

[t, y]=ode45(@fce, time, init);
[m, n]=size(y);
wT=y(m, 1);

function dydt=fce(t, y)
A=20; B=65; gamma=2; alpha=0.04; r=0.1;
w=y(1);
c=y(2);
yy=(t >= A)*(t <= B);
dydt=[r*w+yy-c;-c/gamma*(alpha-r)];
end

end
```

Table 1 The MATLAB function for the system (11) on the left and for the system (12) on the right.

3.2 Optimal consumption or life-cycle model

The simplified model for optimal consumption is shortly presented here. Consider a consumer, see [11], [4], who plans his consumption from the present time, $t = 0$, until time T, $T > 0$. For further consideration let $c(t)$ denote the consumption of the consumer at time t and $y(t)$ denote his exogenous income. For instance it can be assumed that $y(t) = 1$ for $t \in [A,B] \subset [0,T]$ and $y(t) = 0$ otherwise. It means that the interval $[A,B]$ could represent a productive part of the consumer life. If $w(t)$ denote wealth of the consumer at time t then

$$\dot{w}(t) = r \cdot w(t) + y(t) - c(t), \quad w(0) = w(T) = 0,$$

where r is the constant real interest rate. It means that it is expected that the wealth at the beginning and at the end of the planning period is zero. Consumer wants to maximize the discounted future utility

$$\max \int_0^T e^{-\alpha t} u(c(t))dt,$$

where $\alpha > 0$ and u is a utility function. The Hamiltonian of the problem is

$$H(t,w(t),c(t),\lambda(t)) = e^{-\alpha t}u(c(t)) + \lambda(t)(r \cdot w(t) + y(t) - c(t)),$$

where the adjoint function λ is the solution to the equation

$$\dot{\lambda}(t) = -r\lambda(t).$$

As $c > 0$, the maximum principle yields $H_c(t,\widehat{w}(t),\widehat{c}(t),\lambda(t)) = e^{-\alpha t}u'(\widehat{c}(t)) - \lambda(t) = 0$ and thus $\lambda(t) = e^{-\alpha t}u'(\widehat{c}(t))$. To abbreviate the notation we omit the symbol $\widehat{}$ above the variables w and c. If we substitute λ by $e^{-\alpha t}u'(c)$, and $\dot{\lambda}$ by $-\alpha e^{-\alpha}u'(c) + e^{-\alpha t}u''(c)\dot{c}$ in the adjoint equation, we finally obtain the following boundary value problem

$$\begin{aligned} \dot{w}(t) &= r \cdot w(t) + y(t) - c(t), \quad w(0) = w(T) = 0, \\ \dot{c}(t) &= \tfrac{u'(c(t))}{u''(c(t))}(\alpha - r). \end{aligned} \tag{12}$$

We assume CRRA utility function given by the formula $u(c) = (c^{1-\gamma} - 1)/(1-\gamma)$, for more details [1], for the numerical experiment. Then $u'(c)/u''(c) = -c/\gamma$. The selected values of parameters for simulation are as follows

$T = 80$, $r = 0.10$, $\alpha = 0.04$, $A = 20$, $B = 65$ and $\gamma = 2$. The MATLAB implementation of this system that returns the boudary condition $c(T)$ and that allows to seek the appropriate initial condition $c(0)$ can be found in the Table 1, on the right side. Now, the command `[c0,fval] = fzero(@wealth, [0.05 0.1])` can be used to find the correct initial value $c(0)$ as an inner element of the interval $[0.05, 0.1]$. The result that was found is $c(0) \doteq 0.09528$ with the terminal value $w(T) = 6.38400 \cdot 10^{-9}$.

4 Conclusion

This paper introduced essential shooting method implementation steps, a basic numerical method for solution of two-point boundary value problem, in the context of numerical solution to optimal control problems. The method was successfully tested on two problems. One of them was a typical economics model for optimal consumption. Unfortunately it was not possible to present a graphical representation of the solutions that were found due to scope of the article. In the future work it is necessary to accomplish more tests and to deal with more demanding problems. For instance, it is necessary to deal with bang-bang control problems or problems with singular controls or problems with state constraints or other possible numerical methods, cf. [9].

Acknowledgements

Support of the Specific research project of the Faculty of Informatics and Management of University of Hradec Králové is kindly acknowledged. Special thanks belong to prof. Antonín Slabý and Ing. Kateřina Frončková for their valuable recommendations.

References

[1] Barro, R. & Sala-i-Martin, X. (2004). *Economic Growth*. Cambridge MA: MIT Press.
[2] Gander, W., Gander, M.J. & Kwok, F. (2010). *Scientific Computing, An Introduction using Maple and MATLAB.* Cham, Heidelberg: Springer.
[3] Grass, D. et. al. (2008). *Optimal Control of Nonlinear Processes, With Applications in Drugs, Corruption, and Terror*. Berlin, Heidelberg: Springer.
[4] Judd K.L. (1998). *Numerical Methods in Economics*. Cambridge MA: MIT Press.
[5] Chiang, A.C. (2000). *Elements of dynamical optimization*. Illinois: Waveland Press.
[6] Pontryagin, L.S. (1962). *Mathematical Theory of Optimal Processes*. New York: John Wiley & Sons.
[7] Pražák, P. (2004). On the Vidale - Wolfe Advertising Model, In Bauer, L. (Ed.), *Mathematical Methods in Economics 2004* (pp. 269 - 274). MU Brno.
[8] Pražák P. (2011). On Necessary Transversality Condition for Infinite Horizon Optimal Control Problems, In M. Dlouhý, V. Skočdopolová (Eds.), *Proceedings of 29th International Conference on Mathematical Methods in Economics 2011 - part II* (pp. 575-580). Professional Publishing.
[9] Pražák P. & Frončková, K. (2020). Discrete Model of Optimal Growth on a Finite Time Horizon as a Boundary Value Problem, In Kapounek, S., Vranova, H. (Eds.), *38th International Conference on Mathematical Methods in Economics, Conference Proceedings* (pp. 467-473). Mendel University in Brno.
[10] Seierstad, A., Sydsæter, K. (1987). *Optimal Control Theory with Economic Application. Amsterdam*: Elsevier,
[11] Sydsaeter, K., Hammond, P., Seierstad, A. & Strøm, A. (2005). *Further Mathematics for Economic Analysis*. Harlow: Prentice Hall.
[12] Sethi, S.P., Thompson, G.L. (2003). *Optimal Control Theory, Applications to Management Science and Economics.* Boston, Dordrecht, London: Kluwer Academic Publishers.
[13] Tu, P. N. V. (1994). *Dynamical Systems, An Introduction with Applications in Economics and Biology*. (2nd Edition), Berlin, Heidelberg: Springer.

Is the Czech Capital Market Weak Form Efficient?

Pavla Říhová[1], Milan Svoboda[2]

Abstract. This study deals with the short-term prediction of share prices in the Czech stock market. A stochastic model based on the analysis of simple Markov chains was used for the short-term prediction of share prices. Buy and sell signals were generated on the basis of this prediction. The prediction model is considered to be successful if trading with the use of this model outperforms the market; in other words, it yields higher returns than the passive "Buy and Hold" investment strategy. The study was performed using daily data from the Czech stock market for the 14-year period, from the beginning of 2006 until the end of 2019, i.e., approximately 3,500 trading days. The study results have shown that stocks that are traded in higher turnovers outperformed the Buy and Hold strategy over the period under review. Conversely, stocks that are traded daily in low turnovers do not outperform the Buy and Hold strategy. Thus, it seems that stocks traded in large turnovers do not behave efficiently. Stocks traded in small turnovers tend to behave as weak form efficient.

Keywords: Markov chain analysis, Efficient Market Hypothesis, share price prediction

JEL Classification: C22, G17
AMS Classification: 90C40, 91G15

1 Introduction

This empirical study is based on the long-term research and deals with the short-term prediction of share prices in the Czech stock market. The study works on the assumption that the share price on the stock exchange is formed in a continuous manner as a result of the mutual interaction between demand and supply. The supply and demand are generated by different types of traders (long-term investors and speculators). These traders have different time frames, use different methods for predicting the future development of share prices, evaluate market information in a different way, have different risk aversion and different amounts of capital at their disposal. The bid or demand for a given stock may be created not only because of a market participant's subjective view that the stock is undervalued or overvalued, but also for many other reasons, such as the need to raise money, to modify an investment strategy, to force purchases or sales in trades with borrowed assets, to buy stocks back by the joint stock company itself, etc. The simultaneous interaction of these numerous factors results in the fact that the share price can be regarded as a random variable that fluctuates around its fair price.

Another premise is that the share price moves in short-term trends and during the course of such trends, the share price accumulates a certain profit or loss relative to the price at the beginning of the trend. The greater the change is, the greater is the probability of a change in the trend. The key question is how high the accumulated loss or profit must be for the trend to change with a sufficiently high probability. A simple Markov chains analysis (MCA) is used to model the probability of a trend change.

The predictability of stock markets is in direct conflict with the Efficient Market Hypothesis (EMH). The basic ideas of the EMH were formulated by Fama [1]. According to the EMH, stock prices are unpredictable and therefore efficient. It implies that the market responds immediately to any new information. This information cannot be predicted, it comes to the market randomly, and hence the change of the exchange rate is also random, and the exchange rates take the so-called "random walks". Above-average profits cannot be made in efficient markets and other approaches are ineffective according to this theory. Roberts [5] then distinguishes three forms of efficiency: a weak form, a semi-strong form, and a strong form of efficiency. A weak form EMH means that the rate includes all the information from historical data, which means that the historical data cannot be used to predict market developments. The semi-strong form EMH is a state in which the price incorporates both historical data and all publicly available information, and also the methods of fundamental analysis fail in this form. The strong form

[1] University of West Bohemia in Pilsen/Faculty of Economics, Department of Economics and Quantitative Methods, Univerzitní 22, 306 14 Plzeň, divisova@kem.zcu.cz.
[2] University of West Bohemia in Pilsen/Faculty of Economics, Department of Economics and Quantitative Methods, Univerzitní 22, 306 14 Plzeň, svobodm@kem.zcu.cz.

EMH is when even non-public (insider) information is included in the price. Thus, even insider information is worthless in a strong form efficient market and does not yield above-average returns.

Initially, the EMH was widely accepted by academia. Later, however, studies questioning the EMH were published. For example, Shiller [6] points out the higher volatility of share prices than that which can be explained by dividend volatility. Haugen [3] believes that markets overreact to unexpected information and calls it over reactive capital market. The Czech stock market was also examined. A weak form of efficiency is predominantly investigated in the Czech stock market. The work of Filáček et al. [2] tends to the view that the Czech stock market behaves inefficiently. On the contrary, Hájek [4] is more inclined to think that the stock market is weak form efficient, and thus the methods of technical analysis do not yield above-average returns.

Thus, this study aims to indicate whether the short-term prediction of share prices using the MC analysis is successful and therefore the Czech stock market is weakly efficient. It will be considered successful if the trading based on this prediction outperforms the passive Buy and Hold strategy.

2 Data

The company called Patria Direct is the data source. The study was conducted using shares traded in the prestigious market segments of the Prague Stock Exchange over a fourteen-year period, from 2nd January 2006 to 2nd January 2020, i.e., for approximately 3,500 trading days. Throughout that period, only the following shares were traded in those segments: ČEZ (CEZ), Komerční banka (KB), O2CR (O2), Philip Morris ČR (PM). The study also included the shares of Pegas (PG), which were listed for trading only at the end of 2006. The shares have been included since the beginning of 2007. The daily opening and closing prices, daily trading volumes and dividends paid are available for all the shares mentioned above. KB shares were split during the stated period. In the case of O2, the majority owner of the company split into two entities. Those events have been dealt with as follows.

The KB share split took place on 12th May 2016 in the ratio of 5:1 (the shareholder received five new shares for one existing share). To maintain continuity, the dividends and share price before 12th May 2016 were divided by five. In the case of O2, the change of the majority owner was announced in November 2013 and the free float was reduced to approximately 10% following the subsequent mandatory takeover bid. Then, on 1st June 2015, O2 split into two companies, O2 and CETIN. The closing price of the O2 share on the day before the split was CZK 177.6. The shareholder received one new O2 share and one CETIN share for one original O2 share. On the first day after the split, the price of the CETIN share was CZK 133.5 and the price of the O2 share was CZK 69.2. At the time of the demerger, it was already known in the market that the majority shareholder had planned to squeeze out minority shareholders of CETIN shares and the shares would later be withdrawn from the market. The data continuity is maintained in the following way. The CETIN shares are sold at the price of CZK 133.5 on the very first day after the split and the proceeds are then treated in the same way as a dividend, i.e., they are used to buy O2 shares in accordance with the business strategy.

Figure 1 shows the price development of shares with dividend reinvestment after tax. Lines on Figure 1 are ordered according to the value of capital at the end of the period under review. This means that the lowest value of capital (1.34) is represented by the CEZ shares and the highest (3.79) by the O2 shares.

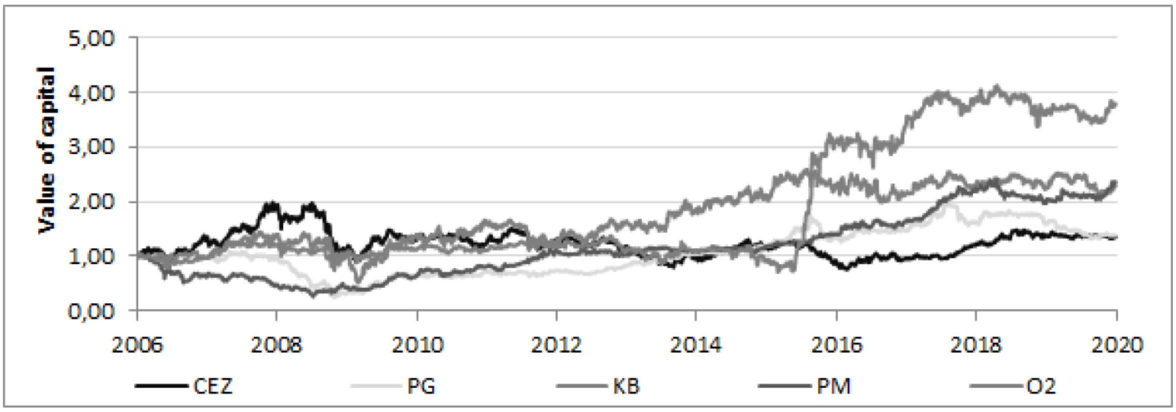


Figure 1 Value of capital

The steep increase in the appreciation of the O2 shares in the second half of 2015 is worth noticing in Figure 1. The increase was caused by ending the uncertainty that had been evident in the market since the information about the change of majority ownership leaked to the market. Figure 2 shows the 20-day moving averages of daily trading turnovers. It is evident that the PM and PG shares have lower turnovers by an order of magnitude than the KB and CEZ shares. The O2 shares were initially traded at similar turnovers to the KB shares, but after changing majority ownership and reducing free float, they are traded at similar turnovers to PM and PG.

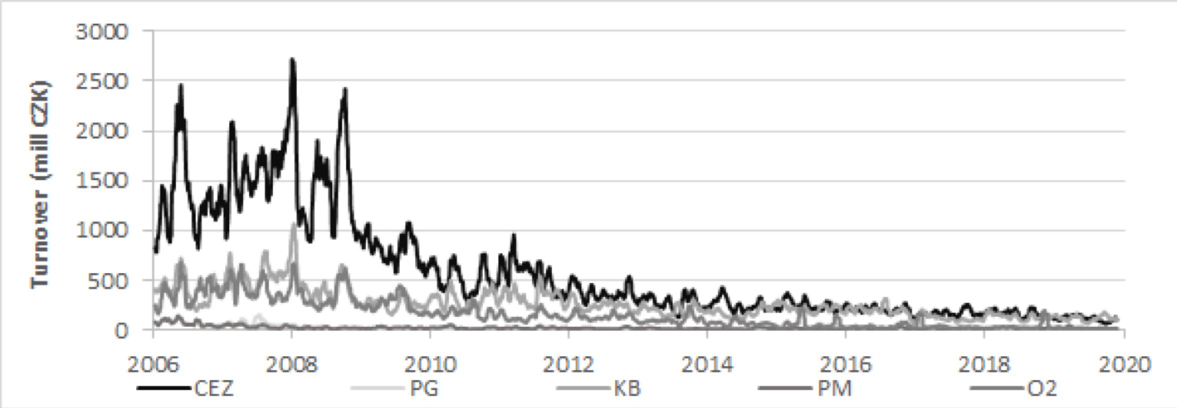


Figure 2 Turnover (millions CZK)

3 Methodology

As was already mentioned in the introduction, this study is based on the long-term research. Therefore, the methodology may overlap with that published in other papers by the team of authors.

To transform the share prices into a suitable MC, we use the method described in Svoboda and Gangur [7] in this study. The state space with eight states is used here. It is defined by multiples of the moving standard deviation and the cumulative change in the share price.

3.1 State space

The cumulative change in the share price is denoted by Y_t (Y_t expressed as a percentage is y_t). Y_t is interpreted as short basis indices of daily closing prices, where the basic period is the day of the trend change, i.e., the transition from a decline to an increase or vice versa. The duration of the trend is determined by the number of consecutive rising or falling closing prices. The calculation of Y_t, the cumulative price change, is formally described by (1)

$$Y_t = Y_{t-1}\frac{P_t}{P_{t-1}} \text{ if } \left(\ldots \le P_{t-2} \le P_{t-1} \le P_t\right) \text{ or } \left(\ldots \ge P_{t-2} \ge P_{t-1} \ge P_t\right) \text{ else } Y_t = \frac{P_t}{P_{t-1}}, \tag{1}$$

where P_t is the closing daily price at time t, P_{t-1} is the closing daily price at time t-1 a P_{t-2} is the closing daily price at time t-2. The moving standard deviation of length n is denoted by $s_{t,n}$ and is calculated according to formula (2)

$$s_{t,n} = \sqrt{\frac{1}{n}\sum_{i=0}^{n-1}\left(x_{t-i} - \bar{x}_{t,n}\right)^2}, \tag{2}$$

where x_{t-i} is the daily percentage change in the share price at time t-i and $\bar{x}_{t,n}$ is moving average length n at time t. The general model of state space is defined as follows:

D_4: $y_t < -3\Delta_t$	G_1: $0 \le y_t < 1\Delta_t$
D_3: $-3\Delta_t \le y_t < -2\Delta_t$	G_2: $1\Delta_t \le y_t < 2\Delta_t$
D_2: $-2\Delta_t \le y_t < -1\Delta_t$	G_3: $2\Delta_t \le y_t < 3\Delta_t$
D_1: $-1\Delta_t \le y_t < 0$	G_4: $3\Delta_t \le y_t$,

where $\Delta_t = k{\cdot}s_{t,n}$ k is the multiple of the moving standard deviation. Thus, the state space model is unambiguously determined by parameters k and n. The states where the share price falls are denoted by D_i. D_1 is the state with the minimum price decline and, conversely, D_4 is the state with the maximum price decline. The states where the share price increases are denoted by G_i. G_1 is the state with the minimum price increase and, conversely, G_4 is the state with the maximum price increase.

3.2 Trading rules

Buy signals are generated when the share price sufficiently falls and sell signals are generated when the share price sufficiently rises. D_3 and D_4 are considered to be the states with a sufficient decline, these states generate buy signals and sell signals are generated by the G_3 and G_4 states. Combining the buy and sell states results in 4 trading strategies: D_3 - G_3, D_3 - G_4, D_4 - G_3 and D_4 - G_4. The D_3 - G_3 strategy means that the D_3 state generates buy orders (if we do not hold the shares anymore) and the G_3 state generates sell orders (if we hold the shares). In other words, we divide the capital we want to invest in a particular stock into quarters and we trade with each quarter using one of the strategies. Trading is always carried out by the following rules:

- one deal (transaction) is understood as the purchase and subsequent sale of a share,
- two consecutive purchases are not possible,
- if a buy or sell signal is generated on a day, the deal is realized at the next day's opening price,
- the entire capital is always invested, so in theory even parts of shares can be bought,
- no transaction fees are taken into account,
- dividends and other income (hereinafter referred to as dividends), in the event that we are entitled to them, are reinvested after tax.

The invested capital value is calculated according to the following formula (3)

$$C_n = C_0 \prod_{i=1}^{n} \frac{S_i + d_i}{B_i}, \tag{3}$$

where $C_0 = 0.2500$ is the initial value of capital, C_n is the value of capital after the n-th transaction, S_i is the sales price in the i-th transaction, d_i represents after-tax dividends if there was a relevant date during the i-th transaction, B_i is the purchase price in the i-th transaction. The total value of capital C is given by the sum of the capital of individual trading strategies see formula (4).

$$C = C_{D_3-G_3} + C_{D_3-G_4} + C_{D_4-G_3} + C_{D_4-G_4}, \tag{4}$$

where C_{Di-Gj} is the achieved appreciation of a trading strategy in which a buy signal generates the D_i state and a sell signal generates the G_j state.

4 Results and discussion

In total, 80 state space models were calculated for the observed stocks. The k parameter gradually took values from 0.5 to 2.0 in increments of 0.1. Five lengths of moving standard deviation were tested for each value of the k parameter, with the n parameter taking values from 10 to 30 in increments of 5. The best results were obtained when the n parameter was equal to 20 or 25 in combination with the k parameter values of 1.1 and 1.2. The resulting returns for individual shares for parameters k = 1.2, n = 20 and the passive Buy and Hold strategy are presented in Table 1. The table shows the results for the five-year sliding windows as well as the result for the entire 14-year period. The length of five years for the sliding windows was chosen because the minimum recommended investment periods for investing in stocks is five years. The values in bold are those when the active trading based on the MC prediction outperformed the B&H.

Years	**CEZ**		**PG**		**KB**		**PM**		**O2**	
	C	B&H	C	B&H	C	B&H	C	B&H	V C	B&H
2015 - 19	**31.1%**	15.3%	-11.1%	16.3%	**26.4%**	7.3%	47.3%	98.8%	20.2%	301.2%
2014 - 18	**49.4%**	41.3%	16.6%	51.9%	**56.1%**	20.0%	35.8%	82.2%	-4.1%	227.0%
2013 - 17	**50.3%**	2.1%	24.4%	91.3%	**80.8%**	41.1%	48.9%	118.4%	-2.3%	238.8%
2012 - 16	**39.2%**	-25.8%	38.7%	97.5%	**119.6%**	65.8%	37.9%	48.1%	2.1%	182.1%
2011 - 15	**20.1%**	-28.5%	32.6%	85.7%	**95.5%**	39.3%	62.2%	71.9%	9.9%	178.8%
2010 - 14	**46.8%**	-12.0%	47.9%	82.5%	**133.7%**	54.7%	52.0%	79.5%	**1.4%**	-17.5%
2009 - 13	**35.0%**	-15.2%	55.0%	209.1%	**129.4%**	83.1%	48.0%	149.8%	**32.1%**	7.7%
2008 - 12	**9.6%**	-38.3%	-23.6%	-15.6%	**97.1%**	18.5%	16.7%	111.2%	**27.0%**	-4.8%
2007 - 11	**41.6%**	-3.7%	**-17.9%**	-25.9%	**70.5%**	34.1%	11.4%	75.6%	**28.7%**	23.2%

2006 - 10	**69.1%**	26.1%			**95.8%**	61.5%	-38.4%	-18.1%	**26.8%**	12.0%
2006 - 19	**188.2%**	34.8%	-5.9%	37.9%	**337.4%**	128.3%	35.9%	136.2%	36.1%	279.2%

Table 1 Yield

Table 1 clearly shows that in the case of KB and CEZ shares, the active trading based on the MC prediction outperformed the passive B&H strategy for each five-year period. Conversely, in the case of PG and PM shares, the trading based on the MC prediction outperformed the B&H strategy in only one five-year period. This exception is the very first period in the case of PG shares. In the case of O2 shares, the MC prediction was successful until 2014. After 2015, it was no longer successful. If we examine the daily trading turnovers, the idea that the success of the prediction model is related to the trading turnover comes into consideration. The KB and CEZ shares have higher-order trading turnovers than the PM and PG shares. As already mentioned in the case of PG shares, the passive strategy was outperformed only in the first five-year period. The fact that in 2007 and 2008 the daily trading turnover of PG shares was in the higher-order tens of millions is worth noticing, and since 2009 the trading turnovers have only been in units of millions of CZK. It was especially in those first two years when the active trading based on the MC prediction significantly outperformed the passive strategy of B&H. The active trading lost this "edge" in the following years. However, by that time the shares were traded in much lower turnovers. Until 2014, the O2 shares were traded in similar turnovers as the KB shares, and the trading outperformed the B&H strategy in line with the MC prediction. After the change of majority owner and the mandatory share buyback offer, which many shareholders took advantage of, the free float decreased. The trading turnovers dropped substantially. The O2 shares began to be traded in the similar turnovers as the PM and PG shares and the prediction model ceased to be successful. On the other hand, it must be mentioned that in 2015 the period of uncertainty about the new majority shareholder's plans ended and the share price rose sharply. All sliding windows containing the year of 2015 are distorted by this one-off event.

Figure 3 shows the trend in the ratio of the appreciation achieved by the active trading to the appreciation by the B&H strategy. The drop in the ratio for the O2 shares in the second half of 2015 should be noted. The steep peaks in 2008 are caused by greater slumps in prices to some extent, with the capital value in the B&H strategy reacting immediately to price changes, but the capital value in the active trading strategy reacting with a time lag. The value of capital is not recalculated every day, but only at the moment of selling the shares, as shown in formula (3).

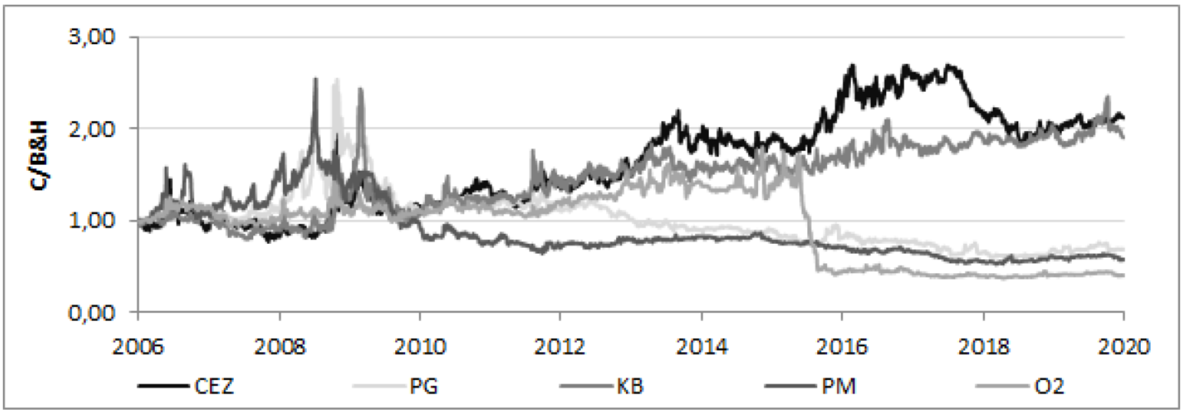


Figure 3 Ratio C/B&H

5 Conclusion

The answer to the question whether the Czech stock market is weakly efficient is not definite. Some shares have managed to outperform the passive Buy and Hold strategy throughout the 14-year period under review, while others have not. In the case of shares that are traded in higher-order turnovers, the trading based on the prediction model was successful and outperformed the passive B&H strategy in each five-year period. Conversely, in the case of shares traded in low turnovers the B&H strategy was outperformed in only one five-year period. This, of course, does not necessarily mean that these shares behave efficiently. Only the prediction model did not work as it should.

This ambiguity is in accordance with the already mentioned works of Filáček et al. [2] or Hájek [4]. The work of Filáček et al. [2] inclined to the opinion that the Czech stock market is behaving inefficiently. On the contrary, Hájek [4] pointed out that the Czech stock market was approaching a weak form of efficiency. The efficiency of the Czech stock market was also examined by other authors. Some have concluded that the Czech stock market is weakly efficient, some considered the Czech stock market to be inefficient. However, these studies have related to the Czech stock market development at the end of the 1990s and the beginning of the new millennium.

According to the authors of this study, a partial explanation for this situation could be short-term traders - speculators. These speculators prefer highly liquid shares that are traded in large turnovers. They often trade margin (financial leverage) and make profits with relatively small price changes. If the share price rises, they start selling their shares, thereby causing a trend reversal. They mainly use technical analysis (TA) tools to forecast the share price. As they use similar tools, they act in concert and thus create a self-fulfilling prophecy. Shares that are traded in small turnovers in the market are bought by long-term investors. These investors pursue long-term profits. They use fundamental analysis tools; they do not use TA tools. Since these traders do not use TA, the TA tools do not work for these shares. However, this hypothesis would need to be studied for a longer period of time.

The authors will continue their research. They will certainly try to confirm these results using longer time series. The only problem is that the Czech stock market is small, and some shares are traded for a short period of time. Nowadays, the O2 and PG shares are no longer traded on the stock market. Applying these methods to foreign stock markets, both developed and emerging ones, will be the next research direction.

Acknowledgements

The research project was supported by Grant SGS-2021-022 of the Faculty of Economics, University of West Bohemia in Pilsen.

References

[1] Fama, E. F. (1965). The behavior of stock market prices. *The Journal of Business*, 38(1), 34-105. https://doi.org/10.1086/294743.

[2] Filáček, J., Kapička, M., & Vošvrda, M. (1998). Testování hypotézy efektivního trhu na BCPP. *Czech Journal of Economics and Finance (Finance a úvěr)*, 48(9), 554-566.

[3] Haugen, R. A. (1999). *The new finance: The case against efficient markets (2nd edition).* New Jersey: Prentice Hall.

[4] Hájek, J. (2007). Czech capital market weak-form efficiency, selected issuses. *Prague Economic Papers*, 16(4), 303-318. https://doi.org/10.18267/j.pep.310

[5] Roberts, H. W. (1967). *Statistical versus clinical prediction in the stock market*. Unpublished manuscript. University of Chicago.

[6] Shiller, R. J. (1981). Do stock prices move too much to be justified by subsequent chance in dividends? *The American Economic Review*, 71(3), 421–436. https://doi.org/10.3386/w0456.

[7] Svoboda, M., & Gangur, M. (2017). Using of Markov chains with varying state space for predicting short-term of the share price movements. In P. Pražák (Ed.), *35th International Conference Mathematical Methods in Economics, Conference Proceedings* (pp. 749-754). Hradec Králové: Gaudeamus.

Robust Optimization Approach in Travelling Salesman Problem with Service Time

Tereza Sedlářová Nehézová[1], Robert Hlavatý[2]

Abstract. The travelling salesman problem has been addressed by researchers over again in the past decades. The main focus of the problem is computing a Hamiltonian path over a given set of vertices with respect to edge costs, while other criteria and aspects are considered concerning the needs of practitioners. In our contribution, we consider a travelling salesman problem with a service time needed when each vertex is visited. The expected service time may vary depending on the time of the day. We propose an optimisation model for minimising the total journey length, including the service time. Consequently, we consider that both service time on vertices and travelling time between vertices may be uncertain. The Gamma-robustness approach is used in order to deal with the uncertainty issues and propose an optimal travelling strategy.

Keywords: robust optimization, travelling salesman problem, uncertainty

JEL Classification: C61
AMS Classification: 90C05, 90C08, 90C11

1 Introduction

The classical traveling salesman problem (TSP) is well known NP-Hard problem that has been studied extensively in last years, for instance see [3], [14] or [16] for a recent review. It is defined as a problem of finding the shortest way in set of nodes while visiting each of them only once [3]. It has a wide area of applications in various sectors such as telecommunications or IT. Until today no efficient algorithm has been developed for its general solution. Comparison of different algorithms for solving TSP has been made in [1]. From the linear programming view, we can name work of Dantzig, Fulkerson and Johnson [9] and their solution for set of 49 cities. Recently, there has been use of neural networks [17], which has made a great step forward in solving TSP and can effectively improve the accuracy of the approximate solution. Also, the machine learning has been in use for TSP and in instances with 100 cities the solutions provided are roughly 1,33% away from optimal solution [10].

Lots of variations of TSP exists these days [13], each of them is a variant of classical TSP to solve different problems and make it more suitable for use in different applications. Time constraints or dependencies are widely used e.g. in Yuan et al. [19] or Gendreau et al. [11] in case of Traveling Salesman Problem with Time Windows. The problem is defined on a directed graph and for each edge is assigned time interval representing a time window, in which the vertex must be visited. Also, each node has defined service time. Another modification of TSP is the Time-dependent Traveling Salesman problem (TDTSP), that seeks a solution for case where edge-costs depend on the time the edge is entered [2].

Most of the models mentioned above consider all its parameters precisely known in advance. That is unfortunately not the case in most of real-life situations, so there is a need for a tool that allows us to consider deviations in data values and is immune to data-side uncertainty. First author who addressed such issue in case of linear optimization was Soyster [18], and for further description of robust optimization (RO) see [5]. The important step in robust optimization is construction of uncertainty set, which can be made by different approaches, including interval, ellipsoidal, and polyhedral [4]. In recent years has appeared new approach towards modelling uncertainty that integrates big data into RO [6].

In this paper we discover possibilities of using TSP in such cases that require service time when each vertex is visited, and the length of that service time is considered uncertain. Further, those uncertainties are dependent on the order of each vertex. We propose a modified robust TSP approach based on Γ robustness introduced by Bertsimas and Sim in [7], that describes the uncertainty as symmetric deviations from given nominal value. This approach also allows to control the number of deviation coefficients by parameter Γ and thus test different scenarios with multiple levels of uncertainty.

[1] Czech University of Life Sciences, Kamýcká 129, Prague, Czech Republic, nehezova@pef.czu.cz.
[2] Czech University of Life Sciences, Kamýcká 129, Prague, Czech Republic, hlavaty@pef.czu.cz.

2 Materials and methods

Integer programming formulation of Traveling Salesman Problem is shown, as well as its robust counterpart based on Γ robustness approach.

2.1 Traveling Salesman Problem (TSP) formulation

The following integer programming formulation of TSP is further assumed:

$$z = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}$$

s.t.

$$\sum_{j=1}^{n} x_{ij} = 1, i = 1,2, \dots, n$$

$$\sum_{i=1}^{n} x_{ij} = 1, j = 1,2, \dots, n \tag{1}$$

$$u_i - u_j + n x_{ij} \le n - 1, i = 1,2, \dots, n; j = 2, \dots, n; i \ne j$$

$$x_{ij} \epsilon \{0; 1\}, i = 1,2, \dots, n; j = 1,2, \dots, n$$

where number of nodes to be visited is denoted by n, the cost for each edge is represented by c_{ij}, x_{ij} is binary variable, value of 1 states that the edge between the nodes i a j is used, 0 if not, u is an auxiliary variable and it is part of constraint that prevents the creation of partial cycles and also denotes the serial number of each vertex in the optimal travelling sequence.

2.2 Robust TSP

The robust formulation of TSP that assume the uncertainties in cost coefficients has been made in [15] can be seen below:

$$min\ E$$

s.t.

$$\sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} + \Gamma z + \sum_{(i,j) \in U} p_{ij} \le E$$

$$z + p_{ij} \ge \delta_{ij}^{c} x_{ij}, \forall (i,j) \in U$$

$$\sum_{j=1}^{n} x_{ij} = 1, i = 1,2, \dots, m$$

$$\sum_{i=1}^{m} x_{ij} = 1, j = 1,2, \dots, n \tag{2}$$

$$z \ge 0$$

$$p_{ij} \ge 0, \forall (i,j) \in U$$

$$u_i - u_j + n x_{ij} \le n - 1, i = 1,2, \dots, n; j = 2, \dots, n; i \ne j$$

$$x_{ij} \epsilon \{0; 1\}, i = 1,2, \dots, n; j = 1,2, \dots, n$$

where maximal deviation of c_{ij} is denoted as δ_{ij}^{c}, the level uncertainty is controlled by parameter $\Gamma, 0 \le \Gamma_i \le |U|$ where Γ indicates the maximum number of coefficients c_{ij}, which are expected to deviate by no more than δ_{ij}^{c},

p_{ij} and z are auxiliary variables, U indicates the set of indices (i, j) of those c_{ij} for which the deviation is considered [15].

With this approach, there are assumptions stated in [7] and [8] that must be held:

1. For each c_{ij} is defined its nominal value and maximal deviations δ_{ij}^c.
2. Value c_{ij} belongs to $[c_{ij} - \delta_{ij}^c, c_{ij} + \delta_{ij}^c]$.
3. One is able to define maximum number of coefficients c_{ij} that will simultaneously deviate at most by δ_{ij}^c. This number is denoted by parameter Γ.

3 Results

Neither of TSP formulations mentioned above assume any service time for each vertex. For that purpose, we modify (2), that allows us to include the service time and its uncertainties. The value of the uncertainty depends on the serial number of the node in the resulting path through the set of all nodes. We test 2 different ways of how to define the dependency on the node serial number. For calculations was used Gurobi® 6. 5 solver engine [12].

3.1 Formulation of the model

The way of incorporating the service time to the model is as following:

1) For each node i to which we add an index a we create a fictitious node with added index b, and it is assumed that the nodes ia and ib are topologically the same node.
2) The edge between the nodes ia and ib is denoted by x_{ia_ib}
3) The cost of the edge x_{ia_ib} is equal to the service time in given node and is denoted as c_{ia_ib}
4) The value of δ_{ij}^c depends on the variable u_i.

Then we can modify the model (2) in following way:

$$
\begin{gathered}
min\ E \\
\text{s.t.} \\
\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} + \sum_{i=1}^{n} c_{ia\_ib}x_{ia\_ib} + \Gamma z + \sum_{(ia,ib)\in\bar{U}} p_{ia\_ib} \leq E \\
z + p_{ia\_ib} \geq \delta_{ia\_ib}^c x_{ia\_ib}, \forall(ia, ib) \in \bar{U} \\
\sum_{j=1}^{n} x_{ij} = 1, i = 1, \dots, m \\
\sum_{i=1}^{m} x_{ij} = 1, j = 1, \dots, n \\
x_{ia\_ib} + x_{ib\_ia} = 1, i = 1, \dots, m \\
c_{ia\_j} = c_{ib\_j}, j, \dots, n, i \neq j \\
\delta_{ia\_ib}^c = \frac{u_{ia} + u_{ib}}{2} R, \forall(ia, ib) \in \bar{U} \\
z \geq 0 \\
p_{ia\_ib} \geq 0, \forall(ia, ib) \in \bar{U} \\
u_i - u_j + nx_{ij} \leq n - 1, i = 1,2, \dots, n; j = 2, \dots, n; i \neq j \\
x_{ij} \epsilon \{0; 1\}, i = 1,2, \dots, n; j = 1,2, \dots, n \\
x_{ia\_ib} \epsilon \{0; 1\}, i = 1,2, \dots, n; j = 1,2, \dots, n
\end{gathered}
\tag{3}
$$

where R is any coefficient that suits the needs how to describe the uncertainty, a and b are denotes the original and fictitious nodes. The equation $x_{ia\_ib} + x_{ib\_ia} = 1$ ensures that the edge between ia and ib is used. $c_{ia\_j} = c_{ib\_j}$ statest that the edge cost from ia to j is the same as from ib to j. $\delta^c_{ia\_ib} = \frac{u_{ia}+u_{ib}}{2} R$ ensures that the service time depends on the position of the node in the travelling sequence. We assume that ia and ib correspond to u_i and u_j based on the order in which they are visited.

3.2 Practical example

For illustration how it is possible to implement proposed approach, we assume following example with 6 nodes and with symmetrical matrix of costs and service times:

	1	2	3	4	5	6
1	0	0,68	0,7	0,74	0,86	0,88
2	0,68	0	0,52	0,96	1,52	1,16
3	0,7	0,52	0	0,86	1,6	1,66
4	0,74	0,96	0,86	0	1,56	1,62
5	0,86	1,52	1,6	1,56	0	1,6
6	0,88	1,16	1,66	1,62	1,6	0
Service time	0	0,25	0,25	0,75	0,75	0,5

Table 1 Matrix of costs and service times

For each node i is generated node ib, thus $n = 12$. We have tested different ways how to define value of $\delta^c_{ia\_ib}$. First we define it with direct dependence on coefficient c as:

$$\delta^c_{ia\_ib} = \frac{u_{ia} + u_{ib}}{2} \cdot H \cdot c_{ia\_ib}, \forall (ia, ib) \in \bar{U} \tag{4}$$

where H can be any decimal number, in our case $H = 0{,}1$.

Then we define it as value regardless of the c as:

$$\delta^c_{ia\_ib} = \frac{u_{ia} + u_{ib}}{2} \cdot H, \forall (ia, ib) \in \bar{U} \tag{5}$$

Results for tested scenarios are as follows:

Γ	*OF value*	*1.* *Order of the nodes*	*OF value*	*2.* *Order of the nodes*
0	8,24	$x_1 \to x_5 \to x_6 \to x_2 \to x_3 \to x_4$	8,34	$x_1 \to x_6 \to x_2 \to x_3 \to x_4 \to x_5$
1	9,02		10,1	
2	9,8		12,2	
3	10,6		13,4	
4	10,6		15	$x_1 \to x_5 \to x_6 \to x_2 \to x_3 \to x_4$
5	10,8		16,3	$x_1 \to x_4 \to x_3 \to x_2 \to x_6 \to x_5$
6	11,3	$x_1 \to x_5 \to x_6 \to x_2 \to x_4 \to x_3$	17,6	

Table 2 Results for first set of scenarios

Clearly the second way has a greater impact to the solution, thus we can see higher increase in value of objective function, and more different orders of the given nodes. In those 2 scenarios, the value of $\delta^c_{ia\_ib}$ simply depends on the serial number of nodes, which is not practical in all applications. There are situations where simple linear growth is not sufficient solution to describe the relationship between the uncertainty and the order of the node.

We propose another way of how to define the uncertainty for the service time by using modular arithmetic:

$$\delta^c_{ij} = (\frac{u_{ia} + u_{ib}}{2} mod\ \frac{n}{2}) \cdot H \cdot c_{ia_b}, \forall (ia, ib) \in \bar{U} \tag{6}$$

This way can be helpful in scenarios where after a certain number of visited nodes there is a decrease or restart of the deviations. E.g., part of the nodes is visited in the morning, the remaining part in the afternoon. In the morning the service times can be increasing to certain point, then return to zero and start grow again. In our case the n= 12, and we assume that after visiting half of the nodes the deviations go again from zero. Thus, we assume mod 6. Results can be seen in table below:

x	OF value	Order of the nodes
0	8,24	$x_1 \to x_5 \to x_6 \to x_2 \to x_3 \to x_4$
1	8,57	
2	8,90	
3	9,24	
4	9,57	
5	9,71	$x_1 \to x_6 \to x_5 \to x_4 \to x_3 \to x_2$
6	9,45	$x_1 \to x_6 \to x_2 \to x_3 \to x_4 \to x_5$

Table 3 Results for scenarios with use of modulo

3.3 Discussion

As can be seen in the results above, the uncertainty affects the solution. Further, the level of solution changes depends on how the uncertainty is defined. As can be seen in table 2, the second scenario that generates more significant changes in objective function values generates more different tours than the first scenario. Moreover, in the table 3 can be seen even greater variability in the resulting order of nodes. Thus, one must be able to choose the accurate way of defining the uncertainty.

Regarding time dependence, we can compare our approach with time dependent TSP used in [2]. There, the time influence is aimed towards the cost of the arc itself, as it depends on its position in the tour while all the data are certain. In our approach, the arc costs stay the same, only the deviations are dependent on the time/node position. Also, the time dependency is derived from an auxiliary variable u, rather than from position of the node in relation to source and terminal node, as it is done in [2]. In case of our formulation, all solution times were under 5 seconds (running on configuration 11th Gen Intel(R) Core (TM) i5-1135G7 @ 2.40GHz - 2.42 GHz, 8 GB RAM).

Disadvantage of our approach may be that we define another for each already existing node that is assumed to have service time. This leads to up to twice the number of nodes compared to the original dimension of the problem. This is not the most efficient way in case of problem like TSP, that is NP-Hard. Thus, our approach is rather usable for smaller problems.

4 Conclusion

We have shown how to incorporate service times into the Traveling Salesman Problem with dependency on the serial number of the nodes. Service time is a common occurrence in logistics, where the TSP has wide range of applications. Moreover, these service times are often uncertain and can vary, depending on, for example, the time of day. We proposed 2 different ways of how to define the time dependency of the uncertainty in service times by use of the order of each node. Our approach does not need any sufficient modifications of the original Robust TSP formulation and it is possible to use tools that are standard for solving integer programming problems.

Acknowledgements

This paper was supported by Internal Grant Agency (IGA) of FEM CZU, project No. 2021B0003.

References

[1] Abdulkarim, H., & Alshammari, I. F. (2015). *Comparison of Algorithms for Solving Traveling Salesman Problem.* International Journal of Engineering and Advanced Technology, ISSN, 2249 – 8958.

[2] Abeledo, H., Fukasawa, R., Pessoa, A., & Uchoa, E. (2013). The time dependent traveling salesman problem: polyhedra and algorithm. *Mathematical Programming Computation*, 5(1), 27–55. https://doi.org/10.1007/s12532-012-0047-y

[3] Applegate, David L. (2006). *The traveling salesman problem: a computational study*. Princeton: Princeton University Press

[4] Ben-Tal, A., & Nemirovski, A. (2000). *Robust Solution of Linear Programming Problems Contaminated with Uncertain Data*. Mathematical Programming, 88(3): 411-424.

[5] Ben-Tal, A. & Nemirovski, A. (1999). *Robust solutions of uncertain linear programs. Operations Research* Letters 25, 1-13.

[6] Bertsimas, D., Gupta, V., & Kallus, N. (2017). Data-driven robust optimization. *Mathematical Programming*, 167(2), 235–292. https://doi.org/10.1007/s10107-017-1125-8

[7] Bertsimas, D. & Sim, M. (2004). *The price of robustness*. Operations Research (52), 35-53.

[8] D. Bertsimas & M. Sim. (2003). Robust discrete optimization and network flows. *Mathematical Programming*, vol. 98, no. 1–3, pp. 49–71, Sep. 2003, doi: 10.1007/s10107-003-0396-4.

[9] Dantzig, G., R. Fulkerson & S. Johnson (1954). Solution of a Large-Scale Traveling-Salesman Problem. *Journal of the Operations Research Society of America*, 2(4), 394-410.

[10] Francois, A., Cappart, Q. & Rousseau, L.-M., (2019). How to Evaluate Machine Learning Approaches for Combinatorial Optimization: Application to the Travelling Salesman Problem. *Computing Research Repository.* Available at: http://arxiv.org/abs/1909.13121

[11] Gendreau, M., Hertz, A., Laporte, G., & Stan, M. (1998). A generalized insertion heuristic for the traveling salesman problem with time windows. *Operations Research*, 46(3), 330–335. https://doi.org/10.1287/opre.46.3.330

[12] Gurobi Optimization, LLC. (2020) *Gurobi Optimizer Reference Manual*; Gurobi Optimization, LLC: Beaverton, OR, USA. 2020.

[13] Gutin, G., Punnen, A., Barvinok, A., Gimadi, E., & Serdyukov, A. (2001). *The Traveling Salesman Problem and Its Variations*.

[14] Lawler, E. L., Lenstra, an K., Kan, A. H. G. R., & Shmoys, D. B. (1991). *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization*.

[15] Nehézová, T. (2019). *Robust optimization approach in travelling salesman problem*. 37th International Conference on Mathematical Methods in Economics 2019, 535–540. https://mme2019.ef.jcu.cz/files/conference_proceedings.pdf

[16] Osaba, E., Yang, X.-S., & Del Ser, J. (2020). *Traveling salesman problem: a perspective review of recent research and new results with bio-inspired metaheuristics*. In Nature-Inspired Computation and Swarm Intelligence (pp. 135–164). Elsevier. https://doi.org/10.1016/B978-0-12-819714-1.00020-8

[17] Shi, Y., & Zhang, Y. (2022). *The neural network methods for solving Traveling Salesman Problem*. Procedia Computer Science, 199, 681–686. https://doi.org/10.1016/j.procs.2022.01.084

[18] Soyster, A.L.: Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, 21, (1973), 1154-1157.

[19] Yuan, Y., Cattaruzza, D., Ogier, M., & Semet, F. (2020). A branch-and-cut algorithm for the generalized traveling salesman problem with time windows. *European Journal of Operational Research*, 286(3), 849–866. https://doi.org/10.1016/j.ejor.2020.04.024

Influence of the Inclusion of the Dominated Alternative on the Final Arrangement of Alternatives in Selected MCDA Methods

Jana Sekničková[1], Martina Kuncová[2]

Abstract. Multi-criteria decision making (MCDM) usually covers the multi-criteria evaluation of alternative methods, sometimes called multi-criteria decision analysis (MCDA) and multi-criteria programming. MCDA methods have been developed to help the decision-maker to find a solution among a lot of different alternatives on the basis of several selected criteria. The goal may lie in sorting alternatives into acceptable and unacceptable, alternatives' ranking, or in the choice of the best alternative. An alternative that is dominated, meaning there is another alternative better or equally evaluated in all criteria, cannot be the best in the best choice problem, however, it can be at the forefront when the aim is alternatives' ranking, and it can also influence the final arrangement. The aim of this article is to analyze the effect of the inclusion of dominated alternatives on the results. WSA and TOPSIS methods using criteria weights were selected for the analysis in this paper.

Keywords: multi-criteria evaluation of alternatives, ranking problems, dominated alternatives

JEL Classification: C44, C61
AMS Classification: 90B50, 90C29

1 Introduction

Multi-criteria decision-making (MCDM) methods, especially multi-criteria decision analysis (MCDA) methods, are used in many areas where it is necessary to compare multiple alternatives according to multiple criteria. The main aim of the MCDA as the discrete decision-making problem could be to find the efficient (or the best) alternatives according to several selected criteria, to separate the alternatives into acceptable and unacceptable, or to find the complete ranking of alternatives [5]. If we are looking for the winner (the best alternative), the so-called dominated alternatives can be omitted from the data set as they can never be in the first place; i.e. alternatives for which there is another alternative having better (or the same and better) values for all criteria as the dominated alternative. Sometimes it is better not to look for the best alternative only, but also for the rank of all alternatives. But omitting the dominated alternatives may affect the overall ranking of the other, non-dominated alternatives. Therefore, in this article, we focus on selected MCDA methods to determine whether the omission of dominated alternatives may change the final order. A model comparing the regions of the Czech Republic from the point of view of 5 criteria was used as an example.

Other articles also address the issue of dominated alternatives. Huber et al. [6] focused on asymmetrically dominated alternatives and the impact of adding these alternatives to the model. An asymmetrically dominated alternative is dominated by one item in the set of alternatives but not by another. Adding such an alternative can change the probability of the better, so-called dominant, alternative, being chosen. Xu and Xia [14] described the process of identifying and eliminating the dominated alternatives in multi-criteria decision-making under uncertainty with fuzzy numbers for the alternatives' evaluation. Wang [13] studied the ranking irregularities when ELECTRE II or ELECTRE III methods were used – he showed that these methods do allow some types of ranking irregularities to happen. He also found out that the common practice of dropping dominated alternatives may lead to misleading conclusions if the analysis involves regret or rejoicing effects.

Since we did not find many other articles that would address the impact of the removal of dominated alternatives on the final ranking, we decided to contribute to this topic ourselves. For the analysis, we chose among the methods resulting in a complete order of alternatives. Finally, we decided to show 2 methods: one method using the

[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

principle of utility maximization (WSA method) and one method using the principle of minimizing the distance from the ideal solution (TOPSIS method). However, we would get similar results also when analyzing the methods of the PROMETHEE and ELECTRE classes.

2 Methods and data

MCDA methods are widely spread to help the decision-maker to find the best alternative or the ranking of alternatives [7]. The main areas of the methods' usage can be found in [11]. For the solution, it is necessary to know the list of alternatives $(a_1, a_2, \dots, a_p)$, the list of criteria $(f_1, f_2, \dots, f_k)$, and the matrix $\mathbf{Y} = (y_{ij})$ of the values of each alternative a_i and criterion f_j. In addition to this information, the decision-maker should also specify the order or the weights of the criteria (usually stated as v_j for $j = 1, \dots, k$) [5]. In this article, the methods with criteria weights are analyzed – they are WSA and TOPSIS. The influence of inclusion or removal of the dominated alternatives is demonstrated in a small example of the comparison of 14 Czech regions by 5 criteria with several different weight vectors. Data for the analysis were taken from the Czech statistical office [3]. They describe the selected economic characteristics of the 14 Czech regions.

2.1 Data used for the analysis

Table 1 summarizes the data used for the calculations; 14 Czech regions were compared with respect to available data from 2019. For the comparison, 5 criteria are used, all are maximized (economic activity, average wage, income per capita, consumption per capita, and investments per capita). This data was described in [8]; another comparison in the years 2012-2018 via the PROMETHEE method is in [10]. The green cells in Table 1 show the best values and the red cells the worst values. For the examples, we use four weight vectors of the criteria that prefer the selected criteria in different ways: $\mathbf{v}^{(1)} = (0.2, 0.2, 0.2, 0.2, 0.2)$, $\mathbf{v}^{(2)} = (0.20, 0.17, 0.24, 0.16, 0.23)$, $\mathbf{v}^{(3)} = (0.125, 0.125, 0.5, 0.125, 0.125)$ and $\mathbf{v}^{(4)} = (0.19, 0.19, 0.24, 0.19, 0.19)$.

Region	income per capita	consumption per capita	investments	economic activity	average wage
	1 000 CZK / cap.	1 000 CZK / cap.	1 000 CZK / cap.	%	CZK
Hl. m. Praha	364.25	51.02	9.91	65.37	45 928
Středočeský	283.84	36.42	13.91	62.17	36 960
Jihočeský	247.29	42.68	14.39	58.97	32 707
Plzeňský	265.02	44.36	13.26	60.76	35 208
Karlovarský	242.84	43.58	11.35	62.33	31 651
Ústecký	234.55	41.45	7.90	57.28	33 188
Liberecký	245.50	40.26	11.14	57.62	34 226
Královéhradecký	263.40	40.59	13.49	59.58	34 343
Pardubický	250.79	39.98	13.23	60.15	32 607
Vysočina	261.03	41.44	15.45	59.54	33 443
Jihomoravský	262.02	39.38	12.54	60.13	35 356
Olomoucký	238.77	41.10	11.87	58.81	32 668
Zlínský	247.03	37.38	10.04	58.40	32 688
Moravskoslezský	239.28	40.90	9.98	59.03	32 845

Table 1 – Data for the comparison [3]

2.2 Dominance and non-dominance

Theoretically, without loss of generality, let us assume that all criteria are maximizing since each minimization function (and thus the minimization criterion) can be transformed to the maximization function, either by multiplying by minus one (or by a similar linear transformation) or by using an inverse value [9].

An alternative a_d is dominated by alternative a_b (where a_b denotes better alternative, so-called dominant) if and only if $y_{dj} \le y_{bj}$ for all criteria f_j, $j = 1, \dots, k$ and there exists any criterion f_q, such that $y_{dq} > y_{bq}$. This dominated alternative is not evaluated better than a_b according to any criterion and the reasonable method does not mark such alternative as the winner. Moreover, a reasonable method places the dominated alternative in the ranking after the dominant one.

In our data file, we can see (Table 1) that Ústecký, Liberecký, Pardubický, Jihomoravský, Zlínský, and Moravskoslezský regions are dominated, for example, all by the Plzeňský region. The Plzeňský region is dominant and each MCDM method should place it in front of these 6 dominated regions.

2.3 Removal of dominated alternative in WSA method

The Weighted Sum Approach method (WSA) belongs to the utility maximization type of methods [9]. For the data, the normalization formula (1) is applied. For each criterion f_j symbol A_j^+ denotes the highest value of this criterion, $A_j^+ = \max_i y_{ij}$, and A_j^- denotes the lowest value of this criterion, $A_j^- = \min_i y_{ij}$. According to the data y_{ij} for each alternative a_i and each criterion f_j the normalized values r_{ij} are calculated [1]:

$$r_{ij} = \frac{y_{ij} - A_j^-}{A_j^+ - A_j^-}. \quad (1)$$

The final ranking is based on the utility (2) and the higher value is the better:

$$u(a_i) = \sum_{j=1}^{k} v_j r_{ij}, \; \forall\, i = 1, \cdots, p. \quad (2)$$

It can be easily shown that the removal of a dominated alternative (denoted e.g. a_d) from the list of alternatives may or may not change the final arrangement of the alternatives by WSA. If alternative a_d is dominated by alternative a_b (where a_b denotes better alternative, so called dominant) then $y_{dj} \le y_{bj}$ for all criteria $j = 1, \dots, k$. As formulas (1) and (2) hold, $r_{dj} \le r_{bj}$ for all $j = 1, \dots, k$ and for non-negative weights v_j $u(a_d) \le u(a_b)$. Therefore, the dominated alternative a_d cannot be the winner (as the utility of a_b is higher).

For any two alternatives a_s and a_t where a_s is better by WSA than a_t holds:

$$u(a_s) \ge u(a_t) \iff \sum_{j=1}^{k} v_j \frac{y_{sj}}{A_j^+ - A_j^-} \ge \sum_{j=1}^{k} v_j \frac{y_{tj}}{A_j^+ - A_j^-}. \quad (3)$$

Note that the value of A_j^+ is independent on the dominated alternative a_d and the values y_{sj}, y_{tj} and the weights v_j are fixed and unchangeable by removing any alternative from the set.

Let us assume two non-dominated alternatives a_s and a_t for which relation (3) holds, i.e. a_s is preferred to a_t because it has higher utility in the whole data set. Their relative rank does not change if relation (3) holds even after removing the dominated alternative. A change of utilities and therefore a change of the final arrangement can only occur by changing the value of A_j^-. Therefore, utilities of other alternatives after the removal of the dominated alternative a_d do not change if the value of A_j^- does not change.

The value of A_j^- depends on the negative-ideal (also called non-ideal or nadir) alternative. If $y_{dj} \ne A_j^-$ for all $j = 1, \dots, k$ then r_{ij} in formula (1) and also $u(a_i)$ in formula (2) remain the same values after the removal of the dominated alternative from the data set. The arrangement of alternatives does not change in such a case.

If $y_{dj} = A_j^-$ for any $j = 1, \dots, k$ then r_{ij} and also $u(a_i)$ can change the value after the removal of the dominated alternative. A new negative-ideal value $A_j^{-(new)} \ge A_j^-$ for all criteria $j = 1, \dots, k$ and so for each alternative a_i:

$$u^{(new)}(a_i) = \sum_{j=1}^{k} v_j \frac{y_{ij}}{A_j^+ - A_j^- - \delta_j} - \sum_{j=1}^{k} v_j \frac{A_j^- + \delta_j}{A_j^+ - A_j^- - \delta_j} \quad (4)$$

where $\delta_j = A_j^- - A_j^{-(new)} \ge 0$ and the second sum is independent of alternative a_i. After data analysis, we can see that $u^{(new)}(a_i) \le u(a_i)$ for each alternative a_i, since $A_j^{-(new)} \ge A_j^-$. But the arrangement of alternatives does not change if and only if formula (5) holds:

$$u^{(new)}(a_s) \ge u^{(new)}(a_t) \iff \sum_{j=1}^{k} v_j \frac{y_{sj}}{A_j^+ - A_j^- - \delta_j} \ge \sum_{j=1}^{k} v_j \frac{y_{tj}}{A_j^+ - A_j^- - \delta_j}. \quad (5)$$

2.4 Removal of dominated alternative in TOPSIS method

TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) method [9] ranks the alternatives according to the maximal relative ratio of their distance to the negative-ideal alternative:

TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) method [9] ranks the alternatives according to the maximal relative ratio of their distance to the negative-ideal alternative:

$$c_i = \frac{d_i^-}{d_i^+ + d_i^-}, \quad \forall\, i = 1, \dots, p, \quad (6)$$

where $d_i^+ = \sqrt{\sum_{j=1}^{k} (w_{ij} - A_j^+)^2}$ and $d_i^- = \sqrt{\sum_{j=1}^{k} (w_{ij} - A_j^-)^2}$, $\forall\, i = 1, \dots, p$. Symbol $w_{ij} = v_j \cdot r_{ij}$ denotes an element of weighted decision matrix $\mathbf{W}$, where

$$r_{ij} = \frac{y_{ij}}{\sqrt{\sum_{i=1}^{p} y_{ij}^2}}, \quad \forall\, i = 1, \dots, p, \quad j = 1, \dots, k, \quad (7)$$

is the normalized value of criterion f_j for alternative a_i. Symbols A_j^+ and A_j^- denote ideal and negative-ideal values of the matrix **W** and they are calculated as $A_j^+ = \max_i w_{ij}$ and $A_j^- = \min_i w_{ij}$, $\forall\ j = 1, \ldots, k$.

In contrast to WSA, in the TOPSIS method, the removal of any alternative (a_i) from the file affects all normalized values (unless the criterion value of omitted alternative is zero), regardless of whether the removed alternative is dominated or not. Each alternative affects the sum in the denominator in relation (7). From this formula we can see that the removal of an alternative will decrease (more accurately non-increase) the value of $\sum_{i=1}^{p} y_{ij}^2$ because the number of additions will decrease. Therefore, all normalized values in (7) will non-decrease: $r_{ij}^{(new)} \geq r_{ij}$.

The removal of dominated alternatives will not affect the location of ideal values in the weighted normalized matrix **W** (although for the presented reason, it will affect the values themselves). The increase in normalized values will necessarily cause an increase in the ideal values in the weighted criterion matrix. Because dominated alternatives often affect negative-ideal values, the removal of these alternatives affects not only negative-ideal values but also their location in the weighted criterion matrix. The increase in normalized values will also necessarily cause an increase in the negative-ideal values in the weighted criterion matrix **W**. Removal of the dominated alternatives will also cause deviations from the ideal values to increase. If removing these alternatives does not change the negative-ideal values, the deviations from the negative-ideal values will also increase. However, these may decrease if the negative-ideal alternative is changed. In the final result, after removing the dominated alternatives from the set, the relative ratio of the distance to the negative-ideal alternative may decrease or increase, and therefore the final arrangement may or may not change. Thus, removing the alternatives may significantly affect the winning compromise alternative.

3 Results of an experiment

In our analysis, we show that the removal of dominated alternatives from a file or adding dominated alternatives to a data set can affect their final arrangement (see Table 2). However, it depends on the method used.

3.1 WSA

After WSA analysis with the first weight vector $\mathbf{v}^{(1)} = (0.2, 0.2, 0.2, 0.2, 0.2)$, which has equal weights of all criteria, all six dominated regions are placed in the last positions (places 9 to 14) – see Table 2. However, it is not a rule that the method places all dominated alternatives at the end. In our data set for the second weight vector $\mathbf{v}^{(2)} = (0.20, 0.17, 0.24, 0.16, 0.23)$, which does not differ from the original by more than 4 % in each criterion, the dominated Pardubický region is ranked better (8th) than the non-dominated Karlovarský region (9th), since formula (3) holds.

WSA	weight vector $\mathbf{v}^{(1)} = (0.2, 0.2, 0.2, 0.2, 0.2)$						$\mathbf{v}^{(3)} = (0.125, 0.125, 0.5, 0.125, 0.125)$					
	full data set			after dominated removal			full data set			after domin. removal		
	utility	rank (all)	rank (ND)	utility	rank change	utility change	utility	rank (all)	rank (ND)	utility	rank change	utility change
Hl. m. Praha	0.8534	1	1	0.8000	1	0.0534	0.6334	1	1	0.5000	3	0.1334
Středočeský	0.4306	3	3	0.3864	2	0.0442	0.5678	3	3	0.5124	2	0.0554
Jihočeský	0.3339	6	6	0.2696	6	0.0643	0.5311	5	5	0.4717	4	0.0594
Plzeňský	0.4335	2	2	0.3718	3	0.0617	0.5371	4	4	0.4590	5	0.0781
Karlovarský	0.3271	8	8	0.2549	8	0.0722	0.3757	9	8	0.2565	8	0.1192
Ústecký	0.0904	14	D	D	D	D	0.0565	14	D	D	D	D
Liberecký	0.1999	11	D	D	D	D	0.2862	11	D	D	D	D
Královéhradecký	0.3443	5	5	0.2770	5	0.0673	0.4929	6	6	0.4155	6	0.0775
Pardubický	0.2993	9	D	D	D	D	0.4517	7	D	D	D	D
Vysočina	0.3906	4	4	0.3416	4	0.0490	0.6191	2	2	0.5885	1	0.0306
Jihomoravský	0.3283	7	7	0.2553	7	0.0730	0.4359	8	7	0.3377	7	0.0981
Olomoucký	0.2278	10	D	D	D	D	0.3394	10	D	D	D	D
Zlínský	0.1315	13	D	D	D	D	0.1888	13	D	D	D	D
Moravskoslezský	0.1841	12	D	D	D	D	0.2186	12	D	D	D	D

Table 2 Results of WSA for dominated (D) and non-dominated (ND) alternatives with weights $\mathbf{v}^{(1)}$ and $\mathbf{v}^{(3)}$

From the dataset (in Table 1), we see that the ideal values are given by the non-dominated alternatives, as well as negative-ideal values for consumption (36.42 for the Středočeský region) and average wage (31 651 for the Karlovarský region). On the contrary, the negative-ideal value of income per capita (234.55) was given by the dominated Ústecký region, as well as for investments (7.90) and economic activity (57.28) – see Table 1. If the Ústecký region remains in the set while the other dominated alternatives are omitted, the utilities, the winner and the final ranking of the non-dominated alternatives will not change. If the Ústecký region is omitted together with the other dominated alternatives, the negative-ideal values of these three criteria increase to 242.84, 9.91, and 58.97. These changes will then cause a decrease in the overall utility of all non-dominated alternatives after the removal of dominated alternatives. We can see that after removing the dominated alternatives (including the Ústecký region) for weight vector $\mathbf{v}(^1)$ all utilities will decrease (Table 2), but the winner will remain the same (Hl. m. Praha). However, the rankings for 2nd and 3rd place will change, since formula (3) holds but formula (5) does not. If weight vector $\mathbf{v}(^2)$ is used, the utilities will again decrease, the winner will also remain the same, but the final ranking of the non-dominated alternatives will not change. It is therefore obvious that the final arrangement may or may not change.

From the results presented, it may appear that removing the dominated alternatives does not affect the winner. However, the derivation in section 2.3 shows that this is not true. Assume now a weight vector $\mathbf{v}^{(3)} = (0.125, 0.125, 0.5, 0.125, 0.125)$. After removing all dominated alternatives, of course, the utilities will decrease and the arrangement of the first five alternatives will change (hl. m. Praha will even end up third after the Vysočina and the Středočeský regions) – see Table 2. Changing the negative-ideal alternative in combination with the weight vector causes relation (5) does not hold and the arrangement changes.

3.2 TOPSIS

The results of the TOPSIS method show, among other things, that the dominated alternatives do not have to be placed in the last places in the final order, as for the weight vector $\mathbf{v}^{(1)}$ the Pardubický region (dominated) is in the 7th place and the non-dominated Karlovarský region is on the 10th place. The ideal values are stated by the non-dominated Hl. m. Praha region (for all criteria except investments) and the Vysočina region (for investments). The negative-ideal values are given by the Ústecký region (for income per capita, investments and economic activity), the Středočeský region (for consumption per capita) and the Karlovarský region (for an average wage). Therefore, in the weighted criterion matrix $\mathbf{W}$ after the removal of dominated regions, the ideal values will be stated by the Hl. m. Praha region and the Vysočina region. The negative-ideal values will be given by the Karlovarský region (income per capita and average wage), the Středočeský region (consumption per capita), the Hl. m. Praha region (investments) and the Jihočeský region (economic activity). We see a clear increase in all ideal values of weighted criterion matrix $\mathbf{W}$ and we also see that in all cases the distance from the ideal alternative has increased. In the case of the negative-ideal alternative, some distances increased (e.g. the Hl. m. Praha region) and some decreased (e.g. the Středočeský or the Plzeňský region). In all cases, the relative ratios (6) have decreased, but the overall order does not change. Note that the dominated Pardubický region was placed by TOPSIS at 7th place and the dominated Olomoucký region was the 9th, that is why we see the movement in order of the non-dominated Karlovarský region from the 10th to 8th position, but it is because of the removal of the dominated regions. However, the relative ranking of the non-dominated regions with weights $\mathbf{v}^{(1)}$ does not change after removing the dominated alternatives. But if we change the weights vector to $\mathbf{v}^{(2)}$, the order of the 4th and 5th places will change – see Table 3 (as well as the order of the 7th and 8th places for the $\mathbf{v}^{(3)}$ vector) and removing the dominated alternatives will affect the final order, even though the winner will be the same.

TOPSIS	weight vector $\mathbf{v}^{(2)} = (0.2, 0.17, 0.24, 0.16, 0.23)$						weight vector $\mathbf{v}^{(4)} = (0.19, 0.19, 0.24, 0.19, 0.19)$					
	full data set			after dominated removal			full data set			after dominated removal		
	distance	rank (all)	rank (ND)	distance	rank change	distance change	distance	rank (all)	rank (ND)	distance	rank change	distance change
Hl. m. Praha	0.5886	1	1	0.5814	1	0.0072	0.5742	2	2	0.5653	1	0.0090
Středočeský	0.5439	3	3	0.4521	3	0.0918	0.5466	3	3	0.4507	3	0.0959
Jihočeský	0.4973	5	5	0.4036	4	0.0938	0.5202	4	4	0.4271	4	0.0931
Plzeňský	0.4985	4	4	0.4013	5	0.0973	0.5182	5	5	0.4214	5	0.0968
Karlovarský	0.3228	10	8	0.2090	8	0.1139	0.3445	10	8	0.2338	8	0.1107
Ústecký	0.1017	14	D	D	D	D	0.1110	14	D	D	D	D
Liberecký	0.3077	11	D	D	D	D	0.3165	11	D	D	D	D
Královéhradecký	0.4816	6	6	0.3751	6	0.1065	0.4975	6	6	0.3896	6	0.1079
Pardubický	0.4345	7	D	D	D	D	0.4529	7	D	D	D	D
Vysočina	0.5553	2	2	0.4723	2	0.0831	0.5762	1	1	0.4937	2	0.0825
Jihomoravský	0.4312	8	7	0.3148	7	0.1164	0.4405	8	7	0.3195	7	0.1210
Olomoucký	0.3438	9	D	D	D	D	0.3595	9	D	D	D	D
Zlínský	0.2018	13	D	D	D	D	0.2075	13	D	D	D	D
Moravskoslezský	0.2095	12	D	D	D	D	0.2203	12	D	D	D	D

Table 3 Results of TOPSIS for dominated (D) and non-dominated (ND) alternatives with weights $\mathbf{v}^{(2)}$ and $\mathbf{v}^{(4)}$

In some cases, however, removing the dominated alternatives can affect not only the ranking of the alternatives but also the winner itself. Assume a weight vector $\mathbf{v}^{(4)} = (0.19, 0.19, 0.24, 0.19, 0.19)$, which again does not differ from $\mathbf{v}^{(1)}$ by more than 4% in each criterion. When it is used, the winner is the Vysočina region ($c_i = 0.5762$), while after the removal of the dominated alternatives, the winner is Hl. m. Praha region ($c_i = 0.5653$), which was originally second ($c_i = 0.5742$). This is evidence that also in the TOPSIS method the removal of dominated alternatives from the set can significantly affect not only the arrangement but also the winning compromise alternative.

4 Conclusions

In this paper, we have shown that the inclusion of dominated alternatives in the data set (as well as their removal) can affect the final arrangement of non-dominated alternatives. This fact is fairly well known in analyses to find the final arrangement dominated alternatives are not commonly removed from the data set.

Often, however, in the search for only the winning alternative, dominated alternatives are removed from the set since they cannot win. As we have shown on real data, this removal can be a gross error as it can affect not only the final arrangement but also the winning alternative. The results would be much more pronounced when using artificially created data for this issue.

We have reached the same results when analyzing with the PROMETHEE method, and the methods of the ELECTRE class also give analogous results. Methods based on pairwise comparison of alternatives are relatively sensitive to this fact. Also, De Keyser and Peters [4] pointed out the rank reversal occurrences in the results of the PROMETHEE I method and Mareschal et al. [12] also mentioned the rank reversal problem in PROMETHEE II.

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] Alinezhad, A. & Khalili, J. (2019). *New Methods and Applications in Multiple Attribute Decision Making (MADM)*. Springer Nature Switzerland AG.

[2] Brans JP. & Mareschal B. (2005). Promethee Methods. In: Multiple Criteria Decision Analysis: State of the Art Surveys. *International Series in Operations Research & Management Science*, vol. 78. Springer, New York, NY. https://doi.org/10.1007/0-387-23081-5_5

[3] Czech Statistical Office, (2020) [online], available at https://www.czso.cz/csu/czso/ [Accessed 17 Jan 2020]

[4] De Keyser, W. & Peeters. P. (1996). A note on the use of PROMETHEE multicriteria methods. *European Journal of Operational Research*, 89(3), 457-461, 1996.

[5] Figueira, J., Greco, S. & Ehrgott M. (2005). *Multiple Criteria Decision Analysis - State of the Art Surveys*. New York : Springer Science + Business Media Inc.

[6] Huber, J., Payne, J.W. & Puto, Ch. (1982). Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis. *Journal of Consumer Research*. 9(1), 90–98. https://doi.org/10.1086/208899

[7] Ishizaka, A. & Nemery, P. (2013). *Multi-Criteria Decision Analysis*. UK: John Wiley & Sons, Ltd.

[8] Kuncova, M. & Seknickova, J. (2018). Multicriteria Evaluation of the Czech Regions from the Selected Economic Activity Aspects Point of View. *Proceedings of the International Scientific Conference Quantitative Methods in Economics (Multiple Criteria Decision Making XIX)* 2018 (pp. 193-200), Tren.Teplice.

[9] Kuncova, M. & Seknickova, J. (2020). Influence of the Different Transformation of the Minimization Criteria on the Result – the Case of WSA, TOPSIS and ARAS Methods. *International Conference on Mathematical Methods in Economics 2020 (MME 2020)* 2020 (pp.332-338), Brno.

[10] Kuncova, M. & Seknickova, J. (2020). Visual PROMETHEE Usage for the Multicriteria Evaluation of the Czech Regions from the Selected Economic Activity Aspects Point of View. *Quantitative Methods in Economics (Multiple Criteria Decision Making XXI)*, 2020 (pp. 182-189), Púchov.

[11] Mardani, A., Jusoh, A., Nor, K.MD., Khalifah, Z., Zakwan, N. & Valipour, A. (2015). Multiple criteria decision-making techniques and their applications – a review of the literature from 2000 to 2014. *Economic Research-Ekonomska Istraživanja*, 28(1), 516-571

[12] Mareschal, B., De Smet, Y. & Nemery de Bellevaux. P. (2008). Rank Reversal in the PROMETHEE II Method: Some New Results. *Proceedings of the IEEE 2008 International Conference on Industrial Engineering and Engineering Management* (pp.959-963), Singapore.

[13] Wang, X. (2007). *Study of ranking irregularities when evaluating alternatives by using some ELECTRE methods and a proposed new MCDM method based on regret and rejoicing.* LSU Master's Theses. 860. https://digitalcommons.lsu.edu/gradschool_theses/860

[14] Xu, Z. & Xia, M. (2012). Identifying and eliminating dominated alternatives in multi-attribute decision making with intuitionistic fuzzy information. *Applied Soft Computing*, 2(2012), 1451-1456.

Application of a System Dynamics Model of Recovery

Anna Selivanova[1] [2]

Abstract. This work focuses on the application of a recovery model to inhabited areas around a selected nuclear power plant (NPP). The mathematical model includes dosimetry calculations and allows to perform economic analyses for cases of decontamination of urban/agricultural objects after deposition of radionuclides. Costs of items required for the decontamination process were set up in accordance with actual conditions in the Czech Republic. The model was designed using the System Dynamics approach and consists of dynamic sequences of selected countermeasures. In order to estimate possible deposition patterns, transport of an atmospheric radionuclide release was simulated using the JRODOS tool. Within the modelling, historical meteodata provided by the Czech Hydrometeorological Institute were used. The release composition and its duration (source term) were selected according to the JRODOS library. Results of simulations in JRODOS were used as input data (affected areas and corresponding surface activities) for the System Dynamics model of recovery. Applying the decontamination model to the NPP of interest, proposed scenarios were simulated in the Vensim software and mutually compared. Based on the simulation results, the cost-benefit analysis of each scenario was carried out.

Keywords: System Dynamics, simulation, Nuclear Power Plant, recovery, decontamination, countermeasure, radiation, atmospheric release

JEL Classification: C63, Q51
AMS Classification: 90B99

1 Introduction

Since the Chernobyl disaster in 1986 and the Fukushima accident in 2011, a wide range of various decontamination methods was developed, having been summarized e.g. in handbooks [8], [16]. Based on practical knowledge, possible countermeasure strategies could be proposed using specific software solutions, e.g. ERMIN [9]. Nevertheless, such computer programs are mainly dedicated to dosimetry issues, while inbuilt economical calculations provide very basic evaluations only. Although rough economical estimations are proper in case of a shortage of time, very complex tasks require detailed analyses within the preparation phase. According to ICRP [6], costs of protective options (i.e. recovery strategies) can be compared with costs of radiological health detriment. Hence, in order to perform in-depth economical assessments, physical calculations are necessary. Therefore, the System Dynamics approach seems to be very convenient for the task of recovery after radioactive contamination of inhabited areas. Originally, the recovery model was applied to a very improbable case of the hypothetically contaminated grassed meadow with parking lots around, located in Prague [11]–[13]. The presented results are newly related to the vicinity of a real NPP and are based on detailed analyses, using JRODOS. Owing to available historical meteodata and source terms for a selected reactor type (from the JRODOS library), surroundings of the NPP Dukovany were used to test the recovery model. The probability of the accidents on the NPP is negligible, hence, the NPP was chosen for demonstration purposes only. The results were obtained using a new version of the model. The recent updates were implemented in the dosimetry part, in the costs estimates (e.g. health detriment of workers was added) and in the scenario set-ups (e.g. re-demarcation of decontaminated areas was newly included).

2 Materials and Methods

2.1 System Dynamics Approach

Complex tasks can be presented as difficult systems with a large amount of various parameters and variables with many interrelations between them. Such interconnections very often lead to the non-linear behavior of the overall system with specific patterns. These specific behavior patterns have a source in the typical structure called feedback loops [14]. Considering radioactive contamination, exponential time-dependent behavior can be observed

[1] Faculty of Economics and Management, Czech University of Life Sciences Prague, Kamýcká 129, Prague, 165 00, Czech Republic, selivanova@pef.czu.cz,
[2] National Radiation Protection Institute (SÚRO), Bartoškova 1450/28, Prague, 140 00, Czech Republic.

e.g. in case of radioactive decay, being an example of a negative feedback loop [2]. Hence, the System Dynamics methods can be applied to such tasks. In order to depict relations between variables of the model and to reveal feedback loops, causal loop diagrams (CLD) are used. Nevertheless, this type of diagrams does not allow a direct conversion to the set of differential equations [14]. The structure of the model (as well as the set of the corresponding equations) can be described using stock and flows diagrams (SFD), where definite integrals are presented as stocks ("boxes") and rates of their increase/decrease are shown as flows ("pipes") [14].

2.2 Affected areas

Owing to the necessity of estimations of affected areas and initial contamination in the surroundings of the NPP, simulations in the JRODOS tool were performed [9]. In order to perform simulations, a dataset with weather records in 2011–2013 (wind speed, wind direction, precipitation etc.) provided by the Czech Hydrometeorological Institute (CHMI) was used. In the dataset, the most frequent parameters of the wind speed and the wind direction were investigated. Moreover, time sequences with the highest precipitation rates were preferred due to the expected higher levels of subsequent contamination patterns [7]. Afterwards, the source term for the most severe accident for the reactor type VVER 440/213 was adopted from the JRODOS library (as the most conservative approach). The prognosis duration was set to 48 h. Thereafter, scenarios with the same source term and different meteosequences were simulated, covering the overall emergency planning zone (EPZ). Obtained maps of surface deposition were merged and averaged. Owing to longer half-lives, surface activities of ^{134}Cs and ^{137}Cs only were used as input levels for recovery planning. According to the results, the EPZ sector #7 with the highest activities was selected. In order to test the model, segment #7-10 (one of the most affected areas) was chosen. Owing to prevailing landscape types of the segment, fields and main highways only were anticipated.

2.3 Physics description

The most recent version of the recovery model created in the Vensim software [17] consists of two sub-models dedicated to decontamination of a large grassed area and streets. Contrary to previous versions [11]–[13], the decrease of surface activity contains two exponential components for **both types** of objects [4]. Employed equations were similar for ^{137}Cs and ^{134}Cs. Considering open grassed surfaces, the modified equation was used [4]:

$$A_{grass} = A_{0,grass}\, e^{-\lambda_r t} \left(0.575\, e^{-\lambda_{g1} t} + 0.425\, e^{-\lambda_{g2} t}\right), \tag{1}$$

where $A_{0,grass}$ is the initial activity on the grassed surface, λ_r is the decay constant (30.05 years for ^{137}Cs and 2.06 years for ^{134}Cs), λ_{g1} and λ_{g2} are weathering rates for half-lives of 3.3 years and 21 years respectively. For the recovery process, an additional decontamination rate should be added [1].

In case of streets, the equation was adopted [4]:

$$A_{street} = 0.5\, A_{0,grass}\, e^{-\lambda_r t} \left(0.7\, e^{-\lambda_{s1} t} + 0.3\, e^{-\lambda_{s2} t}\right), \tag{2}$$

where $A_{0,grasss}$ is the initial activity on the reference grassed surface, λ_r is the same decay constant, λ_{s1} and λ_{s2} are weathering rates (for half-lives of 120 days, resp. 3 years).

2.4 Health impacts

On the basis of selection of the EPZ segment, two localities, Tulešice and Vémyslice, were analyzed. In accordance with the Czech Statistical Office (CZSO), a number of all inhabitants was equal 896 persons [5]. For estimations of health impacts, surface activities (in $MBq \cdot m^{-2}$) were converted to dose rates (in $mSv \cdot h^{-1}$). Hence, equations (1) and (2) were integrated over 1 year [6], using conversion factors and other given parameters [12], [13]. Afterwards, obtained effective doses (in mSv) were summed up and then recalculated to collective effective doses (CED), assuming the irradiated group (896 persons). Using a coefficient for accidents of 2.5 mln. CZK/Sv [15], CEDs were expressed as costs of the health detriment (in CZK). In order to estimate benefits of implemented countermeasures, the financial expression of CEDs saved up after decontamination was used [15].

2.5 Countermeasure techniques

For fields, the grass removal and soil stripping were considered, while for main highways (asphalt surfaces), the high-pressure washing only was employed [8], [16]. As well as in the earlier version of the model [12], the demarcation of the overall segment was assumed. Hence, three recovery scenarios were simulated: 1) demarcation only (the reference scenario), 2) demarcation -> grass removal -> high-pressure washing and 3) demarcation -> grass removal -> soil stripping -> high-pressure washing. All scenarios contained decontamination of workers and vehicles and the waste handling process [11]. Due to testing purposes, the best values of decontamination parameters of the model were newly set. Total costs consisted of labor costs, costs of water and fuel consumption, costs of personal protective equipment, auxiliary tools/materials and consumption of large agricultural machinery (fixed capital) [10]. Newly, costs of the health detriment of workers were added to the total costs.

3 Results and Discussion

3.1 Affected areas

In accordance with the wind rose graph (Figure 1, a), the most frequent origin of winds corresponded to the north-west. Obtained results agreed in general with the weather in the Dukovany location for the longer period of 9 years [3]. Therefore, considering prevailing winds, opposite sectors of the EPZ located to the southeast were expected to demonstrate higher fallout patterns after atmospheric releases of radionuclides. Simulated maps of surface deposition demonstrated contamination in all sectors of the EPZ (Figure 1, b). The most common surface activities of both cesium isotopes were in a range of 0.1–1 MBq·m^{-2}. The highest contamination levels were most often equal to 1–100 MBq·m^{-2}, being observed mainly in sectors #6–8 (southeast). Hence, the results of simulations were in a good agreement with the wind rose and were used for the subsequent recovery planning. On the grounds of the performed evaluations, the inhabited segment #7-10 (sector #7) with higher levels of surface activities was chosen to test the recovery model. For fields in the segment, surface activities were equal to 16 MBq·m^{-2} (^{134}Cs) and 9 MBq·m^{-2} (^{137}Cs), while in case of highways, the deposition was about 8 MBq·m^{-2} (^{134}Cs) and 5 MBq·m^{-2} (^{137}Cs).

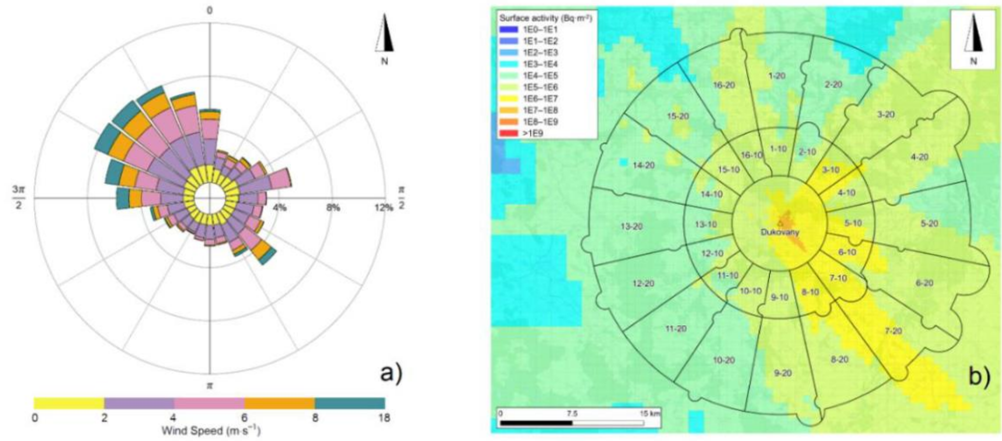


Figure 1 a) Wind rose for Dukovany location, b) Map of surface deposition of ^{137}Cs in EPZ

3.2 Effective doses

The model of recovery created in the Vensim software contains 40 layers with various calculations. Based on the Vensim simulations, non-linear behavior of the dose accumulation was observed in all scenarios (Figure 2), being in a good agreement with the analytical equations (1) and (2). The corresponding layer with the SFD for doses from ^{137}Cs is shown in Figure 3. The layer contains at least six feedback loops and presents physical processes of radioactive decay, weathering and decontamination itself. Afterwards, all total doses are summarized in Table 1.

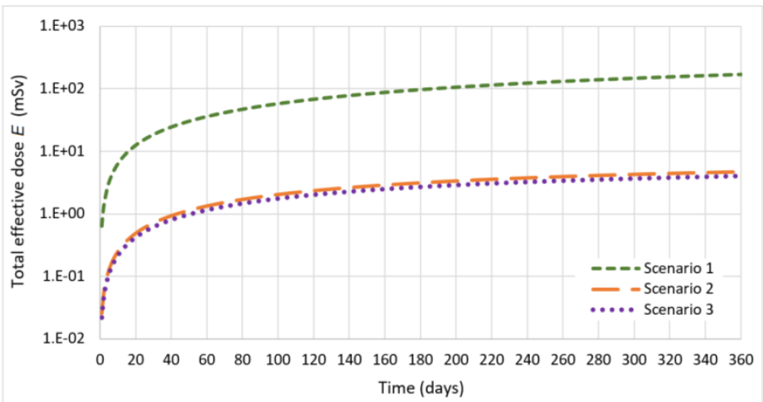


Figure 2 Effective dose *E* after implementation of recovery scenarios proposed

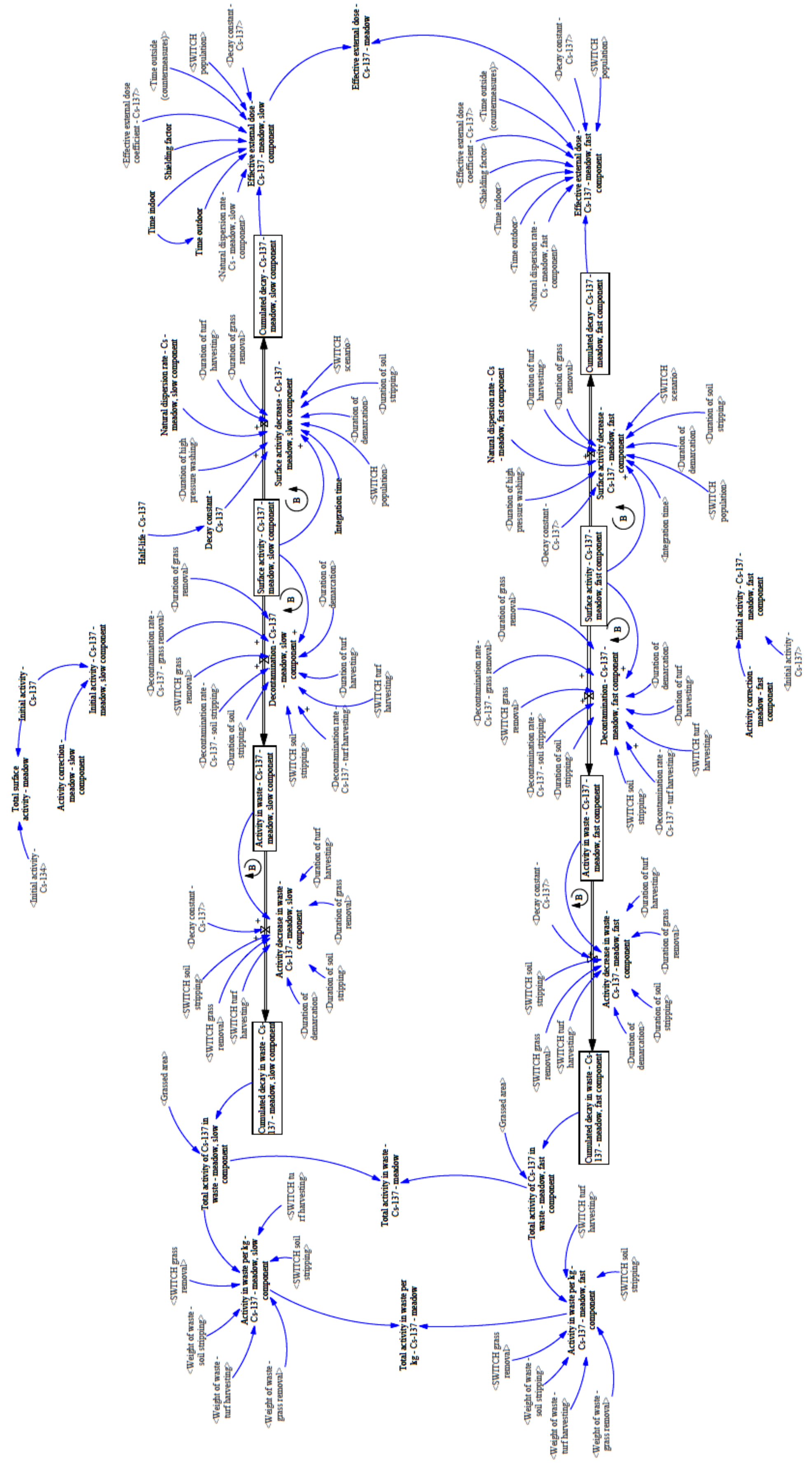


Figure 3 Stock and Flow Diagram for assessments of dose obtained from ^{137}Cs on grassed surfaces

Object	Scenario 1 – *E* (mSv)	Scenario 2 – *E* (mSv)	Scenario 3 – *E* (mSv)
Fields	140	0.1	0.0
Highways	31	4.6	4.1
Overall site	171	4.8	4.1

Table 1 Annual effective doses E from affected objects after implementation of recovery scenarios proposed

Scenario 1

According to Table 1, for the reference scenario (no decontamination), the total annual dose from fields and highways was equal to 140 mSv and 31 mSv, or 171 mSv for the overall site. Obtained results were significantly higher than the level of 20 mSv for the accident exposure of inhabitants and the limit of 1 mSv per year [15], requiring recovery strategies.

Scenario 2

In case of the grass removal, the effective dose from fields was approximately 0.1 mSv (Table 1), being substantially below the annual limit of 1 mSv for population [15]. However, such an enormous decrease was caused by the best decontamination factor achievable several days after deposition only [8]. For highways, the total dose from roads after the high-pressure washing was roughly 5 mSv contrary to 31 mSv in Scenario 1 (Table 1). Hence, the technique seemed to be also suitable for recovery of the site with parameters anticipated.

Scenario 3

Considering the soil stripping of field areas, the total dose was almost zero (Table 1). Nonetheless, this countermeasure followed the grass removal, where the dose < 20 mSv had been already obtained. In accordance with the principle of ALARA – "As Low As Reasonably Achievable" [6], the technically and time demanding soil stripping can be therefore substituted with other methods. In case of highways, the annual dose after high-pressure washing was 4 mSv. Thereafter, such decontamination method was very efficient, as well as in Scenario 2.

3.3 Cost-Benefit Analysis

Based on Table 2, benefits for Scenarios 2 and 3 were almost the same (up to 375 mln. CZK). Such similar values were caused by the grass cutting, where the most of the deposited radioactivity was removed. The most expensive scenario (261 mln. CZK) was Scenario 3 (soil stripping). Scenario 2 (grass removal) costed 75 mln. CZK only, while the cost of Scenario 1 (no decontamination) was 16 mln. CZK. Hence, benefits were higher than the costs both for Scenario 2 and 3. Nevertheless, other findings should be also reflected during the decision-making process. According to the ALARA principle, the soil stripping scenario was superfluous and should be rejected under conditions considered because of the highest costs and final effective doses comparable with Scenario 2. Afterwards, Scenario 2 could be implemented, being the most suitable sequence of countermeasures selected.

Scenario	Benefits (mln. CZK)	Costs (mln. CZK)
1	–	16
2	373	75
3	375	261

Table 2 Costs and benefits for recovery scenarios proposed

3.4 Cost comparison

Assuming costs per unit of contaminated area, costs of the grass removal and the soil stripping were roughly 4 CZK·m^{-2} and 22 CZK·m^{-2}, while the high-pressure washing cost was 34 CZK·m^{-2}. Considering the Fukushima clean-up and e.g. the soil stripping for grassed areas > 1000 m^2 [16], corresponding costs of recovery were 290–710 Yen·m^{-2}, or 69–170 CZK·m^{-2}, being several times higher than the simulated costs. In comparison with the previous research [12], the discrepancy between simulated and real data was newly observed. However, in the most recent model, the best parameters were utilized, especially for the waste production, instead of the very conservative approach employed earlier. According to [8], [12], the waste handling can significantly affect the total costs of recovery. Therefore, considering worse values of the waste production, higher simulated recovery costs for fields could be expected. For high-pressure washing in Fukushima [16], costs were 150–1150 JPY·m^{-2}, or 34–276 CZK·m^{-2}. Hence, for this technique, the simulated cost was in the same interval and corresponded to real costs.

4 Conclusion

Using the JRODOS tool, areas affected after the atmospheric accident release of radionuclides were estimated. Thereafter, the System Dynamics model of recovery was applied to the selected region. The model contained three recovery scenarios simulated in the Vensim software. Based on the results of the cost-benefit analyses, the most suitable scenario was selected for the site and conditions anticipated. Moreover, obtained costs of chosen countermeasures were comparable with the Fukushima clean-up costs.

Acknowledgements

The research was supported by the project 2021B0003 "System approach to cost-benefit analysis of recovery of habitation and adjacent areas after a nuclear or radiation accident" of The Internal Grant Agency of Faculty of Economics and Management CZU Prague (IGA FEM).

References

[1] Ahn, J., Carson, C., Jensen, M. & Juraku, K. (2014). *Reflections on the Fukushima Daiichi Nuclear Accident: toward social-scientific literacy and engineering resilience*. Springer Berlin Heidelberg.

[2] Albin, S. (2001). *Generic Structures: First-Order Linear Positive Feedback*. static.clexchange.org/ftp/documents/roadmaps/RM4/D-4475-2.pdf. [1 March 2022].

[3] Amec Foster Wheeler s.r.o. (2017). *New nuclear source at the Dukovany site* (Environmental Report). CENIA - Czech Environmental Information Agency.

[4] Andersson, K. G & Roed, J. (2006). Estimation of doses received in a dry-contaminated residential area in the Bryansk region, Russia, since the Chernobyl accident. *Journal of Environmental Radioactivity*, 85(2–3), 228–240.

[5] CZSO. (2020). *Věkové složení a pohyb obyvatelstva v Jihomoravském kraji – 2020* [*Age structure and population movement in the South Moravian Region – 2020*] https://www.czso.cz/csu/czso/vekove-slozeni- a-pohyb-obyvatelstva-v-jihomoravskem-kraji. [3 March 2022].

[6] ICRP. (2006). *ICRP Publication 101. Assessing Dose of the Representative Person for the Purpose of Radiation Protection of the Public and The Optimisation of Radiological Protection: Broadening the Process* (Annals of the ICRP). Elsevier Ltd.

[7] Mattsson, S., Vesanen, R. (1988). Patterns of chernobyl fallout in relation to local weather conditions. *Environment International*, 14(2), 177–180.

[8] Nisbet, A., Brown, J., Cabianca, T., Jones, A. L., Andersson, K. G., Hänninen, R., Ikäheimonen, T., Kirchner, G., Bertsch, V. & Hiete, M. (2010). *Generic handbook for assisting in the management of contaminated inhabited areas in Europe following a radiological emergency* (Version 2).

[9] RODOS team. (2019). *JRODOS: Realtime Online DecisiOn Support system*.

[10] Roed, J., Andersson, K. G., Barkovsky, A. N., Fogh, C. L. & Mishine, A. S. (1998). *Mechanical decontamination tests in areas affected by the Chernobyl accident*. Risø National Laboratory.

[11] Selivanova, A. (2020a). Creation of a Model of Waste Handling Within Recovery After a Nuclear Accident Using the System Dynamics Approach. *ICRP International Conference on Recovery After Nuclear Accidents. Radiological Protection Lessons from Fukushima and Beyond*, 1-4.12.2020, on-line.

[12] Selivanova, A. (2020b). Sensitivity Analysis of a System Dynamics Decontamination Model. *38th International Conference on Mathematical Methods in Economics (MME 2020) – Conference Proceedings*. 9-11.9.2020, Brno, Czech Republic, 523–529. Mendel University in Brno.

[13] Selivanova, A. (2021). The System Dynamics approach to creation of a recovery model of an urban object. *39th International Conference on Mathematical Methods in Economics (MME 2021) – Conference Proceedings.* 8-10.9.2021, Prague, Czech Republic, 429–434. Czech University of Life Sciences Prague.

[14] Sterman, J. D. (2000). *Business dynamics: systems thinking and modeling for a complex world*. Irwin/McGraw-Hill.

[15] SÚJB. (2016). *Implementing decree 422/2016 Coll., on Radiation Protection and Security of a Radioactive Source*. State Office for Nuclear Safety.

[16] U.S. EPA. (2016). *Current and Emerging Post-Fukushima Technologies, and Techniques, and Practices for Wide Area Radiological Survey, Remediation, and Waste Management*. Office of Research and Development, Homeland Security Research Center.

[17] Ventana Systems. (2021). Vensim. *Vensim software*.

A Sightseeing Tour That Maximizes Efficiency

Ondřej Sokol[1], Vladimír Holý[2], Jan Pelikán[3]

Abstract. A modification of the traveling salesman problem that finds the optimal route by maximizing the ratio between the time spent at places of interest and the time spent moving (denoted as the time efficiency), under constraints of a set of mandatory places, a minimum number of visited optional places, and a maximum total time, is proposed. The model can be linearized using the Charnes–Cooper transformation, transforming the model to linear integer programming problem.

Keywords: vehicle routing problem with profits, orienteering problem, integer programming

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Our goal is to find the perfect daily sightseeing tour of a given city. As there can be many interesting places, it is necessary to choose only a few that will be visited. Even more, our travelers are quite lazy and do not wish to simply see as much as possible within the time frame of the day but rather to balance the time they spend admiring places of interest and the time it takes them to move between the places.

For this purpose, we follow [9] and propose a modification of the traveling salesman problem that finds the optimal route by maximizing the ratio between the time spent at places of interest and the time spent moving (denoted as the time efficiency), under constraints of a set of mandatory places, a minimum number of optional places visited, and a maximum total time. The objective function defined in this way is non-linear but can be linearized using the Charnes–Cooper transformation. The resulting model is thus an integer linear program.

Our problem falls into the class of vehicle routing problems with profits. A straightforward application is in tourism recommender systems (see e.g. 7). Similar problems are know as the orienteering problem, the selective traveling salesman problem, the maximum collection problem, the bank robber problem, the profitable tour problem, or the prize-collecting traveling salesman problem. For a survey of the related literature, see [10], [1], [6], and [8].

2 Model formulation

In the model, the nodes of the graph represent individual places of interests. The set of n nodes is further divided into two subsets: $\mathcal{M}$ consists of mandatory nodes and $\mathcal{O}$ consists of optional. Parameters of the model are following

- n, number of nodes,
- d_{ij}, travel time between node i and j,
- c_i, time spent in node i,
- W, time limit of the route,
- C, minimal number of nodes to visit,

and variables of the model are

- x_{ij}, binary variables, equals 1 if a tourist travels from node i to j,
- u_i, variables in anti-cyclic constraints.

[1] Prague University of Economics and Business, Faculty of Informatics and Statistics, Department of Econometrics, Prague, Czech Republic, ondrej.sokol@vse.cz.
[2] Prague University of Economics and Business, Faculty of Informatics and Statistics, Department of Econometrics, Prague, Czech Republic , vladimir.holy@vse.cz
[3] Prague University of Economics and Business, Faculty of Informatics and Statistics, Department of Econometrics, Prague, Czech Republic , pelikan@vse.cz

The model is then as follows

$$f(x) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} c_i x_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij} x_{ij}} \to \max \tag{1}$$

subject to

$$\sum_{i=1}^{n} x_{ij} = \sum_{i=1}^{n} x_{ji}, \qquad j = 1,\dots,n, \tag{2}$$

$$u_i + d_{ij} + c_j - W(1 - x_{ij}) \le u_j, \qquad i = 1,\dots,n,\ j = 2,\dots,n,\ i \ne j, \tag{3}$$

$$u_i + d_{i1} \le W, \qquad i = 2,\dots,n, \tag{4}$$

$$\sum_{i=1}^{n}\sum_{j=2}^{n} x_{ij} \ge C, \tag{5}$$

$$\sum_{i=1}^{n} x_{ij} = 1, \qquad j \in \mathcal{M}, \tag{6}$$

$$x_{ij} \in \{0,1\}. \tag{7}$$

The objective function (1) maximizes the ratio of time spent in places of interests to time spent on the move.

The first three sets of constraints are the same as for the standard traveling salesman problem. Constraint (2) ensures that if tourist enters a node, then it also has to leave it. Constraint (3) is the Miller-Tucker-Zemlin anti-cyclical constraint (see 5). Constraint (4) set the time limit of trip, e.g. tourist must return to the starting point by given time W.

Constraint (6) guarantees that the tourist visits mandatory nodes and constraint (5) states the minimal number of visited nodes, e.g. tourist must visit at least Cp laces of interests. Finally, in constraint (7) is declared that variables x_{ij} must be binary.

As the objective function (1) is linear-fractional function, the model is non-linear. However, we can use the Charnes–Cooper transformation (see 3, 4, 2) and transform the model to linear integer programming problem. The linear integer problem is then as follows:

$$f(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i y_{ij} \to \max \tag{8}$$

subject to

$$\sum_{i=1}^{n} y_{ij} = \sum_{i=1}^{n} y_{ji} \qquad j = 1,\dots,n, \tag{9}$$

$$v_i + \frac{d_{ij}}{4} t + c_j t - W(t - y_{ij}) \le v_j \qquad i = 1,\dots,n,\ j = 2,\dots,n,\ i \ne j, \tag{10}$$

$$v_j \le Wt \qquad j = 2,\dots,n, \tag{11}$$

$$\sum_{i=1}^{n} y_{ij} = t \qquad j \in \mathcal{M}, \tag{12}$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij} y_{ij} = 1, \tag{13}$$

$$-M(1 - x_{ij}) \le y_{ij} - t \le M(1 - x_{ij}) \qquad i = 1,\dots,n,\ j = 2,\dots,n,\ i \ne j, \tag{14}$$

$$-Mx_{ij} \le y_{ij} \le Mx_{ij} \qquad i = 1,\dots,n,\ j = 2,\dots,n,\ i \ne j. \tag{15}$$

$$x_{ij} \in \{0,1\} \qquad i = 1,\dots,n,\ j = 1,\dots,n. \tag{16}$$

where

$$y_{ij} = \frac{x_{ij}}{\sum_{\ell=1}^{n}\sum_{k=1}^{n} d_{k\ell} x_{j\ell}}, \tag{17}$$

$$v_j = \frac{u_j}{\sum_{i=1}^{n} d_{ij} x_{ij}}, \tag{18}$$

$$t = \frac{1}{\sum_{j=1}^{n}\sum_{i=1}^{n} d_{ij} x_{ij}} \tag{19}$$

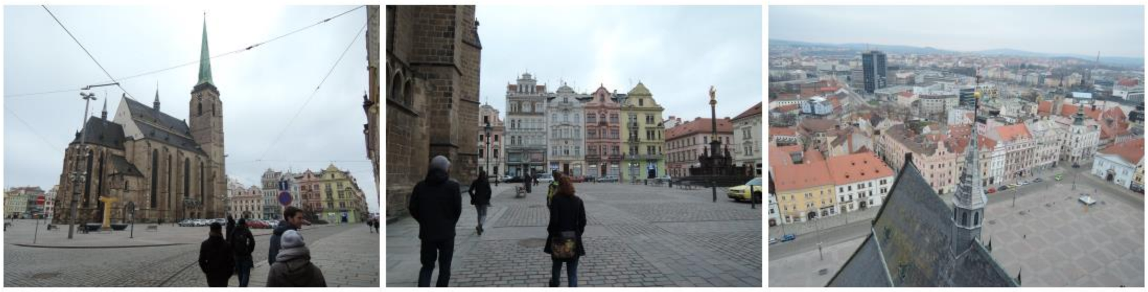


Figure 1 The Square of the Republic. Photographed by Tomáš Koudelka.

Table 1 Places of interest in Pilsen, Czech Republic.

No.	Place of Interest	Duration (in minutes)	Plus Code
1	Pilsen Main Railway Station	-	P9VQ+85
2	Cathedral of St. Bartholomew	45	P9WH+X2
3	Marian Column	15	P9XG+5W
4	Renaissance Town Hall	30	P9XH+63
5	Puppet Museum	45	P9WG+HQ
6	Pilsen Historical Underground	45	P9XJ+G7
7	Brewery Museum in Pilsen	45	P9XJ+FC
8	Museum of West Bohemia	60	P9VH+XM
9	Great Synagogue	30	P9WF+M4
10	J. K. Tyl Theater	30	P9WF+5C
11	Thank You America Memorial	15	P9VF+H3
12	General Patton Memorial Museum	45	P9XF+XC
13	Pilsner Urquell Brewery	90	P9WP+PW
14	Techmania Science Center	180	P9R6+7R
15	Pilsen Zoo	180	Q954+8F
16	DinoPark	120	Q965+23

and M is some large number.

The constraints (14) and (15) are used to enforce the correct relationship between binary variables x_{ij} and y_{ij}. If $y_{ij} = t$, then x_{ij} must be assigned value of 1. Similarly, if $y_{ij} = 0$, then x_{ij} can be 0.

3 Case study

In the illustrative application, the model is used to plan a trip to sights and other interesting places in Pilsen. Pilsen is a historic city in the west of the Czech Republic with many attractions for its visitors, see Figure 1.

For simplicity, we have selected 15 most known sights and places of interests – then the size of this problem is so small that common commercial solvers like Gurobi or CPLEX can find the optimal solution in seconds or minutes. We assume that the tourist arrives (and leaves) the city by train, therefore the trip starts and ends at the train station. The places of interests are listed in Table 1. The goal is to find a tour in which the mandatory places and at least 5 optional places are visited, the time limit of 9 hours is met, and the ratio of time spent at the city's attractions to time spent moving is maximized. In order to transform distance bewtween places to travel time, we assume average walking speed 4 km per hour and thus to get the travel time the distance is divided by 4 (if the time is measured in hours).

Tour 1. In the first case, the mandatory places are Cathedral of St. Bartholomew (no. 2) and Great Synagogue (no. 9) and a following route is found: 1-13-7-6-4-2-3-12-9-10-5-8-1. The overall efficiency of the trip is then 8.4. Thus, 8.4 times more time is spent at places of interests than walking.

Tour 2. In the second case, we change the mandatory visit only to Pilsner Urquell Brewery (no. 13). The optimal route is 1-13-7-6-4-3-2-5-8-1 with the efficiency of 9.16. On this route the cathedral is visited, similar to previous case, but the synagogue is not.

In Figure 2 are shown both optimal tours and the layout of places of interests.

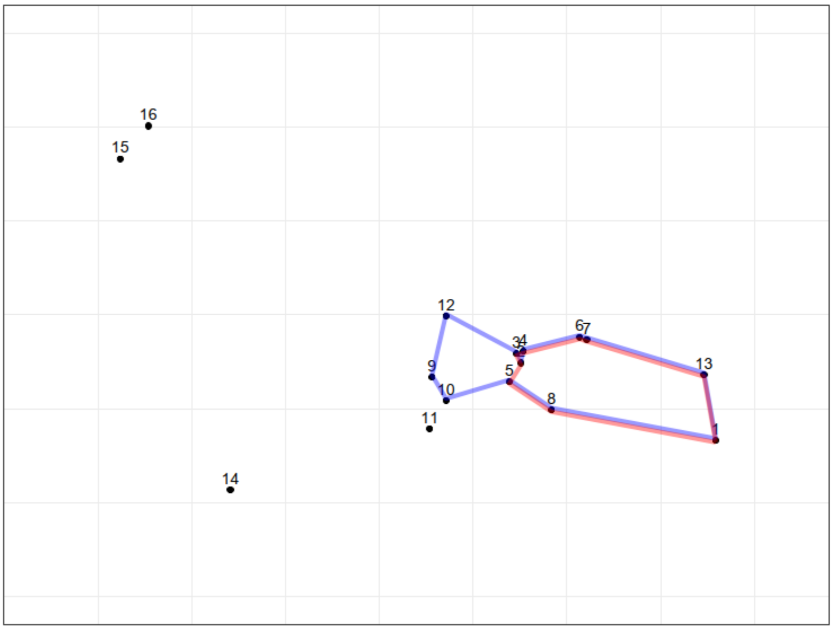


Figure 2 The Graph of Pilsen, Czech Republic, with Tour 1 (blue) and Tour 2 (red).

4 Discussion and future research

Although the problem is linear (albeit integer), compared to the standard traveling salesman model, the process of finding the optimum is, in our experience, significantly more demanding.

The problem is NP-hard (although no proof is presented here). A fast and effective heuristic is therefore needed for larger instances of the problem. Our future research will focus on developing a simple constructive heuristic as well as modifying a suitable meta-heuristic such as ant colony optimization or simulated annealing.

The model could also be extended to be able to find multiple routes, i.e. a plan for several days. Another direction of our future research lies in accommodating ratings of individual places. This can be achieved e.g. in the form of weighted times in the objective function.

Acknowledgements

The work was supported by the Czech Science Foundation project 22-19353S and by the Prague University of Economics and Business project IGA F4/27/2020.

References

[1] Archetti, Claudia, Speranza, M. Grazia, & Vigo, Daniele. 2014. Vehicle Routing Problems with Profits. Chap. 10, pages 273–297 of: Toth, Paolo, & Vigo, Daniele (eds), *Vehicle Routing: Problems, Methods, and Applications*, second edn. Philadelphia: Society for Industrial and Applied Mathematics.

[2] Barros, Ana Isabel. 1998. Discrete and Fractional Programming Techniques for Location Models. First edn. *Combinatorial Optimization*, vol. 3. Boston: Springer.

[3] Charnes, A., & Cooper, W. W. 1962. Programming with Linear Fractional Functionals. *Naval Research Logistics*, 9(3-4), 181–186.

[4] Charnes, A., & Cooper, W. W. 1973. Explicit General Solution in Linear Fractional Programming. *Naval Research Logistics*, 20(3), 449–467.

[5] Desrochers, Martin, & Laporte, Gilbert. 1991. Improvements and Extensions to the Miller-Tucker-Zemlin Subtour Elimination Constraints. *Operations Research Letters*, 10(1), 27–36.
[6] Gavalas, Damianos, Konstantopoulos, Charalampos, Mastakas, Konstantinos, & Pantziou, Grammati. 2014a. A Survey on Algorithmic Approaches for Solving Tourist Trip Design Problems. *Journal of Heuristics*, 20(3), 291–328.
[7] Gavalas, Damianos, Konstantopoulos, Charalampos, Mastakas, Konstantinos, & Pantziou, Grammati. 2014b. Mobile Recommender Systems in Tourism. *Journal of Network and Computer Applications*, 39(1), 319–333.
[8] Gunawan, Aldy, Lau, Hoong Chuin, & Vansteenwegen, Pieter. 2016. Orienteering Problem: A Survey of Recent Variants, Solution Approaches and Applications. *European Journal of Operational Research*, 255(2), 315–332.
[9] Pelikán, Jan, & Jablonský, Josef. 2020. Nonlinear Vehicle Routing Problem. Pages 593–597 of: Jedlička, Pavel, Marešová, Petra, Firlej, Krzysztof, & Soukal, Ivan (eds), *Proceedings of the 18th International Scientific Conference Hradec Economic Days 2020*, vol. 10. Hradec Králové: University of Hradec Králové.
[10] Vansteenwegen, Pieter, Souffriau, Wouter, & Oudheusden, Dirk Van. 2011. The Orienteering Problem: A Survey. *European Journal of Operational Research*, 209(1), 1–10.

Secondary Education Efficiency in Selected Countries

Barbora Staňková[1], Hana Stojanová[2]

Abstract. At present, it is assumed that the skills of employees depend, among other things, on having a completed education. The major part of the population over the age of 15 consists of people with at least a secondary education. This article focuses on evaluating the efficiency of secondary education, which then affects the skills of the majority of employees in selected countries. The data envelopment analysis method was used to evaluate this efficiency. The compiled model is based on three input and two output variables. Inputs include public funding of secondary education, the number of teachers and the average wage. The output variables are represented by the number of young people employed and the number of students in secondary education. The radial model with variable returns to scale has been selected to calculate the efficiency score in 2018 and 2019. The empirical results show that countries such as Austria, Germany, Belgium, Portugal, Norway and Italy generally achieve the lowest efficiency values. In contrast, some economically weaker countries are achieving 100% efficiency. These countries included, for example, the Slovak Republic and Israel.

Keywords: data envelopment analysis, efficiency, linear programming, secondary education

JEL Classification: C44, H52, I21
AMS Classification: 90B50, 90C08

1 Introduction

In the context of the Industry 4.0 technological revolution, education is gaining greater and greater prominence. As a result, the education sector is under pressure to improve quality. Article [22] argues that education brings with it profits in the form of growth in labour productivity and thus in economic growth. In many countries, education is at least partially publicly funded. This significant share of public spending is one of the reasons why it makes sense to address the efficiency of education. In their article, [25] examined variants for financing secondary and higher education. The number of private schools is also gradually increasing in secondary education. Although these schools do not draw as large an amount from the public budget as state schools, this is not a significant expense for the state budget. At this time, it is a significant advantage to have a university degree, but there are still many professions for which a secondary education on its own is sufficient. Graduate unemployment has fallen in many countries in the last few years. This decreasing trend is due not only to the ageing of the population but also due to the fact that younger generations have more and more technical knowledge. As mentioned, [16], employers value graduates with a knowledge of foreign languages and computer skills. In many studies, such as [20], [21] the authors point out that computer skills are a great advantage for graduates who are just entering the labour market and therefore have the disadvantage of not having experience.

During the Industry 4.0 technical revolution, demand for good computer skills is becoming more common. The applicability of graduates is also currently a much-discussed topic, on which work has also been done [17]. One of the methods used after measuring efficiency is the non-parametric data envelopment analysis (DEA) method. This method is very popular for measuring efficiency in many areas. Applications can be found, for example, in construction [12] and travel [14].

In their research, the authors [7] addressed the efficiency of education in selected countries around the world. Their research was as input variables the share of students and teachers as well as expenditure on educational institutions as a percentage of GDP. [9] in his article uses the DEA model to evaluate the efficiency of the entire Tunisian education system. Based on his study, he concluded that the inefficiency of the system is mainly due to the inefficiency of tertiary education. The comparison of secondary and tertiary education clearly resulted in secondary education as the level with significantly higher efficiency. The issue of efficiency in tertiary education is very well mapped, see [10], [24] and [9]. There are significantly fewer studies dealing with the efficiency of secondary

[1] Mendel University in Brno, Department of Statistics and Operation Analysis, Zemědělská 1, 613 00 Brno, Czech Republic, xjanouskov7@mendelu.cz..
[2] Mendel University in Brno, Department of Management, Zemědělská 1, 613 00 Brno, Czech Republic, hana.stojanova@mendelu.cz.

education. [18] in their article examined various measures for secondary schools and their impact on efficiency. The efficiency of secondary education was also addressed in their article [19]. Based on data from more than 20,000 students, they tried to estimate the external efficiency of secondary schools. The result of their study was that efficiency is greatly affected by the school, more than by the education system. The efficiency of secondary education in Slovakia has been investigated by the authors in [6]. They selected 26 schools from a single region as a research sample. They used the DEA method to measure efficiency, and the result was that state schools achieved higher efficiencies than church or private schools. [4] sees successful completion of secondary education as important, because graduating from a well-qualified school is key to entering the labour market or tertiary education. The authors included variables such as PISA test results in the DEA model; this is the ratio between the number of students and teachers and the annual number of hours spent at school.

The main aim of this article is to evaluate the efficiency of secondary education in selected countries in 2018 and 2019. The efficiency of education can be viewed from different perspectives. In this article, we address the efficiency of education in terms of the applicability of graduates in connection with the financing of education.

2 Material and Methods

Five variables were selected for the efficiency analysis. The data were drawn from the OECD database. The average wages of teachers employed in secondary education (in EUR), the number of teachers in secondary education (in thousands) and education spending (in millions of EUR) were chosen as input variables. The output for the model was selected indicators the number of employed young people (in thousands) and the number of secondary education students (in thousands). The basic statistical characteristics for the data are in Table 1. The analysis was performed on a sample of 21 countries for which we had complete information for both the years being analysed. Due to the large range of analyses, only aggregated values for all countries are in Table 1.

Year	Variable	Type	Min	Max	Mean	Stand. Dev.	Median
2018	Education spending	Input	425.8	377905.0	29108.2	78110.9	11000.0
	Number of employed people	Output	96.4	56348.1	6648.0	11626.9	1642.0
	Average wage	Input	19060.1	117069.8	46699.6	20926.0	45156.0
	Number of teachers	Input	13.6	877.1	141.0	199.1	41.8
	Number of students	Output	79.72	54667.0	6908.6	11915.0	1600.9
2019	Education spending	Input	464.7	393470.0	44772.0	85063.0	10448.0
	Number of employed people	Output	813.6	55821.6	7305.0	12286.5	1842.5
	Average wage	Input	19836.6	83080.6	43885.0	15111.3	45072.9
	Number of teachers	Input	21.2	868.8	149.7	203.5	44.3
	Number of students	Output	756.5	55318.5	7596.9	12177.4	1902.5

Table 1 Basic descriptive characteristics of the variables used in individual years. Education spending in EUR, average wage in EUR, number of employed people and number of teachers in secondary education in thousands, number of students in secondary education in thousands.

If we analyse the values from Table 1, we find that the lowest average wage in both years is in the Slovak Republic, the lowest percentage of employed students is in Turkey and for other indicators is based as the countries with the lowest values of Iceland. On the other hand, the country with the highest values for the number of students, employed people, teachers and education spending is the United States. If we look at the education spending indicator as a percentage of GDP, then Belgium would have the largest expenditure, with this indicator being 2.5% of GDP. At the other end of the series would be Greece and Poland with a value of 1.5% of GDP.

The DEA method was chosen to calculate the efficiency. The input orientation of the model was chosen during modelling and variable returns to scale (i.e. the BCC model) were assumed, similar to [10], [12] and [14].

Based on the input matrix (*X*) and output matrix (*Y*) the efficiency for each country *c* can be calculated by solving the model *n* times:

$$\min_{\theta,\lambda} \theta \tag{1}$$

$$\theta x_c - X\lambda \geq 0$$
$$Y\lambda \geq y_c$$
$$e\lambda = 1$$
$$\lambda \geq 0,$$

where e is a row vector with all element's unity and λ is a column vector with all elements non-negative $\lambda = (\lambda_1, \dots, \lambda_n)^T$, θ is a scalar.

Similar to [12] and [13], we used the Malmquist index to calculate the change in efficiency of unit c from period 1 to period 2, more precisely its partial component, the so-called catch-up effect (*CE*):

$$\text{CE} = \frac{\text{Efficiency of } (x_c, y_c)^2 \text{ with respect to period 2 frontier}}{\text{Efficiency of } (x_c, y_c)^1 \text{ with respect to period 1 frontier}} \tag{2}$$

To compile Malmquist index (*MI*), it is necessary to construct a total of four DEA models. *MI* can be defined as the geometric mean of two efficiency ratios (δ), 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

$$MI = \left[\frac{\delta^1((x_c, y_c)^2)}{\delta^1((x_c, y_c)^1)} * \frac{\delta^2((x_c, y_c)^2)}{\delta^2((x_c, y_c)^1)}\right]^{1/2} \tag{3}$$

This index was constructed using the DEA model according to Formula 1. Technical details about the DEA method can be found in [3]. The results are from the DEA Solver Pro program.

3 Results and discussion

Figure 1 below shows the efficiency score of secondary education for selected countries for the years 2018 and 2019. Countries that reach a value of 1 can be considered as efficient. The upper part of Figure 1 shows the efficiency rate in 2018. The average efficiency in secondary education in that year was 81.5%. The median also reaches similar values (81.6%). We can therefore say that half of the countries had efficiency above 81%. In the V4 countries the average efficiency was around 98%. In the European Union overall, the average efficiency in 2018 is lower, around 75%. In that year, Israel, Luxembourg, Slovenia, Poland, the Slovak Republic, Turkey, the United Kingdom, and the United States of America all achieved full efficiency. By contrast, Austria (50.3%) and Germany (50.4%) had the lowest efficiency in secondary education.

The lower part of Figure 1 also shows the efficiency score for 2019. The same countries extended by Finland and Greece, are achieving full efficiency in this year. The average efficiency in the countries being studied increased slightly to 81.7%. In the V4 countries it is 98% and in the EU countries the value is again lower, namely 75%. The lowest efficiency is to be found again in Austria (54.9%) and Germany (50.9%). These two countries and Belgium had efficiency less than 60%.

From the results we can see that in the V4 countries the average efficiency is higher than the average efficiency of all selected countries in both years. The V4 countries apply a non-standard mixed model of secondary education ([5]). These countries have a common post-war history, which was also reflected in education, and these countries in Europe are among those with the highest percentage of people with a secondary education.

A common feature of the years under analysis was, among other things, that economically developed European countries such as Germany and Austria failed in the efficiency evaluation of secondary education and were in the last two places in the overall ranking. Austria and Germany use the differentiated secondary education model (see in [5]) in their education systems. This model is characterized by the division of students in lower secondary education into vocational, technical, and general education schools. Thanks to this early differentiation in these countries, a higher number of teachers per student is needed. In 2019, there were fewer than 26 students per teacher in secondary education in these countries. There are relatively more teachers, which is associated with a higher need for spending on their salaries. By contrast, Israel and the UK have over 73 students per teacher, so the need for public spending is lower and these states are among the efficient countries. The results of the authors in [2] also matched our results. According to their study, the technical efficiency of secondary education varies significantly across the vast majority of EU and OECD countries. They recommend rationalizing public spending on secondary education and redirecting funds to tertiary education.

A similarity with our results can also be found in [22], where Latvia, Lithuania, Romania and the Czech Republic ranked best. Similar results are confirmed by [1]. Similar to our results, [7] also concluded that more economically developed countries such as Germany and Austria are much less efficient than, for example, the V4 countries. In contrast to our research, the authors [15] examined the efficiency of tertiary education. However, their efficiency models confirm that the Czech Republic can be ranked among the most efficient countries in terms of efficiency.

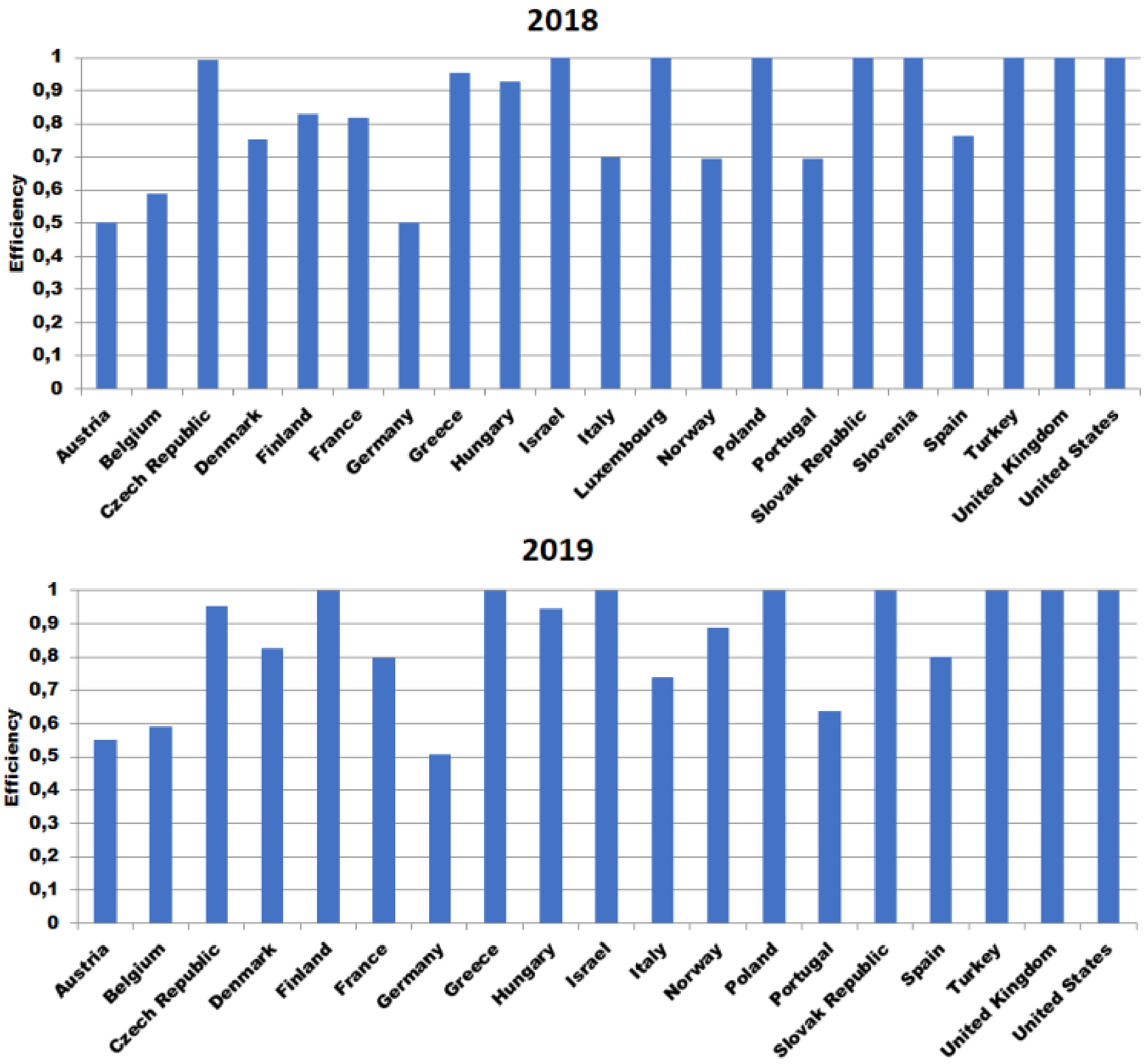


Figure 1 The resulting efficiency score for each country in individual years.

Figure 2 shows the results of the efficiency evaluation for individual countries in 2018–2019. The Catch-up effect tells us how efficiency has changed at the same production limit and under the same technological conditions. On the other hand, the frontier shift tells us how the limits of production possibilities have changed. The Malmquist index combines both parts. If the value of the index is equal to one, then there has been no change in the overall situation in a given country. If the value of the index is higher (or lower) than one, then the situation in the given country has improved (or worsened).

The results of the analysis and decomposition of the Malmquist index can be divided into three categories. The first category includes countries such as Belgium and the USA, where there are no significant changes from 2018 to 2019 both in terms of technical efficiency and production capacity limits, and thus no changes in the value of the Malmquist index. A more interesting trend is represented by the countries in the second group (category), such as France, Germany, Italy, Poland, Spain, Turkey, and the United Kingdom. In these countries, there were no significant changes in the value of their individual efficiency during the period under review, but due to the decline in the frontier, the overall situation worsened, as the Malmquist index itself is lower than one. The amount of education spending in these countries increased significantly when expressed as education spending per student in secondary education. The same situation can be observed in France, Germany, Italy, Mexico, Spain, Turkey and the UK. The last category consists of countries where, on the contrary, efficiency has improved, production opportunities have improved, and the overall situation has improved. These are Austria, Denmark, Finland, Greece, Hungary and Slovakia. Here we also observe a large effect on overall efficiency from the decline in education spending. In these countries, there has been a decrease in education spending per student, leading to lower inputs and improving overall efficiency in secondary education.

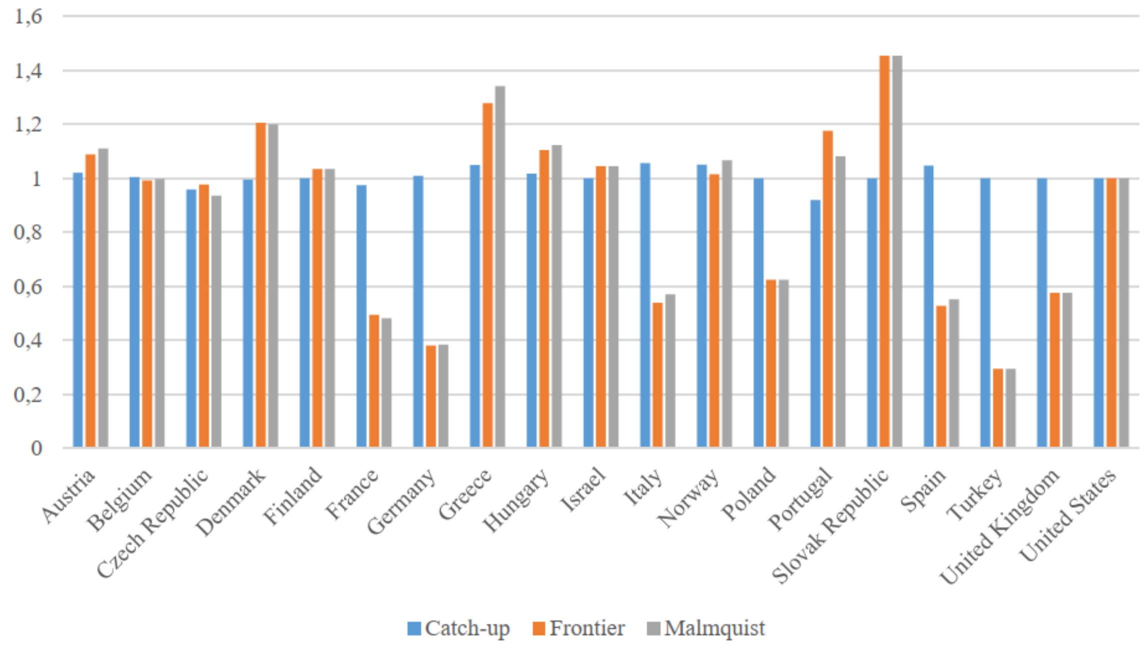


Figure 2 Malmquist index for each country.

4 Conclusion

The results of this article show that even economically developed countries such as Germany or Austria, when compared with other selected countries, do not perform well in terms of the efficiency of secondary education. This is due to the alternatively differentiated model of secondary education. Even the V4 countries, which are influenced by a common history, have better average results than these developed countries. Overall efficiency according to the MI improved only in Austria, Denmark, Greece, Hungary and Slovakia, mainly due to a reduction in education spending. Another group of countries is France, Germany, Italy, Poland, Spain, Turkey and the United Kingdom, which did not do poorly due to the catch-up effect, but the decline in frontier efficiency prevailed, so the MI decreased. Also, an increase in education expenditure could be observed. For other countries, we could not find any significant dramatic changes in efficiency over time.

Acknowledgements

This article was supported by the grant No. IGA-PEF-TP-22-004 of the Grant Agency IGA PEF MENDELU.

References

[1] Ahec Sonje, A., Deskar-Skrbic, M. & Sonje, V. (2018). Efficiency of public expenditure on education: comparing Croatia with other NMS. *International Technology, Education and Development Conference Valencia*, Spain, 12, 1–14.

[2] Aristovnik, A. & Obadic, A. (2014) Measuring relative efficiency of secondary education in selected EU and OECD countries: The case of Slovenia and Croatia. *Technological and economic development of economy*, 20(3), 419–433.

[3] Cooper, W. W., Seiford, L. M. & Tone, K. (2007). *Data envelopment analysis: A comprehensive text with models, applications, references and DEA-solver software*. 2nd edition. New York: Springer Science & Business Media.
[4] Dančáková, D., Glova, J. & Andrejovska, A. (2021). The robust efficiency estimation in lower secondary education: Cross-Country evidence. *Mathematics*, 9(24), 3249
[5] Eurostat, 2018. Strategie EU pro mládež na období 2010–2018. [on-line] Available at: ec.europa.eu/youth/policy/youth-strategy/strategy-2010-2018_ en [Accessed 2020, January 15].
[6] Halaskova, R., Měřičková, B. M. & Halaskova, M. (2022). Efficiency of public and private service delivery: The case of secondary education. *Journal on efficiency and responsibility in education and science*, 15(1), 33–46.
[7] Jelic, O. & Kedzo, M. (2018). Efficiency vs effectiveness: an analysis of tertiary education across Europe. *Public sector economics*, 42(4), 381–414.
[8] Kim, M. H., Lee, I. &Oh, S. (2016). Measuring efficiency of higher education using DEA. *International Journal of u- and e-Service, Science and Technology*, 9(5), 321–328.
[9] López-Torres, L. & Prior, D. (2020). Long-term efficiency of public service provision in a context of budget restrictions. An application to the education sector. *Socio-Economic Planning Sciences*, 100946.
[10] Mašková, K. & Blašková, V. (2021). Efficiency of tertiary education in EU countries. *Mathematical Methods in Economics 2021: Conference Proceedings*. Praha: ČZU, 312–317.
[11] Ramzi, S. (2019). Modeling the education supply chain with network DEA model: The case of Tunisia. *Journal of quantitative economics*, 17(3), 525–540.
[12] Staňková, M. & Hampel, D. (2018). Efficiency Comparison in the Development of Building Projects Sector. *Mathematical Methods in Economics 2018: Conference Proceedings*. MatfyzPress: Praha, 503–508.
[13] 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), doi: 10.5552/crojfe.2022.1541.
[14] 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.
[15] Stefanova, K. & Velichkov, N. (2020). Analysis of the Efficiency of Tertiary Education Expenditure in European Union Member States from Central and Eastern Europe: An Efficiency Frontier Approach. *South-Eastern Europe Journal of Economics* 1, 115–128.
[16] Stojanová, H., Blašková, V., Tomšík, P. & Tesařová, E. (2015). Specification and Characteristic of Generation Y in the Sphere of Work Attitude. *DIEM 2015: Innovation, Leadership & Entrepreneurship*, 565– 579.
[17] Stojanová, H. & Blašková, V. (2014). The significance of the chosen field of study, depending on the difficulty of finding a job. *INTED2014 Proceedings. Valencia*, Spain: IATED, 4002–4012.
[18] Thanassoulis, E., De Witte, K., Johnes, J., Johnes, G., Karagiannis, G. & Portela, C. S. (2016). Applications of data envelopment analysis in education. *Data envelopment analysis: A handbook of empirical studies and applications*, 238, 367–438.
[19] Timmermans, A. C., Rekers-Mombarg, L., T., M. & Vreeburg, B., A., N., M. (2016). External efficiency of schools in pre-vocational secondary education. *Pedagogische studien*, 93(1), 15–37.
[20] Trčka, L. (2014). Vzdělávací procesy – výzkum potřeb a zkušeností zaměstnavatelů absolventů. *Trendy ekonomiky a managementu*, 8 (19), 63–69.
[21] Úlovec, M. (2014). *Potřeby zaměstnavatelů a připravenost absolventů škol – komparační analýza*. Praha: NÚV. 35p
[22] Urbánek, V. (2007). *Financování vysokého školství*. Praha: Nakladatelství Oeconomica
[23] Villano, R. A. & Tran, C. D. T. T. (2021). Survey on technical efficiency in higher education: A metafractional regression analysis. *Pacific economic review*, 26(1), 110–135.
[24] Yotova, L. & Stefanova, K. (2017). Efficiency of Tertiary Education Expenditure in CEE Countries: Data Envelopment Analysis. *Economic Alternatives*, 3, 352–364.
[25] Zámková, M. & Blašková, V. (2013). Comparing the views on tuition fee introduction of Brno university students. *Efficiency and Responsibility in Education 2013*. Praha: CULS, 671–679.

Modelling the Process of Selecting Satisfactory Methods for Controlling System Tasks

Petr Suchánek[1], Robert Bucki[2]

Abstract. The paper highlights the problem of the selection of the most appropriate methods for process control in the logistics system. Representative methods most often implemented to improve the functioning of such systems were selected for the analysis. It was assumed that for input data selected randomly, manually or taken from a ready file, it is possible to carry out a simulation process whose task is to find the best method to improve the previously programmed process flow. Due to the fact that each method optimizing the process has both advantages and disadvantages, they had to be taken into account when looking for a satisfactory solution. For the problem formulated in this way, the mathematical model containing the necessary specification assumptions was presented, and then a system design equipped with a pseudocode enabling the creation of the adequate process simulator was presented. The simulation process was carried out for real input data. The resulting data was thoroughly analyzed and adequate conclusions were drawn.

Keywords: process control, mathematical model, simulation, optimization, satisfactory solution

JEL Classification: C61, C63
AMS Classification: 00A72, 68U01, 90B06

1 Introduction

Optimization actions are all actions that are taken to reduce the number of deficiencies or complaints, increase production efficiency, increase sales volume, reduce operating costs, etc. [3]. Regardless of the assumptions on which they are based, all process management concepts indicate the continuous improvement of processes as a necessity. They unanimously indicate that the lack of improvement activities can lead to disturbances in the management system. The continuous improvement of the processes is an advanced way of applying the principles of quality management which takes into account the creation of an appropriate organizational culture that promotes active and constant search for improvement opportunities [8]. There are many concepts, philosophies and programs for improving business processes. The basic starting point for business process optimization is process mining and the creation of process models [5]. A number of methods can be used to optimize business processes, e.g. the Five Diamond Method [4], Context-Aware BPM Method Assessment and Selection (CAMAS) [2], methods based on big data [6], a self-organizing evolutionary method [1] and many others. A lot of attention in companies is paid to the optimization of logistics processes as it is considered that the continued improvement of the selected areas in company activities leads to an increase in competitiveness on the contemporary global business market [7]. Another important thing that should be taken into account is the flow of information in business systems and the need to look for methods leading to the minimization of the time and costs of these information transfers [9]. The method of logistics planning is a technique of organization and rules of conduct, purposefully selected for the subject and purpose of planning, aimed at making planning decisions. It should be used consciously and precisely adjusted to the requirements of the conditions and planning areas. The goal of the article is to present the method of analyzing the way of determining the optimal approach of selecting a universal course of action that improves the business process so that it can be implemented to satisfy a wider sequence of customer orders.

2 Mathematical model

Let us assume the order to make is illustrated by the matrix of orders (1):

$$Z^k = \left[z^k_{m,n}\right], m = 1, \dots, M, n = 1, \dots, N, k = 0,1, \dots, K, \tag{1}$$

[1] Silesian University in Opava, School of Business Administration in Karvina, Department of Informatics and Mathematics, Univerzitní náměstí 1934/3, 73340 Karviná, Czech Republic, suchanek@opf.slu.cz.
[2] Wyższa Szkoła Informatyki i Zarządzania w Bielsku-Białej, Institute of Management and Information Technology, Bielsko-Biała, Legionów 81, 43-302 Bielsko-Biała, Poland, bucki.robert@wp.pl.

where: $z_{m,n}^{k}$ - the number of pieces of the n-th product to be made for the m-th customer at the k-th stage of ready input data. Let us introduce the matrix of times of making orders (2):

$$T^{0} = \left[\tau_{m,n}^{0}\right], m = 1, \dots, M, n = 1, \dots, N, \tag{2}$$

where: $\tau_{m,n}^{0}$ - the time of making the n-th product for the m-th customer without implementing any improvement method. Let us introduce the matrix of unit costs of making orders (3):

$$C^{0} = \left[c_{m,n}^{0}\right], m = 1, \dots, M, n = 1, \dots, N, \tag{3}$$

where: $c_{m,n}^{0}$ - the unit cost of making the n-th product for the m-th customer without implementing any improvement method. The cost of making a certain number of units of the n-th product for the m-th customer without implementing any improvement method can be calculated in the following way (4):

$$C_{m,n}^{0/k} = z_{m,n}^{k} \cdot \tau_{m,n}^{0} \cdot c_{m,n}^{0} \tag{4}$$

Subsequently, the cost of making all order matrix elements by the system without implementing any improvement method is as follows (5):

$$C_{M,N}^{0/k} = \sum_{\mathrm{m}=1}^{\mathrm{M}} \sum_{\mathrm{n}=1}^{\mathrm{N}} z_{m,n}^{k} \cdot \tau_{m,n}^{0} \cdot c_{m,n}^{0} \tag{5}$$

Let us introduce the vector of improvement methods (6):

$$\Psi = [\psi^{\alpha}], \alpha = 1, \dots, A \tag{6}$$

where: ψ_{α} - the α-th improvement method. Let us introduce the matrix of times of making orders by means of improving methods (7):

$$T^{\alpha} = \left[\tau_{m,n}^{\alpha}\right], \alpha = 1, \dots, A, m = 1, \dots, M, n = 1, \dots, N \tag{7}$$

where: $\tau_{m,n}^{\alpha}$ - the time of making the n-th product for the m-th customer with the use of the α-th method. Let us introduce the matrix of unit costs of making orders by means of given improvement methods (8):

$$C^{\alpha} = \left[c_{m,n}^{\alpha}\right], \alpha = 1, \dots, A, m = 1, \dots, M, n = 1, \dots, N \tag{8}$$

where: $c_{m,n}^{\alpha}$ - the unit cost of making the n-th product for the m-th customer with the use of the α-th method. Moreover, there is a need to introduce benefit times as well as disadvantage times of each method which affects it while it is implemented. These times are correlated with adequate benefit unit costs and disadvantage unit costs. The matrix of benefit times for methods is introduced (9):

$$T_{m,n}^{\alpha/\beta} = \left[\tau_{m,n}^{\alpha/\beta}\right], \alpha = 1, \dots, A, \beta = 1, \dots, B, m = 1, \dots, M, n = 1, \dots, N \tag{9}$$

where: $\tau_{m,n}^{\alpha/\beta}$ - the β-th benefit time of making the n-th product for the m-th customer with the use of the α-th method. Let us introduce the matrix of unit costs of benefits (10):

$$C_{m,n}^{\alpha/\beta} = \left[c_{m,n}^{\alpha/\beta}\right], \alpha = 1, \dots, A, \beta = 1, \dots, B, m = 1, \dots, M, n = 1, \dots, N \tag{10}$$

where $c_{m,n}^{\alpha/\beta}$ - the β-th benefit unit cost of making the n-th product for the m-th customer with the use of the α-th method. Subsequently, the matrix of disadvantage times for methods is introduced (11):

$$T_{m,n}^{\alpha/\gamma} = \left[\tau_{m,n}^{\alpha/\gamma}\right], \alpha = 1, \dots, A, \gamma = 1, \dots, \Gamma, m = 1, \dots, M, n = 1, \dots, N \tag{11}$$

where: $\tau_{m,n}^{\alpha/\gamma}$ - the γ-th disadvantage time of making the n-th product for the m-th customer with the use of the α-th method. Let us introduce the matrix of unit costs of disadvantages (12):

$$C_{m,n}^{\alpha/\gamma} = \left[c_{m,n}^{\alpha/\gamma}\right], \alpha = 1, \dots, A, \gamma = 1, \dots, \Gamma,\ m = 1, \dots, M, n = 1, \dots, N \tag{12}$$

where $c_{m,n}^{\alpha/\gamma}$ - the γ-th disadvantage unit cost of making the n-th product for the m-th customer with the use of the α-th method. The matrix of costs of making the n-th product for the m-th customer with the use of the α-th method takes the following form (13):

$$C_{m,n}^{\alpha} = \left[\tau_{m,n}^{\alpha} \cdot c_{m,n}^{\alpha} - \sum_{\beta=1}^{B} \tau_{m,n}^{\alpha/\beta} \cdot c_{m,n}^{\alpha/\beta} + \sum_{\gamma=1}^{\Gamma} \tau_{m,n}^{\alpha/\gamma} \cdot c_{m,n}^{\alpha/\gamma}\right] \tag{13}$$

The criterion of making the n-th product for the m-th customer with the use of the α-th method consists in minimizing the overall costs: $Q = C_{m,n}^{\alpha} \to min$ which can be specified as follows (14):

$$C_{m,n}^{\alpha\_min} = minC_{m,n}^{\alpha}, \alpha = 1, \dots, A, m = 1, \dots, M, n = 1, \dots, N \tag{14}$$

On the basis of the above it is possible to calculate the costs of making all products for all customers with the use of the α-th method at the k-th stage as follows (15):

$$C_{M,N}^{\alpha/k} = z_{m,n}^{k} \cdot \sum_{\mathrm{m}=1}^{\mathrm{M}} \sum_{\mathrm{n}=1}^{\mathrm{N}} \left(\tau_{m,n}^{\alpha} \cdot c_{m,n}^{\alpha} - \tau_{m,n}^{\alpha/\beta} \cdot c_{m,n}^{\alpha/\beta} + \tau_{m,n}^{\alpha/\gamma} \cdot c_{m,n}^{\alpha/\gamma}\right) \tag{15}$$

The criterion of making all products for all customers with the use of the α-th method consists in minimizing the overall costs: $Q = C_{M,N}^{\alpha} \rightarrow min$ which can be specified as follows (16):

$$C_{M,N}^{\alpha/k\_min} = min z_{m,n}^{k} \cdot \sum_{\mathrm{m}=1}^{\mathrm{M}} \sum_{\mathrm{n}=1}^{\mathrm{N}} \left(\tau_{m,n}^{\alpha} \cdot c_{m,n}^{\alpha} - \tau_{m,n}^{\alpha/\beta} \cdot c_{m,n}^{\alpha/\beta} + \tau_{m,n}^{\alpha/\gamma} \cdot c_{m,n}^{\alpha/\gamma}\right) \tag{16}$$

The average cost of making the n-th product for the m-th customer with the use of all methods is included in the matrix of average costs as follows (17):

$$C_{m,n}^{av} = \left(\sum_{\alpha=1}^{\mathrm{A}} \tau_{m,n}^{\alpha} \cdot c_{m,n}^{\alpha} - \sum_{\alpha=1}^{\mathrm{A}} \tau_{m,n}^{\alpha/\beta} \cdot c_{m,n}^{\alpha/\beta} + \sum_{\alpha=1}^{\mathrm{A}} \tau_{m,n}^{\alpha/\gamma} \cdot c_{m,n}^{\alpha/\gamma}\right)/A \tag{17}$$

The average cost of making all N products for M customers with the use of all available methods A at the k-th stage is included in the matrix of average costs as follows (18):

$$C_{M,N}^{av/k} = z_{m,n}^{k}\Big(\sum_{\alpha=1}^{A} \sum_{\mathrm{m}=1}^{M} \sum_{\mathrm{n}=1}^{N} \tau_{m,n}^{\alpha} \cdot c_{m,n}^{\alpha} - \sum_{\alpha=1}^{A} \sum_{\mathrm{m}=1}^{M} \sum_{\mathrm{n}=1}^{N} \tau_{m,n}^{\alpha/\beta} \cdot c_{m,n}^{\alpha/\beta} + \sum_{\alpha=1}^{A} \sum_{\mathrm{m}=1}^{M} \sum_{\mathrm{n}=1}^{n} \tau_{m,n}^{\alpha/\gamma} \cdot c_{m,n}^{\alpha/\gamma}\Big)/A \tag{18}$$

There is a need to calculate the difference between the average result and the best result in case of making the n-th product for the m-th customer at the k-th stage to evaluate the proposed methods (19):

$$\Delta C_{m,n}^{av(k)} = C_{m,n}^{av(k)} - C_{m,n}^{\alpha\_\mathrm{min}(k)} \tag{19}$$

Consequently, the difference between the average result and the best result in case of making N products for M customers at the k-th stage to evaluate the proposed methods is as follows (20):

$$\Delta C_{M,\mathrm{N}}^{av(k)} = C_{M,\mathrm{N}}^{av(k)} - C_{M,\mathrm{N}}^{\alpha\_\mathrm{min}(k)} \tag{20}$$

It is assumed that once an order matrix is sent to the production system to be made, each order matrix element has to be completed before considering the subsequent awaiting order. Moreover, the production operation can begin on the following condition (21):

$$\sum_{\mathrm{m}=1}^{\mathrm{M}} \sum_{\mathrm{n}=1}^{\mathrm{N}} z_{m,n}^{k} \geq \lambda,\ \lambda = 0,1, \dots, \Lambda,\ \Lambda = M \cdot N \tag{21}$$

where: λ - the number of units of all products enabling starting the production process. Taking into account the above ranges for drawing, it is possible to propose the way of calculating the following variables (22):

$$\lambda = \left(\left(H\left(z_{m,n}^{k}\right) - L\left(z_{m,n}^{k}\right)\right)/\rho\right) \cdot M \cdot N,\ \rho = 1, \dots, P, \tag{22}$$

where: ρ - the coefficient of division of the sampling range. If $\sum_{\mathrm{m}=1}^{\mathrm{M}} \sum_{\mathrm{n}=1}^{\mathrm{N}} z_{m,n}^{k} < \lambda$, then the need to increase the number of order matrix elements arises before beginning the manufacturing process as follows (23):

$$\sum_{\mathrm{m}=1}^{\mathrm{M}} \sum_{\mathrm{n}=1}^{\mathrm{N}} \left(z_{m,n}^{k} + \Delta z_{m,n}^{k}\right) \geq \lambda, \tag{23}$$

where: $\Delta z_{m,n}^{k}$ - the added number of order units of the n-th product pieces to be made for the m-th customer at the k-th stage. At the same time: $z_{m,n}^{k} \leqslant H\left(z_{m,n}^{k}\right) - z_{m,n}^{k}$.
It is justified by the fact that too small an order is not able to provide profits for the proper functioning of the company. At the same time, if $Z^{k} < \lambda$, the order set cannot be accepted and the search for the satisfactory order data meeting the condition $Z^{k} + \Delta Z^{k} \geq \lambda$ continues. The simulation tool implemented for the case study was created on the basis of the following **pseudocode**:

i) Initial data introduction: Ψ, M, N, K, Λ, P, $L(z_{m,n}^{k})$, $H(z_{m,n}^{k})$, $L(\tau_{m,n}^{\alpha})$, $H(\tau_{m,n}^{\alpha})$, $L(c_{m,n}^{\alpha})$, $H(c_{m,n}^{\alpha})$, $L(\tau_{m,n}^{\alpha/\beta})$, $H(\tau_{m,n}^{\alpha/\beta})$, $L(c_{m,n}^{\alpha/\beta})$, $H(c_{m,n}^{\alpha/\beta})$, $L(\tau_{m,n}^{\alpha/\gamma})$, $H(\tau_{m,n}^{\alpha/\gamma})$, $L(c_{m,n}^{\alpha/\gamma})$, $H(c_{m,n}^{\alpha/\gamma})$
ii) $k = 1$
iii) Draw at random: $Z^{k}, T^{\alpha}, C^{\alpha}, T^{\alpha/\beta}, C^{\alpha/\beta}, T^{\alpha/\gamma}, C^{\alpha/\gamma}$.
iv) $\sum_{\mathrm{m}=1}^{\mathrm{M}} \sum_{\mathrm{n}=1}^{\mathrm{N}} z_{m,n}^{k} \geq \lambda$? If YES, go to (*vii*). Otherwise, go to (*v*).
v) $z_{m,n}^{k} + \Delta z_{m,n}^{k}$ and go to (*vi*).
vi) $z_{m,n}^{k} + \Delta z_{m,n}^{k} =: z_{m,n}^{k}$ and go to (*iv*).
vii) $C_{m,n}^{\alpha\_\mathrm{min}/\mathrm{k}}$, $C_{M,N}^{\alpha\_min/k}$, $C_{M,N}^{\alpha v(k)}$, $\alpha = 1, \dots, A$.
viii) Store the results.
ix) $k = K$? If YES go to (*xi*). Otherwise, go to (*x*).

x) $k = k + 1$ and go to (iv).
xi) Adjust a weight to each α-th method, $\alpha = 1, \dots, A$.
xii) Ignite the point analysis of evaluation and selection of the methods.
xiii) Choose the optimal method for all orders Z^k, $k = 1, \dots, K$.
xiv) Present the graph for making all orders Z^k with the use of all available methods A, $k = 1, \dots, K$.

3 The case study

In order to conduct an adequate case study, a system with the following input data was selected for the analysis and successive assessment: M= 5, N= 5, K= 7, A= 10, λ= 1250. Consequently, the study case requires implementing the sample methods as well as the ranges for randomizing input data which is proposed in Table 1. The calculation units for times are seconds, and for unit costs, the contractual unit of account.

Orders		**Method**				**Benefits**				**Disadvantages**			
$L(z^k_{m,n}) = 1$	$L(z^k_{m,n}) = 5$	$c^{\alpha}_{m,n}$		$\tau^{\alpha}_{m,n}$		$c^{\alpha/\beta}_{m,n}$		$\tau^{\alpha/\beta}_{m,n}$		$c^{\alpha/\gamma}_{m,n}$		$\tau^{\alpha/\gamma}_{m,n}$	
Method		L	H	L	H	L	H	L	H	L	H	L	H
TQM	total quality management	7	13	39	165	1	5	1	27	1	5	1	27
BPR	business process reengineering	7	13	39	165	1	5	1	27	1	5	1	27
6σ	six sigma	7	13	39	165	1	5	1	27	1	5	1	27
ABM	activity based management	7	13	39	165	1	5	1	27	1	5	1	27
TBM	time based management	7	13	39	165	1	5	1	27	1	5	1	27
BS	balanced scorecard	7	13	39	165	1	5	1	27	1	5	1	27
ABC	activity based costing	7	13	39	165	1	5	1	27	1	5	1	27
ABB	activity based budgeting	7	13	39	165	1	5	1	27	1	5	1	27
TC	target costing	7	13	39	165	1	5	1	27	1	5	1	27
LM	lean management	7	13	39	165	1	5	1	27	1	5	1	27

Table 1 Range for randomizing input data

In case of not implementing any improvement method, the base initial values for the unit cost and the operation time equal respectively $c^0_{m,n} = 13$ and $\tau^0_{m,n} = 165$. The drawn orders for each Z^k, on the basis of the range stated in Table 1, are presented in Table 2.

Set 1		Total:	**1392**			**Set 2**		Total:	**1276**			**Set 3**		Total:	**1327**		
$z^1_{m,n}$	n=1	n=2	n=3	n=4	n=5	$z^2_{m,n}$	n=1	n=2	n=3	n=4	n=5	$z^2_{m,n}$	n=1	n=2	n=3	n=4	n=5
m=1	95	14	74	26	31	m=1	67	4	68	87	87	m=1	76	95	47	18	6
m=2	76	78	36	58	39	m=2	90	17	21	71	86	m=2	39	97	16	31	62
m=3	98	4	14	9	88	m=3	11	12	71	40	34	m=3	45	58	84	42	73
m=4	55	50	65	87	60	m=4	3	43	99	99	51	m=4	14	69	19	90	60
m=5	57	56	50	92	80	m=5	65	20	73	51	6	m=5	90	12	73	20	91
Set 4		Total:	**1316**			**Set 5**		Total:	**1584**			**Set 6**		Total:	**1322**		
$z^4_{m,n}$	n=1	n=2	n=3	n=4	n=5	$z^5_{m,n}$	n=1	n=2	n=3	n=4	n=5	$z^6_{m,n}$	n=1	n=2	n=3	n=4	n=5
m=1	25	86	69	95	92	m=1	80	64	64	52	83	m=1	34	50	35	76	20
m=2	17	82	65	0	1	m=2	91	20	85	87	89	m=2	52	72	92	81	47
m=3	97	11	68	97	90	m=3	44	76	4	8	88	m=3	19	14	82	1	98
m=4	47	79	1	58	28	m=4	61	88	47	27	98	m=4	63	18	76	83	57
m=5	83	47	18	28	32	m=5	64	87	41	56	80	m=5	45	50	82	22	53
Set 7		Total:	**1417**														
$z^7_{m,n}$	n=1	n=2	n=3	n=4	n=5												
m=1	26	67	87	50	47												
m=2	59	66	23	63	72												
m=3	79	15	91	16	86												
m=4	97	52	74	75	4												
m=5	77	94	9	80	8												

Table 2 Input data sets - orders

For the established input data, the analysis of exemplary methods was carried out in terms of their suitability for minimizing operating costs in the discussed system. The analysis of the results shown in Table 3 was carried out in terms of searching for:

i) the most optimal solution for the production of the n-th type of product for the m-th customer within a certain set of initial data $minC^{\alpha}_{Set\_k\_m,n}$
ii) the most optimal solution for the production of individual orders within each set of orders $minC^{\alpha}_{Set\_k}$
iii) the most optimal solution for the production of all individual orders for all input data sets $min\sum_{k=1}^{K}\sum_{\alpha=1}^{A}C^{\alpha}_{Set_k}$
iv) the most optimal solution taking into account the best solutions for each set of input data $\sum_{k=1}^{K}\sum_{\alpha=1}^{A}minC^{\alpha}_{Set\_k}$

Method	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Σ
Base	2985840	2737020	2846415	2822820	2822820	2835690	3039465	20090070
TQM	2083013	1807475	2124728	1943247	1923211	1842779	2141208	13865661
BPR	2294300	1977885	2106546	2043583	2001190	2151205	2187054	14761763
6б	2058869	1895948	2003591	1942510	2049463	1879119	2094672	13924172
ABM	2117625	1891696	2003444	2058213	2090963	2024545	2087579	14274065
TBM	2097400	2088500	2026425	1961597	2074022	1958028	1967688	14173660
BS	2197656	1804515	2075596	1962149	1906641	1912889	2037091	13896537
ABC	2021694	1945233	2064300	2000776	1861129	1965794	2064308	13923234
ABB	2148110	1896954	2140199	2160379	1918323	1959431	2079486	14302882
TC	2085578	1850473	2133869	1983214	1950053	2071711	1977696	14052594
LM	2063297	1741115	1951172	2114025	1923991	1905968	2127686	13827254
Min	2021694	1741115	1951172	1942510	1861129	1842779	1967688	13328087
Σ	21167542	18899794	20629870	20169693	19698986	19671469	20764468	
ΔΣ	95060	148864	111815	74459	108770	124368	108759	
Av.	2116754	1889979	2062987	2016969	1969899	1967147	2076447	

Table 3 Final results table – total costs

Final results can be presented in the following supplementary way:

i) $minC^{\alpha}_{Set\_1\_1,1} = C^{\mathrm{TQM}}_{Set\_1\_1,1} = 99275; minC^{\alpha}_{Set\_1\_1,2} = C^{66}_{Set\_1\_1,2} = 15022$; etc.
ii) $minC^{\alpha}_{Set\_1} = C^{\mathrm{ABC}}_{Set_1} = 2021694; minC^{\alpha}_{Set\_2} = C^{\mathrm{LM}}_{Set\_2} = 1741115$; etc.
iii) $min\sum_{k=1}^{7}\sum_{\alpha=1}^{10}C^{\alpha}_{Set\_k} = 13827254$
iv) $\sum_{k=1}^{K}\sum_{\alpha=1}^{A}minC^{\alpha}_{Set\_k} = 13328087$

The results presented above induce the possibility of alternating the implementation of the improvement methods for each set of input data characterized by a large number of orders. Moreover, for an exemplary comparison, each method was assigned an identical weight of 0.10 which translated directly into the final result due to the frequency of occurrences of a single method for individual orders. The best result was obtained for the TQM method. However, the number of occurrences of this method does not guarantee obtaining the minimal total cost of making all orders included in the sets of initial data as this study case is satisfied by the LM method as the sample one.

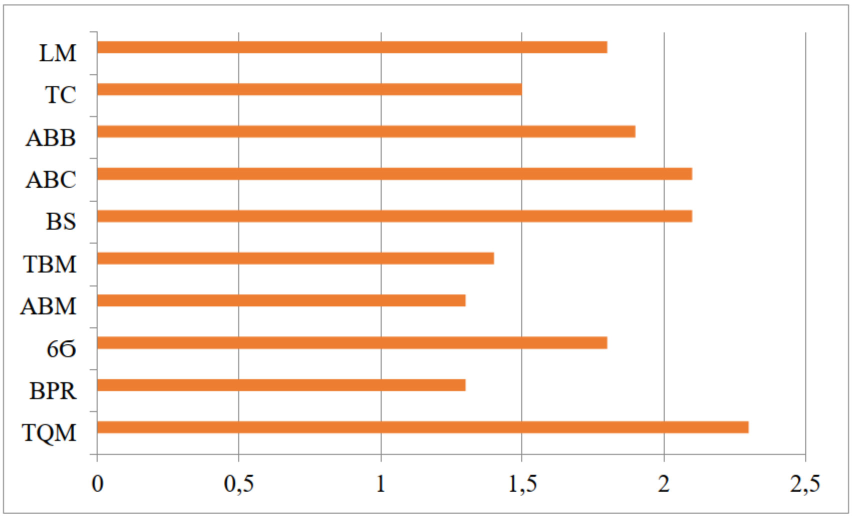


Figure 1 Results of the weighing of the implemented methods

4 Conclusions

The issue of choosing the right methods to improve the functioning of business systems is a big challenge in modern economies. The issue discussed in this article concerns modelling and simulation of the process of selecting the most optimal solution of the improving method for the given input data. It should be noticed that even a slight difference in the values of the input data may lead to the necessity to verify the included method and its possible change for another which is not always possible during a real business activity, however, in the case of a small number of orders, this will not always lead to the satisfactory result. A thorough comparative analysis regarding the possibility of implementing improvement methods indicates which of them should be used in order to minimize the costs of making customers' orders. Further research should focus on searching for a combination of best practices resulting from the implemented improvement methods. This type of approach could lead to the creation of a model that would ultimately make it possible to find a satisfactory solution for a given business system. However, it should be remembered that each individual business system is created on the basis of assumptions that differ from those of other systems, even if they are similar in nature.

Acknowledgements

This paper was supported by the Ministry of Education, Youth and Sports in the Czech Republic within the Institutional Support for Long-term Development of a Research Organization in 2022.

References

[1] Aljarrah, O., J. Li, W.Z. Huang, A. Heryudono & J. Bi, (2019). A Self-Organizing Evolutionary Method to Model and Optimize Correlated Multiresponse Metrics for Additive Manufacturing Processes. *Smart and Sustainable Manufacturing Systems,* 3(2), 190-214.

[2] vom Brocke, J., M.S. Baier, T. Schmiedel, K. Stelzl, M. Roglinger & C. Wehking, (2021). Context-Aware Business Process Management Method Assessment and Selection. *Business & Information Systems Engineering*, 63(5), 533-550.

[3] Chelbi, A. & Rezg, N. (2012). Optimisation methods in production, maintenance and logistics. *International Journal of Production Research*, 50(13), 3515-3516.

[4] Grisold, T., S. Gross, K. Stelzl, J. vom Brocke, J. Mendling, M. Roglinger & M. Rosemann, (2022). The Five Diamond Method for Explorative Business Process Management. *Business & Information Systems Engineering*, 64(2), 149-166.

[5] Ito, S., D. Vymetal & R. Sperka, (2021). Process mining approach to formal business process modelling and verification: a case study. *Journal of Modelling in Management*, 16(2), 602-622.

[6] Li, T.S., L. Xiong, A.Q. Dong, Z.S. Liu & W. Tan, (2019). Optimization method based on big data in business process management. *Cluster Computing - The Journal of Networks Software Tools and Applications*, 22, S5357-S5365.

[7] Prokopenko, O., Y. Kichuk, O. Ptashchenko, I. Yurko & M. Cherkashyna, (2021). Logistics Concepts to Optimise Business Processes. *Estudios de Economia Aplicada*, 39(3).

[8] Shokhnekh, A.V. (2019). The Main Notions, Principles, and Procedures of Strategic Quality Management of Small Business in the System of Digital Economy in View of Risks of Drift: A Cognitive Approach. *International Journal for Quality Research*, 13(3), 655-668.

[9] Suchánek, P. & R. Bucki, (2020). Modelling Timings of the Company's Response to Specific Customer Requirements. *Smart Innovation, Systems and Technologies, Agents and Multi-Agent Systems: Technologies and Applications* 2020, 186, 209-219.

EU Artificial Intelligence Act in Banking Sector: Impact and Implementation

Pavel Šálek[1]

Abstract. The European Union Commission's regulatory proposal on Artificial Intelligence will enter into force within the next two years. The focus of the Act is the introduction of regulatory and legal frameworks for the application of the artificial intelligence across European Union. The Act divides the implementation of AI into four main categories by the risk involved with their possible applications. The risks assessed are mainly opacity, complexity, unpredictability, autonomy, and data. The systems in the low risk category are permitted to use without any restrictions. The next category is AI systems with specific transparency obligations (i.e., chatbots, deep fakes). The high risk category involves systems that are products of other safety regulations (i.e., machinery, toys, and medical care) and systems listed by the European Commission as high risk. The AI systems in this category must be sufficiently transparent to enable users to understand and control how the high-risk AI system produces its output. The last category is systems with unacceptable risk where the application of AI is prohibited. Systems that can cause physical or emotional harm (i.e., social scoring) fall into this category. This article aims to assess the impact of the obligations in different risk categories on the AI systems and discuss the potential explainability and interpretability techniques that can be used to ensure the successful implementation of the Act.

Keywords: AI, artificial intelligence act, explainability, interpretability, regulation, machine learning

JEL Classification: C44
AMS Classification: 90C15

1 Brief history of AI regulation

Legislation and policies around artificial intelligence started to take shape during the presidency of Jean Claude Juncker from 2014 to 2019. The European Union was looking to establish itself as a leader in the digital global economy, and Juncker introduced a collection of policies called The Digital Single Market in 2015. The Digital Single Market strategy contains about 30 initiatives under three pillars which are focused on better access for consumers and businesses to online goods and services across Europe, creating the right conditions for digital networks and services to flourish and maximising the growth potential of our European Digital Economy [10].

The regulations focused exclusively on AI have been published since 2018. In April 2018, the European Commission launched The EU strategy for AI. The strategy introduced three concepts that became the main topics for the European AI strategy [10].

- Firstly, it aims to enhance the usage and adoption of AI across the industries in both the private and public sectors. It sets a target for investments to reach 20 billion euros each year over the next decade.
- Secondly, the strategy prepares for the impact of AI adoption from the social and economic perspective. It deals with one of the most stressed problems regarding the broad implementation of robotics and automation: employment. The steps include labor market transition analysis, social security, and education.
- Thirdly, the strategy creates the ethical and legal framework based on European values. This includes the development of AI ethics guidelines and an idea of AI as a service for human development.

The need for an ethical guide for AI has been stressed in all EU AI policies. The ethics guidelines were developed and published in April 2019 as The Ethics Guidelines for Trustworthy AI. The guidelines emphasize that AI should be used as a mean for human progress and innovation and not as an end of itself [5]. Based on the guidelines, trustworthy AI should be lawful, ethical, and robust [3].

[1] Prague University of Economics and Business, Prague, Czech Republic, xsalp09@vse.cz.

The new President of the European Commission, Ursula von der Leyen (2019-2024), will prepare legislation for a coordinated European approach to the AI. This led to the introduction of a white paper on AI in February 2020. The main idea of the paper is to set a clear regulatory framework that would build trust among consumers and businesses in AI [5]. This white paper was later introduced on the 21st of April 2021 by the European Commission as a The Artificial Intelligence Act (The Act), or, to use its full name, The proposal for a regulation of the European Parliament and the Council laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts [4]. The AI Act looks to introduce the regulatory and legal frameworks for the application of the artificial intelligence across the European Union, harmonize rules for the development, placement on the market, and specific utilization of AI systems and associated risks [11].

2 The AI Act: Scope and framework

There is no unique definition of artificial intelligence that is broadly accepted by the scientific community, and AI is often used as an alternative term that covers multiple computer software or applications developed using various approaches. All of these techniques resemble the process of learning in any way, which can be connected to human intelligence. The proposed definition can be found in the regulation under the Article3(1), and it states [4]:

> "artificial intelligence system (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objec- tives, generate outputs such as content, predictions, recommendations, or decisions influencing the envi- ronments they interact with"

Annex I provides the regulation with a list of various approaches to AI development. These approaches can be viewed as a range of software applications that rely on the knowledge from statistics, machine learning, or logical systems [2]. The groups defined in the Annex I are [4]:

- Machine learning approaches, including supervised, unsupervised, and reinforcement learning, use various methods, including deep learning.
- Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference, deductive engines, (symbolic) reasoning, and expert systems.
- Statistical approaches, Bayesian estimation, search and optimization methods.

The AI Act is the first attempt to establish a broad regulation aimed solely at AI systems. The proposed legal framework focuses on utilizing AI systems in the EU and the risks associated with the implementation. The Commission divided the risk levels of AI systems into four different categories based on their characteristics in opacity, complexity, dependency on data, autonomous behavior, and unpredictability. The AI Act follows a risk-based approach where AI systems are assigned to a concrete level of risk, and each level of risk has its own legal consequences. The risk levels are formalized in [4] as follows:

- Unacceptable risks (Defined in Title II).
- High risks (Defined in Title III).
- Limited risks (Defined in Title IV)
- Minimal risks (Defined in Title IX)

Unacceptable risks

Title II (Article 5) of the AI Act defines the group of AI systems and their application with unacceptable risk. It bans the AI practices, which can prove to be a clear threat to people's safety, livelihoods, and rights [5]. The AI systems in this category are outright banned, and their usage is permitted only under specified circumstances. The AI systems in this category [4] can be listed as:

- AI systems that deploy harmful manipulative 'subliminal techniques' in order to distort a person's behavior in a manner that causes or is likely to cause that or another person physical or psychological harm.
- AI systems that exploit specific vulnerable groups (physical or mental disability) to materially distort the behavior of a person pertaining to that group in a manner that causes or is likely to cause that person or another person physical or psychological harm.
- AI systems by public authorities or on their behalf for the evaluation or classification of the trustworthiness (social scoring purposes) of natural persons over a certain period of time based on their social behavior or known or predicted personal or personality characteristics. With the social score leading to either or both of the following
  - Detrimental or unfavorable treatment of certain natural persons or whole groups thereof in social contexts unrelated to the contexts in which the data was originally generated or collected.

  - Detrimental or unfavorable treatment of certain natural persons or whole groups thereof that is unjustified or disproportionate to their social behavior or its gravity.
- A real-time remote bio-metric identification systems in publicly accessible spaces for law enforcement purposes, except in a limited number of cases

High risk

Title III is by far the longest description of the risk category, and it concerns the AI system that poses a high risk to health, safety, and fundamental rights [11]. The Act distinguishes these systems into two groups [2]:

- High-risk AI systems used as a safety component of a product or as a product falling under Union health and safety harmonization legislation (e.g., toys, aviation, cars, medical devices, lifts)
- High-risk AI systems deployed in eight specific areas identified in Annex III, which the Commission would be empowered to update as necessary by way of a delegated act (Article 7): Bio-metric identification and categorization of natural persons, management and operation of critical infrastructure, education and voca- tional training, employment, worker management, and access to self-employment, access to and enjoyment of essential private and public services and benefits, law enforcement, migration, asylum and border control management, administration of justice and democratic processes.

Any AI products must register the AI systems in a database managed by the Commission before the deployment. From the point of view of conformity assessment, the systems governed by exiting legislation fall under the existing third-party conformity frameworks that already apply. Providers of AI systems currently not governed by EU legislation would have to conduct their own conformity assessment (self-assessment), showing that they comply with the new requirements for high-risk AI systems. Providers established outside the EU have to delegate a representative in the EU to ensure the conformity assessment [2].

There are several other requirements that need to be full-filled in order to deploy or use high risk AI systems in EU. These requirements touch on risk management systems surrounding AI systems, technical documentation, data sets expectations, logging and traceability, human oversight and others [5].

Limited risks

There are specified AI systems that present limited risks. These are defined by Title IX (Article 52). The providers must ensure that the AI systems which are intended to interact with natural persons (such as bots) or bio-metric categorization systems are developed in a way that the users of such systems are informed that they are interacting with an AI system [5]. The provider of the AI system that generates or manipulates image, audio, or video content in order to realistically resemble existing persons, objects, places, or other entities ('deep fake') must label such outcomes properly [4].

Minimal risks

All other AI systems not specified in any other risk category present only low or minimal risk. These systems can be developed and deployed in the EU without additional legal restrictions [2]. The AI act additionally encourages developers of non high risk AI systems to apply the mandatory requirements for high-risk AI systems on a voluntary basis [11].

3 Application of regulation

This chapter will discuss the regulation related to the rising field of model explainability and interpretability. Another interesting point is the application of The AI Act in the banking industry. Although it is not visible at first sight, the regulation impacts the scoring and rating systems used in banks as the proposal stands at the moment.

3.1 Explainability and Interpretability

Under Title II Article 13, The Act specifies that high-risk AI systems must be developed and designed in such a way to be sufficiently transparent in order to ensure the user's ability to interpret and use the system's output. Recent development in explainability and interpretability as a field of research have made significant progress which makes the machine learning algorithms more trustworthy. The following few paragraphs aim to introduce the current techniques in explainability, which will help implement The AI Act in practice. The explainability models are classified into three main categories [8]:

- Intrinsically interpretable techniques.
- Model agnostic technique.
- Example-based methods.

Intrinsically Interpretable Techniques The first possibility of explainability is the Intrinsically interpretable techniques. This category relies on the property of the models that are interpretable in their nature. These models are generally considered interpretable when their coefficients have a clear meaning and practical interpretation.

The interpretability of these models benefits from the linear relationship of the variables, and due to this property, they are widely used in areas of medicine, sociology, science, finance, and many more. Some of the examples are generalized linear model, tree-based models, generalized additive models and more. Besides these, Naive Bayes classifier or K-Nearest neighbors algorithms are included in the intrinsically interpretable models [7].

Unfortunately, as the non-linearity and correlation of features increases, it is not possible to use the Intrinsically interpretable techniques, and the focus has to be turned to more complex machine learning methods.

Model agnostic Model agnostic approaches can be applied to any machine learning or any other black-box model. The techniques used for interpretation are applied after the model has been trained. These methods rely on analyzing the relationships between feature input and output. These methods cannot leverage the model's inner workings, such as weights or structural information. The interpretation of the models using this approach does not decline the predictive power of the model itself as they are applied after training [7]. The five main categories are [8]:

- Visualization-based approaches rely on plots that help understand the features' individual effects on the model output. There are several visualization techniques under this category, such as Partial dependency plots, Individual expectation conditions, or Accumulated individual effects.
- Feature interaction-based algorithms compute the feature interactions with each other. The interaction between two features is computed using the differences in the partial dependencies of pairs of features and the sum of these partial dependencies. The partial dependencies approach is very computationally expensive.
- Global surrogate models aim at interpreting the model with a less complex model. This method is under critique that it cannot be classified as model interpretability per se.
- Local surrogate models focus on specific predictions made by the black box model in order to explain it. The most known models in this category are LIME, which tries to train a local surrogate model to explain local predictions, and models based on Shapley value from game theory.
- Propagation-based methods rely on quantifying the outputs after permutation of the individual features. There are either forward- or back-propagation based methods.

The Shapley value based method produce high quality explanations, however the exact computation of the algorithm can be applied only on the tree-based models (i.e, XGBoost, LightGBM). It is a set of algorithms which leverage the Shapley regression values from cooperative game theory [12] for interpreting predictions. Shapley regression assigns each feature an importance value for a particular prediction to compute the explanation. This value is unified and represents the additive feature attributions. Using the notation from [9], Shapley regression can be formalized as:

$$\Theta_i = \sum_{S \in F \setminus \{i\}} \frac{|S|!(M - |S| - 1)!}{M!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)], \tag{1}$$

where Θ_i is a unified measure of additive feature attributions ($\Theta_i \in \mathbb{R}$) , F is the set of input features , S is a subset of input features and M=|F| is the number of input features. This formula computes the importance of individual features by calculating the difference in contribution when the feature is present in the prediction and when it is omitted.

- $f_{S \cup \{i\}}(x_{S \cup \{i\}})$ is the output when the $i^t h$ feature is present,
- $f_S(x_S)$ is the output when the $i^t h$ feature is not present,
- $\sum_{S \in F \setminus \{i\}} \frac{|S|!(M-|S|-1)!}{M!}$ is the weighted average of all possible subsets of S in F.

The Shapley value based algorithm is the only explanation model with a solid theory. Shapley values are the only values satisfying the property of local accuracy, missingness and consistency. The high quality explanations produced by this algorithms can be used for local explanations [9]. This can prove to be very useful for the transparency expectations by The Act as well as to justify a specific decision made by AI system (i.e., credit rating output).

The advantage of model agnostic methods is that there focus on the interpretation of the whole system. Unfortunately, this can prove to be very complex in computation [7].

Example base methods This method does not explicitly explain the model but highlights its essential elements instead. This approach generates local explanations. The methods are, i.e., Break down or Counterfactual explanation [7].

This is not by far the exhaustive list of all approaches. Many other algorithms are currently used to deal with the interpretability of the black box models. Mainly model-specific methods which focus on specific types of models, i.e., Neural networks or time series [8].

3.2 Banking

The first response from the banking market came on the 28th of September 2021. The response was written by European Banking Federation (EBF). The EBF is concerned that although The AI Act is aimed at creating an environment that encourages investment, aims to strengthen competitiveness, and ensures that AI systems respect fundamental rights and EU values, there is a reasonable risk of creating unjustified barriers or restrictions on the development of AI systems, may have an impact on the potential of research, creates obstacles in applications and therefore implies negative influence on the competitiveness. The regulations need the find balance between the high-risk AI systems requirements and space for innovations [6].

The banking and financial service industry is already subject to a wide range of specific regulations and supervision, which goal is to ensure consumer protection, capital stability, and risk management in the cases of creditworthiness assessment. A significant number of requirements for high-risk scoring models are already addressed in these initiatives and avoiding overlapping or conflicting requirements must be guaranteed. This also applies to the General Data Protection Regulation (GDPR). There are three key messages which are the most concerning for EBF and whose are the definition of AI, the scope, and supervision [6].

The EBF believes that the current definition of high-risk AI systems under The AI act is too broad and can include all types of systems or software applications that do not bear the same risks. EBF states that it is surprising to include rule-based procedures, which have been in use at banks for a long time and have been monitored and approved by the supervisory authorities, are now to be covered by the Act. Mainly because the risks involved and their impact connected to these algorithms are already adequately included in the institution's risk management, raising significant uncertainties. As a result, EFB recommends a more targeted approach to the definition in accordance with the principle: "same activity, same risks, same rules" [6].

The EBF asks for more guidance to clearly identify when the use of AI systems will be considered high risk and, therefore, subject to the requirements set out in the AI Act. More accurately, the EBF asks for replacing the term high-risk AI system with high-risk applications of AI systems, which better refers to a set of applications mentioned in Annex III of the Act. Clarifying the scope of the creditworthiness assessment is another concern of the EBF. The main objective of The Act is to prevent any physical or emotional harm to customers. However, these requirements are already embedded in current banking regulations, and banks have obligations, under strict prudential rules, to be able to measure, monitor, and manage their sources of financial as well as non-financial risk. This includes model, technology, and information security risk management. This approach already sets the highest standard of risk management and control. Furthermore, it applies regardless of the application of AI systems in the process. Only systems used to evaluate the access to credit should be considered as high risk, and AI systems that are used in the other parts of the credit system should be excluded (i.e., collateral valuation, VaR models, anti-fraud or anti-AML models and more).

Additionally, the scope of creditworthiness assessment should not be limited to the financial services but should be applicable to and should be regulated for all providers, irrespective of the sector (i.e., recruitment systems). EBF also stresses the importance of the support customer education and awareness of The AI Act itself in order to address possible myths and doubts. For example, avoiding AI rating systems because they are considered "high risk " and choosing another institution that does not use this technology instead.

The second response was issued on the 29th of December 2021 by European Central Bank (ECB). The findings are very similar to the ones presented by EBF. Additionally, ECB notes that it should be included as a consulting entity and also highlights that its role should be clarified under the proposed regulation, especially in relation to the market surveillance and conformity assessment [1].

4 Conclusion

The Artificial Intelligence Act is the first attempt to harmonize the horizontal regulation of the AI systems by introducing the different risk categories connected to the application of the AI systems and their impact on respect to society. However, the proposed frameworks are not without weaknesses that need to be addressed before the entry into force. The Act has to be adjusted to ensure that overlapping or conflicting requirements with the current regulatory environment in the banking industry are not present in the final proposal. The banking industry also calls for better definitions of high-risk models and their applications.

The Act focuses on the high-risk application of AI systems where it specifies the requirements for these models. One of the requirements is transparency. The field of explainability and interpretability in machine learning will play a crucial part during the implementation of The Act. Several techniques, such as intrinsically interpretable techniques, model agnostic techniques, or example-based methods, can be used to ensure the transparency of AI systems. The Shapley value method is one of the algorithms that can justify an individual decision based on AI system output.

This article has been created as part of a broader initiative focusing on model explainability and its application in the banking regulatory environment.

References

[1] European Central Bank (2021). *OPINION OF THE EUROPEAN CENTRAL BANK of 29 December 2021 on a proposal for a regulation laying down harmonised rules on artificial intelligence*. Frankfurt am Main

[2] Benife Brandoand Tudorache Dragoş (2022). *BRIEFING EU Legislationin Progress Proposal foraregulation of the European Parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain Union legislative acts Committees responsible*. Brussels: European Parliament

[3] European Commission (2019). *Ethics guidelines for trustworthy AI. Independent High-Level Expert Group on Artificial Intelligence set up by the European Commission*. Brussels

[4] European Commission (2021). *The Proposal for a regulation of the European Parliament and the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*. Brussels

[5] Martin Ebers & Veronica R. S. Hoch and Frank Rosenkranz and Hannah Ruschemeier and Björn Steinrötter (2021). *The European Commission's Proposal for an Artificial Intelligence Act—A Critical Assessment by Members of the Robotics and AI Law Society (RAILS).* J, 4, 589–603.

[6] European Banking Federation (2021). *EBF position paper on the EC proposal for a regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. Brussels

[7] Ambreen Hanif (2021). Towards Explainable Artificial Intelligence in Banking and Financial Services. *CoRR*, abs/2112.08441.

[8] Sheikh Rabiul Islam & William Eberle and Sheikh Khaled Ghafoor and Mohiuddin Ahmed (2021). Explainable Artificial Intelligence Approaches: A Survey. *CoRR*, abs/2101.09429.

[9] Andreas Messalas & Yiannis Kanellopoulos and Christos Makris (2019). Model-Agnostic Interpretability with Shapley Values. *10th International Conference on Information, Intelligence, Systems and Applications, IISA 2019.* Patras: Institute of Electrical and Electronics Engineers Inc.

[10] Jędrzej Niklas & Lina Dencik (2020). European artificial intelligence policy: mapping the institutional landscape. *DATAJUSTICE*. *European Research Council* (grant no. 759903).

[11] Michael Veale and Frederik Zuiderveen Borgesius (2021). Demystifying the Draft EU Artificial Intelligence Act. *Computer Law Review International*, 4, 97–112.

[12] Sergiu Hart (1989). Shapley value. J. Eatwell, M. Milgate, & P. Newman, editors, *The New Palgrave: Game Theory*, 210–216. W. W. Norton & Co Inc.

Stakeholders' Support by Agricultural Holdings: Analysis of Determinants by Logit Model

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

Abstract. It is acknowledged by many firms that their survival depends not only on its competitiveness, but also the relation towards various stakeholders that can affect the firm, is important. Therefore, we held a primary survey that analysed the social responsibility in agricultural firms. The aim of the paper is to assess what factors affect how important is for an agricultural holding the supports of the stakeholders.
We used ordered logistic regression model. As explanatory variables were included characteristics of the agricultural holding and of its manager. However, we found out that only legal form of a firm and business target of improving existing products and developing new products were statistically significant determinant of to what extent is the support of stakeholders important.
Despite that we assumed that representants of cooperatives would see the support of stakeholders as more important, our results show that when the company is a cooperative, the odd that the support of stakeholders is very important (over other categories) is lower 0.48 times. When the business target is improving and developing the product the odds that support is "very important" (over other categories) is 25.4 times higher. The importance of stakeholders' support was not statistically significantly determined nor by the sex, education or age of the manager, nor by the size (number of employees) and other business targets of the agricultural holding.

Keywords: agricultural holding, determinant, logistic regression, social responsibility

JEL Classification: C25, Q13
AMS Classification: 62J05

1 Introduction

Social responsibility theme has been discussed for several years with increasing magnitude especially regarding the environmental issues. It is acknowledged by many firms that their survival depends not only on financial competitiveness – "it is equally important that the organization could demonstrate their position in relation to the various interested parties (stakeholders), which are affected by the activities of the enterprise". [6].

Therefore, a concept of social responsibility (CSR) has been proposed and became not only a part of a business, but also a subject of a research. CRS can be defined as integrated corporate responsibilities that encompass the legal, economic, ethical, and philanthropic expectations that society has of firms. This definition is base od Carroll [2], who introduced a pyramid of CRS with 4 components: economic responsibility (be profitable is a base upon which are all others), legal (to obey the law), ethical (to do what is right, just, and fair), philanthropic (contribute resources to the community; improve quality of life). In our research, we focus on the last component – particularly on the relation and contribution to the stakeholders.

It is widely acknowledged nowadays that agricultural production has to comply not only with the economic and legal requirements, but also with the ethical responsibilities [3]. The relation towards the stakeholders is important as they can influence the operation of a firm. On the other hand, those groups are affected by the firm. "Applying Corporate Social Responsibility in agriculture would contribute to improving the image of farmers as perceived by stakeholders, as well as bringing notable economic, social and environmental benefits" [9]. Luhmann and Theuvsen [8] used an explorative factor analysis to identify CSR demands from society on agribusiness companies in Germany. They found out that there were three areas of responsibility: economic, internal and external.

This was confirmed by the study of Wicaksono and Setiawan [15], because they found out based on random effect model that government, foreign shareholders, and international operations were found to be significant drivers of water disclosure practices of agricultural holdings. Hence, the pressure of stakeholders can lead to more ethical

[1] Institute of Agricultural Economics and Information, Mánesova 75, 120 00 Prague 2, Czech Republic, simpachova.marie@uzei.cz
[2] Institute of Agricultural Economics and Information, Mánesova 75, 120 00 Prague 2, Czech Republic, simpach.ondrej@uzei.cz.

behaviour of the firms. The objectives of the producers can be different. For example, the target of food production was ranked significantly higher by the food system stakeholders than other people in New Hampshire in a research of Wilhelm et al. [16]. Their findings suggest that "there are differences in landscape preferences and perception of ecosystem service benefits between the general public and those who work with or in the food system" [16]. Also while the stakeholders in the Czech Republic would see as the most preferred erosion mitigation measures the growing of appropriate crops and the splitting of large fields into smaller ones, the views of farmers and local leaders are different. "A productivist paradigm and corresponding modulation of erosion is significantly more prevalent among farmers, although not the dominant viewpoint in this group" [13].

The determinants of the scope of socially responsible behavior of firms were examined for example by Udayasankar [12] who found out that firm size is very important factor. "Smaller firms may face fewer pressures compared to large ones and receive little recognition for their CSR because of their lower visibility." [12]. Zhirnov et al. [17] argued that a major reason for the low social responsibility of local agricultural businesses is agricultures' low profitability. Bavorová et al. [1] found out based on ordered logistic regression that farms with the legal form of a production cooperative and those with good economic performance were most likely to conduct socially responsible activities.

Hajdu et al. [4] conducted a survey among farms in Russia and Kazakhstan and found notable positive effects of local labour sourcing, insecure land use conditions and farm size (in terms of land area) on farms' CSR engagement. Also individually owned farms, contrary to corporate farms, tend to be more CSR affine. On the other hand, weak CSR engagement was found among the farms affiliated with agroholdings [4].

The results of a survey of German agribusiness companies was conducted by Heyder and Theuvsen [5]. They used partial least squares method for evaluation of a structural equation model and found that CSR efforts in their sample is significantly influenced by the competitive strategy of a company, number of employees (e.g. the size of a company), and the degree of altruism in the company. Setiawan et al. [10] examined the determinants of environmental performance on the firms listed in the agriculture sector of the Indonesian Stock Exchange. The results of multiple regressions showed that return on equity negatively affects environmental performance while firm age, size, and leverage had a positive effect on environmental performance. Large-scale farms also concerned Jelínek et al. [7].

2 Data and Methods

The aim of the paper is to assess what factors affect how important is for an agricultural holding the support of the stakeholders. A primary survey took place among representants (mainly managers / directors) of Czech agricultural firms in 2018 within internal research project IVP 1117/2018. We received answers on the questionnaire from 133 respondents.

2.1 Importance of support of stakeholders

The respondents were assessing how important was a support of the stakeholders – the general public and local communities – for their company on a scale: very important, rather important, rather unimportant, fully unimportant. This variable was considered as a proxy of CSR level. We assumed that if the support of public and local communities was important that the level of CRS of a firm is higher. The scope of CSR activities depends on certain factors – features of the manager and characteristics of the firm. The managers were asked about their sex, education and age. We can assume that women are more caring and therefore support of stakeholders would be more important for them. Similarly, people with higher education may perceive CRS as more important. Also, age of a manager can have an influence, but it is not clear whether older or younger managers would implement CRS more. Average age was 50.8 years. Two dummy variables were included in the model – lower and upper age intervals. Middle age interval was not included, and other categories were compared to it. Analyzed characteristics of the firm were their legal form and number of employees. Many researches proclaimed (e.g. Bavorová et al. [1]) that cooperatives stress the CRS more than other legal forms. Therefore, joint-stock companies and limited liability companies were merged together and compared to the cooperatives. Also, the larger is the firm (in terms of the number of employees), the more probable that it will carry out CRS activities.

Also, the business strategy was examined. Particularly, we asked the managers what the main target of the business was. The reason was that social responsibility can influence the competitiveness of a company in a positive way in a long run but requires financial resources. "Fulfilling social responsibility will never be free lunch, but in the long run, social responsibility has positive effects on enterprise competitiveness" [14].

We included this variable among explanatory variables to analyses whether firms who stated that their target is sustainable and long-term profit making are aware that CRS is part of the process of achieving this goal. Also, we wanted to see whether if the main target of a company is the improving existing products and developing new products; satisfaction of customer needs; good name of the company; and environmental protection are significant determinants of the relation toward stakeholders. Four dummy variables were included in the model, each one for each business target. Only target of environmental protection was not included, and other categories were compared to it. Description of a sample and the character of the variables is given below in Table 1.

Variable	**Descriptive statistics**		**Inclusion in a model**
importance of support of stakeholders (explained variable)	12.0%	very important	0
	44.4%	rather important	1
	41.4%	rather unimportant	2
	2.3%	fully unimportant	3
sex of the respondent	82.0%	male	0
	18.0%	female	1
education of the respondent	33.8%	high school, graduated	0
	66.2%	university	1
age of the respondent (21–40 years)	21.1%	21–40 years	1, 0
age of the respondent (41–60 years)	63.9%	41–60 years	1, 0
age of the respondent (61–80 years)	15.0%	61–80 years	not included (baseline)
legal form	42.9% + 19.5%	Joint-stock company + Limited liability company	0
	37.6%	Cooperative	1
number of employees	4	Minimum	real value
	40	Median	
	47.9	Average	
	350	Maximum	
business target (profit)	64.7%	sustainable and long-term profit making	1, 0
business target (customer)	3.8%	satisfaction of customer needs	1, 0
business target (product)	2.3%	improving existing products and developing new products	1, 0
business target (name)	27.1%	good name of the company	1, 0
business target (environment)	2.3%	environmental protection	not included (baseline)

Table 1 Statistical description of the variables; own elaboration

2.2 Methods

The variables were included in an ordered logistic regression model (ologit). This type of model is used when the explained variable is an ordered categorical outcome. In our case, it is an answer on the question: “How important is the support of the stakeholders (general public and local communities) by your company?”. There are 4 answers: 0 – very important, 1 – rather important, 2 – rather unimportant, 3 – fully unimportant.

Explained variables sex, education, and legal form are of dummy character (i.e. take values of 0, 1). Legal form takes value 0 when it is joint-stock company or limited liability company and value 1 when it is cooperative. Only number of employees is real number. There are three age categories, so only two of them (21–40 years and 41–60 years) were included in a model to avoid multicollinearity. A variable business target takes five values: (1) sustainable and long-term profit making, (2) satisfaction of customer needs, (3) improving existing products and developing new products, (4) good name of the company, and (5) environmental protection. Therefore, four dummy variables (except the target environment) were included into the model.

Unlike in multinomial logistic regression, in ordered logistic regression is assumed that there is an order to the categories of the outcome variable. The model examines the log-odds (a ratio of expected number of successes to each failure) that are computed as (1). There are 4 answers to the question, so there are 3 logarithms of the odds. When the answer is “very important”, than the outcome is 0 and probability is calculated as logarithm of probability of outcome 0 (very important) divided by the sum of probability of other options (outcome 1 – rather important, outcome 2 – rather unimportant, outcome 3 – fully unimportant), etc. The data has to meet the proportional odds assumption – “the relationship between each pair of outcome groups is the same” [11].

Answer	Probability calculation, outcome
very important	$\ln\left(\dfrac{p_0}{p_1+p_2+p_3}\right),\quad 0$
very important + rather important	$\ln\left(\dfrac{p_0+p_1}{p_2+p_3}\right),\quad 1$
very important + rather important + rather unimportant	$\ln\left(\dfrac{p_0+p_1+p_2}{p_3}\right),\quad 2$

(1)

where p is probability function for y, i.e. p_1 is probability of outcome 0 – very important, p_2 is probability of outcome 1 – rather important, p_3 is probability of 2 – rather unimportant and p_4 of 3 – fully unimportant. y is the explained variable taking value 0, 1, 2, 3. The functional relation between explained variable and explanatory variable is given as (2).

$$\mathbf{y}^* = \mathbf{x}^T\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2}$$

where $\mathbf{y}^*$ is a vector of the dependent variable (the level of agreement on question "How important is the support of the stakeholders (general public and local communities) by your company?"). Its size is 133 x 1. $\mathbf{x}$ is a matrix of independent variables (listed in Table 1) with size 133 x 10 (or 10 x 133 after transposition), $\boldsymbol{\beta}$ is the vector of regression coefficients (size 10 x 1) that will be estimated and $\boldsymbol{\varepsilon}$ is the vector of error terms with size 133 x 1.

We observed the categories of response on the question where the parameters are the externally imposed endpoints of this categories. Then the ordered logit technique will use the observations on y, which are a form of censored data on y^*, to fit the parameter vector (3).

$$y = \begin{cases} 0 & \text{if} \quad y^* \le \mu_1, \\ 1 & \text{if} \quad \mu_1 < y^* \le \mu_2, \\ 2 & \text{if} \quad \mu_2 < y^* \le \mu_3, \\ 3 & \text{if} \quad \mu_3 < y^* \end{cases} \tag{3}$$

Model was estimated in Stata 15.1 using maximum likelihood method. Then, the statistical significance of each parameter was tested using one-sample z-test (we have large sample, so we can assume normality and use tabled value from normal distribution rather than from Student t distribution). The model as a whole was tested using likelihood ratio (LR) χ^2 test.

3 Results and Discussion

The support of stakeholders was rather important for 44,4% of respondents, while for 41,4% it was rather unimportant. Majority of representatives of the agricultural holdings were men (82,0%) with university degree (66,2%) in age between 41–60 years. Majority of farms had legal form of joint stock company (42,9%). There were 37,6% of cooperatives. Average firm had almost 48 employees. Regarding the main target of the business, mostly the respondents answered that it is sustainable and long-term profit making (64.7%). Then it was important to maintain a good name of the company (27.1%). Other targets were negligible (3.8% answered that satisfaction of customer needs is important for them and 2.3% that improving existing products and developing new products or that environmental protection).

The variables listed in Table 1 were included into ordered logistic regression. Results of the estimation are displayed at Table 2. LR test of the model had critical value $\chi^{2\,[10]} = 9.52$, so the p-value was too high (0.48) and exceeded the 0.05 level of significance, so the model as a whole is statistically insignificant. There were only 133 observations, so only 2 parameters were statistically significant – legal form of a company and target "improving existing products and developing new products" – on the level of significance 0.1.

Coefficient for legal form is interpreted that the unit change in the legal form (i.e. when it is a cooperative contra the situation when the firm is joint-stock company or limited liability company) results in decrease of log of the odds by 0.737. It is more convenient to interpret the odds ratio which is and exponent of the coefficient (e^β). The odds of ticking particular answer on the question about support of the stakeholders for cooperatives is 0.48 times lower than when the firm has different legal form. More concretely one unit increase in legal form (e.g. when it is cooperative instead of the joint-stock company or limited liability company), the odds of “very important” versus the “rather important”, “rather unimportant” and “fully unimportant” answers on the question about the stakeholders' support are 0.48 times lower, given the other variables are held constant in the model. If the legal form of the farm is a cooperative, the log odds that the support general public and local stakeholders is very important is lower than in other categories. When the objective improving existing products and developing new products is the main target of the business, the odds that the support of stakeholders is “very important” (over “rather important” or “rather” or “fully unimportant”) is 25.4 times higher.

Our results can be compared with Bavorová et al. [1]. They found a positive relationship between the support of social and technical rural infrastructure and farm size in Altai Kray in Russia. Contrary to this, Udayaskar [12] found that the relation between firm size and CSR participation is U-shaped as very small and very large firms are equally motivated to participate in CSR while medium-sized firms are the least motivated. While Heyder and Theuvsen [5] found out that number of employees significantly influenced the CSR efforts of the German agribusiness companies in their sample, our results shows opposite. In our case, the number of employees as a proxy for the size of a firm was not statistically significant determinant of the importance of stakeholders’ support.

In our research, farm manager’s age was found to be statistically insignificant (all age categories). Also, Bavorová et al. [1] who used the age of a manager as a proxy for socialist experience found out that it does not positively affect social responsibility. Regarding the legal form of the agricultural holding, Bavorová et al. [1] concluded that cooperative legal structure increases social responsibility. Because Heyder and Theuvsen [5] suggest, based on the findings of their model, that CSR efforts significantly enhance corporate reputation, we supposed that the business target “maintain a good name of a company” would be linked to CRS. However, it was statistically insignificant.

	Odds ratio (e^β)	**Std. err.**	**Coef. (β)**	**Std. err.**	**Z-value**	**P-value**
sex	1.086	0.493	0.082	0.454	0.180	0.856
education	0.640	0.242	-0.447	0.378	-1.180	0.237
age (21–40 years)	1.046	0.610	0.045	0.583	0.080	0.939
age (41–60 years)	1.042	0.496	0.041	0.476	0.090	0.932
legal form	*0.478*	*0.180*	*-0.737*	*0.376*	*-1.960*	*0.050*
number of employees	1.024	0.284	0.024	0.278	0.090	0.932
target (profit)	3.356	4.080	1.211	1.216	1.000	0.319
target (customer)	1.684	2.362	0.521	1.402	0.370	0.710
target (product)	*25.238*	*47.623*	*3.228*	*1.887*	*1.710*	*0.087*
target (name)	3.753	4.712	1.323	1.256	1.050	0.292

Table 2 Estimated parameters of ordered logistic regression; own elaboration

4 Conclusion

The aim of the paper was to assess what factors affects how important is for an agricultural holding the supports of the stakeholders. Majority out of 133 respondents considered the support of general public and local communities as rather important (44,4%) or rather unimportant (41,4%). As factors that can influence it were included characteristics of the agricultural holding and of its representant (manager) in ordered logistic regression model. We found out that only legal form of a firm and target of improving and developing of the product were statistically significant determinants of whether the support of stakeholders is important. When the company is a cooperative, the odds for “very important” versus the “rather important”, “rather” or “fully unimportant” stakeholders' support are 0.48 times lower. When the business target is enhancing the product the odds that support is “very important” (over “rather important” or “rather” or “fully unimportant”) is 25.4 times higher.

On the other hand, sex, education and age of the farm’s representant were not statistically significant. It seems that the personal features of the person that is the head of the company does not influence the level of CRS. Maybe value orientation (the degree of altruism) is more important. One reason for that the results are not significant in many cases is low number of observations. The other is that we examined the importance of general public and local communities for the managers of agricultural holdings. We did not distinguish between those two groups. However, Bavorová et al. [1] found out that “local stakeholders (community, state) rather than international ones are important”. So, the choice of a proxy of social responsibility could also influence the results. The challenge for future research is to examine also financial performance of the firms (as did for example Heyder and Theuvsen [5]) and its relation towards the scope of stakeholders' support.

Acknowledgements

This paper was created within the project of the Ministry of Agriculture of the Czech Republic, institutional support MZE-RO0918, Internal Research Project of Institute of Agricultural Economics and Information no. 1117/2022.

References

[1] Bavorová, M., Bednaříková, Z., Ponkina E.V. & Visser O. (2021). Agribusiness social responsibility in emerging economies: effects of legal structure, economic performance and managers' motivations. *Journal of Cleaner Production*, 289, 125157. https://doi.org/10.1016/j.jclepro.2020.125157

[2] Carroll, A. (1991). The Pyramid of Corporate Social Responsibility: Toward the moral management of organizational stakeholders. *Business Horizons*, 34(4), 39–48. https://doi.org/10.1016/0007-6813(91)90005-g

[3] de Olde, E. M., & Valentinov, V. (2019). The moral complexity of agriculture: A challenge for corporate social responsibility. *Journal of Agricultural and Environmental Ethics*, 32(3), 413–430. https://doi.org/10.1007/s10806-019-09782-3

[4] Hajdu, A., Gagalyuk, T., Bukin, E. & Petrick, M. (2021). Determinants of corporate social responsibility among farms in Russia and Kazakhstan: a multilevel approach using survey data. *International Food and Agribusiness Management Review*, 24(4), 697–716. https://doi.org/10.22434/IFAMR2020.0197.

[5] Heyder, M. & Theuvsen, L. (2013). Determinants and Effects of Corporate Social Responsibility in German Agribusiness: A PLS Model. *Agribusiness*, 28(4), 400–420, https://doi.org/10.1002/agr.21305

[6] Hornungová, J. & Pavláková Dočekalová, M. (2015). Social responsibility and its importance to the performance management of business companies: Actual situation in the Czech Republic. In Pastuszkova, E. et al. (Eds.), *Proceedings of the 7th International Scientific Conference on Finance and Performance of Firms in Science, Education and Practice* (pp. 416–430). Zlín: Tomas Bata Univ, Fac Management & Econ

[7] Jelínek, L., Medonos, T., Foltýn, I., Humpál, J. & Boyinová, S. (2018). Social and Economic Consequences of Direct Payments Capping and Redistribution on the Czech Large-scale Farms, 162nd EAAE Seminar The evaluation of new CAP instruments: Lessons learned and the road ahead, 1–15.

[8] Luhmann, H., Theuvsen, L. (2017). Corporate Social Responsibility: Exploring a Framework for the Agribusiness Sector. *Journal of Agricultural Environmental Ethics*, 30, 241–253. https://doi.org/10.1007/s10806- 017-9665-8

[9] Mazur-Wierzbicka, E. (2015). The application of Corporate Social Responsibility in European Agriculture. *Miscellanea Geographica*, 19(1), 19–23. https://doi.org/10.1515/mgrsd-2015-0001

[10] Setiawan, D., Asrihapsari, A., Maisaroh, S. and Widawati, M. W. (2021). Determinants of environmental performance: Evidence from the agriculture industry in Indonesian Stock Exchange. *The 8th International Conference on Sustainable Agriculture and Environment IOP Conf. Series: Earth and Environmental Science 905* (012006). Surakarta, Indonesia: IOP Publishing. https://doi.org/10.1088/1755-1315/905/1/012006

[11] UCLA (2006). Ordered Logistic Regression | Stata Data Analysis Examples. [on-line] UCLA. Retrieved March 23, 2022, from https://stats.oarc.ucla.edu/stata/dae/ordered-logistic-regression

[12] Udayasankar (2008). Corporate Social Responsibility and Firm Size. *Journal of Business Ethics*, 83(2), 167– 175.

[13] Vávra, J. et al. (2019). Socio-economic context of soil erosion: A comparative local stakeholders' case study from traditional agricultural region in the Czech Republic. *Land Use Policy*, 16(84), 127-137.

[14] Wang, D. M. & Yang, L. (2013). An Economic Analysis of Interaction between Social Responsibility and Competitiveness in Agribusiness. In Hu, J. (Eds.), *Proceedings of the 3rd International Conference on Applied Social Science*, (pp. 574). Taipei: Hong Kong Educ Soc.

[15] Wicaksono, A. P. & Setiawan, D. (2022). Water disclosure in the agriculture industry: Does stakeholder influence matter? *Journal of Cleaner Production*, 18, 337, 130605.

[16] Wilhelm, J. A., Smith, R. G., Jolejole-Foreman, M. C. & Hurley, S. Resident and stakeholder perceptions of ecosystem services associated with agricultural landscapes in New Hampshire. *Ecosystem Services*, 45, 101153. https://doi.org/10.1016/j.ecoser.2020.101153.

[17] Zhirnov, A. V., Lapshin, Y. A. & Lapshina, G. V. (2013). Experience of Implementation of Social Responsibility in Agribusiness in Ulyanovsk Region. *International Technical and Economic Journal* 4, 23–27.

The Effect of Business Clustering on Scale Efficiency

Eva Štichhauerová[1], Miroslav Žižka[2]

Abstract. This paper examines the impact of an organised and natural cluster on the scale efficiency of firms in the engineering industry. Three groups of firms were included in the research. The first group consisted of member firms of the Czech Machinery Cluster, representing one of the oldest organised clusters in the Czech Republic. In the second group, firms of the natural cluster were represented. These companies operate in the same industry and the same region as the Czech Machinery Cluster but are not members of it. The third control group represents machinery firms from other regions of the Czech Republic. Scale efficiency was examined using CRS and VRS DEA models for the period 2009 to 2019. The results showed that firms achieved the best scale efficiency in the cluster organisation. Member firms of this organised cluster operate under constant or increasing returns to scale. On the contrary, there was no significant difference in scale efficiency between firms in the natural cluster and other firms. Firms in these two clusters predominantly operate under conditions of decreasing returns to scale, i.e. they do not use their inputs efficiently enough.

Keywords: Data envelopment analysis, returns to scale, scale efficiency, technical efficiency, cluster organisation, natural cluster.

JEL Classification: C61, L25, L64.
AMS Classification: 90B90, 90C90

1 Introduction

Firms join together in clusters because such cooperation brings them a number of positive externalities. The geographical proximity of firms in clusters allows for the sharing of R&D results, more effortless transfer of tacit knowledge, access to skilled labour or joint purchasing of specific inputs. Important microeconomic externalities also include the increase in returns to scale resulting from knowledge spillovers [9].

The aim of the paper is to find out whether the participation of companies in one of the oldest Czech clusters had a positive impact on their scale efficiency (SE). The Czech Machinery Cluster, which was chosen for the research, was founded in 2003. That is, even before the beginning of the cluster support programme, which has been implemented since 2004 within the framework of the operational programme Industry and Enterprise. The establishment of the Czech Machinery Cluster was the result of a cluster initiative. Historically, however, a natural engineering cluster existed on the territory of the Moravian-Silesian Region much earlier. A natural cluster can be characterised as a network of companies within engineering and related industries or within educational and re- search institutions. After 2003, some of these organisations joined together to form a cluster organisation. Others operate independently in the region and only cooperate with the firms in the cluster on the basis of supplier-custo- mer relations. The term cluster can be understood in two ways. Firstly, as a cluster organisation, which is a legal entity that manages and promotes the joint activities of its members. Or secondly, as a natural cluster that exists independently of the cluster organisation and represents an informal network of cooperating institutions in a given industry. In our paper, we examine both two variants of the cluster. In addition, there are other firms operating in the sector (engineering) in other regions that are outside the reach of the positive externalities of both types of clusters.

Logically, it can be assumed that positive externalities should manifest themselves most strongly in firms in a cluster organisation, where the interaction between firms and institutions is the closest. However, firms located in the same region should also benefit from this environment, albeit with a smaller effect. The weakest SE should be observed for other geographically scattered firms from different regions.

[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

Returns to scale are a characteristic of the firm's production function in the long run. They indicate how the firm's output will change if all inputs are increased in exactly the same proportion. If a proportional increase in all inputs leads to a supra-proportional increase in output, this is increasing returns to scale (IRS). If a proportional increase in all inputs causes an increase in output in the same proportion, we are talking about constant returns to scale (CRS). Finally, if an increase in all inputs leads to a proportionally smaller increase in output, we are talking about decreasing returns to scale (DRS). Mathematically, if we multiply all inputs by a constant $k > 1$, the above variations of returns to scale can be expressed by relations (1) to (3) [4].

$$f(k\boldsymbol{x}) > kf(x) \rightarrow IRS \tag{1}$$

$$f(k\boldsymbol{x}) = kf(x) \rightarrow CRS \tag{2}$$

$$f(k\boldsymbol{x}) < kf(x) \rightarrow DRS \tag{3}$$

SE measures the extent to which a firm deviates from its optimal scale size, that is, from the point on the cost curve where constant returns to scale exist [7]. Approximately SE can be found as the ratio of the average product of the firm lying on the efficient frontier in VRS conditions and the average product at the technically optimal scale point in CRS conditions [8]. If the firm is scale inefficient, at the same time, the tendency of average production costs to increase can be observed [7]. The efficiency of a firm can therefore be improved by changing the scale of its activities. This means, for example, for a given combination of inputs, changing the scale of output. A firm achieves its optimal output scale if it operates under CRS conditions [1]. On the other hand, a company in an IRS situation is too small. However, the cause of the IRS situation may also be the indivisibility of some inputs and the high specialisation of workers. Conversely, a firm in a DRS environment achieves an unnecessarily large scale of output relative to its optimal size. This may result from low management efficiency and cumbersome communication between top management and regular employees [5]. A comparison of SE and pure technical efficiency (PTE) allows the identification of the primary source of inefficiency of the firm. It may be a technical problem related to the quantity and combination of inputs and outputs or a problem in the overall operational concept of the firm. Based on the SE analysis, it is then possible to determine the necessary extent of scale change [6].

The following procedure based on the CRS and VRS models can be used to determine the type of returns to scale [11]. Assume that λ_j is the weight of the peer unit in the assessment of unit 0. We calculate the overall technical efficiency (OTE) score θ^* according to the CRS model (4) and also determine the optimal values of λ_j^*. We then determine the pure technical efficiency (PTE) score θ^* according to the VRS model. The two models differ only in the sum condition (5) for the VRS model. Relationship (4) applies to the input-oriented model used in our research.

$$\begin{gathered}\theta^* = \min \theta \\ \text{subject to} \\ \sum_{j=1}^{n} \lambda_j x_{ij} \leq \theta x_{i0} \quad i = 1, 2, \dots, m; \\ \sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{r0} \quad r = 1, 2, \dots, s; \\ \lambda_j \geq 0 \quad j = 1, 2, \dots n\end{gathered} \tag{4}$$

$$\sum_{j=1}^{n} \lambda_j = 1 \tag{5}$$

where: n … number of units assessed,

x_{i0} … i-th input of the assessed unit 0,

y_{r0} ... j-th output of the assessed unit 0.

In the next step, we can determine the type of returns to scale.

The ratio of OTE according to the CRS model and PTE according to the VRS model gives the SE; see the ratio (6) [1]. If the SE is equal to one, the unit is in the optimal size range. If the SE is less than one, the unit is too large or too small, or scale-inefficient [2]. If a unit is efficient under the CRS model, it is certainly efficient under the VRS model. If a unit is efficient under the VRS model, then all variants of returns to scale are possible (CRS, IRS, DRS)

$$SE = \frac{\theta^*{}_{CRS}}{\theta^*{}_{VRS}} \tag{6}$$

Units that are in the CRS area must meet the assumption of equal efficiency scores under the CRS and VRS models, regardless of the value of the sum (7). If the efficiency score values are different and condition (8) holds, the unit is in the IRS area. If the scores are different and condition (9) holds, the unit is in the DRS area [11].

$$\sum_{j}^{n} \lambda_j^* \tag{7}$$

$$\theta^*{}_{CRS} \neq \theta^*{}_{VRS} \; and \sum_{j}^{n} \lambda_j^* < 1 \rightarrow IRS \tag{8}$$

$$\theta^*{}_{CRS} \neq \theta^*{}_{VRS} \; and \sum_{j}^{n} \lambda_j^* > 1 \rightarrow DRS \tag{9}$$

Alternatively, the type of returns to scale can be determined as a proportion of technical efficiency according to CRS and NIRS (non-increasing returns to scale) DEA models. If relationship (10) holds, then a DMU operates under IRS. If relation (11) holds, a DMU operates under DRS [2].

$$\frac{\theta^*{}_{CRS}}{\theta^*{}_{NIRS}} = 1 \tag{10}$$

$$\frac{\theta^*{}_{CRS}}{\theta^*{}_{NIRS}} < 1 \tag{11}$$

3 Data and methodology

The research was carried out in firms operating in NACE 251 (Manufacture of structural metal products) and NACE 28 (Manufacture of machinery and equipment). Firms in these industries were divided into three groups: C, N, and O. Group C consists of 4 firms - members of the Czech Machinery Cluster. Group N consists of 120 firms in the natural cluster. These firms operate in the Moravian-Silesian, Olomouc and South Moravian regions, i.e. in the same area as the members of the Czech Machinery Cluster. However, they are not members of an institutionalised cluster. The last group includes 118 firms from other regions of the Czech Republic. The research included all firms from the above industries for which it was possible to obtain financial statements for the period 2009-2019. The source of accounting data was the MagnusWeb database [3]. Each firm must be part of only one group. The division of firms into three groups may affect the significance of the analysis results, as the sample size is reduced. We attempted to reduce this limitation by using panel data. However, the aim of the analysis was to determine whether belonging to one of the above groups has an effect on overall, pure or scale efficiencies.

The OTE scores for the above companies were calculated using the DEA CRS model and the PTE scores using the DEA VRS model. Both models were input-oriented; see relationships (4) and (5). MaxDEA 7 Ultra software was used as the computational tool. The model inputs were equity and liabilities. The outputs were revenues from their own products with services and economic value added. These four variables were obtained from the MagnusWeb database [3]. According to relation (6), the SE value for each firm was calculated. Subsequently, the average OTE and PTE scores and average SE values were determined for all groups of firms. The differences between the average scores of OTE, PTE, and SE values were tested using the Games-Howell multiple range test with Statgraphics Centurion 18 software.

In order to determine the type of returns to scale, dual models were solved in the next step, and lambda values were determined. Following the algorithm described in chapter 2 (see relationships 7 to 9), it was determined whether the firm operates in CRS, DRS or IRS conditions. A chi-square test of independence of categorical data combined with a mosaic plot was used to examine the dependence of the type of returns to scale on the firm's group affiliation. All tests were performed at a significance level of alpha 5%.

4 Research Results

Table 1 shows the trend in average OTE scores (according to the CRS model) in individual years for each group of firms in relation to their membership or non-membership in clusters. It is clear from the table that the OTE of the cluster member firms was better than that of the other two groups over the whole period. However, the average OTE of the group of firms in the natural cluster was not better than that of the other firms outside the cluster for most of the period. The only exception was the years 2010 to 2012.

Through the application of the Games-Howel test, the statistical significance of the positive difference in the OTE of the member firms of the cluster organisation compared to the firms in the natural cluster was verified for the whole period under study. The validity of the relationship C > N is indicated by a solid circle. However, a significant relationship C > O was only demonstrated in 2014 (the sign of the solid square).

Group	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Mean
C	0.57	0.67	0.70	0.76	0.79	0.85	0.68	0.74	0.63	0.77	0.78	0.72
N	0.30	0.41	0.40	0.38	0.39	0.35	0.42	0.37	0.36	0.40	0.35	0.38
O	0.36	0.35	0.36	0.36	0.44	0.46	0.49	0.48	0.51	0.46	0.48	0.43
G-H	●	●	●	●	●	●■	●	●	●	●	●	●

Table 1 Average OTE scores by a group of companies by year (CSR Model)

Similarly, table 2 shows the average PTE scores of the three groups of firms (according to the VRS model). To a large extent, the results are analogous to the CRS model. Throughout the period, firms in the cluster organisation achieved the highest efficiency.

Group	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Mean
C	0.71	0.75	0.71	0.84	0.94	0.99	0.89	0.91	0.90	0.93	0.94	0.86
N	0.50	0.57	0.56	0.59	0.62	0.55	0.59	0.57	0.52	0.58	0.62	0.57
O	0.52	0.53	0.52	0.53	0.56	0.54	0.61	0.60	0.63	0.56	0.54	0.56
G-H	●	●	●	●	●■	●■	●■	●■	●	●■	●■	.

Table 2 Average PTE scores by a group of companies by year (VRS Model)

Table 3 shows the trend in average SE across firm groups. Over the whole period, the value of SE was the best in the group of cluster members. Statistical significance of the positive difference of the SE level in the group of member firms compared to firms in the natural cluster was demonstrated especially in the first half of the period under study (2009-2014). Compared to other firms outside the cluster, a significant difference was proven only in 2010-2011. However, the values of SE in the group of cluster organisation firms fluctuated quite significantly and, at the end of the period, were below the values in the group of other firms (statistically significant difference, empty square sign).

Group	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Mean
C	0.82	0.91	0.97	0.92	0.82	0.85	0.74	0.79	0.68	0.81	0.82	0.83
N	0.66	0.76	0.77	0.72	0.70	0.70	0.69	0.76	0.80	0.75	0.68	0.72
O	0.64	0.62	0.68	0.68	0.73	0.81	0.79	0.83	0.81	0.84	0.90	0.75
G-H	●	●■	●■	●	●	●	.	.	.		□	.

Table 3 Average SE scores by a group of companies by year

The following table 4 shows the shares of firms in each group by type of returns to scale. In the group of cluster organisation firms, half of the firms were in the optimal production scale at the end of the period. The other half were operating under conditions of increasing returns to scale. It can be concluded that there was some slight improvement in SE in this group. However, the number of firms in this group is small. Contingency tables were constructed for all periods, where the group was a row variable, and the range return type was a column variable.

The chi-square test's assumption was not met (some expected cell counts in all contingency tables were less than 5 because of the low number of firms from the cluster organisation). Therefore, the evaluation each year was first performed on the basis of a mosaic plot that included all three groups of firms. It was clear that the type of returns to scale depended on the group of firms. In all years studied, the structure of firms from the cluster organisation differed significantly from those in the natural cluster and the other firms in terms of the types of returns to scale.

In the group of firms in the natural cluster and the group of other firms, the share of firms with CRS was in the minority and practically constant (about 7%) throughout the whole period under review. At the same time, the majority of firms in both groups were enterprises operating under DRS conditions. The differences between these two groups were then shown to be statistically significant in 2009, 2012, 2014-2015 and 2018. Then, the share of enterprises operating in DRS was significantly higher in the natural cluster compared to the share in the group of other enterprises. Logically, on the contrary, the percentage of firms operating with IRS was significantly lower in the natural cluster. Otherwise, there was no significant shift in SE in either group over the whole period.

Group	Share	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**Mean**
	% CRS	0.25	0.25	0.50	0.50	0.50	0.50	0.25	0.25	0.25	0.50	0.50	0.39
C	% DRS	0.00	0.00	0.00	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02
	% IRS	0.75	0.75	0.50	0.25	0.50	0.50	0.75	0.75	0.75	0.50	0.50	0.59
	% CRS	0.06	0.10	0.08	0.08	0.07	0.05	0.06	0.05	0.03	0.08	0.05	0.07
N	% DRS	0.74	0.65	0.58	0.74	0.78	0.75	0.80	0.72	0.83	0.80	0.80	0.74
	% IRS	0.20	0.25	0.34	0.18	0.16	0.20	0.14	0.23	0.14	0.12	0.15	0.19
	% CRS	0.09	0.06	0.07	0.05	0.09	0.08	0.10	0.06	0.06	0.07	0.08	0.07
O	% DRS	0.58	0.78	0.60	0.58	0.70	0.58	0.59	0.64	0.86	0.62	0.72	0.66
	% IRS	0.32	0.16	0.33	0.36	0.20	0.34	0.31	0.30	0.08	0.31	0.19	0.27

Table 4 Share of firms by type of returns to scale in each group of companies

5 Conclusions

The research intended to determine the effect of targeted and natural clustering of engineering firms on their SE compared to non-clustered firms. The research results showed that members of the organised cluster achieved on average higher PTE and OTE compared to firms in the natural cluster. In some years, they also had significantly better efficiency compared to non-clustered firms. However, the average SE of the members of the organised cluster was higher only in the first half of the period.

On the contrary, the assumption of higher SE of engineering firms in the natural cluster compared to other non-clustered firms was not confirmed. This means that the positive externalities arising from the natural location and collaboration of firms in a region are not strong enough to be reflected in the SE of firms. Engineering firms in the natural cluster did not even achieve higher levels of technical efficiency compared to firms in other regions.

Most firms in the natural cluster, like most other firms, are in a condition of decreasing returns to scale. This means that an increase in output requires a higher than directly proportional growth in all inputs and logically also a higher than directly proportional growth in costs. The curve of average costs is rising. These firms have thus, to some extent, already exhausted their further growth opportunities.

On the other hand, firms in a cluster organisation operate either under constant returns to scale or increasing returns to scale. In the former case, the combination of inputs is thus optimally exploited. In the latter case, output increases faster than the equivalent growth in inputs. The curve of average costs falls. These firms can thus use their inputs efficiently to grow further. From this perspective, the cluster initiative aimed at establishing a cluster organisation appears to be successful.

In conclusion, the existence of an institutionalised cluster organisation brings higher SE to its member firms. However, a similar assumption was no longer confirmed for firms in a natural cluster. This may also be due to the characteristics of the engineering industries studied. The industries are not highly specialised. This means that the production characteristics of firms in the natural cluster and in other regions may not be fundamentally different. On the contrary, closer cooperation between firms in a cluster organisation and the resulting positive externalities may be the factor that increases both technical and scale efficiencies of the participating firms.

In terms of the limitations of the research, it is important to note the limited sample of firms, which was negatively affected by the poor availability of financial statements. Further research can be oriented towards other industries where organised and natural clusters also exist.

Acknowledgements

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

References

[1] Anang, B. T. (2021). Assessing technical and scale efficiency of groundnut production in Tolon district of northern Ghana: A nonparametric approach. *Journal of Agriculture and Food Research*, 4, 100149. https://doi.org/10.1016/j.jafr.2021.100149.

[2] Bielik, P. & Rajčániová, M. (2004). Scale efficiency of agricultural enterprises in Slovakia. *Agricultural Economics*, 50, 331-335.

[3] Bisnode: *Magnusweb: Komplexní informace o firmách v ČR a SR* [online]. Bisnode ČR, Prague, 2022 [cit. 2022-04-22]. Available at: https://magnusweb.bisnode.cz.

[4] Coelli, T. J., Prasada Rao, D. S., O´Donell, Ch. J. & Battese, G. E. (2005). *An Introduction to Efficiency and Productivity Analysis*. New York: Springer Science.

[5] Kirigia, J. M. & Asbu, E. Z. (2013). Technical and scale efficiency of public community hospitals in Eritrea: an exploratory study. *Health Economics Review*, 3, 1-16.

[6] Lee, Ch. Ch. (2009). Analysis of overall technical efficiency, pure technical efficiency and scale efficiency in the medium-sized audit firms. *Expert Systems with Applications*, 36, 11156-11171. https://doi.org/10.1016/j.eswa.2009.02.092.

[7] Majumdar, S. K. & Chang, H. H. (1996). Scale Efficiencies in US Telecommunications: An Empirical Investigation. *Managerial and Decision Economics*, 17, 303-318.

[8] Ngueyen, T. T. T., Le, H. H., Ho, T. M. H., Dogot, T., Burny, P., Bui, T. N. & Lebailly, P. (2020). Efficiency Analysis of the Progress of Orange Farms in Tuyen Quang Province, Vietnam towards Sustainable Development. *Sustainability*, 12, 3170. https://doi.org/10.3390/su12083170.

[9] Potter, A. & Watts, H. D. (2011). Evolutionary agglomeration theory: Increasing returns, diminishing returns, and the industry life cycle. *Journal of Economic Geography*, 11, 417-455. https://doi.org/10/cj7xkc.

[10] Seiford, L. M. & Zhu, J. (1999). An investigation of returns to scale in data envelopment analysis. *Omega, International Journal of Management Science*, 27, 1-11.

[11] 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.

Design of Optimization Model for Network Coordination of Public Transport Connections with Periodically Alternating Headway

Dušan Teichmann[1], Denisa Mocková[2], Michal Dorda[3], Pavel Edvard Vančura[4], Ivana Olivková[5], Vojtěch Graf [6]

Abstract. The basic aspect of public transport in terms of its use must be its attractiveness for passengers. If it is not possible to provide a direct connection for each passenger, discomfort must be minimized for passengers who will have to change during the transport process. Their discomfort can be reduced by shortening the time they must wait for the consequent connection. Reducing the waiting time during transfers can be achieved mainly by a higher level of network coordination of connections. The given problem can be solved as an optimization task. The motivation for the article is the specific nature of the timing of connections in the Prague public transport system, where coordination between subway and bus lines is required on weekends. For the bus lines, the so-called periodically alternating headway is applied on weekends. The connections are run at 7- and 8-minute headways, and this sequence is repeated regularly throughout the weekend. The optimization criterion is the total time loss of transferring passengers in all transfer nodes. The value of the total time loss is minimized. The computational experiment was realized in the Xpress-IVE optimization software.

Keywords: Public transport, network coordination, periodically alternating headway, optimization

JEL Classification: C44
AMS Classification: 90C11

1 Introduction - motivation to solve the problem

Time coordination of connections in transfer nodes is one of the basic tasks that public transport operators must deal with. If no transfer links are ensured between the connections in the transfer nodes, then transferring passengers incur time losses. A public transport system, with great time losses of transferring passengers, loses its conmetitiveness as it ceases to be attractive to passengers.

In public transport systems with many transfers, network coordination of connections is important. By network coordination we mean coordination between connections which is performed simultaneously in multiple transfer nodes. Network coordination is a complicated optimization problem that has many specifics. For example, it is necessary to preserve time continuity of connections of the same lines. It is therefore necessary to respect the driving times of the connections between the individual transfer nodes. Further it is often necessary to obey defined headways between the connections of the individual lines. These headways may be different for the individual lines. In addition, on some lines, there may be so-called periodically alternating headways between the connections. This means that a sequence of several headways of different lengths may occur between the connections of a line – they are regularly repeated, for example the sequence 7, 8, 7, 8 min, etc. Each additional operation specificity must be incorporated in the optimization approach to obtain relevant results.

[1] CTU in Prague, Faculty of Transportation Sciences, Department of Logistics and Management of Transport, Prague 1, Konviktská 20, 110 00, teichdus@fd.cvut.cz.
[2] CTU in Prague, Faculty of Transportation Sciences, Department of Logistics and Management of Transport, Prague 1, Konviktská 20, 110 00, mockova@fd.cvut.cz.
[3] VSB – Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, Ostrava – Poruba, 17. listopadu, 15/2172, 708 33, michal.dorda@vsb.cz.
[4] CTU in Prague, Faculty of Transportation Sciences, Department of Logistics and Management of Transport, Prague 1, Konviktská 20, 110 00, vancupav@fd.cvut.cz.
[5] VSB – Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, Ostrava – Poruba, 17. listopadu, 15/2172, 708 33, ivana.olivkova@vsb.cz.
[6] VSB – Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, Ostrava – Poruba, 17. listopadu, 15/2172, 708 33, vojtech.graf@vsb.cz.

2 State of the art

Approaches to coordination of connections in public transport are proposed in the literature both at the tactical and operational level. The tactical level of coordination between connections is realized primarily through synchronization of time positions of connections in timetables. At the tactical level, it is possible to implement coordination throughout the planning period [7], some authors emphasize coordination of the first connections [6] or the last connections [4] in a coordination time span. Examples of approaches for solving the tactical level of time coordination have recently been published, for example, in [2], where mixed integer programming in combination with a heuristic method is used. An important aspect addressed in coordination of timetables is preservation of their resilience, as stated, for example, in article [3]. In this context, it is interesting to mention an approach based on dynamic traffic management published in article [1]. In the article, coordination of transfers is realized by adjusting driving times while preserving the same level of energy consumption. In addition to mathematical programming and heuristic methods, methods used to coordinate connections also include different metaheuristic methods, see for example in [5]. Metaheuristic approaches include both classical metaheuristic approaches (simulated annealing) and metaheuristic approaches based on evolutionary principles (genetic algorithms).

Time coordination of connections at the operational level of management takes place especially in cases of connection delays in a line network. One of coordination tools for these purposes is offered in article [13]. This tool is based on redirection of passenger flows in cases when transfers in the original transfer nodes cannot be realized. Recently published approaches also involve telematics systems in coordination issues. An example of these approaches was published, for example, in article [8], in which the issue of connection coordination is addressed through intervehicle communications.

In the presented article we follow up on our previous articles [9], [10], [11] and [12].

3 Mathematical model

Before designing a mathematical model of network coordination with the application of periodically alternating headways it is necessary to:

1. determine the so-called coordination period, i.e., the time period in which the time coordination will be implemented in the transfer nodes,

2. determine the number of connections included in the coordination period,

3. ensure time continuity of connections on their routes (there are no undesired idle times for the connections serving individual lines in the transfer nodes),

4. ensure that coordination is realized within the same coordination period of the given time span (usually a day).

The procedure for determining the coordination period will be described on an example. Consider a case in which we have the headway of 15 minutes on the line passengers transfer from (the inbound line) and 12 minutes on the line passengers transfer to (the outbound line). When we consider the arrival and departure times of the first connections at time 00 (in minutes) of a specific hour, then the connections of the inbound line arrive at times 00, 15, 30, 45 minutes and the connections of the outbound line depart at times 00, 12, 24, 36, 48 minutes. The coordination period will have a length corresponding to the least common multiple of the values of both headways, i.e., in this case the coordination period will be 60 minutes.

Let us determine the number of the connections subject to coordination. Assume that the basic time unit is 1 minute. While preserving the request on the base time unit of 1 minute, the maximum time offset for the connections of the inbound line for our example is 14 minutes and the maximum time offset for the connections of the outbound line is 11 minutes. The latest possible time positions of the last connection of the inbound line and the outbound line result from the given headways. In the case of the last connection of the inbound line, this time will be 59 minutes (45 + 14 = 59). In the case of the last connection of the outbound line, this time will also be 59 minutes (48 + 11 = 59). However, when coordinating the connections of the inbound and outbound lines, a combination of time shifts, in which the time shift of the inbound line connections will be such that it would not be possible to transfer from the last inbound line connection to the last outbound line connection, may prove advantageous. In our example this would occur in a case where the time shift of the connections of the inbound line would be 14 minutes (the individual connections would arrive at times 14, 29, 44 and 59 minutes) and the time shift of the connections of the outbound line would be 0 minutes (the connections would depart at times 00, 12, 24, 36, 48 minutes). This may be caused, for example, by a request to prefer transfer links in other transfer nodes.

In general, the number of the outbound connections of the line $j \in L_u$ must be determined so that it is possible to ensure transfers from the last connection of each inbound line scheduled in their latest possible time positions and by considering the transfer times.

Preserving time continuity of the connections of the individual lines is realized by considering the driving times between individual transfer nodes.

When a time position of any connection regarding a transfer node falls outside the coordination period, it is necessary to convert the time position into the original coordination period. The requirement to preserve the same coordination period is important so as not to lose continuity of the time span to be coordinated. The conversion of the connection to the original coordination period is done by subtracting the value of the coordination period from the time representing the service time of the transfer node which has fallen outside the coordination period. For greater driving times between different transfer nodes, it may happen that it is necessary to subtract a multiple of the coordination period higher than 1.

The whole process of the conversion of the time position, which is outside the coordination period, into the coordination period can be shown on an example. Assume a coordination period of 30 minutes and a pair of transfer nodes with a driving time of 21 minutes between them. Consider 3 connections that depart from the first transfer node at times 07, 17, and 27 minutes. Thus, the same connections will serve the second transfer node at times 28, 38, and 48 minutes. However, times 38 and 48 minutes are outside the coordination period. To convert the corresponding times into the coordination period, we must subtract the value of the coordination period (30 minutes) from both times, and we get times 08 and 18 minutes. The first service time of the second transfer node which equals to 28 minutes does not need to be changed because it has fallen into the coordination period. Thus, the second transfer node will be served at times 08, 18 and 28 minutes in the coordination period. Although the second node is actually served by different connections, the time positions of the second node service correspond to the real time positions when the second transfer node is served within the same coordination period.

In the model, which will be presented in the article, only lines with at least 2 transfer nodes on their routes are included. It does not make sense to include lines with single transfer node on their route in the coordination process, as their timetable can be coordinated individually. The results of network coordination will be used in the process of the individual coordination.

Let a set of transfer nodes U and a set of lines L be defined. For each transfer node $u \in U$, a set of lines L_u is given – connections of these lines should be coordinated in the transfer node $u \in U$. Furthermore, for each line $i \in L$ a set of directions Si, in which the connections are run, is defined. In the following text we assume that no line is circular, i.e., the sets of directions are the same for all lines – there are two, so it is possible to omit the index in case of the sets S_i and we can use the simplified notation S. For each line $i \in L$ and each direction $l \in S$, a set of connections P_{il} subject to coordination is defined.

Each coordination request in the solved network is defined by an ordered seven $[u; i; l; k, j; s; f]$, where $u \in U$, $i \in L_u$, $l \in S$, $k \in P_{il}$, $j \in L_u$, $s \in S$, and $f \in R^+$. The first number in the ordered seven identifies the transfer node, the second number represents the line, the third number represents the direction, and the fourth number represents the connection from which the transfer is requested. The fifth and sixth number represent the line and its direction to which the transfer link is requested. It is logical that it must hold that $i \neq j$, because the transfer between connections of the same line is not relevant in practice. However, in case of the directions, it can be true that $l = s$, since it may happen that the transfer between different lines going in the same direction may be required. The last number in the ordered seven represents the volume of passengers transferring in the node $u \in U$ from the connection $k \in Pil$ of the line $i \in Lu$ going in the direction $l \in S$ to the connections of the line $j \in Lu$ running in the direction $s \in S$.

For each pair of lines $i \in Lu$ and $j \in L_u$, where $u \in U$, the value of the minimal transfer time $tprestuij$ is defined (it is assumed that the value of the minimal transfer time between the connections of the coordinated lines does not depend on the connections subject to coordination) and the volume of passengers f_{uilkjs} transferring (for the selected coordination period) from the connection $k \in P_{il}$ of the line $i \in L_u$ going in the direction $l \in S$ to the connections of the line $j \in L_u$ going in the direction $s \in S$.

For each line $i \in L_u$ serving the transfer node $u \in U$ in the direction $l \in S$, the headway T_i and the earliest possible service time by two connections t_{uil1} and t_{uil2} of the same line, where $t_{uilk} = t_{uil1} + (k - 1) \cdot Ti$, are defined. The headway applied on the line can be constant, but also periodically alternating (several values of the headway which alternate each other – for example 7 and 8 minutes). In the case of the periodically alternating headway on the line $i \in L$, the value T_i represents the so-called basic headway, which is the minimum of the alternating head- ways.

The task is to decide on the time offsets of the connections in individual directions coordinated in the respective transfer nodes so that the time offsets of the connections of individual lines going in the same direction are constant (to preserve the headways of the individual lines) and the total time loss of all transferring passengers is as minimal as possible.

To set the values of t_{ujsk}, we can proceed as follows. First, set the times for the first connections serving the individual transfer nodes for each line. In the last transfer node of the given direction, in which coordination of the connection of the line $j \in L$ in the direction $s \in S$ is realized, set the time t_{ujs1} (the time when the transfer node is served by the first connection) to 0. Times of the subsequent connections serving the same line in the same direction in the coordination period must always be increased by the required headway. In the next step, we must set the earliest possible service times of the transfer node which is served as the last in the given direction. However, the procedure differs for the lines with the periodically alternating headway and for the lines with the constant headway. In the case of the lines with the constant headway, the earliest possible service times are determined by simply adding the headway. The procedure for the periodically alternating headway will be described further in the article.

To model the decision, we must introduce variables into the optimization problem:

1. x_{il} is a non-negative variable modeling the time shift of all the connections of the line $i \in L_u$ in the direction $l \in S$ calculated from their earliest possible time positions,
2. h_{uipkjs} is a non-negative variable modeling the time loss of each passenger transferring in the transfer node $u \in U$ from the connection $k \in P_i$ of the line $i \in L_u$ going in the direction $p \in S$ to the soonest connection of the line $j \in L_u$ going in the direction $s \in S$,
3. $z_{uipkjsl}$ is an auxiliary binary variable modeling the creation of a transfer link between the connection $k \in P_i$ of the line $i \in L_u$ going in the direction $p \in S$ and the connection $l \in P_j$ of the line $j \in L_u$ going in the direction $s \in S$ via the transfer node $u \in U$.

However, when the periodically alternating headway should be applied, it is necessary to ensure that the individual headway values are repeated periodically. This can be achieved by using auxiliary binary variables. Let us assume that the headway can take two periodically alternating values, which differ from each other by 1. We introduce two auxiliary binary variables v_{il} and w_{il} into the model for each line $i \in L$ with the periodically alternating headway and each direction $l \in S$. The binary variables will allow the increase of the values of the so-called basic headway of the line $i \in L$ in the direction $l \in S$. If such headway is applied on the line $i \in L$, we must define the basic headway, which is the lowest value of the possible headway values. Thus, if, for example, the periodically alternating headway of 7, 8, 7, 8,… minutes is applied for the line, the basic headway has the value which equals to $min\{7;8\} = 7$. The values of the binary variables for the given line and its individual directions are then added to the basic headway, which results in the periodically alternating values of the headway.

The use of the variables v_{il}, w_{il} for the connections of the line $i \in L$ going in the direction $l \in S$ is explained in case of a transfer node, for which it holds that $t_{uil1} = 0$. To achieve the periodically alternating headway, the value t_{uil2} equals to $t_{uil2} = t_{uil1} + T_i + v_{il}$. If it holds that $v_{il} = 0$ after the optimization calculation, then the headway between the first and the second connection in the last transfer node will be equal to the value of the basic headway T_i. If it holds that $v_{il} = 1$ after the optimization calculation, then the headway between the first and second connection in the last transfer node will be equal to the value of the basic headway T_i increased by 1. The value of the variable v_{il} is thus intended to get the increased value of the basic headway between the first and the second connection. The variable w_{il} is applied analogously. This is intended to achieve the increased value of the basic headway between the second and third connection. In the model, it must be assured that just one of the variables takes value 1. The headways between the subsequent pairs of the connections must be set using the binary variables so that the required headways alternate. So in our example it will be $t_{uil3} = t_{uil1} + T_i + v_{il} + w_{il}$, $t_{uil4} = t_{uil1} + T_i + 2 \cdot v_{il} + w_{il}$, $t_{uil5} = t_{uil1} + T_i + 2 \cdot v_{il} + 2 \cdot w_{il}$ and so on.

The earliest possible service times of the previous transfer nodes on the same line route must be set so that time continuity of the connections serving all the transfer nodes is preserved if the periodically alternating headway is applied. Let us demonstrate the procedure again on an example. Consider a line with the alternating headways of 7 and 8 minutes which goes via two transfer nodes A and B. The value of the basic headway is therefore T_i =7 minutes. Consider the direction in which transfer node A is served as first, and transfer node B is served as second. Let the driving time from node A to node B be 13 minutes and the coordination period be 30 minutes. Node B is served as the second (last), so we must set the earliest possible service times for it firstly. The service times will correspond to the values of t_{uil1}, t_{uil2}, t_{uil3}, t_{uil4} and t_{uil5} explained in the previous paragraph. These will be the $t_{uil1} = 0$, $t_{uil2} = 7 + v_{il}$, $t_{uil3} = 14 + v_{il} + w_{il}$, $t_{uil4} = 21 + 2 \cdot v_{il} + w_{il}$ and $t_{uil5} = 28 + 2 \cdot v_{il} + 2 \cdot w_{il}$. In the next step, we must project these times into transfer node A. Considering the driving time, which is 13 minutes, we can say that for the coordination period of 30 minutes, node A will be served firstly by the connection that serves node B at time $14 + v_{il} + w_{il}$. When this connection serves node B at time $14 + v_{il} + w_{il}$, it must serve node A at time $(14 - 13) + v_{il} + w_{il} = 1 + v_{il} + w_{il}$. Based on this value we can determine the remaining values of the earliest possible service times of node A for the line $i \in L$. For the second connection serving node A, the earliest possible service time equals to $8 + 2 \cdot v_{il} + w_{il}$, for the third connection to $15 + 2 \cdot v_{il} + 2 \cdot w_{il}$, for the fourth connection to $22 + 3 \cdot v_{il} + 2 \cdot w_{il}$ and finally for the fifth connection to $29 + 3 \cdot v_{il} + 3 \cdot w_{il}$.

To simplify the mathematical notation of an optimization criterion, we introduce an incidence matrix $\boldsymbol{A}$. If it is required to coordinate in the node $u \in U$ the connections of the line $i \in L_u$ going in the direction $l \in S$ with the connections of the line $j \in L_u$ going in the direction $s \in S$, then $a_{uiljs} = 1$, otherwise $a_{uiljs} = 0$.

Also, to simplify the notation of the model, let us define $\tau_{uilk}(v, w)$ to denote the earliest possible service times of the transfer node $u \in U$ by the connection $k \in P_{il}$ of the line $i \in L_u$ going in the direction $l \in S$. In case of the periodically alternating headways on the line, it holds that $v_{il} \in \{0; 1\}$ and $w_{il} \in \{0; 1\}$, in case of the constant headway on the line it holds that $v_{il} = 0$ and $w_{il} = 0$.

The mathematical model of the solved problem can be written in the following form:

$$min\, f(x, h, z, v, w) = \sum_{u \in U} \sum_{i \in L_u} \sum_{l \in S} \sum_{k \in P_{il}} \sum_{\substack{j \in L_u \\ j \neq i}} \sum_{s \in S} a_{uiljs} \cdot f_{uilkjs} \cdot h_{uilkjs} \tag{1}$$

subject to:

$$\left[\tau_{ujsp}(v, w) + x_{js}\right] - \left[\tau_{uilk}(v, w) + x_{il}\right] \geq M \cdot \left(z_{uilkjsp} - 1\right) \quad \text{for } u \in U, i \in L_u, j \in L_u, j \neq i, \ l \in S, s \in S, k \in P_{il}, p \in P_{js}, \ a_{uiljs} = 1 \tag{2}$$

$$\left[\tau_{ujsp}(v, w) + x_{js}\right] - \left[\tau_{uilk}(v, w) + x_{il}\right] \leq h_{uilkjs} + M \cdot \left(1 - z_{uilkjsp}\right) \quad \text{for } u \in U, i \in L_u, j \in L_u, j \neq i, \ l \in S, s \in S, k \in P_{il}, p \in P_{js}, \ a_{uiljs} = 1 \tag{3}$$

$$\sum_{p \in P_{js}} z_{uilkjsp} = 1 \quad \text{for } u \in U, i \in L_u, j \in L_u, j \neq i, \ l \in S, s \in S, k \in P_{il}, \ a_{uiljs} = 1 \tag{4}$$

$$x_{il} \leq a_{il} \quad \text{for } i \in L_u \text{ and } l \in S \tag{5}$$

$$x_{il} \in R_0^+ \quad \text{for } i \in L_u \text{ and } l \in S \tag{6}$$

$$h_{uilkjs} \in R_0^+ \quad \text{for } u \in U, i \in L_u, j \in L_u, j \neq i, \ l \in S, s \in S, k \in P_{il}, a_{uiljs} = 1 \tag{7}$$

$$z_{uilkjsp} \in \{0; 1\} \quad \text{for } u \in U, i \in L_u, j \in L_u, j \neq i, \ l \in S, s \in S, k \in P_{il}, p \in P_{js}, \ a_{uiljs} = 1 \tag{8}$$

$$v_{il} \in \{0; 1\} \quad \text{for } i \in L_u \text{ and } l \in S \tag{9}$$

$$w_{il} \in \{0; 1\} \quad \text{for } i \in L_u \text{ and } l \in S \tag{10}$$

Function (1) represents the optimization criterion – the total time loss of all passengers transferring in all transfer nodes. The group of constraints (2) ensures that in case of time infeasibility of the time positions of the connections of the coordinated lines going in the directions affected by the coordination, no transfer link is created. The group of constraints (3) quantifies the time losses of transferring passengers generated by the created transfer links. The group of constraints (4) ensures the creation of requested transfer links. Constraints (5) model that, in case of any time shifts of the connections generated to reduce the total time loss, the maximum values of the allowable shifts are not exceeded. Constraints (6) – (10) define domains of definition of all the variables used in the model.

4 Optimization experiments

Functionality of the proposed model was tested in a fragment of the network of the Prague Public Transport Company – see Figure 1, in which this fragment is shown.

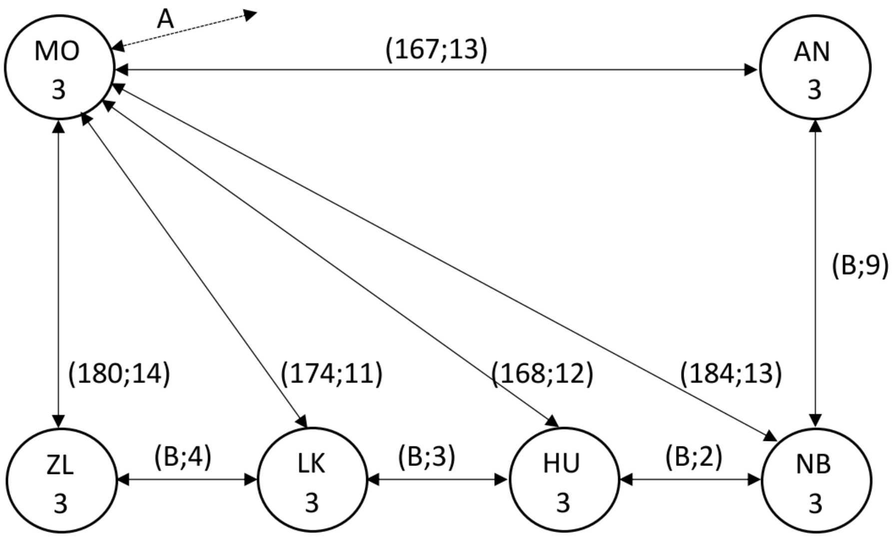


Figure 1 The network subject to coordination

Coordination in the given public transport network was requested between connections of subway and bus lines in both directions. Vertices of the graph represent the transfer nodes in the network. Symbols MO, AN, etc. represent the names of the transfer nodes (see Table 1), the numbers below them represent the transfer times between the connections of coordinated lines.

Notation	MO	ZL	LK	HU	NB	AN
Transfer node	Motol	Zličín	Luka	Hůrka	Nové Butovice	Anděl

Table 1 The transfer nodes in the network

Edges in Figure 1 represent individual sections of the routes of the coordinated lines important from the point of view of network coordination. To simplify the scheme, the pairs of the oriented edges are always merged into one bidirectionally oriented edge. An edge weight (except for the edge with weight A) is represented by two values. The first value represents the line, the second value represents the driving time between the vertices. When a letter is in the first position of the edge weight, it is a subway line. When a number is in the first position, it is a bus line. The driving times between pairs of the transfer nodes are the same in both directions. Thus, for example, between the transfer nodes Motol (MO) and Anděl (AN) there is a route of bus line 167 and the driving time between the nodes in both directions is 13 minutes. The edge with weight A has only a symbolic meaning. It represents a subway line that is included in the coordination model in only one transfer node. Therefore, it is not terminated by vertices on both sides. The same importance is assumed for all transfer links (their weight is set to 1).

Input data to the experiment (in addition to information shown in Figure 1) is summarized in Table 2. Based on the values of the headways given in Table 2 it holds for the coordination period (the least common multiple) that $LCM(7{,}5;\ 6;\ 7+8) = 30$ minutes. The results of the optimization calculation are summarized in Table 3. The optimization experiments were run on a PC with a Processor AMD Ryzen 7 2700X Eight-Core 3.70 GHz and 32 GB of RAM. The calculation time was 0.5 s.

Line	A	B	167	168	174	180	184
Headway [min]	7,5	6	7 and 8				
Number of inbound connections	4	5	5				
Number of outbound connections	5	6	6				

Table 2 Input data

Transfer node	Transfer link line/direction from	to	Arrival time after co-ordination	Departure time after co-ordination	Total waiting time [min]
MO	A/2	167/1	5.5; 13; 20.5; 28	2; 9; 17; 24; 32	3
	A/2	168/1	5.5; 13; 20.5; 28	3; 10; 18; 25; 33	7
	A/2	174/2	5.5; 13; 20.5; 28	4; 12; 19; 27; 34	13
	A/2	180/2	5.5; 13; 20.5; 28	1; 9; 16; 24; 31	1
	A/2	184/2	5.5; 13; 20.5; 28	1; 9; 16; 24; 31	1
	167/2	A/1	0; 8; 15; 23	3.5; 11; 18.5; 26; 33.5	1
	168/1	A/1	0; 8; 15; 23	3.5; 11; 18.5; 26; 33.5	1
	174/1	A/1	0; 7; 15; 22	3.5; 11; 18.5; 26; 33.5	3
	180/1	A/1	0; 8; 15; 23	3.5; 11; 18.5; 26; 33.5	1
	184/1	A/1	0; 8; 15; 23	3.5; 11; 18.5; 26; 33.5	1
ZL	B/2	180/1	1; 7; 13; 19; 25	1; 9; 16; 24; 31	16
	180/2	B/1	0; 8; 15; 23	0; 6; 12; 18; 24; 30	8
LU	B/2	174/1	3; 9; 15; 21; 27	4; 11; 19; 26; 34	16
	B/2	174/2	3; 9; 15; 21; 27	0; 8; 15; 23; 30	14
	174/1	B/1	4; 11; 19; 26	4; 10; 16; 22; 28	8
	174/2	B/1	0; 8; 15; 23	4; 10; 16; 22; 28	10
HU	168/2	B/1	0; 7; 15; 22	1; 7; 13; 19; 25; 31	8
	B/2	168/1	6; 12; 18; 24; 30	3; 11; 18; 26; 33	16
NB	B/2	184/1	4; 10; 16; 22; 28	1; 8; 16; 23; 31	14
	B/2	184/2	4; 10; 16; 22; 28	0; 8; 15; 23; 30	19
	184/1	B/1	1; 8; 16; 23	3; 9; 15; 21; 27; 33	12
	184/2	B/1	0; 8; 15; 23	3; 9; 15; 21; 27; 33	8
AN	B/1	167/2	0; 6; 12; 18; 24	2; 10; 17; 25; 32	19
	B/2	167/2	1; 7; 13; 19; 25	2; 10; 17; 25; 32	14
	167/1	B/2	0; 7; 15; 22	1; 7; 13; 19; 25; 31	8
∑					222

Table 3 Results of the optimization calculation

The first column in Table 3 contains the transfer nodes. The second column shows the line and the direction from which the passengers transfer in the transfer node. The third column presents the line and the direction to which the passengers transfer in the transfer node. The fourth column presents the time positions of the individual inbound connections, which arrive in the transfer node. The fifth column shows the time positions of the individual out-bound connections, which depart from the transfer node. The values in the last column represent the total time loss generated in the transfer node (note that the volumes of the transferring passengers were set to 1). Let us demonstrate the calculation of the total time loss on an example for the first row of Table 3. The passengers, who change in the Motol transfer node (MO) from the connections of subway line A going in direction 2 (A / 2) to the bus connections of line 167 going in direction 1 (167/1), arrive in the transfer node at times 5.5; 13; 20.5; 28 minutes. The transfer time for Motol is 3 minutes. That means the transferring passengers come to the bus stop of line 167, from which the connections depart in direction 1, at times 8.5; 16; 23.5; 31 minutes. The connections of line 167 in direction 1 depart at times 2; 9; 17; 24; 32 minutes. The passenger arriving in the transfer node by subway line A in direction 2 at time 5.5 minutes will therefore wait for the soonest connection of line 167 in direction 1 which departs at time 9 minutes. The time loss will therefore be 0.5 minutes. The passenger who arrives in the transfer node by subway line A in direction 2 at time 13 minutes will therefore wait for the soonest connection of line 167 in direction 1 which departs at time 17. The time loss will therefore be 1 minute. The passenger who arrives at time 20.5 minutes will therefore wait for the soonest connection which departs at time 24 minutes. The time loss will therefore be 0.5 minutes. Finally, the passenger arriving at time 28 minutes will therefore wait for the soonest connection which departs at time 32 minutes. The time loss will therefore be 1 minute. In this case, the total time loss will be 3 minutes, which corresponds to the value given in the last column.

5 Conclusions

The presented article deals with the issue of time coordination of connections in transfer nodes of public transport system networks. The mathematical model is designated for cases when heterogeneous headways are applied on the individual lines and, in addition, the model can be used even in cases when so-called periodically alternating headways are applied – this is assured by means of auxiliary binary variables used in the model. This is the main novelty of the presented approach. Computational experiments with the proposed model were carried out in conditions of the real public transport network of the Prague Public Transport Company. The computational experiments proved full functionality of the proposed model.

Acknowledgements

This work was supported by project CK01000043 System for supporting network time coordination of connections at interchange nodes.

References

[1] Albrecht, T., & Oettich, S. (2002). A new integrated approach to dynamic schedule synchronization and energy-saving train control. WIT *Transactions on The Built Environment*, 61.

[2] Chu, J. C., Korsesthakarn, K., Hsu, Y. T., & Wu, H. Y. (2019). Models and a solution algorithm for planning transfer synchronization of bus timetables. *Transportation Research Part E: Logistics and Transportation Review*, 131, 247-266.

[3] Fretter, C., Krumov, L., Weihe, K., Müller-Hannemann, M., & Hütt, M. T. (2010). Phase synchronization in railway timetables. *The European Physical Journal B*, 77(2), 281-289.

[4] Kang, L., Wu, J., Sun, H., Zhu, X., & Gao, Z. (2015). A case study on the coordination of last trains for the Beijing subway network. *Transportation Research Part B: Methodological*, 72, 112-127.

[5] Kwan, C. M., & Chang, C. S. (2008). Timetable synchronization of mass rapid transit system using multi-objective evolutionary approach. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 38(5), 636-648.

[6] Li, X., Lu, L., Zheng, P., & Huang, Z. (2019). Timetable coordination of the first trains for subway network with maximum passenger perceived transfer quality. *IEEE Access*, 7, 52042-52051.

[7] Liu, T., Cats, O., & Gkiotsalitis, K. (2021). A review of public transport transfer coordination at the tactical planning phase. *Transportation Research Part C: Emerging Technologies*, 133, 103450.

[8] Liu, T., Ceder, A., Ma, J., & Guan, W. (2014). Synchronizing public transport transfers by using intervehicle communication scheme: Case study. *Transportation Research Record*, 2417(1), 78-91.

[9] Teichmann, D., Dorda, M., Mocková, P., Olivková, I., Graf, V. & Dvořáčková, A. (2020). Of Approach to Minimize the Number of Shunting Tracks at Turning Loops. In *Proceedings of the International Scientific Conference Quantitative Methods in Economics Multiple Criteria Decision Making XX* (pp. 333-343). Bratislava: The Slovak Society for Operations Research.

[10] Teichmann, D., Dorda, M., Mocková, D. Vančura, P., E., Graf, V. & Olivková, I. (2020). Minimising Number of Shunting Tracks at Tram Turning Loops. In *Proceedings of 38th International Conference on Mathematical Methods in Economics (MME 2020)* (pp. 585-591). Brno, Mendel University in Brno.

[11] Teichmann, D., Dorda, M., Mocková, D. Vančura, P., E., Graf, V. & Olivková, I. (2021). Combined Time Coordination of Connections in Public Transport. In *Proceedings of 39th International Conference on Mathematical Methods in Economics (MME 2021)* (pp. 487-492). Prague, Czech University of Life Sciences.

[12] Teichmann, D., Mocková, D., Dorda, M., Vančura, P., E., Olivková, I. & Graf, V. (2022). Design of Optimization Model for Network Coordination of Transfers Between Connections - Line Network with Application of Homogeneous Headway. In *Proceedings of the International Scientific Conference Quantitative Methods in Economics Multiple Criteria Decision Making XXI* (pp. 188-195). Bratislava: The Slovak Society for Operations Research.

[13] Wu, W., Liu, R., Jin, W., & Ma, C. (2019). Stochastic bus schedule coordination considering demand assignment and rerouting of passengers. *Transportation Research Part B: Methodological*, 121, 275-303.

Measuring the Efficiency of English Football Clubs: Empirical Evidence from Professional Football

Michal Tomíček[1]

Abstract. The aim of the paper is to propose a method for evaluating the performance of football clubs based on the method of Data Envelopment Analysis. The article also deals with the impact of the reprisals caused by the Covid-19 pandemic, which limited fans' access to the stadiums as well as the football competitions themselves. Professional English football clubs playing in the English Premier League were selected for empirical analysis. The performance evaluation of English football clubs is an important reference for team managers, the management of individual clubs and the selected competition as a whole. One of the two most frequently used DEA models, the CCR model, was used to analyze the relative efficiency of football clubs. The study focuses on clubs operating in the highest English football competition during the seasons 2017/18 to 2020/21.

Keywords: Data Envelopment Analysis, European Football, Efficiency, English Premier League

JEL Classification: C10, L83, C67, C44
AMS Classification: 90B90, 90C90

1 Introduction

The world is slowly returning to normal, and sport is no exception. Especially in Europe, football is at the forefront of popular sports. Fans from Lisbon in the west to Turkey in the east were looking forward to relaunching their national football competitions. The same is true of the so-called cradle of football, England.

Sports organizations try to evaluate its performance: its weaknesses and strengths. Currently, success in the professional football league is linked to successful training and management of the entire team. Efficiency, however, goes beyond this kind of corporate-level view. That is where the board of directors and, crucially, the owner are willing to see the club's performance charts and curves rise. In the business world, comparison with competitors is important to objectively evaluate the direction of the company. One of the most popular methods for this kind of benchmark analysis is Data Envelopment Analysis (DEA).

The presented paper focuses on measuring the performance of clubs in the English Premier League, officially The Football Association Premier League Limited, the top football competition in England. The period under study is the 2017/18 to 2020/21 seasons, these are four consecutive seasons. The comparison covers the seasons before, during and after the Covid-19 restrictions mainly on the number of spectators in the stadiums and the closure of teams to the so-called closed bubbles.

A specific feature of European football competitions is the closed nature of the different competition levels. Football clubs are not free to move between levels, but unlike professional sports competitions in the United States which are usually closed for relegation or promotion, there are clear rules for doing so at the end of the season based on position in the table. Newly formed clubs cannot immediately join the top competitions, they have to go through a long path from the lowest competitions to the professional leagues. Therefore, it is often more profitable for investors to buy an already established club with a history, a fan base and sporting and personnel facilities than to start a club from scratch. At the same time, a single investor may not own more than one football club participating in European cups. This is a safeguard against the oligopolisation of professional European football.

The performance of sports companies has been addressed by a number of authors. Most of them use only sports statistics to assess club efficiency, while the little rest combine sports and economic indicators to achieve more accurate efficiency results.

Palafox-Alcantar and Vargas-Hernández [9] measure the payroll efficiency of 32 National Football League (NFL) teams in the 2014 season using data wrapper analysis. When analyzing American sporting competitions, time

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

series can be better used as the set of units of observation is almost constant over time. However, the paper was focused on analyzing the one season only.

In their research, Barros and Douvis [2] estimate productivity changes using DEA analysis applied to a representative sample of football clubs operating in two South European countries: Greece and Portugal. They rank football clubs according to their productivity changes between the 1999/20 and 2002/03 seasons. They conclude that some clubs experienced productivity growth while others experienced productivity decline. In the evaluation of the conclusions of the work, they mainly mention the clubs that participated in all the seasons studied. Petrovic Djordjevic [10] uses a non-parametric output-oriented DEA model and analyzes the technical efficiency of national football teams in the 2010 World Cup qualification. The DEA model has a two-stage structure. In the first stage inputs are used to produce outputs which then become inputs to the second stage.

Halkos and Tzeremes [5] use the DEA method to compare the actual level of market value of football clubs and their performance. The research shows that the level of market value of football clubs has a negative effect on their performance, more accurately that high value of football clubs does not guarantee higher performance.

Arabzad et al. [1] use the DEA model to identify the best footballers in the English Premier League. A different type of DEA model is used to rank selected players. The proposed approach is investigated in the English Premier League 2010/11 season. The findings show that players Wayne Rooney, Didier Drogba and Carlos Tévez are ranked first, second and third respectively.

Jardin [7] evaluates the efficiency of French football clubs between 2004 and 2007 using DEA. Then the Malmquist indexes are used for the studies of the dynamics of clubs' performance. The first source of inefficiency in the Ligue 1 is linked to size problems and over-investments. The best teams in competition or most profitable clubs are not the most efficient units in the sample. The Kang's study [8] measured the relative efficiency and productivity change of Korean professional sports teams using the DEA model and the Malmquist Index for 2006-2009 with similar results as Jardin's [7].

2 Methods applied

In the main part of the paper the method of DEA, a quantitative non-parametric method, is used. Basic DEA models are divided into input-oriented and output-oriented models. Using input-oriented models, it is possible to estimate the degree of technical efficiency, which determines the reduction of input indicators, so that the unit becomes technically efficient with unchanged output. Efficient units get a score of 1, inefficient ones get a lower score in the interval ⟨0; 1) [3; 4].

The CCR (Charnes, Cooper and Rhodes) model considers constant returns to scale and estimates Overall Technical Efficiency (OTE), which consists of two parts, Pure Technical Efficiency (PTE) and Scale Efficiency (SE). Scale efficiency is then determined by the ratio of OTE and PTE and indicates the extent to which the unit can improve its efficiency by changing its size [3]. Together with the BCC (Banker, Charnes and Cooper) model, the CCR model forms two basic (mathematically simplest) DEA models [6].

The OSDEA-GUI (Open-source DEA Graphic User Interface) software was used for all the calculations related to DEA. The data for each season were entered into the software separately in the form of csv files with inputs and outputs in separate columns. The efficiency limit defines the maximum output combinations that can be selected for a given set of inputs. Assuming a set of n Decision making units ($DMUs$), each DMU_j ($j = 1, \dots, n$) uses m inputs x_{ij} ($i = 1, 2, \dots, m$) to produce s outputs y_{rj} ($r = 1, 2, \dots, s$). Input-oriented models with constant returns to scale can be formulated according to equations (1, 2) to minimize inputs while keeping outputs at their current level [6].

$$\min \theta_q$$

$$S.t. \sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = \theta_q x_{iq} \qquad i = 1, 2, \dots, m, \tag{1}$$

$$\sum_{j=1}^{n} \lambda_j y_{kj} + s_k^+ = y_{kq} \qquad k = 1,2,\dots,r$$
$$\lambda_j \geq 0, \qquad j = 1,2,\dots,n.\ \theta \text{ unrestricted in sign}, \tag{2}$$

where θ_q represents the technical efficiency of entity q and λ_j represents the associated weighting of outputs and inputs of entity j. To find maximal possible inputs excesses and outputs shortfalls while maintaining θ, the slack variables, s_i^- and s_k^+, reflect these excesses and shortfalls, respectively.

The overall technical efficiency of the DMU is measured in relation to the other units analyzed using the efficiency score. The overall level of technical efficiency (OTE_{CCR-I}) is taking on values in the range $\langle 0,1 \rangle$. Technically efficient DMUs achieve efficiency rates of 1, the efficiency rates of inefficient units are less than 1 [3; 6].

The creation of the database is preceded by the collection of data from several sources. In addition to the publicly available information published by the English Football Association itself, the data used in the article also comes from the private database of InStat, a company that analyses sports data. The economic data was obtained from the specialist Transfermarkt server. The complexity of the database used in the paper lies in the combination of these sources and their supplementation with data from the register of companies doing business in the United Kingdom of Great Britain and Northern Ireland. The research process can be divided into the following phases:

1. creation of the list of evaluated companies;
2. collection of sporting and economic metrics:
   a) sports data from InStat database of football players and clubs;
   b) financial data from Transfermarkt server;
   c) attendance data from official English Premier League database and Transfermarkt database.
3. determining the inputs and outputs of the DEA model;
4. determination of technical efficiency values;
5. comparison of the performance of football clubs.

Given the method used and the available data, the input factor is the market value of the team in the season and the average attendance at the team's home stadium. The total market value is only considered for players who have played at least one league game in a given season. The factor of average stadium attendance was used in the article to represent the external football environment. The average attendance factor contains many characteristics that limit its use and power. In future research, the relative attendance to the maximum stadium capacity could be used instead of the average value.

On the output side of the model, there are more factors: goals scored, challenges won, supersaves and accurate key passes. Outputs of the model represent the game characteristics for the main football roles on the pitch, i.e. goalkeepers, defenders, midfielders and forwards. Challenges won consist of the sum of defensive and offensive challenges and thus represent a universal physical component of the factors used. The output factors include the four different components of a football match, namely the offensive and defensive parts, the goalkeeper factor and the creativity factor. The chosen model therefore examines how efficiently clubs can convert the market value of their players and fan support into positive game characteristics in the form of goals scored, challenges won, supersaves and accurate key passes.

3 Research results

The results of the paper are divided into four parts according to the season studied (2017/18, 2018/19, 2019/20 and 2020/21). This is followed by a summary of the results in the seasons studied and an evaluation of the clubs' efficiency in the seasons. The first table (Table 1) shows position in the table (Pos.) and the value of OTE_{CCR-I} of the clubs that participated in a given seasons. The sports successful clubs, which were ranked in the table up to 8th place, did not reach the efficient OTE_{CCR-I} value even once in the examined seasons.

DMU	2017/18		2018/19		2019/20		2020/21	
	Pos.	OTE_{CCR-I}	Pos.	OTE_{CCR-I}	Pos.	OTE_{CCR-I}	Pos.	OTE_{CCR-I}
MCI	1	0.4526	1	0.5503	2	0.6073	1	0.4113
MAN	2	0.2641	6	0.3177	3	0.4081	2	0.4602
TOT	3	0.2709	4	0.4172	6	0.3701	7	0.5232

LIV	4	0.3655	2	0.5292	1	0.5286	3	0.5521
CHE	5	0.3484	3	0.5348	4	0.6012	4	0.4163
ARS	6	0.3392	5	0.4503	8	0.4147	8	0.4201
BUR	7	0.8226	15	0.9717	10	**1.0000**	17	**1.0000**
EVE	8	0.4023	8	0.5142	12	0.4755	10	0.5917
LEI	9	0.5594	9	0.7073	5	0.8343	5	0.5878
NEW	10	0.6325	13	0.8222	13	0.6866	12	0.6627
CRY	11	0.6622	12	0.8020	14	0.8490	14	0.8781
BOU	12	**1.0000**	14	**1.0000**	18	**1.0000**		
WHU	13	0.6942	10	0.4778	16	0.6003	6	0.8190
WAT	14	0.7263	11	**1.0000**	19	**1.0000**		
BRI	15	0.9191	17	0.8877	15	0.8067	16	0.7480
HUD	16	**1.0000**	20	**1.0000**				
SOU	17	0.4970	16	0.7441	11	0.9585	15	**1.0000**
SWA	18	0.9109						
STO	19	0.8112						
WBA	20	0.9203					19	**1.0000**
WOL			7	0.8221	7	0.6988	13	0.9827
CAR			18	**1.0000**				
FUL			19	0.9294			18	**1.0000**
SHU					9	**1.0000**	20	0.9967
AVL					17	0.6571	11	**1.0000**
NOR					20	**1.0000**		
LEE							9	**1.0000**

Table 1 DEA efficiency of Premier League clubs in the seasons 2017/18 to 2020/21

Since the efficient clubs are always ranked 9th to 20th in the table, the result suggests that the team market value of the clubs in the top eight positions in the table is so much higher than the rest of the clubs cannot achieve the efficient $OTE_{CCR\text{-}I}$ value. In the seasons under review, the ratio of the average market values of the Premier League clubs ranked in the Top 8 to those ranked 9th to 20th ranged from 2.6 to 3.6 (see Table 2).

Market Value Average (Th. €)	**2017/18**	**2018/19**	**2019/20**	**2020/21**
Top 8 Clubs	704 550	784 313	681 731	688 588
9^{th} – 20^{th} Clubs	199 916	235 071	219 556	258 750
Ratio	3.5242	3.3365	3.1050	2.6612

Table 2 Market value for clubs according to their position in the table

Of the 17 efficient clubs (see Table 1), there have been a total of six occasions when clubs rated as efficient at the end of the season have been relegated from the top flight of English football. These clubs had the lowest budgets and the market value of their team was also the lowest compared to other Premier League clubs. In each season, there has always been one club newly promoted to the top competition that has managed to finish in the top ten and in two cases even achieve an efficient value of $OTE_{CCR\text{-}I}$. In the 2018/19 season it was Wolverhampton Wanderers F.C. in 7th place (0.8221), in the 2019/20 season it was Sheffield United F.C. in 9th place (1.0000) and in the 2020/21 season it was Leeds United F.C. in 9th place (1.0000).

On the other hand, four clubs failed to keep up (sportingly and financially) with the Premier League clubs and immediately after promotion they left the top competition and were relegated back to the EFL Championship (officially English Football League Championship). In doing so, all these clubs achieved an efficient value of $OTE_{CCR\text{-}I}$.

The four most successful clubs in terms of table position in the seasons under review were Manchester City, Liverpool, Manchester United and Chelsea. Of these most successful English clubs in recent times and in European competitions, only Manchester City achieved an average value of $OTE_{CCR\text{-}I}$ of more than 0.5. 14 clubs participated in all four seasons. Of these, the highest average OTE values were achieved by Burnley (0.9486), Brighton (0.8404), Southampton (0.7999) and Crystal Palace (0.7978).

If we look at the season before (2018/19) the covid pandemic and the season that was affected by the covid pandemic (2019/20), then 10 of the 17 clubs that participated in both of these seasons saw an improvement in their OTE_{CCR-I} values compared to the 2018/19 season. This is due to the combined effects of a reduction in average attendance at large English club's stadiums, a cooling of the market for players and therefore their market value, while maintaining a decent sporting performance of the clubs.

4 Conclusion

This paper measured the efficiency of professional football clubs playing in the English Premier League. For this purpose, a time span of four seasons from 2017/18 to 2020/21 was chosen. The main aim of the paper was to propose an approach to the analysis and evaluation of the best football clubs of the English Premier League. The efficiency of football clubs was measured using a non-parametric DEA method. Total market value of the team and average stadium attendance were chosen as inputs of the model. Average stadium attendance was added to standard models used by several other authors. Output was measured by the total number of goals, challenges won, supersaves and accurate key passes. This particular specification proved to be appropriate for this application (adequate number of efficient DMUs), but it can also be used to analyze the efficiency of other team sports.

Given the current economic and financial situation, there is an increasing need to determine how efficiently a club is using its resources and assets – its players. Based on the seasons analyzed, several conclusions emerge from the research: It turns out that the gap between the top clubs and the rest of the league in terms of the input factor of the total market value of a Premier League club's squad combined with average attendances is too great for the wealthy section of clubs who can afford the best and therefore most valuable football players in the world, and cannot outweigh the better sporting performance of these clubs compared to average or below average English clubs. This leads to the fact that the sport's best Premier League clubs were not rated as efficient in any of the seasons monitored by the chosen DEA model.

As other authors, for example Jardin [7], mentioned in their works, the best clubs from the sport point of view are not the most efficient by the DEA model results.

Acknowledgements

Creation of this paper was supported by the Technical University of Liberec grant no. 21456 "Factors influencing the economic performance of European sports enterprises" (SGS-2022-1017).

References

[1] Arabzad, S. M., Ghorbani, M., & Shahin, A. (2013). Ranking players by DEA the case of English Premier League. *International Journal of Industrial and Systems Engineering*, 15(4), 443.

[2] Barros, C. P., & Douvis, J. (2009). Comparative analysis of football efficiency among two small European countries: Portugal and Greece. *International Journal of Sport Management and Marketing*, 6(2), 183.

[3] Cooper, W. W. (2011). *Handbook on data envelopment analysis*. New York: Springer.

[4] Cooper, W. W., Seiford, L. M. & Tone, K. (2007). *Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software*. New York: Springer.

[5] Halkos, G. & Tzeremes, N. (2013). A two-stage double bootstrap DEA: the case of the top 25 European football clubs' efficiency levels. *Managerial and Decision Economics*, 34(2), 108–115.

[6] Jablonský, J. & Dlouhý, M. (2015). *Modely hodnocení efektivnosti a alokace zdrojů*. Praha: Professional Publishing.

[7] Jardin, M. (2009). *Efficiency of French football clubs and its dynamics*. Munich: Personal RePEc.

[8] Kang, H. J. (2010). Productivity Change and Relative Efficiency of Professional Sport Teams. *The Journal of the Korea Contents Association*. Vol. 10, Issue 10, 456–463.

[9] Palafox-Alcantar, P. G. & Vargas-Hernández, J. G. (2015). Measuring the efficiency of the 32 fran-chises in the NFL during the 2014 season. *Journal of Sports Economics & Management*, 5(1), 37–53.

[10] Petrović Djordjević, D. (2015). Measuring efficiency of football teams by multi-stage DEA model. *Tehnicki vjesnik-Technical Gazette*, 22(3), 763–770.

VaR and CVaR of Czech Financial Assets Returns Using GARCH Models with Heavy Tails Distributions

Quang Van Tran[1], Jiří Málek[2]

Abstract. Value at Risk (VaR) and Conditional Value at Risk (CVaR) are popular measures used to estimate the risk exposure of investments of risky financial assets. Their accurate evaluation is important as they affect the further actions in risk management. As volatility of returns of financial assets exhibit heteroskedastic behavior, GARCH models are often used to capture this property. Further, it is also known that returns of financial assets are often distributed with heavy tails. So far this character is modeled with heavy-tailed distributions. We investigate the ability of the most often used heavy tailed distributions in GARCH and GJR-GARCH models to evaluate how well they can help to properly compute VaR and CVaR on several types of Czech financial time series as index PX, CEZ stock price and exchange rate EURCZK. In the conclusion we offer some inferences for practical implications for the use of heavy-tailed distributions in GARCH models for the stated purpose.

Keywords: VaR and CVaR, GARCH model, Returns of Czech Financial Assets, Comparison

JEL Classification: C13, C46
AMS Classification: 62P05, 91G15

1 Introduction

Value at Risk (VaR) and Conditional Value at Risk (CVaR) are widely used measures to assess the level of risk exposed to entities when investing in risky financial assets. The accurate assessment of risk helps them to effectively perform the risk management. It goes similarly to regulatory bodies in their activities. It is well documented that returns of financial assets are often distributed with heavy tails. They also exhibit heteroskedastic behavior. To deal with the latter in economic and financial time series, the ARCH and later GARCH models were introduced by [4] and [2]. To address the former, various distributions with heavy tails have been used in GARCH models for the error terms. For example, Dyhrberg [3] analyzes the use of t-distribution in a GARCH model for returns of bitcoin, gold and the dollar, Fan et al [5] use GED distribution in a GED-GARCH model to estimate VaR in returns of crude oil price or Kilic [8] introduces the use of NIG distribution into GARCH and FIGARCH models to model conditional volatility in exchange rate. Though attempts to deal with this matter is not infrequent, comparative analyses on their appropriate application to compute VaR and CVaR measures are still rare. To contribute to this debate, we propose an analysis in which we investigate the ability of the most often used heavy tailed distributions as t-distribution, GED distribution and NIG distribution in GARCH and GJR-GARCH models to evaluate how well they can help to properly compute VaR and CVaR. In the analysis we propose a procedure which is consequently verified on three Czech financial time series of daily frequency as Prague stock exchange index PX, CEZ stock price and exchange rate EURCZK in a time span of five years. Based on the results obtained from the analysis we offer some inferences drawn upon from the use of heavy-tailed distributions in GARCH models for computing dynamic VaR and CVaR.

2 Methodology and Data

In this section we will briefly describe the procedure how to compute VaR and CVaR using GARCH model with different types of distribution for the innovations.

[1] Prague University of Economics and Business, Department of Monetary Theory and Policy, nám. W. Churchilla 4, Praha 3, Czech Republic, tran@vse.cz
[2] Prague University of Economics and Business, Department of Banking and Insurance, nám. W. Churchilla 4, Praha 3, Czech Republic, malek@vse.cz

2.1 Value at Risk and Conditional Value at Risk

Value at risk is a metric that measures the risk exposure of of investments firms due to unfavorable market movements. Let a potential loss of an investment be a random variable X and $F_X(.)$ be the cumulative distribution function of X. Then, VaR at level α (VaR_α) is the $(1-\alpha)$-quantile of Y, e.i.,

$$\text{VaR}_\alpha(X) = -\min\{x : F_X(x) \geq \alpha\} = F_Y^{-1}(1-\alpha). \tag{1}$$

An alternative of VaR is the conditional value at risk. If X is the potential loss of an investment at some point in the future and $\alpha \in (0,1)$, then the CVaR is defined as

$$\text{CVaR}_\alpha(X) = \frac{1}{1-\alpha}\int_0^{1-\alpha} \text{VaR}_\gamma(X)\mathrm{d}\gamma. \tag{2}$$

Compared to VaR, CVaR has the advantage that it it a coherent risk measure. A coherent risk measure is the one which the following properties: monotonicity, sub-additivity, homogeneity, and translational invariance.

2.2 GARCH model

An GARCH model of a random variable y_t is a two-part econometric model in which its mean process is defined as

$$y_t = \mu + \epsilon_t, \epsilon_t = z_t\sigma_t, \tag{3}$$

where μ is the mean of y_t, $z_t \sim F(0,1)$ is the standardized residual term. The conditional variance term σ_t^2 follows GARCH process proposed by [2] as

$$\sigma_t^2 = \eta + \sum_{i=1}^{p} \theta_i \sigma_{t-i}^2 + \sum_{j=1}^{q} \lambda_j \epsilon_{t-i}^2, \tag{4}$$

where $\theta_i, i = 1,\dots,p$ and $\lambda_j, j = 1,\dots,q$ are coefficients of the GARCH and ARCH terms, and p and q are their length of lags, respectively. Parameters of a GARCH model are estimated jointly using maximum likelihood estimation method. The shape of the likelihood function depends on the type of distribution of term z_t.

2.3 Heavy tailed distributions

In this analysis, three heavy tailed distribution are used to study: generalized error distribution (GED), Student-t distribution and normal inverse gaussian distribution (NIG). A brief summary of their main characteristics is provided.

Student t distribution: The Student t distribution has three parameters: location μ, scale σ and the number of degrees of freedom ν. Parameter ν just defines the tail property of this distribution. The higher value ν attains, the lower value the kurtosis of this distribution is. Its PDF is

$$\mathrm{f}(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{(x-\mu)^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}} \tag{5}$$

and the corresponding CDF is

$$\mathrm{F}(x) = \frac{1}{2} + \frac{x-\mu}{\sigma}\Gamma\left(\frac{\nu+1}{2}\right)\frac{{}_2\mathrm{F}_1\left(\frac{1}{2}, \frac{\nu+1}{2}; \frac{3}{2}; -\frac{(x-\mu)^2}{\nu\sigma^2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \tag{6}$$

where Γ is the so called gamma function and ${}_2\mathrm{F}_1$ is the hypergeometric function.

Generalized error distribution: It is also called generalized normal distribution which is often used to capture the tail and the peak behavior of a distribution differing from the corresponding normal distribution. It has three parameters: location μ, scale σ and shape ξ. A normal distribution has $\xi = 2$ and a Laplace distribution has $\xi = 1$. Its PDF function is

$$\mathrm{f}(x) = \frac{\xi}{2\sigma\Gamma(1/\xi)} e^{-\left(\frac{|x-\mu|}{\sigma}\right)^\xi}, \tag{7}$$

and the CDF is

$$F(x) = \frac{1}{2} + \operatorname{sign}(x-\mu)\frac{\gamma\left(1/\xi, \left(\frac{|x-\mu|}{\sigma}\right)^{\xi}\right)}{2\Gamma(1/\xi)}, \tag{8}$$

where γ is the lower incomplete gamma function.

Normal Inverse Gaussian distribution: Normal Inverse Gaussian distribution is a special case of generalized hyperbolic distribution which was first introduced by Barndorff-Nielsen (1977), [1]. It has four parameters: location μ, scale σ, tail power α and skewness β. It PDF is

$$f(x) = \frac{\alpha\delta K_1\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right)}{\pi\sqrt{\delta^2 + (x-\mu)^2}} e^{\delta\gamma+\beta(x-\mu)}, \tag{9}$$

where K_1 is the modified Bessel function of the third kind and $\gamma = \sqrt{\alpha^2 - \beta^2}$. Unfortunately, the corresponding CDF of NIG distribution in a closed form is unknown. This distribution is often used instead of the general case as the estimation of parameters of the generalized hyperbolic distribution is complicated by the the multimodal objective function problem. The maximum likelihood estimation procedure is performed as follows:

$$\hat{\theta} = -\arg\min_{\theta\in\Theta} \sum_{i=1}^{n} \ln f(X_i;\theta), \tag{10}$$

where Θ is the set of all admissible parameters and $f(X_i;\theta)$ is the density function of the corresponding distribution.

2.4 Data

For this analysis we use daily data in the span of five years from 2017 to 2022. The dataset consists of Prague Stock Exchange Index PX, CEZ stock price series and EURCZK exchange rate. The number of observations is chosen to have series of returns long enough for a stable numerical MLE estimation procedure. The original series are converted into series of logarithmic returns. Their descriptive statistics are shown in Table 1 and the dynamics of the returns can be seen in Figures 1 – 3. One can observe that CEZ returns are the most volatile and the least volatile ones are of exchange rate EURCZK. Also, all series of returns have higher value of kurtosis compared to the value for a series normally distributed which justifies the use of heavy tailed distributions in the corresponding models.

Characteristic	PX	Returns	CEZ	Returns	EURCZK	returns
Mean	975.03	8.77e-05	513.49	-2.42e-4	25.805	-7.62e-05
Minimum	790.09	-0.0471	364.1	-0.0908	24.218	-0.0171
Maximum	1140	0.0447	710.9	0.0626	27.815	0.0386
Std deviation	61.901	8.77e-3	81.965	0.0143	0.6064	3.18e-3
Skewness	-0.2716	-0.4255	0.1889	-0.4225	1.9721	1.9721
Kurtosis	2.894	5.7424	1.7523	6.7438	4.2687	26.411
Obs.	1302	1301	1302	1301	1302	1301

Table 1 Descriptive statistics of analyzed assets and their returns

3 Experiment and Results

First, theseriesoflogarithmicreturnsareusedtoestimateparametersofGARCHmodels. We use two specifications for GARCH models: the symmetric one GARCH(1,1) and the so called GJR-GARCH(1,1) which is able to capture the leverage effect that is the asset's volatility tendency to be negatively correlated with its returns [6]. Both length of lag is set at 1 following the recommendation suggested by [7]. To investigate the suitability of both specifications and all three heavy tailed distributions, all combinations are estimated for each series. The estimation has been done with our own programs written for this purpose in MATLAB as, to our information, no official econometric software package has provided software for GARCH model with NIG distributed innovations. The numerical estimation procedure is fast and stable with converging results. The estimation results of all models in terms of log-likelihood values are displayed in Table 2. The values in bold is the best model for the corresponding series.

Model	Distribution	PX	CEZ	EURCZK
GARCH(1,1)	GED	4426.4	3785.3	**6039.5**
	t	4411.9	3775.8	6024.4
	NIG	4429.7	**3798.6**	6018.2
GJR-GARCH(1,1)	GED	4427.4	3788.7	6040.7
	t	**4432.9**	3787.7	6030.2
	NIG	4429.8	3798.7	6018.3

Table 2 The log-likelihood values of two GARCH models with three distributions

The results in Table 2 show that across assets the less volatile a series is, the better a model can approximate judging by the log-likelihood values. Across models, their suitability is evaluated either by likelihood ratio test if they are nested or by the Bayes-Schwartz information criterion when they are not nested. Judging by these measures, we can see there is no distribution which is good for all investigated series. Rather, in our analysis, each series has its own best distribution for a GARCH model. The best one for series PX is the GJR-GARCH(1,1) with t-distributed innovations, for series CEZ is the GARCH(1,1) with NIG distributed error terms and for series EURCZK the best one is a GARCH(1,1) model with GED distributed error terms. Further, one can also observe that with NIG distribution which has four parameters and can be asymmetric, then adding the leverage effect term into the GARCH model does not help to increase the goodness of fit as it is already captured by the shape parameters of the distribution. Due to the limited space of the article, we show only the most relevant model for each series. Their parameters are shown in Table 3. The three stars indicate the statistical significance at level 1%.

	PX -GJR-GARCH (t)		CEZ - GARCH (NIG)		EURCZK - GARCH (GED)	
Parameter	Coefficient	SE	Coefficient	SE	Coefficient	SE
$\alpha/\nu/\xi$	7.0998	0.0316***	67.8733	0.3162***	1.2961	0.0301***
β	-	-	-5.7553	0.3061***	-	-
μ	2.917e-4	2.106e-4	9.433e-4	3.498e-4***	-8.608e-5	5.484e-5
η	1.139e-6	4.881e-7***	3.99e-06	3.41e-06	1.09e-07	5.87e-08***
θ	0.8949	0.0188***	0.7864	0.0276***	0.8615	0.0179***
λ	0.0835	0.0013***	0.2135	0.0253***	0.1384	0.0230***
κ	0.0432	0.0220**	-	-	-	-

Table 3 Estimation results of GARCH model for returns of three assets

The results in Table 3 show that the all shape parameters in every model are statistically significant and so are the parameters of GARCH model except the constant in some cases. This justifies the use of GARCH models to model the heteroscedasticity in returns of the series in the analysis. The negative skew can be captured by a GARCH model with leverage effect or by a negative asymmetry parameter in the NIG distribution which is in line with the descriptive statistics. When the skewness is positive, it remains unnoticed. The estimation results then are used to compute the time changing VaR and CVaR of the three series according to (1) and (2).

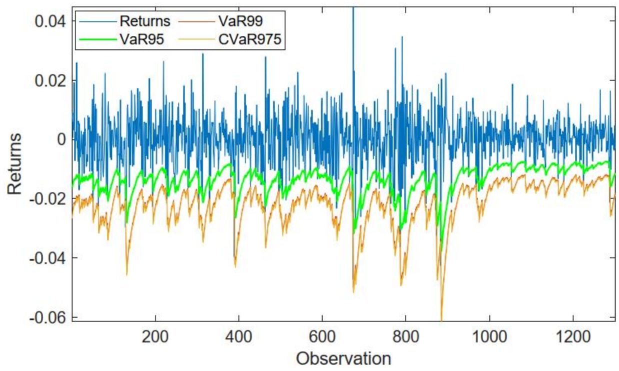


Figure 1 Computed values of VaR95, VaR99 and CVaR975 for PX returns series

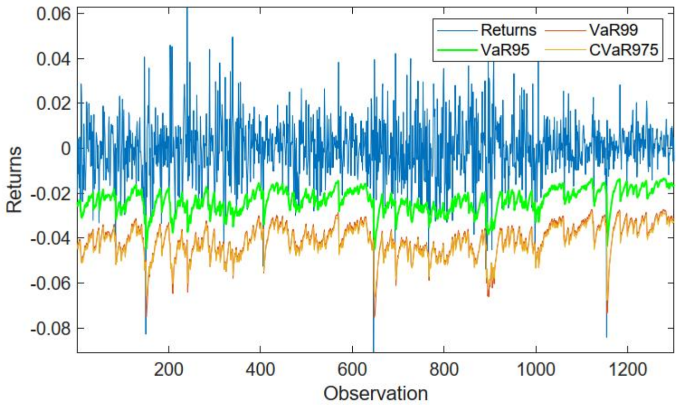


Figure 2 Computed values of VaR95, VaR99 and CVaR975 for CEZ returns series

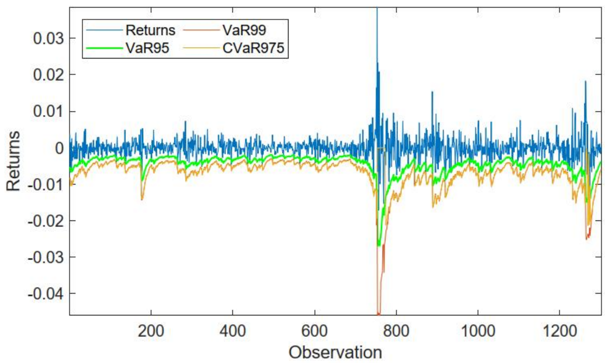


Figure 3 Computed values of VaR95, VaR99 and CVaR975 for EURCZK returns series

First, the conditional time-varying standard deviations are generated using the estimated values of the GARCH models. We use them together with other parameters of the corresponding distribution to compute VaR at $\alpha = 0.95$ (VaR95) and $\alpha = 0.99$ and CVaR at $\alpha = 0.975$ (CVaR975) for each point inside the time span of returns series. Computing VaR and CVaR from a GARCH model with t distributed innovation is easy and fast as CDF and inverse CDF of t distribution exist. Though CDF of GED distribution also exists, it is a bit less comfortable as the evaluation of the gamma incomplete function is slower. If there is no CDF and inverse CDF as in the case of NIG distribution, VaR and CVaR are computed with numerical integration and interpolation if needed. This procedure is numerically stable and relatively accurate, but much slower. The computed values of VaR and CVaR are displayed in Figures 1 – 3. Some of their descriptive statistics are also calculated and shown in Table 4.

	PX			CEZ			EURCZK		
Feature	VaR95	VaR99	CVaR975	VaR95	VaR99	CVaR975	VaR95	VaR99	CVaR975
Mean	-0.0137	-0.0218	-0.0224	-0.0232	-0.0407	-0.0416	-0.0046	-0.0073	-0.0067
Minimum	-0.0377	-0.0597	-0.0614	-0.0453	-0.0756	-0.0677	-0.0268	-0.0460	-0.0213
Maximum	-0.0074	-0.0118	-0.0122	-0.0134	-0.0272	-0.0285	-0.0019	-0.0030	-4.8e-06
Std dev	0.0047	0.0074	0.0076	0.0052	0.0074	0.0067	0.0030	0.0051	0.0032

Table 4 Descriptive statistics of analyzed assets and their returns

The results in Table 3 show that the all shape parameters in every model are statistically significant and so are the parameters of GARCH model except the constant in some cases. This justifies the use of GARCH models to model the heteroscedasticity in returns of the series in the analysis. Further, the negative skewness in returns can be captured by a GARCH model with leverage effect or by a negative shape parameter for asymmetry in the NIG distribution which is in line with the descriptive statistics. When the skewness is positive, it remains unnoticed. Next, the computed values of VaR at $\alpha = 0.95$ and $\alpha = 0.99$ as well as of CVaR at $\alpha = 0.975$ seem to be consistent with the dynamics of returns of the three series in our analysis. This show that the dynamic approach of computing VaR and CVaR with the use of GARCH models is more suitable than the static one used in our previous studies [9]. When it is static, the number occasions exceeding the threshold is roughly α%. The dynamic approach can more precisely identify the risk exposure of investment in these three assets due to taking into account the time variant volatility nature of their returns. The number of occasions may substantially differ from α%. Finally, the computed values of VaR99 and CVaR975 are quite close, but not equal as one can see in Figures 1 – 3. This observation is confirmed by the descriptive statistics of the computed values of VaR99 and CVaR975 shown in Table 4. Computed series of VaR and CVaR are used to investigate the impact of distributional assumption for GARCH models and the role of an extension of the basic GARCH model. We use the two sample t-test for equal mean to do this task. If they have no effect on them, the means should be equal. The benchmark is the specification with best log-likelihood value which is compared to all remaining specifications within one series. The results of the test shows that the inclusion of the GJR term into the basic GARCH model does not improve the values of computed VaR and CVaR (we cannot reject the null hypothesis of equal means of two series), while distributional mispecification in a GARCH model leads to statistically significantly different values of VaR and CVaR on average (the null hypothesis of equal means is rejected and those computed values of VaR and CVaR tend to be overestimated compared to those derived from a GARCH model with "best" distributional assumption).

4 Conclusion

We have performed an investigation in which we analyze the use of two GARCH models with three different types of distribution of the innovation terms to compute the time variant VaR and CVaR of daily returns of the Czech stock market index PX, CEZ stock price and exchange rate EURCZK. The results of this analysis are following. First, we have found that there is no one-fits-all solution in terms of which distribution is the best option to be used in a GARCH model for all three Czech financial time series of returns and distributional misspecification idoes affect the VaR and CVaR values. Next, if there is a negative skewness in the distribution of returns, in a GARCH model it can be captured by a GARCH model with a leverage effect or using a distribution with a shape parameter for the negative skewness. However, it does not improve the computed VaR and CVaR values on average which is in line with the results if [7]. Finally, time variant VaR and CVaR computed with the used of GARCH models would be a preferred approach compared to the static one as it more reliably detect the risk exposure of investment in risky assets. It is possible to include the confidence interval to computed VaR and CVaR series as in [10] which is the subject of a further extension in our future research.

Acknowledgements

The authors of this paper acknowledge the financial support under the framework of Institutional Support IP 100040 of Prague University of Economics and Business and of grant GACR 22-19617S.

References

[1] Barndorff-Nielsen, O. E. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society London* A 353, 401–419.

[2] Bollerslev. T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 107–327.
[3] Dyhrberg, A. H. (2016). Bitcoin, gold and the dollar, a garch volatility analysis. *Finance Research Letters*, 16, 85–92.
[4] Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987–1007.
[5] Fan, Y., Zhang, Y. J., Tsai, H. T., & Wei, Y. M. (2008). Estimating Value at Risk of crude oil price and its spillover effect using the GED-GARCH approach. *Energy Economics*, 30(6), 3156-3171.
[6] Glosten, L. , Jagannathan, R., & Runkle, D. (1993). Relationship between the expected value and volatility of the nominal excess returns on stocks. *Journal of Finance*, 48, 1779–1802.
[7] Hansen P. R. & Linde, A. (2005). A forecast comparison of volatility models: Does anything beat a garch(1,1)? *Journal of Applied Econometrics*, 20 (7), 873–889.
[8] Kiliç, R. (2007). Conditional volatility and distribution of exchange rates: GARCH and FIGARCH models with NIG distribution. *Studies in Nonlinear Dynamics & Econometrics*, 11(3), 1430.
[9] Málek, J. & Van Tran, Q. (2017). Value at Risk calculated with α-stable distribution for Czech stock market index PX. *Theoretical and Practical Aspects of Public Finance* 2017, 134.
[10] Spierdijk, L. (2016). Confidence intervals for ARMA–GARCH value-at-risk: The case of heavy tails and skewness. *Computational statistics & data analysis*, 100, 545-559.

Determinants of International Tourism Inbound Receipts: The Quantile Regression Approach

Petra Vašaničová[1], Sylvia Jenčová[2]

Abstract. Challenges in tourism arising from the COVID-19 pandemic, e.g., a decline in domestic and international travel, are forcing destinations to focus on improving tourism performance. Therefore, it is important for stakeholders to know the key determinants affecting tourism performance. This paper aims to find out whether a country's international tourism inbound receipts are determined by GDP, the number of international arrivals, and travel and tourism competitiveness. The proposed model for 125 countries is specific because we consider conditional quantiles of the dependent variable. The advantage of quantile regression is that 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 contributes to the existing literature that includes TTCI as an independent variable in tourism performance models.

Keywords: tourism, receipts, arrivals, GDP, competitiveness, quantile regression

JEL Classification: L83, Z33, Z31
AMS Classification: 62G08

1 Introduction

The growing importance of the tourism sector is of interest to researchers and policymakers in order to assess and understand the drivers of the sector's performance over time and across countries and destinations. The country's destination management and tourism companies are trying to improve the overall tour-ism sector performance and its partial determinants. Therefore, stakeholders need to know the key factors influencing tourism performance.

Destination performance can be quantified using indicators such as the number of tourist arrivals, the number of nights spent, tourism inbound receipts, average receipts per arrival (per tourist) [5, 13], travel and tourism industry GDP, travel and tourism industry employment [2]. According to [1], the size of the population correlates positively with the number of tourists arrivals. The reason is that the destination's higher population tends to include more friends and relatives, and it connects with the theory that people tend to visit places where they have friends and relatives more than those where they do not. There exists a theory [15] that destinations with larger populations tend to be cheaper and thus attract tourists more.

In terms of the country's economic growth and development, total income may be more important than the number of arrivals [6]. E.g., revenues from international arrivals capture the number of visitors but also reflect the length of their stay and their economic benefits to the country as a destination. This fact is important mainly for countries that see international tourism as an engine of eco-nomic growth or a means of regional development.

When a destination is able to accumulate revenue from visitors, then it is competitive with other destinations. Therefore, competitiveness is also associated with destination performance. This is in line with the main goals of the tourism sector, which are to increase the quality of life, maximize profits and maintain competitiveness. This can be achieved by strengthening the sector growth, by increasing the quality of services and products offered, the quality of stakeholders, and the general macroeconomic situation [5]. The formulation of public tourism policy can be directly based on the identification of the attributes that have the most significant impact on tourism performance.

This paper aims to find out whether a country's international tourism inbound receipts are determined by GDP, the number of international arrivals, and travel and tourism competitiveness. For this purpose, we use quantile regression (QR) and compare the results with ordinary least squares linear model (OLS). The classical linear regression model estimates how, on average, individual independent variables affect a dependent variable. The

[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, sylvia.jencova@unipo.sk.

regression coefficients obtained by quantile regression estimate the change in a given quantile of the explained variable, which is caused by a one-unit change of the independent variable, fixing all the other independent variables.

2 Data and Methodology

2.1 Data

We use indicators from the 2019 Travel and Tourism Competitiveness Report [2] for 125 countries to create a regression model. Specifically, we use international tourism inbound receipts (*ITIR*) (in million USD), the number of international arrivals (*IA*) (in thousand) and Travel and Tourism Competitiveness Index (*TTCI*). Although the Travel and Tourism Competitiveness Report is published for 2019, it uses older indicators because the creation of such a comprehensive indicator as the TTCI takes some time (according to the periodicity of the publication, it is two years). Therefore, the country's gross domestic product (*GDP*) (at constant prices in 2015 in USD) was obtained for 2017 from the World Bank database.

2.2 Quantile Regression

In this Section, we describe basic concept of the quantile regression methodology, according to [7, p. 25]. In the standard linear regression model

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

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

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

where $X_i = \left(X_{i1}, \ldots, X_{ip} \right)^T$ denotes the i-th observation and ρ_τ (defined in [9] 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 [11]:

$$Y_i = \beta_0 + \boldsymbol{\beta}_\tau \mathbf{X}_{i.} + \varepsilon_{i\tau}, \quad i = 1, \ldots, n, \tag{5}$$

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 [4]. The advantage of QR is that it is the most suitable tool for modeling heteroscedastic data [7, p. 25; 9].

To meet the aim of this paper, the model for the OLS is:

$$\ln ITIR_i = \beta_0 + \beta_1 \ln GDP_i + \beta_2 \ln IA_i + \beta_3 TTCI_i + \varepsilon_i, \quad i = 1, \ldots, n \tag{6}$$

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. Given the inherent variability between countries in terms of international arrivals, GDP, and international tourism inbound receipts, a logarithmic transformation of these variables was needed to avoid undesirable heteroskedasticity in OLS. We test the presence of heteroscedasticity by Breusch-Pagan test. To detect multicollinearity, we use variance inflation factor (VIF). 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 [9, 10]. To test whether the slope coefficients of the models are identical, we use ANOVA and the anova.rq package.

3 Results and Discussion

In Table 1, we present the estimates of QR and OLS models. Moreover, we present the ANOVA test detecting that QR estimates significantly differ across quantiles. Figure 1 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 [3]. 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 [17]. As is stated in [3, 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 ITIR values compared to the OLS estimate.

Model	**Dependent variable: ln*ITIR***							
	Independent variables							
Quantile	**Intercept**	***p*-value**	**ln*GDP***	***p*-value**	**ln*IA***	***p*-value**	***TTCI***	***p*-value**
0.05	-7.5970	**0.0091**	0.1055	0.5206	0.5853	0.1517	1.6664	**0.0026**
0.10	-5.9868	**0.0000**	0.1623	0.1740	0.3763	0.1333	1.5050	**0.0010**
0.15	-5.4958	**0.0129**	0.1428	0.2753	0.4696	**0.0224**	1.3478	**0.0012**
0.20	-4.5355	**0.0009**	0.1141	0.1114	0.5001	**0.0000**	1.2550	**0.0000**
0.25	-4.3403	**0.0000**	0.1141	**0.0425**	0.5554	**0.0000**	1.1120	**0.0000**
0.30	-4.2002	**0.0000**	0.1266	**0.0153**	0.5293	**0.0000**	1.0760	**0.0000**
0.35	-4.3353	**0.0000**	0.1443	**0.0000**	0.5461	**0.0000**	0.9757	**0.0000**
0.40	-4.9164	**0.0000**	0.1831	**0.0003**	0.5104	**0.0000**	0.9631	**0.0000**
0.45	-4.5315	**0.0000**	0.1769	**0.0039**	0.5938	**0.0000**	0.7665	**0.0001**
0.50	-4.3459	**0.0000**	0.1810	**0.0018**	0.6535	**0.0000**	0.5891	**0.0010**
0.55	-4.2840	**0.0000**	0.1874	**0.0011**	0.6360	**0.0000**	0.5855	**0.0009**
0.60	-4.1541	**0.0000**	0.1812	**0.0007**	0.6010	**0.0000**	0.6794	**0.0001**
0.65	-5.0174	**0.0000**	0.2160	**0.0000**	0.5665	**0.0000**	0.7760	**0.0000**
0.70	-5.3573	**0.0000**	0.2277	**0.0000**	0.5565	**0.0000**	0.8188	**0.0000**
0.75	-5.5413	**0.0000**	0.2429	**0.0000**	0.5188	**0.0000**	0.8635	**0.0000**
0.80	-5.1960	**0.0000**	0.2280	**0.0006**	0.4834	**0.0000**	0.9603	**0.0000**
0.85	-5.2343	**0.0000**	0.2566	**0.0002**	0.4512	**0.0000**	0.8778	**0.0000**
0.90	-4.0039	**0.0002**	0.2315	**0.0001**	0.4593	**0.0000**	0.7646	**0.0000**
0.95	-3.6299	**0.0032**	0.1989	0.0849	0.5466	**0.0000**	0.7270	**0.0397**
ANOVA *p*-value = 0.0000								
OLS	-6.4694	**0.0000**	0.2328	**0.0001**	0,5257	**0.0000**	1.0376	**0.0000**
VIF			2.6870		3.3190		3.9240	
BP = 8.5075 (p-value = 0.0366); R^2 = 0.8571								

Table 1 Estimates of model parameters

Note: The P-values marked bold indicate the statistical significance at the significance level of 0.05.

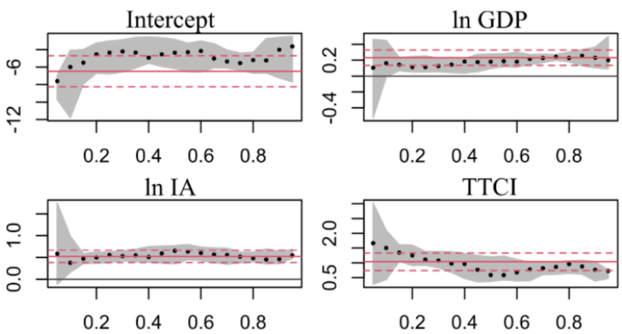


Figure 1 Estimates of model parameters by quantile level

The regression model parameter estimates obtained using OLS were statistically significant for all considered independent variables, multicollinearity was not present (VIF < 10), and the model explained up to 85.71% of the variability of the lnITIR. However, we indicated the presence of heteroskedasticity, which we confirm through the Breuch-Pagan test (BP = 8.5075, p = 0.0366). Therefore, the use of quantile regression is justified.

The results of QR show that lnGDP is not statistically significant from $\tau = 0.05$ to $\tau = 0.20$; lnIA is not statistically significant only for $\tau = 0.05$ and $\tau = 0.10$. We show that a country's international tourism inbound receipts are determined by GDP, the number of international arrivals, and travel and tourism competitiveness. Through quantile regression, we found out which percentiles of lnITIR may be more affected by TTCI (we see high coefficients for low values of quantiles), by lnGDP (we see high coefficients from the 65th percentile), and by lnIA (we see high coefficients around the median).

Table 2 shows the descending order of countries according to the value of lnITIR. Moreover, to better interpretation of QR results, we denote, which country represent one of five quantiles ($\tau = 0.05, 0.25, 0.50, 0.75, 0.95$). To interpret the results, e.g., for the median ($\tau = 0.50$; in Table 2: Island), we see that the change of GDP by 1% will be associated with a 0.18% change in ITIR (fixing all the other independent variables). If the IA increases by 1%, the ITIR will increase by 0.65%. A change in the value of the TTCI indicator by one unit will be associated with an 80.24% change in ITIR.

ID	Country	ID	Country	ID	Country	ID	Country	ID	Country	ID	Country
1	USA	22	NLD	43	HUN	**64**	**ISL ($\tau = 0.50$)**	85	LVA	106	SEN
2	ESP	23	CHE	44	QAT	65	SVK	86	UGA	107	PAK
3	FRA	24	SWE	45	BRA	66	JAM	87	KEN	108	BGD
4	THA	25	KOR	46	IRL	67	SVN	88	SLV	109	MKD
5	GBR	26	POL	47	NOR	68	GEO	89	GHA	110	MDA
6	ITA	27	IDN	48	ARG	69	URY	90	NIC	111	KWT
7	**AUS ($\tau = 0.95$)**	28	BEL	49	COL	70	NGA	91	BOL	112	MLI
8	DEU	29	SAU	50	JOR	71	ROU	92	LAO	113	NAM
9	JPN	30	HRV	51	LUX	72	TZA	93	HND	114	MOZ
10	HKG	31	NZL	52	PAN	73	ALB	94	BWA	115	ZWE
11	CHN	**32**	**RUS ($\tau = 0.75$)**	53	BGR	74	KAZ	**95**	**ZMB ($\tau = 0.25$)**	116	DZA
12	IND	33	VEN	54	LKA	75	OMN	96	NPL	117	GMB
13	TUR	34	ZAF	55	CRI	76	MUS	97	PRY	118	YEM
14	MEX	35	DNK	56	IRN	77	MLT	98	CMR	119	SLE
15	ARE	36	EGY	57	PER	78	EST	99	TTO	**120**	**MWI ($\tau = 0.05$)**
16	AUT	37	LBN	58	BHR	79	GTM	100	RWA	121	TCD
17	CAN	38	MAR	59	KHM	80	SRB	101	ETH	122	LSO
18	SGP	39	DOM	60	CHL	81	LTU	102	CPV	123	MRT
19	MYS	40	PHL	61	FIN	82	TUN	103	KGZ	124	TJK
20	PRT	41	CZE	62	CYP	83	ARM	104	MNG	125	BDI
21	GRC	42	ISR	63	AZE	84	MNE	105	CIV		

Table 2 Order of countries according to the value of ln*ITIR* (descending)

Note: We use a standard defining codes for the names of countries (ISO 3166-1). Cell colors indicates these regions: orange – Asia-Pacific, blue – Europe and Eurasia, yellow – The Americas, green – Middle East and North Africa, violet – Sub-Saharan Africa

In creating the regression model, we also relied on the existing literature on the destination competitiveness. According to [18], there is a presumption that more competitive destinations attract more visitors, and visitors spend more money in such destinations. Then destination GDP and economic growth increase, and thus the economic well-being of the local population increases, too. Authors argue that this may not be true because more visitors to the destination do not necessarily mean more money they spend, nor that more money visitors spend will generate economic growth. Moreover, they state that this may not be the case because more visitors to the destination do not necessarily mean more money they spend, nor that more money visitors spend will generate economic growth. In this paper, we did not consider causal dependencies. We focused on the use of some quantitative indicators of tourism performance from the mentioned study, and we added an indicator measuring the destination competitiveness.

The estimated model confirmed the results of a study by Naude and Saayman [14] that the independent variable GDP is a statistically significant determinant affecting tourism performance. Tourism performance appears to be sensitive to economic growth. QR using data of Travel and Tourism Competitiveness Index and its pillars and subindexes was already realized in [8, 12, 16, 17]. Our results have also contributed to the existing literature, e.g., [1], which includes TTCI as an independent variable in tourism performance models.

4 Conclusion

The proposed model for 125 countries is specific because we consider conditional quantiles when modeling the performance of the tourism sector. The advantage of quantile regression is that 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 has several limitations. In OLS and QR, we do not consider the assumption of the exogeneity of the random error. The presence of a fixed effect of individual countries can be expected. Moreover, random selection cannot be expected, as individual countries will be interconnected by unobserved factors. This analysis could be done on panel data; unfortunately, the methodology for calculating the TTCI changes over the years, and, therefore, it is not possible to use objective data for several years. A more detailed analysis should be done in future research.

Acknowledgements

This paper was supported by the research grant KEGA no. 001PU-4/2022 “Application of Modern Trends in Quantitative Methods in the Teaching of Financial and Managerial Subjects”.

References

[1] Bazargani, R. H. Z., & Kiliç, H. (2021). Tourism competitiveness and tourism sector performance: Empirical insights from new data. *Journal of Hospitality and Tourism Management*, 46, 73–82.
[2] Calderwood, L. U. & Soshkin, M. (2019). *The Travel & Tourism Competitiveness Report 2019. Travel and Tourism at a Tipping Point*. Geneva: World Economic Forum.
[3] 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.
[4] Cupák, A., Pokrivčák, J., & Rizov, M. (2016). Diverzita spotreby potravín na Slovensku. *Politická ekonomie*, 64(5), 608–626.
[5] Hanafiah, M. H., & Zulkifly, M. I. (2019). Tourism destination competitiveness and tourism performance:A secondary data approach. *Competitiveness Review: An International Business Journal*, 29(5), 592–621.
[6] Joshi, O., Poudyal, N. C., & Larson, L. R. (2017). The influence of sociopolitical, natural, and cultural factors on international tourism growth: a cross-country panel analysis. *Environment, Development and Sustainability*, 19(3), 825–838.
[7] 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.
[8] Kalina, J., Vašaničová, P., & Litavcová, E. (2019). Regression Quantiles under Heteroscedasticity and Multicollinearity: Analysis of Travel and Tourism Competitiveness. *Ekonomický časopis*, 67(1), 69–85.
[9] Koenker, R. (2005). *Quantile Regression*. New York: Cambridge University Press.
[10] Koenker, R., Chernozhukov, V., He, X., & Peng, L. (2017). *Handbook of Quantile Regression*. Boca Ranton: CCR Press.
[11] Kováč, Š. (2013). Vybrané faktory predlženosti podnikov v podmienkach SR. *Forum Statisticum Slovacum*, 7, 79–85.
[12] Litavcová, E., Vašaničová, P., Jenčová, S., & Košíková, M. (2019). Travel and tourism competitiveness index 2017: quantile regression approach of enabling environment pillars. In R. Němec & L. Chytilová (Eds.) *Proceedings of the 13th International Conference on Strategic Management and its Support by Information Systems 2019* (pp. 187–195). Ostrava: VŠB - Technická univerzita Ostrava.
[13] Maráková, V., Dyr, T., & Wolak-Tuzimek, A. (2016). Factors of tourism's competitiveness in the European union countries. *E&M Economics and Management*, 19(3), 92–109.
[14] Naudé, W. A., & Saayman, A. (2005). Determinants of tourist arrivals in Africa: a panel data regression analysis. *Tourism economics*, 11(3), 365–391.
[15] Takahashi, K. (2020). Comparing the determinants of tourism demand in Singapure and French Polynesia: appliying the tourism demand Model to Panel Data Analysis. *Tourism analysis*, 25(1), 175–181.
[16] Vašaničová, P., Litavcová, E., Jenčová, S., & Košíková, M. (2017). Dependencies between Travel and Tourism Competitiveness Subindexes: The Robust Quantile Regression Approach. In T. Löster & T.

Pavelka (Eds.) *The 11th International Days of Statistics and Economics Conference Proceedings* (pp. 1729–1739). Slaný: MELANDRIUM.

[17] 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.

[18] Webster, C., & Ivanov, S. (2014). Transforming competitiveness into economic benefits: Does tourism stimulate economic growth in more competitive destinations?. *Tourism Management*, 40, 137–140.

The Optimal Settings of a Genetic Algorithm for Variable Selection in a Non-linear Time Series Model

Lukáš Veverka[1]

Abstract. Common methods for a variable selection in linear regression do not work in models including non-linearity because of the different ranges of the values of the estimated parameters for a single variable. The binary version of a genetic algorithm comes in handy for this purpose. Since the performance of both genetic algorithm and non-linear optimization is sensitive to the setting, it is convenient to look for an optimal setting to reach the best result. However, in the early stages of the model definition, it is practical to limit the estimation time to reveal some misspecifications of the model (e.g. forgotten seasonality). Therefore, the optimal settings for a specific regression case are found with repeated cross-validation selecting various combinations of settings for different time categories.

Keywords: Genetic algorithm, Variable selection, Non-linear optimization

JEL Classification: C22
AMS Classification: 68W50

1 Introduction

Variable selection has been an important topic in econometrics for a long time. There are lots of ways how to approach it and among the most basic ones belongs the general to specific method. It is used by [7] who reduce the number of variables based on various diagnostic tests. A comparison study by [1] shows results of different approaches for a variable selection in various proximity criteria. The work by [4] introduces algorithms based on dropping columns which are run in parallel on a shared-memory machine. It is then extended in the paper [3] where methods avoiding the evaluation of all possible subsets of variables are described. It introduces a possibility to cut certain branches in the regression tree and therefore reduce the computational demands. [5] reveal that when the number of variables is set in advance, then there are algorithms that can outperform the existing branch-and-bound algorithms. However, there are two classes of variable selection methods – sequential testing and information criteria described by [6] who proposes a method belonging to the information criteria class based on a genetic algorithm for variable selection. Genetic algorithms are utilized also by [10] who focuses on non-convex mixed-integer nonlinear programming problems in econometrics. Moreover, the genetic algorithm is not used in econometrics only for variable selection. [8] utilize it in the combination with support vector machines to model the volatility of stock markets.

2 Motivation

Let us study the effect of the media activities of all companies in a selected market segment on the sales of one company. In that case, the number of possible input variables and therefore the degrees of freedom are often greater than the number of observations causing the problem of an under-identified model. This might be solved with shrinkage methods among which belongs the ridge regression. It copes with detecting the most suitable input variables. However, from the formula (1), where y_i is the dependent variable and i stands for the number of observations (in this case every time record), $\beta_0 \dots \beta_j$ are regression parameters to be estimated, x_j are independent variables and λ is the penalization parameter, it is possible to see that the values of β_j play a great role in the penalization.

$$\sum_{i=1}^{T}\left(y_i - \beta_0 + \sum_{j=1}^{N}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{N}\beta_j^2 \rightarrow \text{MIN} \tag{1}$$

Therefore the common approach is to standardize all $x_i j$. However, when it is allowed to estimate more parameters for one variable which can differ in ranges of values significantly (e.g. in a non-linear econometrics model), the

[1] University of Economics, Prague, Department of Econometrics,Winston Churchill Square 4, CZ13067 Prague, Czech Republic, vevl00@vse.cz

ridge regression cannot cope with that anymore. One example of such a situation is an s-shape function described in equation (2).

$$f(x_{it}) = \frac{\beta x_{it}^3}{x_{it}^3 + \rho^3}, \tag{2}$$

where $x \geq 0$ and β imply the direct impact on the dependent variable and therefore the values depend mostly on the range of values of the dependent variable. On the other hand, ρ describes the slope of the s-shape and therefore depends mainly on the values of the independent variable.

3 Methodology

To eliminate the problem of an under-identified model, the binary version of a genetic algorithm is used for the variable selection. A binary genetic algorithm is specific by having only Boolean values in chromosomes. For more information about a textbook treatment of genetic algorithms, see e.g. [2].

The goal of this section is to estimate the parameters of the model and select variables. Note that it is not intended to focus on the correct selection of the response function which is undoubtedly interesting but beyond the scope of this research. In the case of different response functions, there is no limitation in this methodology to switch among them.

Firstly, let us introduce the model for the estimation of the effects of any media activities on sales of a certain company. It includes the lagged effect of communication known as the carry-over effect [9] described in equation (3) and three possible response functions to be set manually depending on the nature of the independent variable

$$x_{it} = \sum_{j=0}^{t} (\phi_i^j z_{i(t-j)}) \tag{3}$$

where ϕ_i represents the carry-over effect and therefore its value must be in the range of $\phi_i \in \langle 0, 1)$, z_{it} stands for media investments and x_{it}, therefore, represents media investments including the carry-over effect and enters the equation (4).

$$y_t = \alpha + \sum_{i=1}^{n} \lambda_i f_i(x_{it}) + u_t \begin{cases} f_i(x_{it}) = \frac{\beta x_{it}^3}{x_{it}^3 + \rho^3} \\ f_i(x_{it}) = \beta x_{it} \\ f_i(x_{it}) = \beta \log(x_{it}) \end{cases}, \tag{4}$$

where y_t represents sales in time t, λ_i is Boolean indicating whether the variable i is included in the model or not and response function $f_i(x_{it})$ is set for each variable i to be one of the possible functions. The non-linear optimization, therefore, incorporates estimation of the parameters: β, ϕ, ρ depending on the chosen response function. The optimization is then nested inside the genetic algorithm which optimizes the Boolean λ_i variable. The loss function of the whole model estimated by these two nested methods describes the equation (5)

$$L(\hat{y}_t, y_t) = \begin{cases} -R_{adj}^2 & \text{if DF} \leq T \\ -\infty & \text{otherwise.} \end{cases}, \tag{5}$$

where DF represent the degrees of freedom that depends on the chosen response function for each variable and T is the number of observations through time. Since both genetic algorithm and non-linear optimization minimize the loss function, the R_{adj}^2 is made to be negative. Moreover, the loss function is yet limited on the adjusted coefficient of determination R_{adj}^2. In future research, it will be transformed to AIC (Akaike information criterion) which is more suitable in the case of variable selection. AIC works based on the likelihood function which will be the next step in the estimation of the standard error and on top of that it allows to switch to a distribution different from normal which is here assumed.

The combination of genetic algorithm and non-linear optimization contains various tuning parameters. In the case of non-linear optimization, the maximal optimization time (in seconds) is considered. As for the genetic algorithm, the population size, elite ratio, and the number of iterations are taken into account.

Since it is hard to reach the optimal solution for a model which has lots of possible combinations of included independent variables and parameters of the nonlinear functions, it is possible to find either an acceptable solution or to let it run for a really long time. Therefore, it is worth finding the optimal combination of tuning parameters when they reach the result fastest. Since the process of any model specification and creation is often interrupted by human mistakes or misspecification of the model due to some missing variables (e.g. seasonality), it is reasonable not to spend much time on the estimation during the first stages. Yet we still want the computation time to be used most effectively and to provide the best results. Therefore, the optimal setting of the tuning parameters is convenient to be determined each time.

To find the optimal settings, the tuning parameters are selected from the grid of different setting combinations. The maximal time of non-linear optimization is considered to be 5, 20, and 60 seconds. The population size is considered to be 10, 25, 50, and 100. The elite ratio is considered to be 5 %, 20 %, and 30 %. The number of iterations is then calculated to not exceed approximately 2.5 hours of the computation time. Each combination of the tuning parameters forms one setting. Therefore, we have 36 settings of the tuning parameters in total. The cross-validation of tuning parameters is repeated ten times to reduce the effect of the random nature of the estimation methods. The whole computation would take in total 900 hours on a single computer. Therefore, the MetaCentrum computational resources were used and the task was split into 10 individual sub-tasks computed by 10 computers in parallel. The necessary time was then reduced to approximately 90 hours.

3.1 Results

Based on the learning curves which record the best value of the loss function for each generation of the genetic algorithm, it is possible to set the optimal setting for each time category. Firstly, it is necessary to average all learning curves with the same setting. It is described in equation (6) which serves mainly as an introduction to the used notation.

$$R_{iG} = \frac{1}{N}\sum_{j=1}^{N} L(\hat{y}_t, y_t)_j \quad ; \forall i \, \forall G, \tag{6}$$

where N is the number of the repetitions of the cross-validation, L is the loss function and R_{iG} is therefore an average loss function for setting i in the G^{th} generation of genetic algorithm.

Then it is necessary to estimate how much time is needed to create the G^{th} generation which depends on the setting i (population size, elite ratio, and maximal optimization time). The time will be denoted as T_{iG} and equation (7) describes its calculation.

$$T_{iG} = \left\lfloor \frac{G}{N}\sum_{j=1}^{N} \frac{t_{ij}}{n_{ij}} \right\rceil \quad ; \forall i \, \forall G, \tag{7}$$

where G is the serial number of the generation (e.g. G=2 for the 2^{nd} generation), t_{ij} is the total running time of the genetic algorithm for setting i in the j^{th} repetition (note that it might slightly differ from the theoretical 2.5 hours), n_{ij} is the number of iterations for setting i in the j^{th} repetition , $\lfloor\rceil$ is a symbol for rounding to the nearest integer and finally T_{iG} is a rounded (or categorized) average time taken to create G^{th} generation for setting i. The plotted outputs of the combination of learning curves are presented in Figure 1 where it is possible to see in total 36 different combinations. Note that the distances between the points represent the time necessary to generate one iteration of the genetic algorithm.

For each unique value of T_{iG} is then found the maximum of R_{iG} which belongs to the time category represented by T_{iG}. For the setting to be optimal and effective, it must be fulfilled that the maximum of R_{iG} in time t is greater than the maximum of R_{iG} in time $t-1$. This condition can eliminate settings for some time categories due to the ineffectiveness related to the certain time. As an example of this serves comparison between Figure 1 and Figure 2, where the last point in Figure 1 took approximately 145 minutes to compute (due to a long time of creating one generation) and it is not shown in Figure 2 as an optimal setting since the value of the loss function is lower than the maximal R_{iG} in the time of approximately 110 minutes.

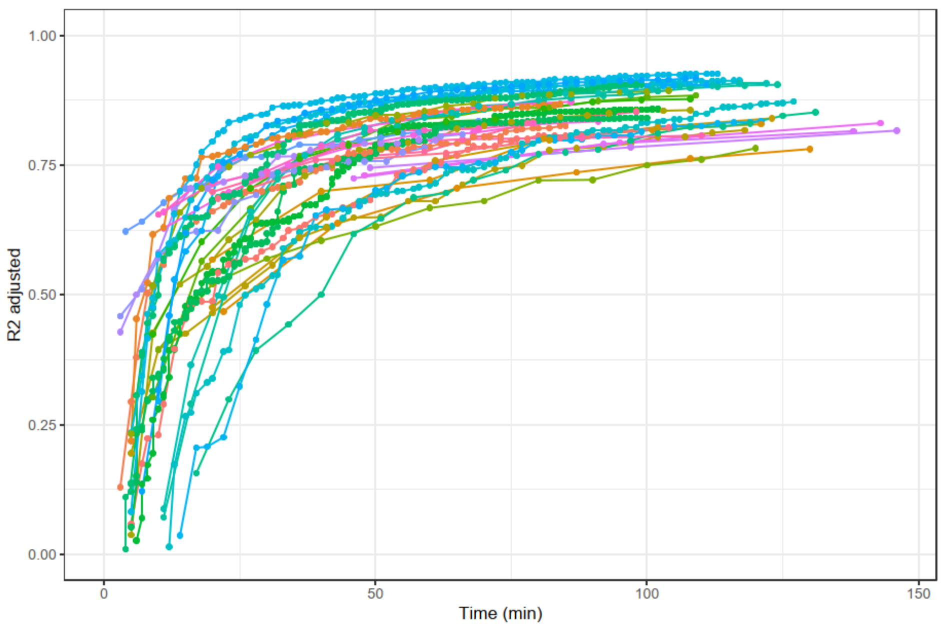


Figure 1: The performance of the estimation depends on the time and the settings of the combination of genetic algorithm and non-linear optimization. Each color represents a different setting.

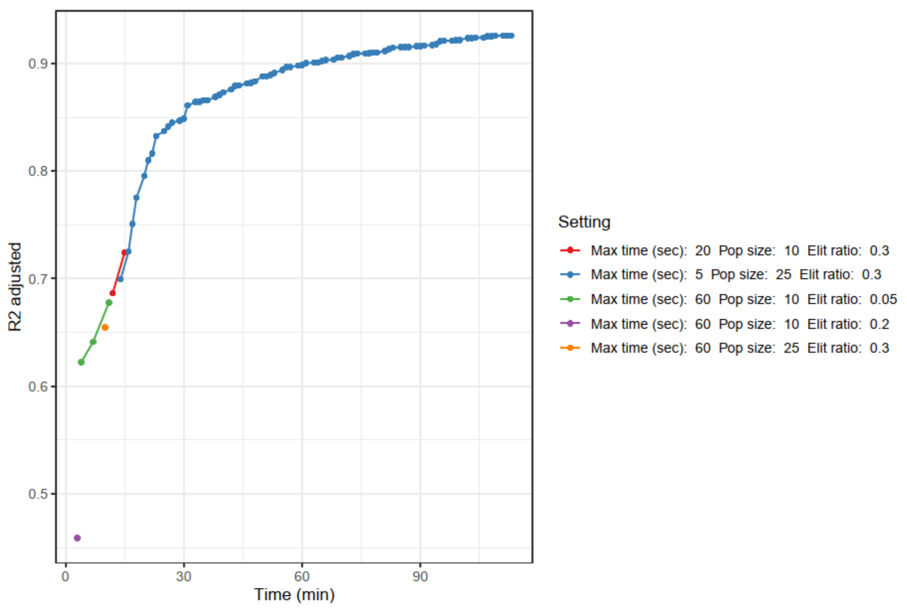


Figure 2: The performance of the optimal settings for each time category depending on time. Each color represents an optimal setting for different time categories.

In Figure 2 it is possible to see 5 different optimal combinations of settings. The best performing setting is the combination of maximal optimization time = 5 seconds, population size = 25 and elite ratio = 30 %. This setting is recommended to run at least for approximately 25 minutes after which it reaches the inflection point of the learning curve. Based on the time we want to run the computation, the optimal settings are selected, and then it is possible to determine the number of iterations needed.

4 Conclusion

The optimal settings of a genetic algorithm for variable selection in a non-linear time series model depend on the time allocated for the computation which especially in the early stages of the model definition and creation might be shorter to just reveal mistakes. Therefore a grid search was performed to select the optimal settings for each time category. Among the optimized tuning parameters belong maximal optimization time, population size, and elite ratio. The number of iterations is derived from the optimal setting and time allocated for the computation. The optimal time, in this case, is approximately 25 minutes after which the model's R^2_{adj} does not increase as rapidly as before the 25 minutes. The optimal combination of settings for this time is maximal optimization time = 5 seconds, population size = 25 and elite ratio = 30 %.

5 Discussion

First of all, it is necessary to state that these results of computation are valid only for this particular regression case. Moreover, the research was done only on one sample of data and might behave differently for different data. It is very probable that the results are data-sensitive which is the biggest cons of this research. Therefore one direction of future research is to perform this analysis on more samples and find a more general conclusion on how to set the tuning parameters.

Another direction of future research is to replace the loss function with AIC instead of R^2_{adj}. It requires the transformation of the estimation method from ordinary least squares to the maximum likelihood function. This brings another advantage of estimating the variance of the population. Moreover, it allows choosing various distributions and not only the normal distribution. The last idea of future research is to select the response functions based on more precise and scientific research which would eliminate another assumption of this model.

Funding

This research was supported by the Internal Grant Agency of the Prague University of Economics and Business under grant F4/27/2020. Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA LM2018140) supported by the Ministry of Education, Youth and Sports of the Czech Republic.

References

[1] Cadima, J., Cerdeira, J. & Minhoto, M. (2004). Computational aspects of algorithms for variable selection in the context of principal components. *Computational Statistics and Data Analysis*. Volume 47. Issue 2 SPEC. ISS. Pages 225–236. ISSN 0167-9473.

[2] Eiben, A. E. & Smith, J. E. (2015). Evolutionary Computing: The Origins. *Introduction to Evolutionary Computing*. Berlin, Heidelberg. Springer. Pages 13–24. ISBN 978-3- 662-44874-8.

[3] Gatu, C.&Kontoghiorghes, E.(2006).Branch-and-boundalgorithmsforcomputingthebest-subsetregression models. *Journal of Computational and Graphical Statistics*. Volume 15. Issue 1. Pages 139–156. ISSN 1061- 8600.

[4] Gatu, C. & Kontoghiorghes, E. (2003). Parallel algorithms for computing all possible subset regression models using the QR decomposition. *Parallel Computing*. Volume 29. Issue 4 SPEC. Pages 505–521. ISSN 0167-8191.

[5] Hofmann, M., Gatu, C. & Kontoghiorghes, E. (2007). Efficient algorithms for computing the best subset regression models for large-scale problems. *Computational Statistics and Data Analysis*. Volume 52. Issue 1. Pages 16–29. ISSN 0167-9473.

[6] Kapetanios, G. (2007). Variable selection in regression models using nonstandard optimisation of information criteria. *Computational Statistics and Data Analysis*. Volume 52. Issue 1. Pages 4–15. ISSN 0167-9473.

[7] Krolzig, H.-M. & Hendry, D. (2001). Computer automation of general-to-specific model selection procedures. *Journal of Economic Dynamics and Control*. Volume 25. Issue 6-7. Pages 831–866. ISSN 0165-1889.

[8] Ou, P. & Wang, H. (2014). Volatility modelling and prediction by hybrid support vector regression with chaotic genetic algorithms. *International Arab Journal of Information Technology*. Volume 11. Issue 3. Pages 287–292. ISSN 1683-3198.

[9] Tull, D. S. (1965). The Carry-Over Effect of Advertising. *Journal of Marketing*. Volume 29. Issue 2. Pages 46–53. ISSN 0022-2429.

[10] Östermark, R. (1999). Solving irregular econometric and mathematical optimization problems with a Genetic Hybrid Algorithm. *Computational Economics*. Volume 13. Issue 2. Pages 103–115. ISSN 0927-7099.

Analysis of Impact of Covariates Entering Stochastic Optimization Problem

Petr Volf[1]

Abstract. In the contribution we study consequences of imperfect information to precision of stochastic optimization solution. In particular, it is assumed that the characteristics of optimization problem are influenced by a set of covariates. This dependence is described via a regression model. Hence, the uncertainty is then caused by statistical estimation of regression parameters. The contribution will analyze several regression model cases, together with their application. Precision of results will be explored, both theoretically as well as with the aid of simulations.

Keywords: stochastic optimization, empirical distribution, regression model, statistical estimation, optimal maintenance

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

1 Introduction

The present paper studies consequences of incomplete information to precision of stochastic optimization result, namely the impact of statistical estimation of involved stochastic characteristics. It is assumed that these characteristics, e.g. the distribution function, depend on a set of covariates. This dependence is described via a certain regression model. Hence, except that the data observation can be incomplete (problem discussed for instance in [7]), an additional uncertainty is caused by estimation of regression parameters, on the basis of observed data sample. The contribution will consider several types of models, for instance a standard linear regression model and the Cox regression model used typically in lifetime studies. Precision of results will be explored with the aid of simulations, convergence of obtained sub-optimal solutions to optimal results will be proved theoretically.

Standardly, the stochastic optimization problem is formulated as a search for solution to

$$\inf_{\boldsymbol{v}} \phi_F(v) = \inf_{\boldsymbol{v}} E_F \varphi(Y, \boldsymbol{v}), \tag{1}$$

where φ is a cost function, $\boldsymbol{v}$ are input variables from certain feasibility set $\boldsymbol{V}$. Further, E_F stands for the expectation under distribution function F, and, finally, Y is a random variable (or vector) possessing this distribution function. A variant formulation may ask for optimization of certain quantile of random criterion $\varphi(Y, \boldsymbol{v})$ instead its expectation, or a combination of several criteria may be considered, thus forming a multi-objective problem. However, often the distribution function F is not known and has to be estimated from observed data. Thus, as the estimate is random (based on available data), obtained solution using estimated distribution instead a "true" one is random as well. Either a parametric form of distribution is assumed, or a non-parametric estimator has to be used. Both these instances were analyzed e.g. in [7] and [8], where the case of incomplete (censored) data has been explored. In fact, the impact of use of empirical distribution instead the exact one has been studied already for a long time, see e.g. Dupačová and Wets (1983), [3].

The present contribution considers just parametric case, on the other hand the precision of estimates is reduced due a presence of covariates. It means that the estimation of corresponding regression parameters causes an additional variability of problem solution.

In the next section some examples of convenient regression models are presented. Then, several assumptions concerning both to cost function φ and the distribution of Y are formulated, in order to prove the main result stating the consistency of solutions based on consistent parameters estimates. Finally, a simple example is solved in detail to show also the behavior of sub-optimal solutions based on data samples of small and medium size.

[1] Institute of Information Theory and Automation Pod vodáarenskou věží 4, Praha 8, Czech Republic,volf@utia.cas.cz

2 Model and its parameters

A parametric model means that the type of distribution $F(y;z,\theta)$ of random variable Y for given value(s) of covariate z (which are from a set $\boldsymbol{Z} \subset R^k$) is known, unknown are the parameters θ, with values from a set $\Theta \in R^m$, containing both the parameters of a "baseline" distribution and parameters characterizing the regression. Let us present here several examples:

1. Standard linear regression model, in which
$$Y = \alpha + \beta \cdot z + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2).$$
Here the "baseline" parameters are α, σ, all parameters then $\theta = (\alpha, \sigma, \beta)$.
Details on the analysis of linear regression model and corresponding confidence intervals and bands can be found elsewhere, e.g. in Lehmann and Romano [6], Ch. 7 and 9.
2. Let us consider Example 1 from above, however with a transformation $T = \exp(Y) = \exp(\beta \cdot z) \cdot \epsilon$, where now random variable ϵ has lognormal distribution with parameters (α, σ^2). Let us denote the distribution function of ϵ as $F_e(.)$, then distribution function of random variable T, for given covariate z, equals
$$F_T(t;z) = P(T \le t; z) = P(\epsilon \le \exp(-\beta \cdot z) \cdot t) = F_e(\exp(-\beta \cdot z) \cdot t).$$
It is seen that the case leads to so called Accelerated Failure Time (AFT) regression model used often in statistical reliability analysis to describe consequences of a load or degradation (expressed via covariates) to lifetime of a device. For more information see e.g. [1].
3. Let the baseline distribution be the Weibull one, with distribution function $F_0(y) = 1 - \exp\{-(y/a)^b\}$ defined for $y > 0$ and let its parameters, both positive, be a, b. The dependence on a covariate z can enter the model via the Cox (proportional hazard) regression model (cf. [4]):
$$F(y;z) = 1 - \exp\{-(y/a)^b \cdot \exp(\beta\, z)\} = 1 - \exp\{-(\frac{y}{a(z)})^b\},$$
where $a(z) = a \cdot \exp\{-\beta\, z/b\}$. Thus, the scale parameter $a(z)$ depends on covariate, while the shape parameter remains b, formally we can write that $\theta = (a, b, \beta)$ is the set of parameters we wish to estimate.

Parameters θ are estimated, as a rule, with the aid of the maximum likelihood estimation (MLE) method. Good properties of the MLE are connected with so called regularity conditions concerning distributions $F(y;z,\theta)$. Their explicit formulation can be found elsewhere in statistical textbooks, for instance in Cox and Hinkley [2], Ch. 9. If they are fulfilled then there exists a consistent sequence of estimates, i.e. such that $\hat{\theta}_N \to \theta_0$ in probability when the data size $N \to \infty$. Here θ_0 denotes the 'true' value of parameter. Further, estimates are asymptotically normal, which means that the distribution of $\sqrt{N}(\hat{\theta}_N - \theta_0)$ tends to normal distribution with zero mean and finite variance given by the inversion of the Fisher information matrix, $I^{-1}(\theta_0)$. In practice the MLE is based on the data $\{y_i, z_i\}, i = 1, ..., N$ and maximizes the log-likelihood function

$$\ln L_N(\theta) = \sum_{i=1}^{N} \ln f(y_i; z_i, \theta))$$

over θ from certain set Θ. Consistent estimate of the Fisher information is then obtained as

$$I_N(\theta) = -\frac{1}{N}\left(\frac{\mathrm{d}(\ln L_N(\theta))^2}{\mathrm{d}^2\theta}\right).$$

We are interested in a question whether (and under which assumptions) the consistency of estimate of θ ensures already the consistency of solutions computed as optimal with respect to estimated parameters instead to 'true' ones. In other words, whether optimal values (both ϕ and $\boldsymbol{v}$) of problem (1) based on estimated parameters $\hat{\theta}_N$, i.e. on the distribution function estimated as $F(y;z,\hat{\theta}_N)$, converge to ϕ^* and $\boldsymbol{v}^*$ optimal for the case when the distribution function $F(y;z,\theta)$ is known.

Naturally, the solution depends on the covariate z, i.e. $\phi^* = \phi^*(z)$ and $\boldsymbol{v}^* = \boldsymbol{v}^*(z)$. In fact, we are interested in optimal solution for each particular value of z, while estimates of parameters are computed from values available in our sample of data $\{y_i,\, z_i,\, i = 1, ..., N\}$. In reality, this data sample is random and of a limited size, therefore it offers just limited and random information for identification of optimized system. That is why in the final example such data are generated repeatedly. With their help the variability of estimates and, consequently, of solutions based on these estimates, is studied. On the other hand, the following theoretical result relies on the consistency of estimates, i.e. on their convergence to 'true' parameters values, when the size of data grows. It is proved that then also corresponding sub-optimal solutions converge to the optimal one.

The setting outlined above can also be interpreted in the sense that the regression parameter is a “nuisance” parameter, its estimation increases the uncertainty of baseline parameters estimates. Further, the confidence of estimates depends also on the covariate design, not only on the data size. In fact, the impact of covariate design has already been explored sufficiently elsewhere, cf. again remarks on design of experiments in [6].

3 Consistency of optimum

As in the following derivation of the main result the value of covariate z is taken as fixed, therefore it is omitted. In other words, following statements hold for any consistent sequence of estimates $\hat{\theta}_N$ and for these $z \in \boldsymbol{Z}$ for which the following assumptions are fulfilled. Let us first introduce some notation:

$$\phi(v,\theta) = \int_{-\infty}^{\infty} \varphi(y,v) f(y,\theta)\mathrm{d}y, \;\; \phi_N(v) = \phi(v,\hat{\theta}_N), \;\; \phi_F(v) = \phi(v,\theta_0),$$

$$v_F^* = \arg\min_v \phi_F(v), \;\; \phi_F^* = \phi_F(v_F^*), \;\; v_N^* = \arg\min_v \phi(v,\hat{\theta}_N), \;\; \phi_N^* = \phi(v_N^*,\hat{\theta}_N).$$

Further, let us assume that there exists a sequence of MLE $\hat{\theta}_N$ tending in probability to 'true' parameter value θ_0. If we use Taylor expansion at θ_0, denoting $f'(y,\theta) = \frac{\partial f(y,\theta)}{\partial \theta}$, we obtain that

$$\phi(v,\hat{\theta}_N) - \phi(v,\theta_0) = \int_{-\infty}^{\infty} \varphi(y,v) f(y,\hat{\theta}_N)\mathrm{d}y - \int_{-\infty}^{\infty} \varphi(y,v) f(y,\theta_0)\mathrm{d}y =$$

$$= \int_{-\infty}^{\infty} \varphi(y,v) f'(y,\hat{\theta}_N)\mathrm{d}y \cdot (\hat{\theta}_N - \theta_0), \tag{2}$$

where, for sufficiently large N, $\hat{\theta}_N$ is arbitrarily (in the sense of convergence in probability) close to θ_0.

Let us formulate several assumptions; first three concern general properties of optimization criterion, they are formulated already in [5] in order to ensure closeness of solutions for sufficiently close (in the sense of L_1 norm) distribution functions. The other assumptions are inspired by (2).

A1. Variable $v \in \boldsymbol{V}$, where $\boldsymbol{V}$ is a compact set in R_1.
A2. Functions $\varphi(y,v)$ are continuous in v on $\boldsymbol{V}$, uniformly w.r. to $y \in R_1$.
A3. $E_F\varphi(Y,v)$ are finite for all $v \in \boldsymbol{V}$.
A4. The distribution of r.v. Y is continuous, with density function $f(y;\theta)$ and its derivative $f'(y,\theta) = \partial f(y,\theta)/\partial\theta$ exists on Θ.
A5. There exists a compact neighborhood $\boldsymbol{O}$ of θ_0 and a positive number $K < \infty$ such that $|\phi'(v,\theta)| = |\int_{-\infty}^{\infty} \varphi(y,v) f'(y,\theta)dy| \leq K$, for each $v \in \boldsymbol{V}$ and $\theta \in \boldsymbol{O}$.

Theorem 1. *Let $\hat{\theta}_N$ be a consistent in probability sequence of estimates of θ_0, further let assumptions A1 – A5 hold. Then, for $N \to \infty$,*
1. *$\phi_F^* = lim\, \phi_N^*$ in probability,*
2. *There exists a sub-sequence $v_{N,k}^* \subset \{v_N^*\}$, $k = 1, 2, ...$ such that it converges (even a.s.), when $k \to \infty$, $lim\, v_{N,k}^* = v_0^*$ and $v_0^* \in \{arg\, min\, \phi_F(v)\}$.*

Proof. Let us divide the proof to several steps:

i) For each sequence of consistent estimates $\hat{\theta}_N$ it holds that each $\phi_N^* \leq \phi_N(v^*)$ and that also $|\phi_N(v^*) - \phi_F^*| \to 0$ in P (due to A5). Then both $\liminf \phi_N^* \leq \phi_F^*$ and $\limsup \phi_N^* \leq \phi_F^*$, this must hold a.s. Denote $\underline{\phi} = \liminf \phi_N^*$. We want to prove that $\underline{\phi} = \phi_F^*$ a.s. (notice that $\underline{\phi}$ is random, defined a.s., while ϕ_F^* is a constant.

ii) First, there exists a sub-sequence of indices $\{N1\} \subset \{1, 2, ...\}$ such that $\underline{\phi} = \lim_{N1\to\infty} \phi_{N1}^*$ a.s. (because a sequence converging in P has a sub-sequence converging a.s.), with corresponding sequence of solutions v_{N1}^*. Then, due compactness of $\boldsymbol{V}$, there is another sequence $\{N2\} \subset \{N1\}$ such that there exists a.s. $\lim_{N2\to\infty} v_{N2}^* = \overline{v} \in \boldsymbol{V}$. Both $\underline{\phi}$ and $\overline{v}$ are defined a.s., to show their existence was a sense of this point.

iii) From A2 and ii) it follows that also $\lim_{N2\to\infty} \phi_{N2}(\overline{v}) = \underline{\phi}$, a.s.

iv) Now we wish to prove that (at least in probability) $\lim_{N\to\infty} \phi_N(\overline{v}) = \phi_F(\overline{v})$. It follows directly from (2) and assumption A5.

v) When combining iii) and iv), it is seen that $\phi_F(\overline{v}) = \underline{\phi}$ a.s. Simultaneously, we have that $\underline{\phi} \leq \phi_F^*$. Therefore $\overline{v}$ is also a solution, i.e. $\overline{v} \in \{\arg\min \phi_F(v)\}$ a.s. If the set $\{\arg\min \phi_F(v)\}$ is just one point, then $\overline{v}$ coincides a.s. with it.

$\square$

Theorem 1 concerns the case when estimated parameters are used for determining an optimal solution w.r. to estimated distribution, and shows that $\phi_{N,k}^* = \phi(v_{N,k}^*, \hat{\theta}_{N,k}) \to \phi_F^*$ in probability. However, in practice, we derive v_N^* optimal with respect to estimated parameter $\hat{\theta}_N$ and then we use it in the setting governed by the underlying 'true' distribution F. That is why the result is in fact $\phi(v_N^*, \theta_0)$. The next proposition therefore proves the convergence of these sub-optimal solutions to the optimum.

Corollary. Under assumptions and with notations of Theorem 1, it holds that, in probability,

$$\lim_{k\to\infty} \phi(v_{N,k}^*, \theta_0) = \phi_F^*.$$

Proof. From (2) and A5 it follows that $|\phi(v_{N,k}^*, \hat{\theta}_{N,k}| - \phi(v_{N,k}^*, \theta_0)) \leq K \cdot |\hat{\theta}_{N,k} - \theta_0|$. As A5 holds uniformly for each $v \in \boldsymbol{V}$ and $\hat{\theta}_{N,k} \to \theta_0$ in probability and, simultaneously, Theorem 1 claims that $\phi(v_{N,k}^*, \hat{\theta}_{N,k}) \to \phi_F^*$, then the proposition is proved.

$\square$

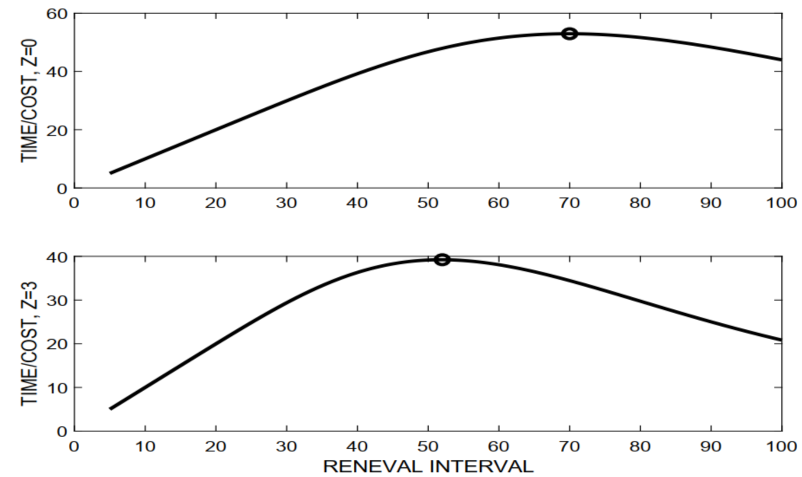


Figure 1 Functions $\phi_N(\tau)$ and their maximal values at optimal τ^*, when $z = 0$ (above) and $z = 3$ (below).

4 Example

Let us consider the following rather simple example of optimization problem: A component of a machine has its time to failure T given (modeled) by a continuous-type probability distribution with distribution function, density, survival function $F,\ f,\ \bar{F} = 1 - F$, respectively. The cost of repair after failure is C_1, the cost of preventive repair is $C_2 < C_1$. For the simplicity we assume that only complete repairs, 'renewals', are provided, i.e. after each repair the component is new (exchanged) or as new. Let τ be the time from renewal to preventive repair, we wish to select an optimal value of τ.

Let us, as a random criterion function, consider the mean time of component availability to the unit of cost,

$$\varphi(T,\tau) = \frac{T}{C_1} \quad \text{if} \quad T \le \tau, \quad \varphi(T,\tau) = \frac{\tau}{C_2} \quad \text{if} \quad T > \tau.$$

Our task is to find optimal τ from a reasonable closed interval $\boldsymbol{T}$, i.e. to maximize

$$\phi_F(\tau) = E_F\varphi(T,\tau) = \int_0^\tau \frac{t}{C_1}\mathrm{d}F(t) + \frac{\tau}{C_2}\overline{F}(\tau). \tag{3}$$

In such a simple case the optimal solution τ^* can be found directly, by solving equation $\mathrm{d}\phi_F(\tau)/\mathrm{d}\tau = 0$. In our case

$$\frac{\mathrm{d}\phi_F(\tau)}{\mathrm{d}\tau} = \frac{\tau}{C_1}f(\tau) + \frac{1}{C_2}\Big(\overline{F}(\tau) - \tau f(\tau)\Big). \tag{4}$$

In the sequel the lifetime distribution will be specified and we shall compare the deterministic solution provided F (i.e. its parameters) is known, and the variability of 'sub-solutions' in cases when the parameters are estimated. Namely, let us consider a case outlined in Example 2 of Section 2: Let the distribution of T be lognormal fulfilling the relation

$$\ln T = \alpha + \beta \cdot z + \varepsilon, \ \ \varepsilon \sim N(0, \sigma^2), \tag{5}$$

with parameters $\alpha = 4.5$, $\beta = -0.1$, $\sigma = 0.4$. Further, let 1-dimensional covariate z has values distributed more or less uniformly in $(-5, 5)$. For $z = 0$ we have the "baseline" lognormal distribution with parameters α, σ, it has approximately the expectation $\mathrm{E}T_0 = 97.51$, the standard deviation $\mathrm{std}T_0 = 40.62$, while for instance for $z = 3$ these characteristics are $\mathrm{E}T_3 = 72.24$, $\mathrm{std}T_3 = 30.09$.

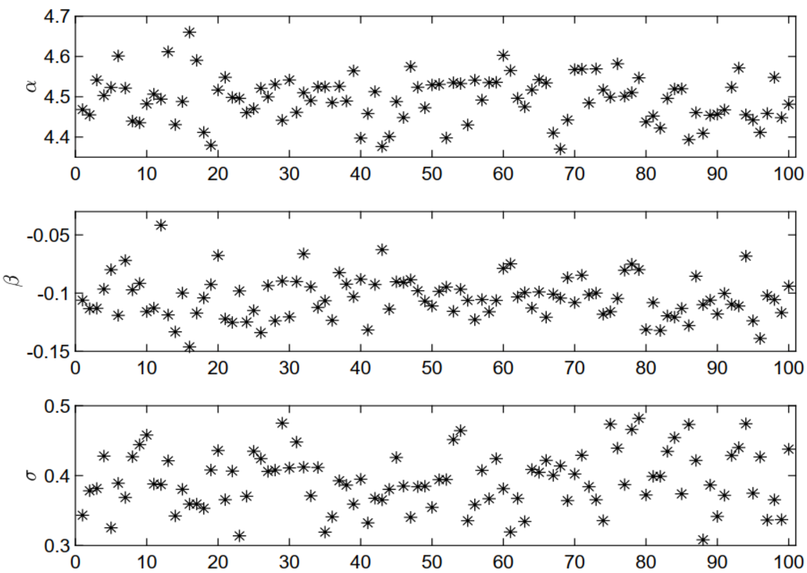


Figure 2 Estimates of parameters from 100 sets of data of size $N = 50$. Above α, then β, below σ.

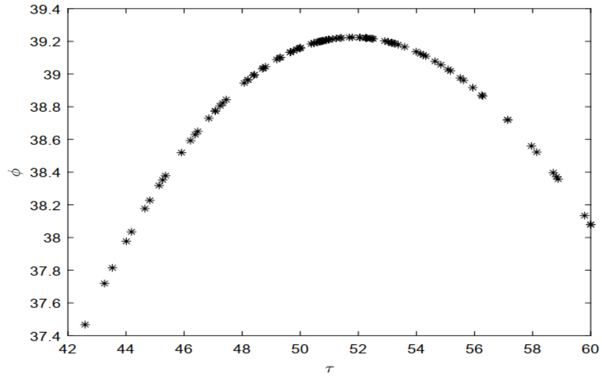


Figure 3 Values of ϕ achieved for parameters estimated from data of size $N = 50$, 100 times.

Further, let the costs be $C_1 = 10$ and $C_2 = 1$. Then, when all parameters are known, optimal values of inter-maintenance time τ and achieved maximum of $\phi(\tau)$ are $\tau_0^* = 69.95$, $\phi_0^* = 52.94$ and $\tau_3^* = 51.82$, $\phi_3^* = 39.22$ for $z = 0$, $z = 3$, respectively. These solutions were obtained from (3) and (4) and are displayed in Figure 1.

Now, we shall generate data of extent N=50 from the correct model, i.e. first the values of covariate z_i, $i = 1, ..., N$, will be generated uniformly in $(-5, 5)$ and then to each z_i one value T_i will be generated from (5). Further, the parameters will be estimated from these data. Finally, we select one value of covariate (namely $z = 3$ to have a comparison with corresponding correct τ^* and ϕ^* from above) and compute optimal values of τ_3 along (4), taking estimated parameters instead 'true' ones. However, then these sub-optimal τ have to be inserted to (3) together with correct F, i.e. given by correct values of parameters. Hence, obtained values of ϕ represent in fact a consequence of taking 'optimal' τ computed along (4) with estimated f, F, and then inserted to (3) with true both f and F,. The result is, naturally, smaller than maximal possible ϕ^*. This procedure will be repeated 100 times to obtain a representation of parameters estimates (this is shown in Figure 2) and corresponding sub-optimal solutions, presented in Figure 3. It is seen how these values copy the curve from Figure 1 (for $z = 3$) and are more or less close to optimal point.

The same procedure has been then performed for larger data extent, N=200. On Figure 4 it is seen that the uncertainty of estimates is smaller than for N=50, i.e. that estimated values are closer to correct values and, consequently, the sub-optimal solutions based on these estimates are closer to correct solution. This is demonstrated in Figure 5 and its comparison with Figure 3.

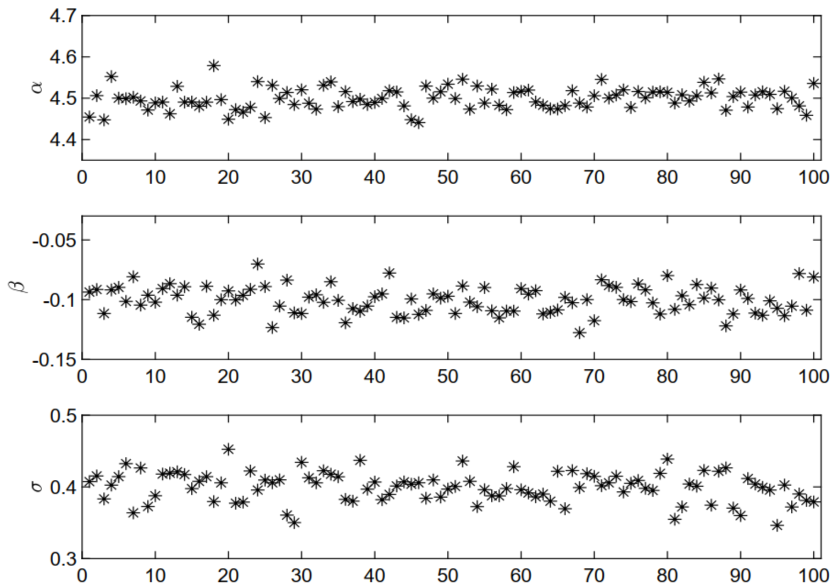


Figure 4 Estimates of parameters from 100 sets of data of size $N = 200$. Above α, then β, below σ.

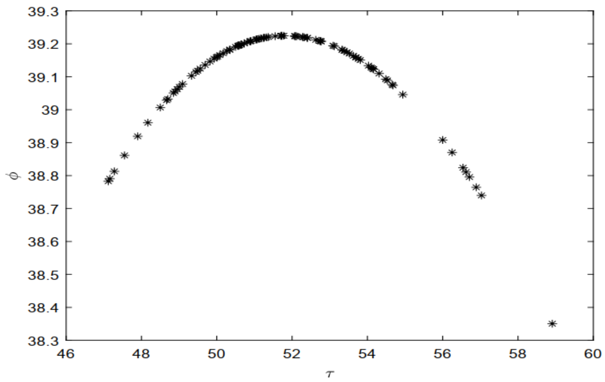


Figure 5 Values of ϕ achieved for parameters estimated from data of size $N = 200$, 100 times.

5 Conclusion

We have studied the impact of presence of covariates to the increase of variability of statistical estimates and, consequently, to the imprecision of solution in a stochastic optimization problem. Asymptotic consistency of solution has been proved, though just in the parametric case. Further, the influence of empirical estimates to optimal solutions has been studied on randomly generated examples. An extension to semiparametric regression models, where the baseline distribution is not specified, could be the further step. Such a generalization may concern both the Cox and AFT models mentioned in the present contribution. Another extension may consider the cases of incomplete data (censored or even truncated observations) encountered quite often in real data studies.

References

[1] Bagdonavicius, V. & Nikulin, M. (2002). *Accelerated Life Models. Modeling and Statistical Analysis*. New York: Chapman & Hall/CRC.

[2] Cox, D.R. & Hinkley, D.V. (1974). *Theoretical Statistics*. London: Chapman & Hall.

[3] Dupačová, J. & Wets, R. (1984). Asymptotic behaviour of statistical estimates and optimal solutions of stochastic optimization problems. *Ann. Statist.*, 16, 1517–1549.

[4] Kalbfleisch, J.D. & Prentice, R.L. (2002). *The Statistical Analysis of Failure Time Data (2nd edition)*. Hoboken, NJ: Wiley Series in Probability and Statistics.

[5] Kaňková, V. (2009). A remark of empirical estimates via economic problems. In H. Brožová (Ed.), *Proceedings of MME 2009* (pp. 169–173). Prague: Czech Univ. of Life Sciences.

[6] Lehmann, E.L. & Romano, J.P. (2005). *Testing Statistical Hypotheses (*3rd edition). New York: Springer.

[7] Volf, P. (2014). On precision of stochastic optimization based on estimates from censored data. *Kybernetika*,50, 297–309.

[8] Volf, P. (2018). On Quantile Optimization Problem with Censored Data. *Kybernetika*, 54, 1156–1166.

The Exact Solution of Travelling Salesman by Mixed Integer Programming in Matlab

Jaromír Zahrádka[1]

Abstract. This contribution comes up with a specific solution of the travelling salesman problem. The driver of hauler has to deliver, using his truck, goods from the depot to n customers. Each customer point of delivery is given by GPS coordinates. The objective of the solution is to select the sequence of delivery points so that firstly the travel distance and subsequently the total travel time are minimal. The driver visits all delivery points and returns to the depot. In this contribution, one general solution is presented using the bound-and-branche method and by using mixed integer linear programming implemented in M-function. The created algorithm can be used in general for any number n of customers.

Keywords: branch-and-bound, linear programming, Matlab, travelling salesman

JEL Classification: C64
AMS Classification: 68W04; 90C11, 05C20

1 The Travelling Salesman Problem

The travelling salesman problem (TSP) and its classical solutions are described e.g. in [1, 2, 4, 5]. Our solution came from the use of integer programming which was published in [7]. In [3] is presented one implementation of the TSP solution with Matlab programming.

1.1 Mathematical Formulation

The TSP can be defined as follows. Let $G_0 = (V, E)$ be a connected directed graph consisting of a set of $n+1$ nodes, seller depot ($i = 0$) and customer locations ($i = 1, \ldots, n$), and a set E of non-negatively weighted arcs between each pairs of corresponding nodes of the graph G_0. For easier reference, let $I = \{1, \ldots, n\}$ be the set of n customers, and $I_0 = \{0\} \cup I$. The constant t_0 means the time-moment when the dealer vehicle leaves the depot. Each customer can be visited only once at any time greater than t_0. The order of the customers visited is not limited, other than by the requirement that the duration of the seller's journey through all customers (terminated by return to the depot) be as short as possible. For each customer $i \in I$ let m_i be the assumed service time associated with the unloading of goods and dealing with the customer.
Let d_{ij} be the length of the path from i - node to j - node for all $i, j \in I_0$. Therefore $\mathbf{D} = \left(d_{ij}\right)_{i,j \in I_0}$ is the non-negative distance matrix. The matrix $\mathbf{D}$ can be, in general, an asymmetric one with zeros on the places of the main diagonal, i.e. $d_{ii} = 0$ for each $i \in I_0$. It is necessary that the triangular inequalities be satisfied for distances among nodes of graph G_0. Instead of the distance matrix $\mathbf{D}$ we will use, for our solution TSP, the time matrix $\mathbf{C} = \left(c_{ij}\right)_{i,j \in I_0}$. Each element c_{ij} represents the pure travelling time of the seller from i node to j one. It is assumed that if the average speed v of the vehicle among each two nodes is used, then the driving time c_{ij} can be expressed $c_{ij} = \dfrac{d_{ij}}{v}$. In this case the travel time c_{ij} is proportional to the distance d_{ij}. We assume that it is given the moment t_0 when the seller's vehicle leaves the depot.

[1]University of Pardubice, Department Mathematics and Physics, Studentská 95, 53210 Pardubice, jaromir.zahradka@upce.cz

1.2 Mathematical Solution

The core of the practical TSP solution is to find the one cycle in the graph G_0 which includes all nodes of the graph and which gives the shortest total driving time. For this purpose, integer variables x_{ij} for $i, j \in I_0$ are introduced, which can only take the values 0 or 1. The variables x_{ij} are called binary variables. Value $x_{ij} = 1$ means that the arc from node i to j is included in the cycle and value $x_{ij} = 0$ means that the corresponding arc is not included. For systemic reason variables, x_{ii} are used but all are fixed by the value zero, i.e. $x_{ii} = 0$, for each $i \in I_0$. Variables x_{ij} are elements of a matrix $\mathbf{X} = (x_{ij})_{i,j \in I_0}$. The number of flow variables x_{ij} is $(n+1)^2$.

In our work we use other specific non-integer variables t_i, for each $i \in I$. Each t_i indicates the moment when the seller leaves the i 's customer location. By using variables t_i, it is guaranteed that the solution will be correct with all nodes during only one cycle in the graph G_0. The variables t_i are included as n elements of the vector $\mathbf{t} = (t_1, t_2, \ldots, t_n)$. The number of all flow variables is $(n+1)^2 + n$.

The solution of TSP is realized like the optimal solution of a mixed-integer linear programming problem:

$$\min_{(\mathbf{X},\mathbf{t})} \left\{ \sum_{i,j=0}^{n} c_{ij} \cdot x_{ij} + \sum_{i=1}^{n} \frac{1}{n \cdot u} \cdot t_i \right\} \quad \text{subject to} \tag{1}$$

$$x_{ij},\ i, j \in I_0 \text{ are binary, } x_{ij} \in \{0,1\} \tag{2}$$

$$(c_{ij} + u - t_0 - c_{0j})\, x_{ij} + t_i - t_j \le u - t_0 - c_{0j} - m_j, \quad i, j \in I,\ i \ne j \tag{3}$$

$$c_{0j}\, x_{0j} - t_j \le -t_0 - m_j, \qquad j \in I \tag{4}$$

$$\sum_{j \in I_0} x_{ij} = 1, \qquad i \in I_0 \tag{5}$$

$$\sum_{i \in I_0} x_{ij} = 1, \qquad j \in I_0 \tag{6}$$

$$x_{ii} = 0, \qquad i \in I_0 \tag{7}$$

$$0 \le x_{ij} \le 1, \qquad i, j \in I_0 \tag{8}$$

$$t_0 + c_{0j} + m_j \le t_j \le u, \qquad j \in I \tag{9}$$

In the expressed model (1) is minimized the linear optimization function

$$\sum_{i,j=0}^{n} c_{ij} \cdot x_{ij} + \sum_{i=1}^{n} \frac{1}{n \cdot u} \cdot t_i \tag{10}$$

The main part $\sum_{i,j=0}^{n} c_{ij} \cdot x_{ij}$ of the optimized function guarantees finding the cycle which takes the minimum amount of time. Due to the assumed constant average speed v, the total travel length is also minimal. In the second part $\sum_{i=1}^{n} \frac{1}{n \cdot u} \cdot t_i$ of the optimized function (10) is used the value of the constant u, which is defined as follows:

$$u = t_0 + \sum_{i=1}^{n} m_i + \sum_{j=0}^{n} \max_{i \in I_0} c_{ij} \tag{11}$$

The value of u guarantees, with respect to the expected values of t_i, that the coefficients of flow variables t_i in the optimized function (10) are so small that they do not change the optimal solution for flow variables x_{ij}, while the time variables t_i are minimized. The use of terms with flow variables t_i in the optimization function (10) is necessary. If these are not included, a solution could be generated with some values of t_i greather than necessary. Our model prefers, from two shortest cycles (with opposite directions), the one that gives a smaller sum of t_i.

Constraint (3) defines $n(n-1)$ conditions between flow variables x_{ij} and departure times t_i, t_j, for $i, j \in I$. In the case $x_{ij} = 1$, the inequality (3) expresses the relationship $t_j \geq t_0 + c_{0j} + m_j + t_i - u$. Due to the large enough value of u, the right side of inequality (3) can be only non-positive and the relationship is satisfied.
In the case $x_{ij} = 1$, the inequality (3) is reduced $t_i + c_{ij} + m_j \leq t_j$, $i, j \in I$. This expresses that the departure time from the node j has to be greater than or equal to the sum of the departure time t_i (from node i), the travelling time c_{ij} (from node i to node j) and the service time m_j in the node j. The created optimization process ensures that, in the case of $x_{ij} = 1$, the condition (3) is satisfied only by the equation $t_i + c_{ij} + m_j = t_j$.

The constraint (4) defines relations between flow variables x_{0j} and t_j, $j \in I$. In the case $x_{0j} = 0$ the inequality expresses the relationship $t_0 + m_j \leq t_j$, $j \in I$. Departure time from the node j is greater than or equal to the sum of departure time t_0 and service time m_j. In the case $x_{0j} = 1$ the inequality (4) expresses the relationship $t_0 + c_{0j} + m_j \leq t_j$. Departure time from the node j is greater than or equal to sum of departure time t_0 from the depot, travelling time c_{0j} from depot to node j and service time m_j.

Statements (5) and (6) declare $2(n+1)$ equation constrains, which express that only one arc leads from each node and only one arc leads to each node. Statement (7) declares that each $x_{ii} = 0$.

The inequalities in (10) declare that the lower and upper bounds of variables x_{ij} are 0 and 1. The inequalities in (11) express the bounds of flow variables (departure times) t_j, $j \in I$.

1.3 Transformation to Matlab

In the Matlab system the index 0 can not to be used, therefore all vector and matrix variables use the smallest index number 1. The distance matrix is transferred to the Matlab environment as matrix `D`, with the row and column indices `i,j = 1,2, … ,n+1`, where each component `D(i,j)` corresponds to the distance $d_{i-1\,j-1}$ of the nodes $i-1$ and $j-1$. Similarly each component `C(i,j)` of the time matrix corresponds to the driving time $c_{i-1\,j-1}$ from the node $i-1$ to the $j-1$ one.

Our created procedure for TSP solving in the Matlab code is included in the M-function *SOLVER_TSP.m* and it is fully listed as an Appendix at the end of the article. The input variables are *n* - number of customers, *D* – distance matrix, *v* – velocity of the vehicle, *t0* – the moment when the seller leaves the depot, and *m* – row vector with customer service duration times. The main output variable is the column vector *X* of flow variables, which is obtained as an output of the optimization via the command *intlinprog*.

The mixed-integer linear programming problem is generally expressed by

$$\min_{X} f^T \cdot X \text{ subject to} \begin{cases} X\,(intcon) \text{ are integers} \\ A \cdot X \leq b \\ A_{eq} \cdot X = b_{eq} \\ l_b \leq X \leq u_b\,. \end{cases} \tag{12}$$

The solver for this problem is the command *X=intlinprog(f,intcon,A,b,Aeq,beq,lb,ub)* in Matlab code (you can see it on the Appendix row No. 55). A more detailed explanation is in the User's Guide [6].

For the solution of TSP via the *intlinprog* command, all flow variables are arranged in a column vector *X* with $(n+1)^2 + n$ components. First $(n+1)^2$ flow variables are integer variables x_{ij}, and each variable x_{ij}, $i, j \in I_0$ is represented by Matlab flow variable *X(i*(n+1)+j+1,1)*. The last n flow variables of *X* are the seller's departure times $t_1, t_2, \ldots, t_n$, and each variable t_i, $i \in I$ is represented by *X((n+1)^2+i,1).*

The objective function of the mixed-integer linear programming problem (12) is, in the Matlab code, expressed like `f'*X`, where `f` is a column vector of coefficients with $(n+1)^2 + n$ components. The first $(n+1)^2$

components are elements of the time matrix `C` so that `f((i-1)*(n+1)+j,1)=C(i,j), i,j` $\in\{1,2,\ldots,n+1\}$.

For the last n components of `f` we use the value $\frac{1}{n\cdot u}$ according to relation (10) (the Appendix, row No. 2).

The vector *intcon* in the command *intlinprog* specifies of flow variables, which are taken integers, *intcon = 1:(n+1)^2*. They are first $(n+1)^2$ flow variables, i.e. variables x_{ij}.

The constraints (3) and (4) give the system of n^2 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 7 in the Appendix. The constraints (5), (6) and (7) 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. 8 to 13 in the Appendix.

The last two input variables of the *intlinprog* command (2) are the column vectors *lb* and *ub* of lower and upper bounds of the flow variables. With respect to the relations (7), (8), (9) the components of vectors *lb* and *ub* are filled by commands on lines No. 14 to 17 in the Appendix

By installation of input variables *f, intcon, A, b, Aeq, beq, lb, ub* in the command *intlinprog*, and running it (the line No. 18), we get the optimal TSP solution, this is the vector of flow variables *X*. The values of the first $(n+1)^2$ variables (component of *X*), which have a value of 1, indicate the arcs that are part of the travel cycle. The last n values of flow variables indicate times when the seller leaves individual customers.

The variables *X(k,1)*, $k\in\{1,2,\ldots,(n+1)^2\}$, which take the value 1, determine the arcs of the shortest cycle. The commands from lines No. 20 to 24 allow the creation of a sequence of cycle nodes, i.e. the *CYCLE* vector. The first item of the *CYCLE* vector is the number 0 – depot, and the other n items are the sequence of customer numbers, and the last item is supplemented by the number 0 with regard to the fact that the seller returns to the depot.

The values of components of *X(k,1)*, $k\in\{(n+1)^2+1,(n+1)^2+2,\ldots,(n+1)^2+n\}$ are the seller's departure times from the customer *k* at the optimal cycle. The vector of the departure times *t*, the time *tRet* of the seller's arrival back to the depot, and the total duration of the seller's business trip *AllWorkTime* are calculated on lines No. 25, 26. On lines No. 27 to 31 is created the vector *tArr* of arrival times to the nodes and the total distance *TotDist* traveled by the seller. The last item of the vector *tArr* means the time *tRet* when the seller returns to the depot. The input variables for the M-function *SOLVER_TSP* and its execution have to be done using a startup M-script that contains commands for drawing the output circle (Figure 1). The startup script is not listed in this article.

2 Illustrative Example

To illustrate the program we have created, we assume a seller and twelve customers. The GPS coordinates of the seller's depot are $E_0=14.068^\circ$ (the eastern longitude) and $N_0=49.427^\circ$ (the northern latitude). The GPS coordinates E_i, N_i and the service times m_i of customers you can find in Table 1.

	Customer											
i	1	2	3	4	5	6	7	8	9	10	11	12
E_i (°)	14.436	14.174	14.026	14.955	14.431	14.962	14.762	14.007	14.680	14.706	14.645	14.552
N_i (°)	49.221	49.452	49.017	49.266	49.358	49.090	49.168	49.094	49.161	49.202	49.274	49.024
m_i (min)	16	13	13	13	12	20	20	19	18	12	14	12

Table 1 The GPS coordinates and the service times of the customers

The distance between two customer locations (nodes) is taken as their orthonormal distance on the Earth sphere multiplied by a factor of 1.25. The orthonormal distance is calculated with a sphere radius R = 6371 km (mean radius of the Earth). All distances are included in the symmetric distance matrix $\mathbf{D}$ in Table 2.

By running the function *SOLVER_TSP.m* with the above chosen parameters, the optimal solution was found. The shortest cycle is given with a node sequence 0-2-5-1-11-4-6-7-10-9-12-3-8-0 and is drawn in Figure 1. Due to the symmetry of the matrix $\mathbf{D}$, there is another solution that gives the same minimal travel distance and min. driving time. This is the opposite directed cycle 0-8-3-12-9-10-7-6-4-11-1-5-2-0, but its sum of departure times is greather.

Distance (km)		j												
		0	1	2	3	4	5	6	7	8	9	10	11	12
i	0	0	58.19	15.11	55.59	125.18	51.30	132.28	102.57	45.70	92.30	93.71	82.80	86.44
	1	58.19	0	47.90	63.27	72.39	18.45	75.20	45.87	62.03	34.86	37.62	29.91	31.03
	2	15.11	47.90	0	62.14	111.40	37.90	119.87	90.22	53.55	80.50	81.24	69.71	77.98
	3	55.59	63.27	62.14	0	133.40	72.66	130.47	104.30	10.71	92.95	97.74	92.74	73.12
	4	125.18	72.39	111.40	133.40	0	73.88	23.65	29.88	133.78	40.74	35.66	43.10	64.77
	5	51.30	18.45	37.90	72.66	73.88	0	82.13	52.63	68.84	43.59	43.61	31.82	47.99
	6	132.28	75.20	119.86	130.46	23.65	82.13	0	29.71	132.74	40.34	38.63	50.52	57.67
	7	102.57	45.87	90.22	104.30	29.88	52.63	29.71	0	105.41	11.43	9.02	21.62	35.03
	8	45.70	62.03	53.55	10.71	133.78	68.83	132.74	105.41	0	93.98	98.24	91.93	76.34
	9	92.30	34.86	80.50	92.95	40.74	43.59	40.34	11.44	93.98	0	6.59	15.96	25.61
	10	93.71	37.62	81.24	97.74	35.66	43.60	38.63	9.03	98.24	6.59	0	12.87	32.11
	11	82.80	29.91	69.71	92.73	43.10	31.81	50.52	21.62	91.93	15.95	12.87	0	36.03
	12	86.44	31.03	77.98	73.12	64.77	47.99	57.67	35.03	76.34	25.61	32.11	36.03	0

Table 2 The distance matrix $\mathbf{D}$

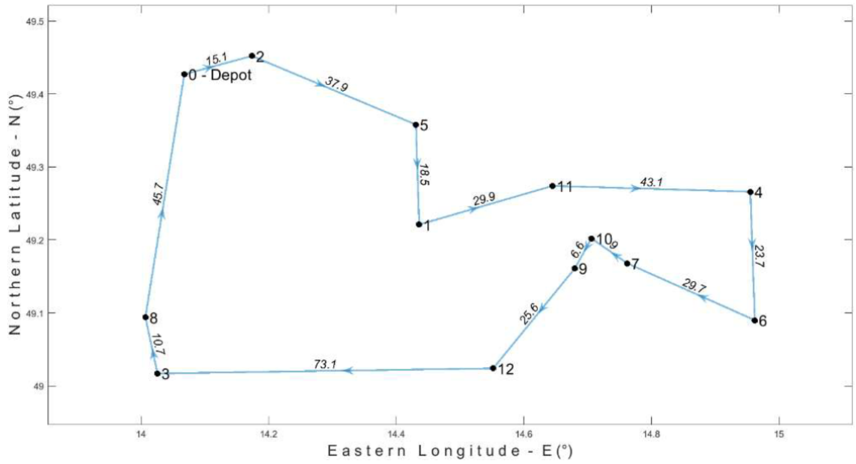


Figure 1 The minimal length cycle of the seller around all customers

The calculated seller's departure times t_{dep_i} from customers, and arrival times t_{arr_i} to customers are in Table 3. The total travelled distance by the seller vehicle is 368.60 km and the total time of a seller's trip is 9 h 10 min.

Times	Depot	Customers ranking in minimal cycle											Depot
i	0	2	5	1	11	4	6	7	10	9	12	3	8	0
t_{arr_i}	-	4:15	5:06	5:36	6:22	7:19	7:56	8:45	9:14	9:33	10:17	11:42	12:05	13:10
t_{dep_i}	4:00	4:28	5:18	5:52	6:36	7:32	8:16	9:05	9:26	9:51	10:29	11:55	12:24	-

Table 3 The arrive and depart times of the seller

3 Conclusion

This paper proposes a practical solution of the travelling salesman problem for any number n-customers in Matlab code. The TSP is formulated as a mixed-integer linear programming problem with a new approach, which respects the given matrix of distances and service duration times of customers, and the constant speed of the seller's movement. The solution lies in minimizing of the seller's trip duration that leads across all customers. The constant speed of seller's movement is assumed, therefore the total distance travelled is also the minimum. The created objective function guarantees that the total travelled distance and the total travelled time of the seller are minimal.

The main result of this article is the creation of the M-function (Appendix) which allows to solve the TSP generally for any number of n customers. The created M-script is practically usable on a common personal computer for up to 30 customers. For 30 customers, the calculation takes less than 60 minutes, and for up to 20 customers, the calculation takes less then 40 seconds. M-script was successfully tested for a maximum of 40 customers. The optimal solution of travelling salesman problem ensures the shortest travel distance and shortest duration of the business trip, and thus the best solution in terms of economic costs for the implementation of the business trip.

Acknowledgements

The paper was supported by institutional support of University of Pardubice.

References

[1] Bentley, J. J. (1992). Fast Algorithms for Geometric Travelling Salesman Problems. ORSA *Journal on Computing*, (4) 4.

[2] Gavendra, G. at al. (2010). Travelling Salesman Problem. In *Tech, Proceedings of the Theory and Applications. Rieka*.

[3] Gradle, K. P. & Mulley, Y. U. (2015). Travelling Salesman with MATLAB programming. International *Journal of Advances in Applied Mathematics and Mechanics*, (3) 2, 258-266.

[4] Gutin, G. & Punnen, A. P. (2007). *The Travelling Salesman Problem and Its Variations*. New York: Springer Science+Business Media, LLC.

[5] Jonak, R., Smutný, Z., Simunek, M. & Dolezel, M. (2020). Rout and Travel Time Optimalization for Delivery and Utility Services. *Acta Informatica Pragensia*, (2) 9, 200-209.

[6] Math Works. Inc. (2020). *Optimalization ToolboxTM. User´s Guide.* Natick.

[7] Winston, W. L. (1994). *Operations Research. Applications and Algorithms. Duxbury*: Duxbury Press.

Appendix

```
function [X, CYCLE, TotDist, AllWorkTime, tArr] = SOLVER_TSP(n, D, v, t0, m)
C=D/60; CT=C'; u=t0+sum(max(CT))+sum(m(1:n)); f=[CT(:);ones(n,1)/u/n];
p=(n+1)*(n+1); A=zeros(n^2,p+n); k=0;
for i=1:n; for j=1:n; if i~=j; k=k+1;
A(k,(n+1)*i+1+j)=C(i+1,j+1)+u-t0-C(1,j+1); A(k,p+i)=1; A(k,p+j)=-1;
b(k,1)=u-t0-C(1,j+1)-m(j); end; end; end
for i=1:n; k=k+1; A(k,1+i)=C(1,1+i); A(k,p+i)=-1; b(k,1)=t0-m(i);end
Aeq=zeros(3*n+3,(n+1)^2+n);
for i=1:n+1; for j=1:n+1; Aeq(i,(i-1)*(n+1)+j)=1; end
Aeq(i,(i-1)*(n+1)+i)=0; beq(i,1)=1; end
for i=1:n+; for j=1:n+1; Aeq(n+1+i,(j-1)*(n+1)+i)=1; end
Aeq(n+1+i,(i-1)*(n+1)+i)=0; beq(n+1+i,1)=1; end
for i=1:n+1; Aeq(2*n+2+i,(i-1)*(n+1)+i)=1; beq(2*n+2+i,1)=0; end
lb = zeros(p,1); for i=1:n; lb(p+i,1)=t0+C(1,1+i)+m(i); end
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; ub(p+i,1)=u; end; intcon = 1:p;
X = intlinprog(f,intcon,A,b,Aeq,beq,lb,ub);
X(1:p)=round(X(1:p));
for i=2:n+1
if X(i)==1; CYCLE=0; Nok=2; CYCLE(Nok)=i-1; TEST=i; break; end; end
while TEST~=1; for j=1:n+1
if X((CYCLE(Nok))*(n+1)+j)==1; Nok=Nok+1; CYCLE(Nok)=j-1; TEST=j; break; end
end;end
for i=1:n; t(i)=X(p+i); end
tRet=t(CYCLE(end-1))+C(CYCLE(end-1)+1,1); AllWorkTime=tRet-t0;
tArr=[t0, t(CYCLE(2:end-1))-(m(CYCLE(2:end-1))), tRet],
tArr=hours(tArr), tArr.Format='hh:mm'; TotDist=0;
for i=1:(n+1); for j=1:(n+1)
if X((n+1)*(i-1)+j)==1; TotDist=TotDist+D(i,j); break; end
end; end
```

Waste Collection Vehicle Routing Using Smart Reports on the Utilization of Bin Capacities

František Zapletal[1], Petr Kozel[2], Lucie Chytilová[3]

Abstract. Vehicle routing is a common optimization problem applicable for various kinds of companies and institutions in general. One of its typical applications is a waste collection, which can help not only a company managing the waste logistics but also citizens and other stakeholder groups in cities and villages. Many studies have been devoted to the optimization of the waste collection in the past under different special conditions like time windows, different types of waste and vehicles, etc. In this paper, we focus on the frequency with which bins are serviced. Usually, the frequency is fixed and the bins are serviced on a given day weekly or every second week. This setting, however, does not distinguish to what extent a bin is full, thus its efficiency is disputable. Our aim is to explore whether some kind of direct smart reports from citizens on the utilization of bin capacities can potentially reduce the total costs of waste collection. The proposed model is verified using numerical examples and different scenarios.

Keywords: waste collection, transport network, fixed servicing system, smart servicing system

JEL Classification: C44
AMS Classification: 90C08

1 Introduction

An effort to behave efficiently is natural when dealing with limited amount of resources. The current increasing prices of fuels (and high inflation rate in general) even emphasize this effort. On the other hand, modern technologies give rise to higher requirements on comfort. The concept of a smart city focuses on applying advanced information and communication technologies to improve the living conditions of citizens [1, 2]. This paper focuses on the waste collection and its management using smart tools. For decades, citizens in villages and cities have been used to the fixed schedule of waste collection. It means that their locations have been serviced repeatedly on predetermined days (and in many cases this approach is still applied). Such 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 servicing.

Many studies focused on smart waste collection have already been published in recent years, see the review paper [3]. However, to our best knowledge, they always work with smart sensors, which provide the information about the level of waste in a bin. The studies have proved that such sensors can enable efficient waste collection. On the other hand, this solution is very expensive and it is very hard to imagine that it could be applied, for example, in some small towns or villages. Therefore we seeked for some alternative smart solution, which would increase the comfort for residents but it would also be affordable for local governments or waste companies. The core idea of this research is that a resident would report that his or her bin needs servicing via either some web portal or smartphone application. Such solution is quite cheap and it can also be user-friendly. The main goals of this paper are to:

- design the new model for waste management based on smart reporting by residents;
- check if the model can decrease the volume of kilometers travelled and/or increase the comfort for the residents.

This paper can be regarded as an initial study for our research that checks the feasibility of the idea and provides an interesting experiment.

[1] VŠB - Technical university of Ostrava, Sokolská 33, Ostrava, Czech Republic, frantisek.zapletal@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, lucie.chytilova@vsb.cz

The rest of this paper is organized as follows. Sec. 2 provides the requirements and assumptions of the model. In Sec. 3, the input data for the model verification and the adopted algorithm are described. Sec. 4 is a core section of this paper and contains the experiment on the model using the numerical example and the results obtained. The last section (Sec. 5) concludes this paper.

2 Requirements and assumptions of the model

In order to check whether our idea of the new smart waste collection management can be viable in practice, the following requirements must be met.

A resident should report the need of servicing (this can be done, e.g., using web interface, smartphone application etc.). Moreover, this reporting can be done in one (it is reported that the bin is full) or more stages (it is also reported that the bin is full to some extent). More stages will not be as advantageous as the sensors proposed by [4, 5, 6, 7, 8], but can make the waste management more fluent. It is worth noting that, in practice, some residents would be very hard to convince to use the system (especially older people).

The capacities of the bins as well as the capacity of the collecting truck are known. Only the volume of waste does matter (regardless its weight or type). The distances between all nodes in the network are known.

The waste production is not fixed in time, but it is a random variable with estimated probability distribution. The parameters of this distribution depends on the type of resident. A senior produces definitely different volume of waste than a family or a company. Therefore, it is reasonable to consider various model residents. In practice, the random waste production is dependent on the year season (e.g., in general, July and August are characteristic because of lower waste production volumes due to holidays and summer family vacation).

The proposed model should increase the comfort for residents. Instead of waiting until the (fixed) day of servicing comes, it is guaranteed that the bin must be serviced not later than the n-th day after reporting the full bin. In the case of free capacities of the truck, the bin can be serviced anytime after the first report (that it is not empty) until the mentioned deadline.

To keep the model reasonably easy to optimize, the following assumptions are considered. First, five different groups of residents are taken into account (in ascending order of expected waste volume): pensioners, couples, small families, big families, companies. Each resident unit has a single bin with equal capacity. A resident reports the state of his/her bin twice. First, when it is (approximately) half full, and then when it is full (or almost full). The waste production is considered uniformly distributed with the parameters shown in Tab. 1. A random value from the given distribution is drawn when the bin is empty (at the beginning or after servicing, and also after reporting the half full state). The truck services the system only if at least one bin must be serviced that day. If the quantity of bins which must be serviced on a given day exceeds the capacity of the truck, the truck must go to service more times that day. More details about the adopted algorithm can be found in the next section.

Group name	Code	Minimum [days]	Maximum [days]
Pensioners	1	10	17
Couples	2	8	14
Small families	3	6	11
Big families	4	4	8
Companies	5	2	5

Table 1: Parameters of uniform probability distributions for each group of residents

For the sake of simplicity, the servicing truck has the capacity c (full) bins (or $2 \cdot c$ half full bins), or their combinations. The truck starts and ends at the same location. On every day t, where $t \in T$ the following information are stored:

- Have the truck serviced some locations that day?
- What bins (locations) have been serviced that day?
- Which bins must be serviced that day, what bins are full (and how many days they have already been waiting) and what bins are half full (after the first report)?
- How many kilometers have been passed that day within servicing?

The next section provides the input data used and the algorithm used for calculation.

3 Input data and algorithm

A real transport network is represented by graph $G(V, E, d)$ where $V = 1, ..., m$ is the set of vertices, $E = 1, ...n$ stands for the set of edges, and d_{ij} is the distance between the vertices i and j (in km). The first vertex represents the depot where the vehicles $r \in R$ with the capacity c start and end. Let us assume that the capacity R is sufficient.

The other vertices $j = 2, ..., m$ represent customers (residents) j with the request for servicing b_{xj}^t at time $t \in T$. The customers are in one of the 5 modes x ($x \in X$) described below. Let $z_{xj}^t \in \{0, 1\}$ be the state variable, then the numbers of customers j that are in each mode x at time t are given by the following relations:

$$\sum_{j=2}^{m} z_{xj}^t, \text{ for } x \in X, t \in T. \tag{1}$$

These sets of customers can analogously be denoted as $S_4^t, ..., S_1^t$. For each t, the following equation holds:

$$\sum_{x \in X} \sum_{j=2}^{m} z_{xj}^t + 1 = |V|, \text{ for }, t \in T. \tag{2}$$

For the sake of clarity, 5 modes x can be distinguished for each vertex $j = 2, ..., m$ and time $t \in T$:

- $x = 0$ when the first report has not been done yet (i.e., when the bin is more than half empty);
- $x = 1$ on the day when the first report has come (i.e., the bin is just now half full);
- $x = 2$ on the day when the second report has come (i.e., the bin is just now full/almost full);
- $x = 3$ on the day after the second report has come (i.e., the bin is full/almost full and wait 1 day for servicing);
- $x = 4$ on the second day after the second report has come (i.e., the bin is full/almost full and wait 2 days for servicing).

The number of nodes in the state x is denoted as:

$$\sum_{j=2}^{m} z_{xj}^t, \text{for } t \in T, \text{for } x \in X. \tag{3}$$

The following algorithm is adopted to find the optimal smart waste collection management.

Step 1: (set $t = 0$, $x = 0$ for all nodes) generate the number of days until the report D_{xj}^t for each node $j = 2, ..., m$ following the given distribution, see Tab. 1;
Step 2: increase t by 1 and decrease D_{xj}^t by one for each node if its original value is positive;
Step 3: check if there is some node with $D_{xj}^t = 0$; if not, go to Step 2, continue with Step 4 otherwise;
Step 4: for all nodes with $D_{xj}^t = 0$, increase x by one; if its new value is 1, generate a new number of days until the report D_{xj}^t (i.e., until the bin is full);
Step 5: if there is $\sum_{j=2}^{m} z_{4j}^t > 0$, find the optimal way of servicing (see the algorithm described below), if not, go to Step 2, continue to Step 6 otherwise;
Step 6: generate a new value of $D_{xj}^t = 0$ for all nodes serviced in the previous step, go to Step 2.

The optimal servicing referred to in Step 5 is done in the following way:
Step 1: select all customers from the set S_4^t, add their requests $\sum_{j=2}^{m} b_{4j}^t$ and proceed to step 2.
Step 2: If $\sum_{j=2}^{m} b_{4j}^t = c$, plan the service vehicle route, otherwise go to step 3 or step 4.
Step 3: if $\sum_{j=2}^{m} b_{4j}^t > c$, plan the p route of the service vehicle, where $p = \frac{\sum_{j=2}^{m} b_{4j}^t}{c} + \frac{e}{c}$, where e is a supplement to the full capacity of the service vehicle. It holds that $e \geq 0$ a $e < c$.
Step 4: If $\sum_{j=2}^{m} b_{4j}^t < c$, add to the full capacity of the service vehicle the requirements of customers from the sets S_3^t, S_2^t, S_1^t so, to apply: $c = \sum_{j=2}^{m} b_{4j}^t + e$ and plan the utility vehicle route.

Because it is necessary to plan the route of the service vehicle exclusively for customers from the set S_4^t, then if the capacity of the service vehicle c is fully used to cover their requirements, exactly one route will be planned, or p routes if $\sum_{j=2}^{m} b_{4j}^t = p \cdot c$. Assuming that after covering the requirements of customers from the set S_4^t the capacity of the service vehicle c remains unused, then the requirements of selected customers from the sets S_3^t, S_2^t, S_1^t will be supplemented so that:

$$\sum_{j=2}^{m} b^t_{4j} + \sum_{j\in S^t_{3*}} b^t_{3j} + \sum_{j\in S^t_{2*}} b^t_{2j} + \sum_{j\in S^t_{1*}} b^t_{1j}, \text{for } t \in T, \tag{4}$$

where the sets $S^t_{3*}, S^t_{2*}, S^t_{1*}$ represent subsets of selected customers from the sets S^t_3, S^t_2, S^t_1. The selection of customers takes into account their location concerning the location of customers who are necessarily served. The mathematical model below will be used to determine the selected customer sets, based on the assignment problem.

$$\text{Min} \sum_{i\in I}\sum_{j\in J} f_{ij}\cdot d_{ij} \tag{5}$$

$$\sum_{i\in I} f_{ij} \le 1, \text{for } j \in J \tag{6}$$

$$\sum_{i\in I}\sum_{j\in J} f_{ij}\cdot b_j = c - \sum_{i\in I} b_i \tag{7}$$

$$f_{ij} \in \{0,1\}, \text{for } i \in I, j \in J \tag{8}$$

The set I represents the set of customers that will be serviced necessarily. Theoretically, the following cases are allowed:

$$I = \begin{cases} S^t_4, & \text{if } \sum_{x=3}^{4}\sum_{j=2}^{m} b^t_{xj} > c; \\ S^t_4 + S^t_3, & \text{if } \sum_{x=3}^{4}\sum_{j=2}^{m} b^t_{xj} < c \wedge \sum_{x=2}^{4}\sum_{j=2}^{m} b^t_{xj} > c; \\ S^t_4 + S^t_3 + S^t_2, & \text{if } \sum_{x=2}^{4}\sum_{j=2}^{m} b^t_{xj} < c \wedge \sum_{x=1}^{4}\sum_{j=2}^{m} b^t_{xj} > c. \end{cases}$$

The set J represents the set of customers, from which the optimally placed customers will be selected concerning the customers from the set I. Theoretically, the following cases are allowed:

$$J = \begin{cases} S^t_3 & \text{if } I = S^t_4; \\ S^t_2 & \text{if } I = S^t_4 + S^t_3; \\ S^t_1 & \text{if } I = S^t_4 + S^t_3 + S^t_2. \end{cases}$$

The sum of customer requests included in the set I in the mathematical model (5)–(8) represents the expression $\sum_{i\in I} b_i$, the sum of customer requests included in the set J then the expression $\sum_{j\in J} b_j$. The variable $f_{ij} \in \{0,1\}$ represents the (non) assignment of a customer from the set J to a customer from the set I. The quantity d_{ij} represents the distance in the respective relation ij. The objective function (5) represents the total distance resulting from the assignment of customers in the set J, to customers in the set I and is intended to minimize it. The conditions (6) ensure that each J customer can be assigned to only one I customer, but any J customer can be assigned to any J customer. Conditions (7) ensure that in the relevant sessions ij exactly the number of customers whose sum of requests $\sum_{j\in J} b_j$ is equal to the difference between the capacity of the service vehicle c will be assigned and the sum of the requests already covered $\sum_{i\in I} b_i$. The conditions (8) are mandatory conditions and define the domain of the variables f_{ij}.

4 Results of the experiment

In this section, the results of the experiment with the proposed model of smart dynamic waste collection system are provided. These results are then compared with the usual fixed servicing system when all nodes are serviced at predetermined time intervals. The comparison is done from two main perspectives: the distance travelled and comfort measured in days of waiting for servicing.

The model example on which the computational experiments were performed has these parameters. Graph G(*V,E,d)*, where |V | = 51 peaks, c = 6 units, |R| = 3 vehicles that can be used repeatedly at a given time t, T = 92 days.

4.1 Smart servicing system

The *smart* customer service method is used to serve customers based on the algorithm described in the 3 chapter. Customer requests $j = 2, ..., m$ in appropriate modes $x \in X$ at time points $t \in T$ are $b^t_{4j}, b^t_{3j}, b^t_{2j} = 1$ units and $b^t_{1j} = 0.5$ units. All 50 resident nodes were randomly split into 5 resident groups described in the previous section without sake of generality as follows: the modelled village has 10 pensioners, 12 couples, 15 small families, 8 big families, and 5 companies, see the details on the network in Tab. 1. The parameters b All bins are empty at $t = 1$. A bin must be serviced not later than 2 days after reporting that it is full.

Within the specified period T, it was necessary to plan 36 routes of utility vehicles, corresponding to the time points t = {10,11,15,18,19,21,24,26,28,29,30,34,36,38,40,43,44,47,50,52,55,57,58,60,61,64,66a, 66b,70, 73,75,78,80,83,84,86}. The lengths of individual routes are given in Tab. 2. In total, 1012.82 =. 1013 km will be driven in the specified period of customer service in the specified period T = 92 days.

Route	[km]	Route	[km]	Route	[km]	Route	[km]	Route	[km]	Route	[km]
1	34.83	7	18.16	13	40.35	19	35.90	25	18.21	31	32.34
2	14.08	8	17.67	14	29.14	20	24.47	26	21.64	32	23.1
3	24.82	9	36.12	15	28.96	21	27.21	27	22.03	33	26.83
4	37.84	10	23.63	16	26.37	22	33.92	28	33.33	34	31.71
5	40.32	11	26.83	17	30.21	23	35.88	29	26.49	35	34.1
6	29.71	12	21.94	18	21.58	24	31.08	30	36.98	36	15.04

Table 2: Evaluation of a dynamic way of serving a set of customers

In Fig. 1, the development of mode in time for five randomly selected nodes (one for each category) is shown. The values on the horizontal axis correspond to the days when the nodes were serviced. The differences in fill rates are easily traceable there.

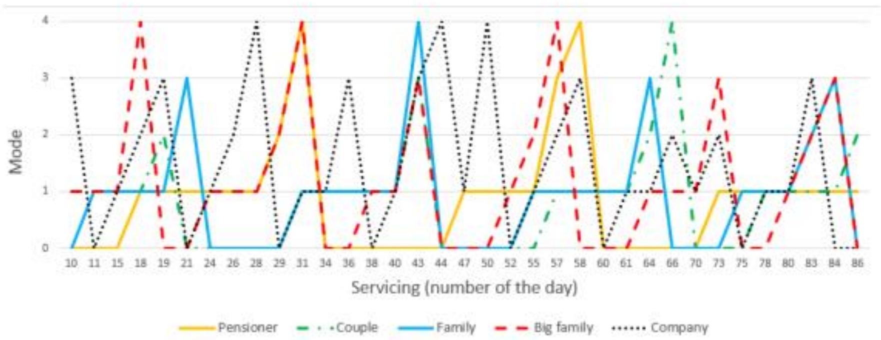


Figure 1: Development of modes over time for 5 selected nodes of different types

4.2 Fixed servicing system

With the static way of serving a set of customers, it will work with a regular service interval of 1x14 days. At time points $t = \{1, 15, 21, 43, 57, 71, 85\}$ all customers located in the individual vertices of the graph will be served. In this approach, the average request size $\phi b_j = 1$ units for $j = 2, ..., m$ will be calculated. Given this average request size ϕb_j, the number of serviced customers $|V| - 1$ and the capacity of the service vehicle c, it is clear that p routes of service vehicles must be implemented at any time t. vehicles. If we accept that $p \geq \frac{(|V|-1)\cdot\phi b_j}{c}$, then after substituting specific values we calculate that $p \geq \frac{50\cdot 1}{6}$, i.e. $p = 9$ routes. Several approaches can be used to determine the optimal sequence of peaks in these 9 routes. One of them is, for example, the *ClusterFirst-RouteSecond* approach, where the initial set of 50 customers is appropriately divided into subsets that match the capacity of the service vehicle. The optimal route is then searched for in these subsets, either using an exact or heuristic algorithm. Within the presented static approach, the so-called p-median approach [10] was used to decompose the initial set of customers, and the exact linear model for finding the minimum Hamiltonian circle [11] was used to find the optimal route. The results are summarized in Tab. 3.

Vehicles	Route 1 [km]	Route 2 [km]	Route 3 [km]	SUM [km]
Vehicle 1	10.52	26.29	4.92	41.73
Vehicle 2	11.96	24.88	7.50	44.34
Vehicle 3	12.98	12.96	13.63	39.57

Table 3: Evaluation of static customer service

Each of the 9 routes was divided among 3 service vehicles. The lengths of individual routes, as well as the total number of kilometres travelled by individual vehicles, are given in Table 3. In total, 125.64 km will be covered with this type of service. Taking into account the service at individual time points t, it can be deduced that, in total, $7 \cdot 125.64 =$. **880** km will be covered by the static method of service of the set of customers in the specified period T = 92 days.

It was shown that the smart system needs by 15% longer travelling distance when considering the servicing every 2 weeks, on the other hand by more than 42% shorter when servicing every week. However, not only an economical point of view is important. The system should also be comfortable for inhabitants who do not want to wait for servicing too long. The results of the analysis on comfort for all considered settings can be found in Tab. 4. The proposed smart model revealed the shortest waiting time on average when comparing with both systems with the fixed servicing (the values for the fixed regime were calculated using the mean value of the filling time shown in Tab. 1. These results confirm that the proposed model has a great potential to bring great benefits for practice: the greater level of comfort for citizens can be expected accompanied with potential reduction of the distance travelled.

[Days out of 90 for all 50 nodes]	New smart model	Fixed schedule 7D (AVG)	Fixed schedule 14D (AVG)
Sum	253	360	2964
Mean	5.06	7.02	59.28

Table 4: Comparison of the proposed smart model and the classical fixed settings (with servicing every 7 and 14 days, respectively)

5 Conclusions

The research was focused on improving the waste collection system using the smart features. The proposed model assumes the existence of a smart application where citizens can report that their bins are full and need servicing. An experiment using the artificial data showed that the proposed model is potentially viable and beneficial. The results proved that a higher level of comfort can be expected through a shorter mean waiting time for service. Moreover, depending on the settings of the system, the total distance travelled could also be reduced. It is worth emphasizing that the presented study is the first step in our research. It has many strong assumptions, which could be relaxed in the future to explore the potential of the presented idea into details.

Acknowledgements

This work was supported by the SGS project No. SP2022/113. This support is gratefully acknowledged.

References

[1] Harrison, C., Eckman, B., Hamilton, R., Hartswick, P., Kalagnanam, J.,Paraszczak, J., & Williams, P.: Foundations for smarter cities. *IBM Journal of Research and Development*, 54(4) (2010), 1–16.

[2] Portmann, E. & Finger, M.: Smart cities – ein Überblick.In *Meier A & Portmann, E., editors,HMD Praxis der Wirtschaftsinformatik* 52(4) 2015, 470–481.

[3] Lu, J. W., Chang, N. B., Liao, L., & Liao, M. Y.: Smart and green urban solid waste collection systems: advances, challenges, and perspectives. *IEEE Systems Journal* 11(4) 2015, 2804–2817.

[4] Popa, C. L., Carutasu, G., Cotet, C. E., Carutasu, N. L., & Dobrescu, T.: Smart city platform development for an automated waste collection system. *Sustainability* 9(11) 2017, 2064.

[5] Anagnostopoulos, T., Kolomvatsos, K., Anagnostopoulos, C., Zaslavsky, A., & Hadjiefthymiades, S.: Assessing dynamic models for high priority waste collection in smart cities. *Journal of Systems and Software* 110 2015, 178-192.

[6] Gutierrez, J. M., Jensen, M., Henius, M., & Riaz, T.: Smart waste collection system based on location intelligence. *Procedia Computer Science* 61 2015, 120–127.
[7] Soh, Z. H. C., Husa, M. A. A. H., Abdullah, S. A. C., & Shafie, M. A.: Smart waste collection monitoring and alert system via IoT. In *2019 IEEE 9th Symposium on Computer Applications & Industrial Electronics (ISCAIE) 2019*, 50–54.
[8] Ramos, T. R. P., de Morais, C. S., & Barbosa-Póvoa, A. P.: The smart waste collection routing problem: Alternative operational management approaches. *Expert Systems with Applications* 103 2018, 146–158.
[9] Fujdiak, R., Masek, P., Mlynek, P., Misurec, J., & Olshannikova, E.: Using genetic algorithm for advanced municipal waste collection in smart city. In *2016 10th International symposium on communication systems, networks and digital signal processing (CSNDSP)*, 2016, 1–6.
[10] Kozel, P., Orlíková, L., Pomp, M., & Michalcová, Š.: Application of the p-Median Approach for a Basic Decomposition of a Set of Vertices to Service Vehicles Routing Design. In *Proceedings of 36th International Conference Mathematical Methods in Economics*, *2018*, 252–257.
[11] Janáček, J. *Matematické programování*. Žilina: Žilinská univerzita, 2003