STATISTIKA

STATISTICS AND ECONOMY JOURNAL

VOL. **103** (4) 2023

CZECH STATISTICAL OFFICE

CONTENTS

ANALYSES

INFORMATION

About Statistika

The journal of Statistika has been published by the Czech Statistical Office since 1964. Its aim is to create a platform enabling national statistical and research institutions to present the progress and results of complex analyses in the economic, environmental, and social spheres. Its mission is to promote the official statistics as a tool supporting the decision making at the level of international organizations, central and local authorities, as well as businesses. We contribute to the world debate and efforts in strengthening the bridge between theory and practice of the official statistics.

Statistika: Statistics and Economy Journal is professional double-blind peer reviewed open access journal included e.g. in the *Web of Science Emerging Sources Citation Index* (since 2016) of the **Web of Science Core Collection** database (**Impact Factor** 2022: **0.2**), in the international citation database of peer-reviewed literature **Scopus** (since 2015, SJR 2022 = 0.179, CiteScore 2022 = 0.6), and others. Since 2011, Statistika has been published quarterly in English only.

Publisher

The Czech Statistical Office is an official national statistical institution of the Czech Republic. The Office´s main goal, as the coordinator of the State Statistical Service, consists in the acquisition of data and the subsequent production of statistical information on social, economic, demographic, and environmental development of the state. Based on the data acquired, the Czech Statistical Office produces a reliable and consistent image of the current society and its developments satisfying various needs of potential users.

Contact us

Journal of Statistika | Czech Statistical Office | Na padesátém 81 | 100 82 Prague 10 | Czech Republic
e-mail: statistika.journal@czso.cz | web: www.czso.cz/statistika_journal

Comparison of (Quasi-)Joblessness in Slovakia and the Czech Republic through the Marginal Means based on Logit Models

Martina Košíková[1] | *University of Economics in Bratislava, Bratislava, Slovakia*
Erik Šoltés[2] | *University of Economics in Bratislava, Bratislava, Slovakia*

Received 26.2.2023, **Accepted** (reviewed) 19.4.2023, **Published** 15.12.2023

Abstract

In the paper, we analyze one of the aspects of measuring poverty and social exclusion in the world within the context of the Europe 2030 Strategy, which is represented by very low work intensity. Using the data obtained from the EU-SILC 2021 statistical survey (for Slovakia and the Czech Republic), we apply logistic regression methods and generalized linear models to quantify the impact of relevant categorical factors on the binary dependent variable very low work intensity of Slovak and Czech households. Based on the obtained results, we then process a comparative analysis, through which we quantify the same and also different features of these countries in terms of (quasi-)joblessness.

Keywords	DOI	JEL code
(Quasi-)joblessness, logistic regression, least squares means, work intensity	*https://doi.org/10.54694/stat.2023.8*	*C12, C21, E24*

INTRODUCTION

Currently, poverty and social exclusion are two of the global problems affecting the quality of life of people around world. Poverty and social exclusion are generally considered a socio-economic problem affecting an individual or a household, and these social phenomena occur to a greater or lesser extent in every country in the world. Šoltés et al. (2018) define that poverty and social exclusion affect people's quality of life and are assessed from several aspects, such as the amount of equivalent disposable income (income poverty), material deprivation or very low work intensity, which is often referred to as (quasi-) joblessness (QJ).[3]

[1] Department of Statistics, Faculty of Economic Informatics, University of Economics in Bratislava, Dolnozemská cesta 1/b, 852 35 Bratislava, Slovakia. E-mail: martina.kosikova@euba.sk.

[2] Department of Statistics, Faculty of Economic Informatics, University of Economics in Bratislava, Dolnozemská cesta 1/b, 852 35 Bratislava, Slovakia. E-mail: erik.soltes@euba.sk. ORCID: <https://orcid.org/0000-0001-8570-6536>.

[3] In the following text, we will also use the abbreviation QJ to denote (quasi-)joblessness.

In the paper, we focus on the risk that a person will live in a household with very low intensity, which is one of the three dimensions of poverty and social exclusion based on the concept of the EU 2030 Strategy. Work intensity is defined as the rate at which work potential is used in the household. The Eurostat methodology (2021) defines that work intensity represents the ratio of the number of months worked by individual household members of productive age (18–64 years) to the total number of months (within the reference period) that household members could have worked (persons of productive age don‘t include students aged 18 to 24). The work intensity indicator takes on values from 0, respectively 0% (none of the household members of productive age worked in the reference period) after 1, or 100% (every household member of productive age worked during the entire reference period). Households in which the work potential was used to less than 20% are considered as households with very low work intensity. Other levels of work intensity are discussed, for example by Eurostat (2022a), or Šoltés and Vojtková (2018). Households with very low work intensity are also referred to as (quasi-) joblessness households. According to Vlačuha and Kováčová (2021), (quasi-)joblesness households have a demonstrably higher risk of poverty than households with high or very high work intensity. This dimension of poverty and social exclusion should be particularly monitored because it significantly affects the poverty and social exclusion of children, which was also confirmed by the authors De Graaf-Zijl and Nolan (2011), who defined in their studies that (quasi-)joblessness in most countries is closely associated with child deprivation.

Many studies have reported that the risk of poverty and social exclusion depends on various measurable factors; for example, according to Duiella and Turrini (2014) these are factors such as gender, age, migration, geographical region. Eurostat (2022b) analyzed the risk of poverty or social exclusion based on socio-economic statistics such as not only gender and age, but also education and economic activity. The authors Johnston and McGauran (2018) in their studies analyzed (quasi-)unemployment depending on the type of households, economic activity and other socio-economic factors (level of education, age, gender, disability, etc.).

The results of Eurostat (2022b) showed that approximately 15.6% of the population in Slovakia was at risk of poverty in 2021 and Slovakia along with the Czech Republic (10.7%) are among the EU countries with the lowest poverty risk rate. However, there is a high risk of poverty in (quasi-)joblessness households in Slovakia (76.5%) and the Czech Republic (58.6%).

In the paper, we analyze the risk that a person will live in a (quasi-)joblessness household in Slovakia and the Czech Republic. The goal is to assess the impact of relevant factors on this risk in both countries individually and at the same time, to compare the results of the analyses. We mainly focus on:

1. assessment of the statistical significance of the impact of selected factors on the probability of living in a (quasi-)joblessness household,
2. identify the categories of which factor between which there are no significant differences and those categories or groups of categories between which there are demonstrable differences in terms of the risk of (quasi-)joblessness,
3. identify the risk g roups of persons in terms of the risk of (quasi-)joblessness,
4. comparison ofprobability of (quasi-)joblessness for different profiles of people determined by the three most fundamental factors.

All the stated goals are implemented specifically for persons living in Slovak households and specifically for persons living in Czech households. The paper further provides a comparison between Slovakia and the Czech Republic according to the mentioned attributes.

1 LITERATURE REVIEW

Poverty is a complex, multidimensional phenomenon that cannot be expressed only on the basis of one aspect or one definition. Poverty is primarily a socio-economic problem that has affected human society since its inception. There are several concepts of poverty that are dealt with in various scientific works. For example, the authors Atkinson et al. (2017), who analyzed poverty indicators based on the EU-SILC survey with macroeconomic analysis of aggregates within the EU countries and found that the EU did not achieve any positive change or progress in achieving the goal in the Europe 2020 Strategy (until 2020 the number of people at risk of poverty and social exclusion was reduced at least by 20 million). Ravallion (2020) stated in his study that over the last 30 years, poverty has decreased globally, which is mainly due to the reduction of absolute poverty in the developing world. Based on Brown and Long (2018), absolute poverty refers to people who lack their basic physical needs. Nolan and Marx (2009) assessed the factors affecting poverty differences in the OECD countries. Lafuente et al. (2020) evaluated the convergence of poverty between the EU countries from the perspective of the three dimensions of poverty and social exclusion in the concept of the Europe 2020 Strategy, finding that convergence in the EU countries exists and takes place based on a long-term process, in which the levels of individual member countries with similar characteristics are gradually approaching the levels of more developed countries.

OECD (2009) confirmed the existence of a close relationship between the level of personal income and the intensity of household work. On the contrary, De Graaf-Zijl and Nolan (2011) stated in their study that joblessness is not directly related to income poverty and pointed out that the risk of (quasi-) joblessness most often occurs among persons with a lower education, a disability or in a single-person household.

The selection of factors affecting (quasi-)joblessness was determined by the results of our previous research (e.g. Šoltés et al., 2022) and studies that analyzed this issue, such as Guio et al. (2022), who dealt with the analysis of the determinants of child deprivation in 31 European countries, or Johnston and McGauran (2018), who assessed the impact of relevant socio-economic factors on very low work intensity. Duiella and Turrini (2014) stated that for a more accurate analysis of (quasi-)joblessness it was important to consider socio-demographic and economic indicators such as age, gender, type of household or education. Since (quasi-) joblessness is closely related to other dimensions of poverty and social exclusion, when choosing potential factors we were also inspired by scientific works that assessed income poverty and material deprivation. Filandri and Strufolino (2019) found that between the indicators affecting the risk of working poverty were a low degree of education and the young age of persons, while households with a larger number of children and a small number of economically active persons are also at risk.

In the paper, we analyze (quasi-)joblessness using logistic regression. Logistic regression is quite popular in the study of poverty and social exclusion, which is confirmed by a number of studies. For example, Dudek and Lisicka (2013) in their study analyzed the incidence of poverty risk through binary logistic regression. Using logistic regression, Rusnak (2012) identified indicators that increase poverty. Saccone and Deaglio (2020) compared poor and developing economies with high-income economies using logit analysis. Gallardo (2020) analyzed multidimensional poverty using probit and logit models. The adequacy of applying logit models in the area of social exclusion was also confirmed by Gao et al. (2022), who used these models to identify poor households, than Stanley et al. (2011) who analyzed the factors that significantly contribute to the increased risk of social exclusion or Bradshaw et al. (2000) who assessed the relationship between poverty and social exclusion through logit models.

As we stated, logistic regression models are relatively extended in poverty and social exclusion research, but their potential is mostly not used to the full extent. A deeper analysis of the impact of individual factors on which dimension of poverty and social exclusion is provided by a contrast analysis, which is bound to such a model. For this reason, in our research on (quasi-)joblessness, we use not only the logit model, but also the contrast analysis, from which the most important conclusions presented in the article stem.

2 METHODS

Logistic regression falls into a broad class of generalized linear models. In logistic regression, the target variable has a binomial distribution, the systematic component is in the form of a linear combination of the explanatory variables, and the linear function is the logit.

The logistic regression model quantifies the log-odds for the "1" level of the binomial dependent variable Y as a function of the explanatory variables. Odds represents the probability ratio p_i, that the observed event occurs ($Y = 1$) against the probability $1 - p_i$, that observed event does not occur ($Y = 0$). Odds is expressed by the relation (1) and the log-odds, also referred to as logit, is given by the relation (2):

$$odds = \frac{p_i}{1 - p_i}, \tag{1}$$

$$logit\ (p_i) = ln\left(\frac{p_i}{1 - p_i}\right). \tag{2}$$

After transforming the qualitative variable Y (0 or 1) into a continuous variable expressed by the logit, the relation between the explanatory variable and the vector of explanatory variables is linear. So, the binary logistic regression model has the form:

$$logit\ (p_i) = ln\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} \dots + \beta_k x_{ik}, \tag{3}$$

where β_j are the unknown model parameters estimated using an iterative maximum likelihood method and x_{ij} are the observed values of the explanatory variables (where $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, k$).

The significance of the logistic regression model is verified using three chi-square tests (likelihood ratio test, score test or Wald test), while Allison (2012) states that in the case of sufficiently large datasets it is not necessary to prefer any of the three tests. To verify the significance of the impact of independent variables on the dependent variable, the Wald test is used:

$$Wald = \hat{\boldsymbol{\beta}}^{\mathrm{T}} \cdot \mathbf{S}_{\mathrm{b}}^{-1} \cdot \hat{\boldsymbol{\beta}}, \tag{4}$$

where $\hat{\boldsymbol{\beta}}$ is the vector of the estimated regression coefficient and $\mathbf{S}_{\mathrm{b}}^{-1}$ is the variance-covariance matrix of the vector $\hat{\boldsymbol{\beta}}$. The Wald statistics has an asymptotically chi-squared distribution. Whereas, in the case of a quantitative variable, the number of degrees of freedom of this chi-square distribution is equal to 1 and in the case of a categorical factor, the number of degrees of freedom is equal to (r – 1), where *r* is the number of categories of the factor whose impact we are verifying.

An advantage of using logistic regression is that regression coefficients can be interpreted through odds ratios. The odds ratio is estimated from the logit model using an exponential transformation of the estimated regression coefficient $\hat{\beta}_j$, as follows

$$\hat{OR}_i = \frac{odds_1}{odds_2} = e^{\hat{\beta}_j}. \tag{5}$$

The odds ratio belonging to an independent continuous numerical variable expresses the extent to which the odd of the observed event changes with a unit increase in this numerical variable, under ceteris paribus conditions. In the case of categorical explanatory variables (categorical factors), the interpretation of the estimated odds ratios depends on the type of coding of the factors. In the case of referential coding, we assess the impact of each non-referential category on odd relative to the reference category. Effect coding compares the effects of individual categories to the mean impact across all categories of a factor.

By analyzing marginal means, we can identify hidden relationships between categories of independent variables in generalized linear models. If the estimated model includes qualitative independent variables that have more than two categories, analysis of marginal means allows comparison of differences in marginal means between pairs of categories. Marginal means, also referred to as LS means (least squares means) or EM means (estimated marginal means), are generally preferred over conditional means because they correct for data unbalance. LS means disregard the frequency of individual levels of the factor and when quantifying the means assign equal weight to each category, while arithmetic mean assigns individual weight to each category of the factor, which is conditioned by its frequency. Based on Suzuki et al. (2019) we note that while arithmetic means are estimated from the data, marginal means are estimated directly from the model. Wang et al. (2018) stated that for unbalanced datasets with more categorical or numerical factors, arithmetic means do not provide an adequate picture of the effect of the dependent variable for a particular factor because they ignore other effects, which may subsequently lead to Simpson's paradox. This paradox points to the importance of not ignoring causal relationships between variables (cause and effect) or confounding effects leading to incorrect conclusions in statistical analyses.

However, if we want to assess the significance of the differences in the marginal mean values of the independent variable between more than two categories at the same time, we use contrast analysis for this purpose. The basis of contrast analysis is general linear hypothesis (GLH) testing. We test hypotheses in the form of general linear hypotheses

$$H_0: c_1\mu_1 + c_2\mu_2 + ... + c_p\mu_p = 0 \,, \tag{6}$$

where μ_j is the marginal mean value of the independent variable for the jth level of the factor whose impact we are assessing (Batzer, 2020). These hypotheses are tested through a linear combination $c_1\beta_1 + c_2\beta_2 + ... + c_p\beta_p$ parameters of the model β_j.

To test GLH, logistic regression uses the Wald statistic, which has an asymptotically chi-square distribution, which has one degree of freedom in the case of testing one linear combination and in the case of a simultaneous test of several linear combinations, the number of degrees of freedom is l, where l is the number of partial hypotheses that enter to the simultaneous test. Littell et al. (2010) provide additional information.

The LOGISTIC and GENMOD procedures in SAS Enterprise Guide statistical software are used in the paper to estimate logit models. The analysis of marginal means is realized in the SAS programming language through the LSMEANS statement in PROC GENMOD and contrast analysis was applied using the CONTRAST statement in PROC LOGISTIC. To identify the risk profiles of people in terms of (quasi-) joblessness in Slovakia and the Czech Republic and for the subsequent comparison between countries, the probabilities of living in a (quasi-)joblessness household were estimated in the paper. Point and interval estimates of the subject probabilities were quantified using the ESTIMATE statement. In more detail about the procedure of adequate testing of marginal means and their estimation by CONTRAST and ESTIMATE statement, e.g. Littell et al. (2010) a SAS Institute Inc. (2018).

3 DATABASE

The analyses presented in the paper are based on the EU-SILC 2021 survey. The input databases were provided by the statistical offices of Slovakia and the Czech Republic and consist of data relating to persons in Slovak and Czech households. The target variable is very low work intensity (VLWI). It is a binary variable with levels of "no" for persons not living in a household with very low work intensity and "yes" for persons living in a household with very low work intensity. The explanatory variables included in the model are listed in Table 1.

Table 1 Information about input factors and their categories

Factor	Levels	Description
EA (Economic Activity)	*Disabled_person* *Inactive_person* *Person_in_household* *Student* *Unemployed* *z_at_Work*	Disabled person, Person unable to work Other inactive person Person in the household Student Unemployed Employed
EDUCATION	*ISCED 0–2* *ISCED 3* *ISCED 4–5* *ISCED 6* *ISCED 7–8*	Primary and Lower secondary education Upper secondary education Post-secondary and Short-cycle tertiary education Bachelor education Master's or Doctorate education
HT (Type of Household)	*1A_0Ch* *1A_1+Ch* *2A(1+R)* *2A_0Ch* *2A_1Ch* *2A_3+Ch* *Other_0Ch* *Other_1+Ch* *z_2A_2Ch*	Household of 1 adult without dependent children Household of 1 adult with at least 1 dependent child Household of 2 adults with at least 1 aged 65+ Household of 2 adults without dependent children Household of 2 adults with 1 dependent child Household of 2 adults with at least 3 dependent children Other household without dependent children Other household with dependent children Household of 2 adults with 2 dependent children
MARITAL STATUS	*Divorced* *Never_married* *Widowed* *z_Married*	Divorced person Single person Widowed person Married person
AGE	*30–40* *40–50* *50+* *z_–30*	People aged between 31 and 40 years People aged between 41 and 50 years People aged 51 and over People under the age of 30
URB (Urbanisation)	*Sparse* *Intermediate* *z_Dense*	Sparsely populated area Intermediate area Densely populated area

Source: EU-SILC 2021 SO SR and CR, own processing

4 RESULTS

4.1 Analysis of the significance of the input factors impact on (quasi-)joblessness in Slovakia and the Czech Republic and modification of input factors

In addition to the factors listed in Section 3, explanatory factors such as gender, region and health status were also included in the analysis, but the method of stepwise elimination showed that these factors did

Table 2 Verification of the factors impact on (quasi-)joblessness

Type 3 Analysis of effects

Effect	DF	Slovakia		Czech Republic	
		Wald Chi-Square	Pr > ChiSq	Wald Chi-Square	Pr > ChiSq
EA	5	548.3650	<.0001	709.0464	<.0001
Education	4	116.9450	<.0001	12.2144	0.0158
HT	8	121.7755	<.0001	172.5083	<.0001
Marital_status	3	6.5516	0.0876	15.7331	0.0013
Age	3	8.2725	0.0407	29.0038	<.0001
Urbanisation	2	28.5390	<.0001	7.9296	0.0190

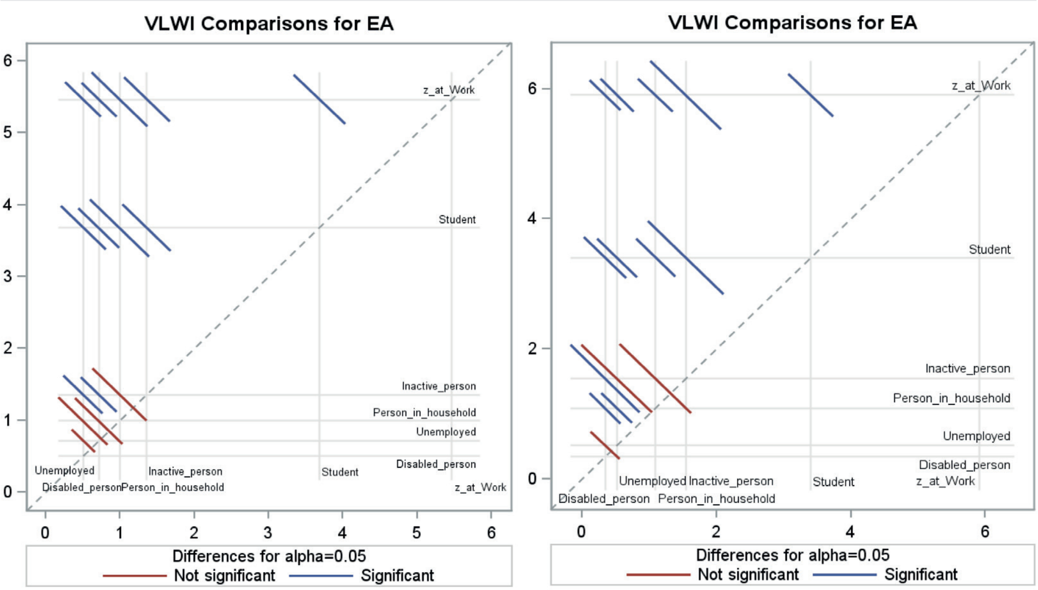
Source: EU-SILC 2021 SO SR and CR, own processing in SAS EG

not have a significant impact on the target variable at the significance level of 0.1. Only those variables listed in Table 1 were included in the logit model separately estimated for Slovakia and the Czech Republic. Table 2 confirms the statistical significance ($\alpha = 0.1$) of the impact of these variables on the target variable.

In both countries, economic activity is a key factor, that significantly affects VLWI. Other factor in the both countries, that fundamentally affects the probability that a person lived in a (quasi-)joblessness household in 2021 is the type of household. Through this variable, we confirm that not only the number of dependent children, but also the number of economically active members in the household has an impact on exclusion from the labor market. In the case of Slovakia, the third most important factor is a person's education, and in the Czech Republic it is age.

An analysis of least squares means performed using the LSMEANS statement in the SAS programming language showed, which categories of individual factors are the most risky and the least risky in terms of (quasi-)joblessness and between which pairs of categories of individual factors there are significant differences in the marginal means of the logit probability of living in a (quasi-)joblessness household. Based on the results of the analyses, we will assess the possibility of clustering the most similar categories into one newly created category and through this we will simplify following results.

Figure 1 Interval estimates of the LS means of the logit of (quasi-)joblessness depending on the status of economic activity (left Slovakia, right the Czech Republic)



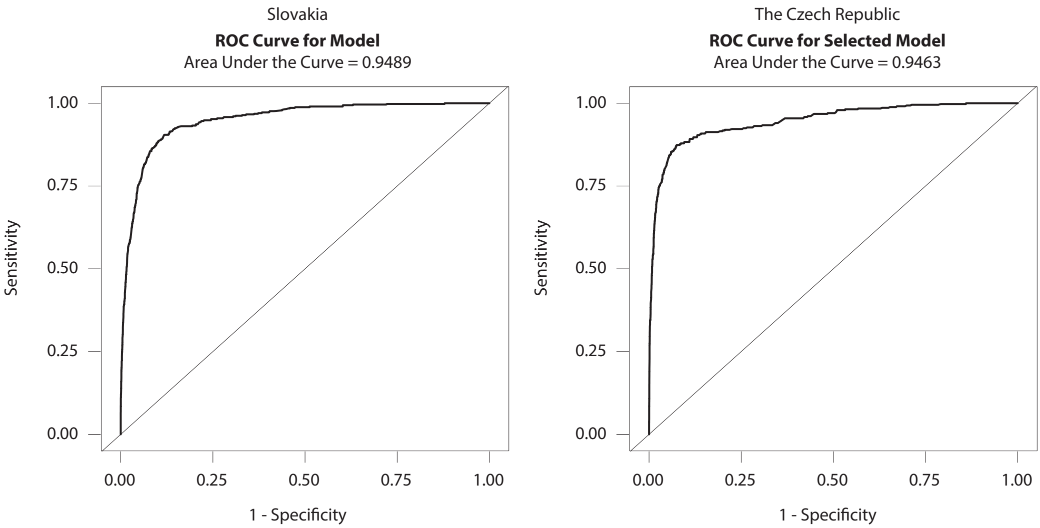
Source: EU-SILC 2021 SO SR and CR, own processing in SAS EG

In the case of several pairs of EA and Education factor categories, statistically significant differences (Figure 1, red lines) in the marginal means of the logit were not confirmed at the 0.05 significance level. In 2021, in both analyzed countries, there was no significant difference between disabled people and unemployed people and between inactive people and people in household (Figure 1). In the next analysis, we used clustered pairs of categories and denoted them DP-UP (Disabled Persons and Unemployed Persons) and IP-PH (Inactive Persons and Persons in Household). We proceeded similarly in the case of the Education factor, in which we clustered the ISCED 3 and ISCED 4-5 categories and the ISCED 6 and ISCED 7-8 categories, between which there was no significant difference. After this modification, the Education factor had 3 categories, namely ISCED 0-2, ISCED 3-5 and ISCED 6-8.

4.2 Quality of logit models and odds ratios estimation of (quasi-)joblessness in Slovakia and the Czech Republic

Adequacy and success of the model in prediction are confirmed by the association measures, namely Sommer's D, Goodman-Kruskal's gamma or the c statistic, whose values are relatively high. The comparison of concordant and discordant pairs also indicates good model quality.

Figure 2 ROC curve for the logit model of Slovakia and the Czech Republic



Source: EU-SILC 2021 SO SR and CR, own processing in SAS EG

The quality of the model is also confirmed by the area under the ROC curve (AUC) characterized by the c statistic, which is 0.9483 in Slovakia and 0.9465 in the Czech Republic (Figure 2).

Economic activity naturally has the greatest impact on the risk of (quasi-)joblessness. In both countries the most at risk are disabled and unemployed persons (DP-UP category). The odds of living in a (quasi-) joblessness household for persons living in Slovakia, who are disabled or unemployed are 127.661 times higher than for employed persons. This odd ratio is almost 2 times higher in the Czech Republic (*OR* = 242.54). Let us note that these results do not indicate that the probability of living in (quasi-) joblessness household in the Czech Republic is greater for disabled and unemployed persons than in Slovakia. This oddity compares the probability of living in a QJ household for disabled and unemployed persons in relation to the probability of living in a QJ household for employed persons. The stated result may be a consequence of the fact that employed persons in Slovakia have a higher probability to live (quasi-)joblessness household than in the Czech Republic. As shown in Figure 3a), in our case the reason for the higher odds ratio in the Czech Republic is precisely this fact.

Other high-risk economic activity statuses in both countries are otherwise inactive people and people in the household. Such people again have a higher odd in the Czech Republic (*OR* = 118.803) than in Slovakia (*OR* = 71.190). Johnston and McGauran (2018) or Šoltés et al. (2022) also confirmed that disabled and unemployed people are the most risky and employed people are the least risky in terms of exclusion from the labor market.

Table 3 Parameter estimates and odds ratios

Analysis of Maximum Likelihood Estimates and Odds Ratio Estimates							
Parameter		**Slovakia**			**Czech Republic**		
		Estimate	**p-value**	**OR**	**Estimate**	**p-value**	**OR**
Effect		–8.6759	<.0001	–	–7.6830	<.0001	–
EA	DP-UP	4.8494	<.0001	127.661	5.4913	<.0001	242.540
	IP-PH	4.2654	<.0001	71.190	4.7776	<.0001	118.803
	Student	1.8584	<.0001	6.413	2.5400	<.0001	12.678
	at_Work	.					
HT	1A_0Ch	2.9550	<.0001	19.203	2.1277	<.0001	8.395
	1A_1+Ch	2.4064	<.0001	11.093	2.6470	<.0001	14.111
	2A(1+R)	3.5837	<.0001	36.006	2.9903	<.0001	19.891
	2A_0Ch	1.3836	<.0001	3.989	0.9384	0.0012	2.556
	2A_1Ch	1.0929	0.0016	2.983	0.3359	0.2165	1.399
	2A_3+Ch	1.1684	0.0013	3.217	0.5148	0.1031	1.673
	Other_0Ch	1.1893	0.0005	3.285	0.7215	0.0164	2.058
	Other_1+Ch	1.0434	0.0011	2.839	–0.3654	0.3387	0.694
	2A_2Ch	.					
ED	ISCED 0-2	2.3241	<.0001	10.217	0.4981	0.0242	1.646
	ISCED 3-5	0.9585	0.0002	2.608	0.0305	0.8946	1.031
	ISCED 6-8	.					
MARITAL STATUS	Divorced	0.3873	0.1420	1.473	–0.3139	0.1987	0.731
	Never_married	0.4531	0.0091	1.573	0.5732	0.0044	1.774
	Widowed	0.1552	0.6854	1.168	–0.1051	0.8384	0.900
	Married	.					
AGE	30–40	0.5021	0.0141	1.652	0.3660	0.1205	1.442
	40–50	0.6038	0.0075	1.829	1.0203	<.0001	2.774
	50+	0.3309	0.1667	1.392	1.4256	<.0001	4.160
	–30	.					
URB	Intermediate	0.5285	0.0044	1.696	–0.3392	0.0414	0.712
	Sparse	–0.1988	0.2827	0.820	–0.4196	0.0129	0.657
	Dense	.					

Source: EU-SILC 2021 SO SR and CR, own processing in SAS EG

The second most affecting factor is the type of household. In both countries, in terms of household type, the most risky persons living in household type 2 adults with one person aged 65+. In Slovakia, the odd of living in a (quasi-)joblessness household for such a person was 36.006 times higher than for a person from a household of 2 adults with 2 dependent children. In the Czech Republic was this odd ratio at the level of 20. Since the work intensity of a household depends on the use of work potential by all adults in the household, in the case of households with one adult person of productive age (1A_1+Ch, 1A_0Ch, 2A(1+R)) there is a threat of very low work intensity per person , while in households with more adults, this risk is shared among more adults, and therefore it is natural that a person living in a household with more adults has a lower risk of living in a QJ household. This assumption is also confirmed by the results of our research shown in Table 3.

Based on education, in both countries, the lowest education level (ISCED 0-2) is the most risky and the tertiary level of education (ISCED 6-8) is the least risky. Odds ratios estimated for individual categories of other factors (Table 3) show that the lowest risk of living in a QJ household is in Slovakia and also in the Czech Republic for never married people, in Slovakia for people aged 40–50 and in the Czech Republic for people aged 50 +. In terms of urbanisation, these are intermediate areas in Slovakia and densely-populated area in the Czech Republic. Compared to odds ratios, the probabilities provide a better idea of the factor impact on the risk of living in a QJ household. Therefore, in the next parts of the paper, we will focus on the estimates of these probabilities.

4.2.1 Probabilities of (quasi-)joblessness in relation to individual factors

Figures 3a to 3f present the estimated probability of QJ in Slovakia and the Czech Republic for individual categories of factors. These probabilities were estimated based on the logit models presented in the previous part of the paper using the ESTIMATE statement in the SAS programming language.

Factor EA

Figure 3a confirms our previous finding that the most risky group of persons in both countries are disabled and unemployed persons, for which we estimated the probability of living in a QJ household at the level of 40.9% (33.7%–48.5%)[4] in Slovakia, or 39.5% (31.2%–48.5%) in the Czech Republic. The next most risky statuses of economic activity are inactive persons and persons in the household. Across these two statuses, we estimated the probability of QJ at the level of 27.9% (20.5%–36.7%) in Slovakia and at the level of 24.2% (17.9%–31.9%) in the Czech Republic. The lowest risk of exclusion from the labor market has employed persons for whom the probability of QJ does not exceed 1% in any of the considered countries. All the above results for individual economic activity statuses are estimated across all categories of individual factors included in the logit models. So, these are the mean QJ probabilities for persons with different economic activity statuses across all levels of education, types of households, marital statuses, age categories and urbanization. In this sense, it is necessary to interpret the QJ probabilities estimated in other parts of this subsection.

Factor HT

Figure 3c confirms the finding we arrived at based on the odds ratios (Table 3) that the greatest risk of QJ in 2021 was for persons living in households with two adults, at least one of whom is 65 or older. Figure 3c confirms (Table 3) that the lowest risk of QJ in 2021 was for persons who live in households of 2 adults of which one is aged 65 and older. In Slovakia, we estimated the probability for this category at the level of 36.9% (24.3%–51.6%) and the Czech Republic by approximately 6 pp lower. The second most risky category in Slovakia was a household of 1 adult, with a probability of 23.8%. In the Czech Republic were people from this type of household in the third place of risk (15.6%), because they were overtaken in the risk of QJ by persons from households of 1adult with at least 3 dependent children (23.7%). People from these 3 types of households, i.e. from households in which there is only 1 adult of productive age, had a significantly higher risk of QJ than persons from other types of households, in which the probability was below 5% in both countries.

Factor Education

The results of our analyses presented in Figure 3b revealed that in 2021 the Education factor had a different effect on the probability of QJ in Slovakia and the Czech Republic and showed that education

4 In the full paper, after the point estimate of probability is present in parentheses the 95% confidence interval.

had a lower impact on the target variable in Slovakia than in the case of the Czech Republic. Although in both countries the most risky persons were those with the lowest level of education (in Slovakia 17.3%–28.4%; in the Czech Republic 6.2%–11.2%), in Slovakia the probability of living in QJ households for these persons was more than 10 pp higher than in the Czech Republic. In general, with the increasing education level of persons, the probability of the risk of QJ decreases. However, in the case of the Czech Republic non-significant difference (p = 0.8946) was confirmed between persons with higher secondary education or post-secondary education (ISCED 3-5: 5.4%) and tertiary education (ISCED 6-8: 5.3%). Economic activity, education and type of household impact risk of QJ to a greater extent than other factors, which is also confirmed by the fact that in the most risky categories of the following factors, the probability did not exceed 12% in any country.

Factor Age
Based on Figure 3d, age causes smaller disparities in the probability of QJ in Slovakia than in the Czech Republic. In Slovakia, the most risky persons are aged 40–50 (9.7%; 6.8%–13.8%), while in the Czech Republic they are aged over 50 (12.0%; 8.3%–17, 0%). While in the Czech Republic the probability of QJ in 2021 increased with increasing age, in Slovakia this pattern was disrupted by the age category over 50 (see Figure 3d).

Factor Marital_Status
The riskiest marital status in 2021 in terms of QJ was never married. In Slovakia, persons with this marital status had a probability of living in a QJ household of 9.4% (7.4%–11.8%) and in the Czech Republic 10.2% (8.0%–12.8%). Married persons (6.12%) were least at risk in Slovakia and divorced persons (4.5%) in the Czech Republic, which is quite strange, since Divorced status was the second most risky in Slovakia. In both countries, there was a statistically non-significant difference in the probability of QJ between widowed and married persons (Slovakia: $p = 0.6854$, Czech Republic: $p = 0.8384$).

Factor Urbanisation
In the Czech Republic, the level of urbanization clearly had the lowest impact on the target variable of all factors, so it was not clear in Slovakia (Figure 3f). In terms of the level of urbanization, in 2021 the least risky areas were sparse populated areas (in Slovakia 5.8% and Czech Republic 5.3%, respectively

Figure 3 Probabilities estimates of living in a QJ household depending on the status of economic activity (a), education (b), type of household (c), age (d), marital status (e) and urbanization (f) for Slovakia (left) and the Czech Republic (right) in 2021

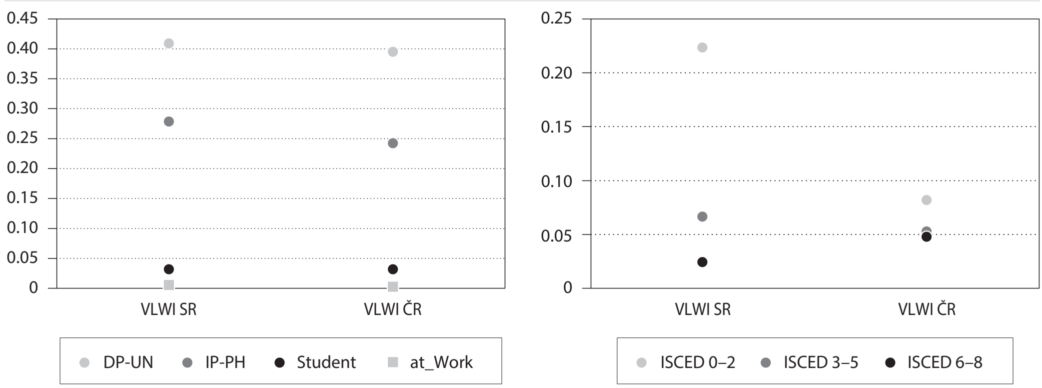
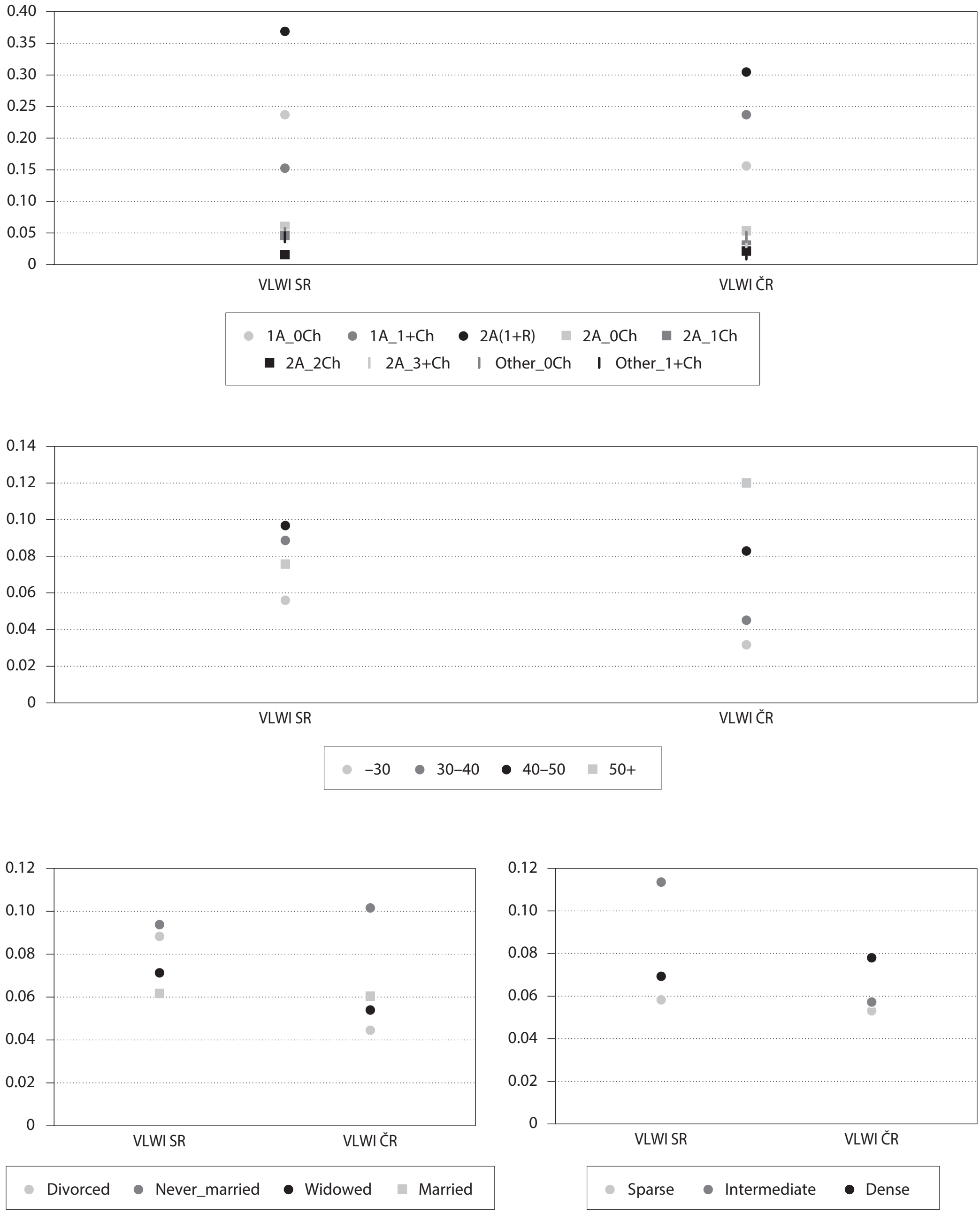


Figure 3 (continuation)



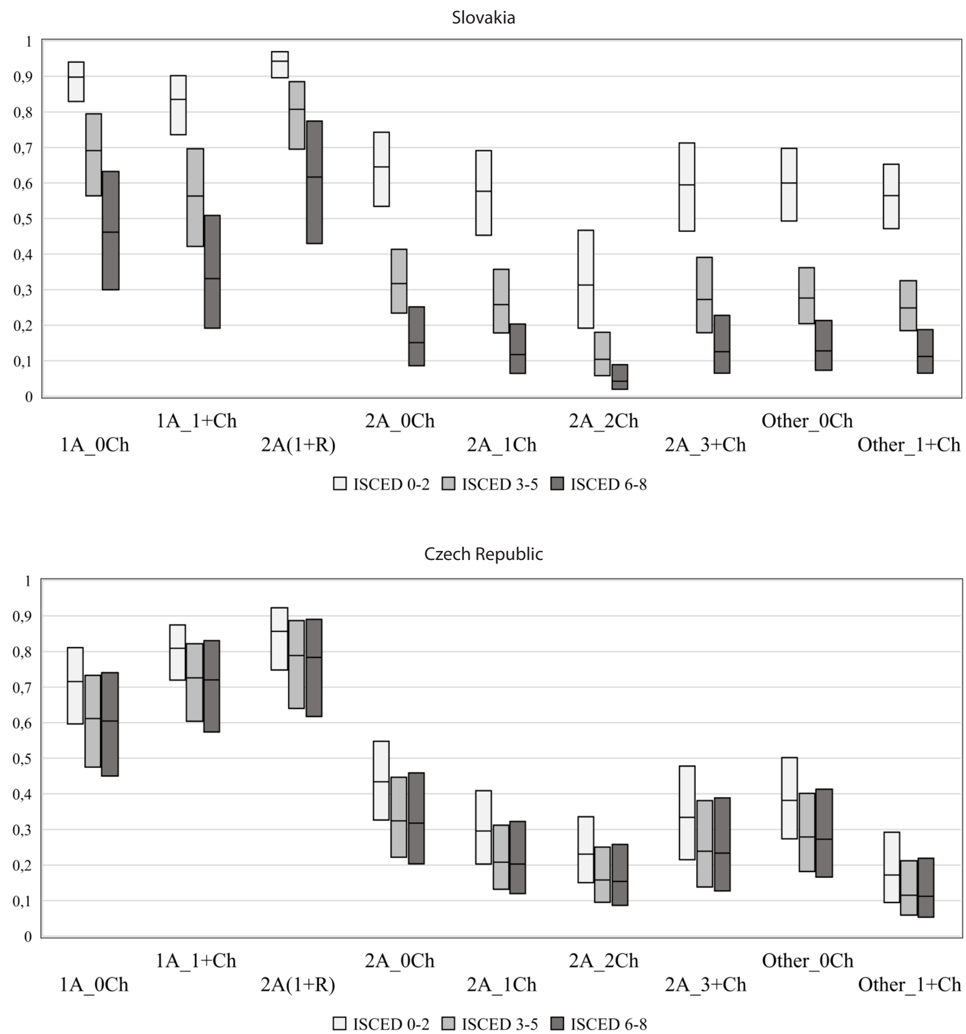
Source: EU-SILC 2021 SO SR and CR, own processing in SAS EG

4.2%–8.0% and 3.7%–5.8%). While intermediate populated areas were the most risky in Slovakia (11.4%, 8.5%–15.0%), in the Czech Republic it was densely populated areas (7.9%; 5.6%–11.0%).

4.2.2 Probabilities of QJ for people profiles determined by their status of economic activity, education and type of household

In the following part of the paper, we estimate the probabilities of QJ depending on the interaction of the three most important factors, namely status of economic activity, type of household in which this person lives and his education. For a better overview, we present the results only for the two most frequent statuses of economic activity, namely the status "unemployed or disabled" and the status "employed", while at the same time it is the most risky and the least risky status. Figure 4 shows the probability estimates for unemployed or disabled people for different profiles of people determined by education and household type. Figure 5 shows estimates of such probabilities, but for people with the status "employed". Both

Figure 4 Point and interval probability estimates of QJ for unemployed and disabled people in 2021 depending on education and household type (Slovakia – left, Czech Republic – right)

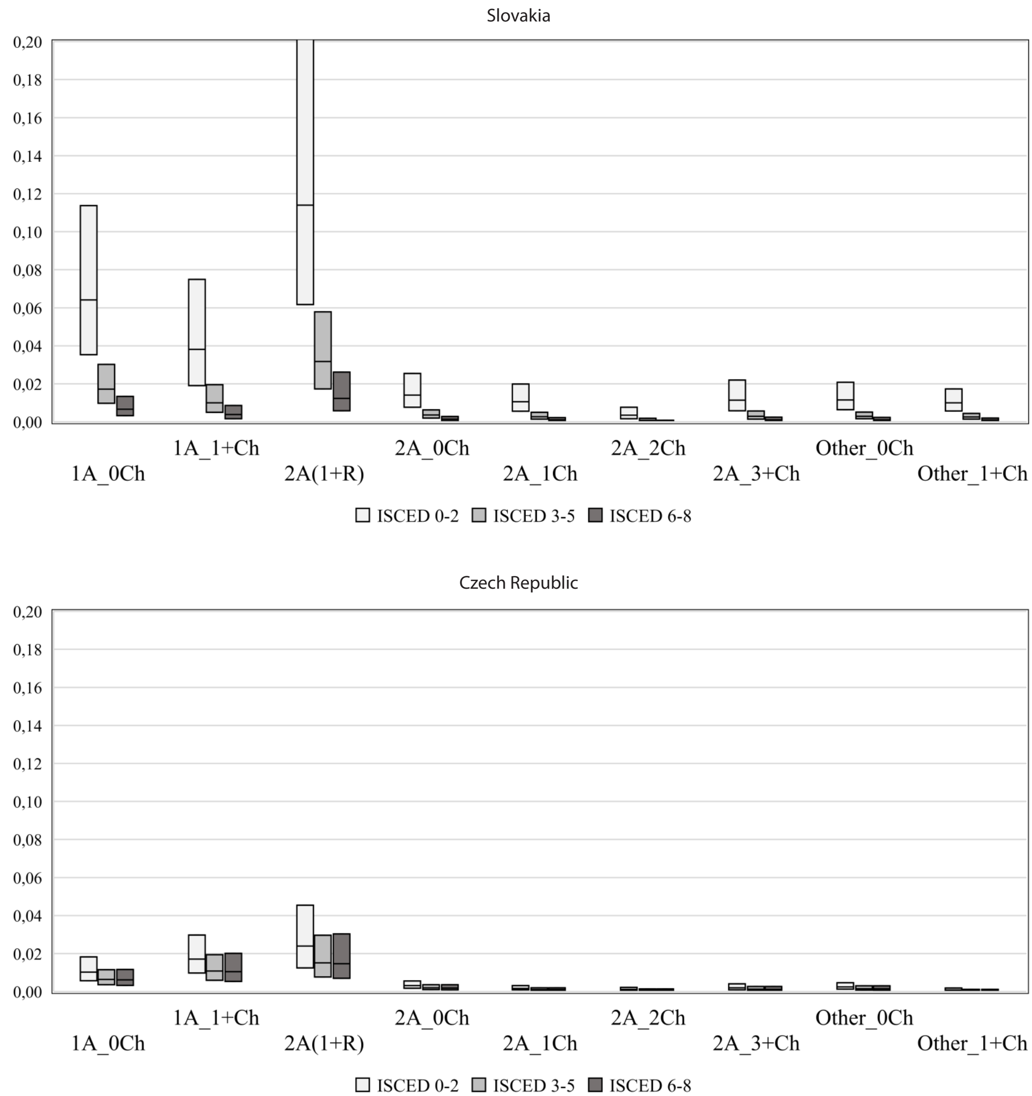


Source: EU-SILC 2021 SO SR and CR, own processing in SAS EG

figures (Figure 4 and Figure 5) provide a comparison of the relevant probabilities between Slovakia and the Czech Republic in 2021.

Among other things, the analyzes presented in the previous part of the article showed that the probability of living in a QJ household decreases with increasing education. Figure 4 confirms that this is the case in all types of households. In terms of a education we can observe more fundamental probability differences in Slovakia. This finding corresponds to the previous finding that in the Czech Republic, education has a smaller impact on the risk of living in a QJ household.

Figure 5 Point and interval probability estimates of QJ for employed persons in 2021 depending on education and type of household (Slovakia – left, Czech Republic – right)



Source: EU-SILC 2021 SO SR and CR, own processing in SAS EG

The highest probability of living in a household with very low work intensity among unemployed and disabled persons is observed in households with 2 adults where one is aged 65+. In 2021, this probability for persons with the lowest level of education (ISCED 0-2) was up to 94.3% (89.6%–96.9%), for persons with higher secondary or post-secondary education (ISCED 3-5) at the level of 80.8% (69.5%–88.5%) and for people with higher education at the level of 61.7% (43.0%–77.4%), if it is Slovakia, respectively 85.6% (74.8%–92.3%) for ISCED 0-2, 78.9% (64.0%–88.9%) for ISCED 3-5 and 78.4% (61.7 %–89.0%) for ISCED 6-8, if it is the Czech Republic.

Other results showed that people living in households of type 1A, 1A_1+Ch and 2A(1+R) had a significally higher probability of living in QJ households in 2021 than people from other types of households. This finding applies to both countries. There are no significant differences in the risk of QJ between people from other types of households. However, the exception is people living in a household of 2 adults with 2 dependent children, who have demonstrably the lowest risk of QJ in Slovakia. In the Czech Republic, people from the household type of 2 adults with 2 dependent children also had a relatively low risk of QJ, but not the lowest.

The probability of QJ for unemployed or disabled persons with low education (ISCED 0-2) was generally above 50% in Slovakia (the exception is the household type 2A_2Ch) and even above 75% in households 1A_1+Ch, 1A, 2A(1+R). In the Czech Republic, this probability for unemployed or disabled persons with the lowest level of education in households 1A_1+Ch, 1A, 2A(1+R) was over 60% in other types of households mostly in the range of 20%-50% (exceptions are household types 2A_2Ch and Other_1+Ch). Employed people (Figure 5) had a considerably lower probability. Even for persons with the lowest education in most types of households did not exceed 2%. The probability of living in a QJ household for employed persons exceeded 2% only in households types 1A_1+Ch and 2A(1+R) in the Czech Republic and in households types 1A_1+Ch, 1A and 2A(1+R) in Slovakia.

DISCUSSION AND CONCLUSION

The presented results are based on logit models and subsequent analysis of marginal means and contrast analysis of QJ in Slovakia and the Czech Republic. By (quasi-)joblessness (QJ) we understand the risk of living in a household that has a very low work intensity. In Slovakia and the Czech Republic, the probability of (quasi-)joblessness is primarily linked to economic activity and type of household, followed by education in Slovakia and age in the Czech Republic, while the authors Verbunt and Guio (2019) reached a similar conclusion for the countries of the European Union. Marital status and urbanization have a smaller, but still significant iimpact on (quasi-)joblessness.

Analyses of individual levels of factors revealed that unemployed and disabled people are most risky in both countries. In Slovakia, the odd of living in a (quasi-)joblessness household for people who are either unemployed or disabled is up to 127.661 times higher than for employed people, while in the Czech Republic this odds ratio is almost 2 times higher. In terms of household type, single-parent and multi-child households are the most risky, which was also confirmed by the findings of the authors Verbunt and Guio (2019) or the authors Filandri and Struffolino (2019), who dealt with another dimension of social exclusion, namely the risk of income poverty. In our analyses, we have revealed that the highest probability of living in a household with very low work intensity among unemployed and disabled people is observed in both countries in households with 2 adults, where one is aged 65+.

For unemployed and disabled people and individual types of households, the probability of very low work intensity decreases with an increase in educational level. However, the results of our analyses showed that education had a much greater impact on QJ in Slovakia in 2021 than in the Czech Republic. Even though we have confirmed in the case of both countries that the persons with the lowest level of education are most risky, in Slovakia the probability of living in QJ households for persons with the lowest educational level was approximately 14 pp higher than in the Czech Republic. The lowest level of education has proven to be a risk factor for poverty and social exclusion in other European Union countries as well, which is also confirmed by the authors Filandri and Struffolino (2019), Dudek and Szczesny (2021).

Our analyses show that the probability of living in a QJ household in the Czech Republic decreases with increasing age. In Slovakia, this pattern of dependence was disturbed by the age category 50+, for which we estimated a lower probability of QJ than in the age categories 30–40 and 40–50. In Slovakia, the most risky persons are aged 40-50 (9.7%), in the Czech Republic it is people over 50 (12.0%). In terms of marital status, never married persons were the most risky in both countries. In Slovakia, never married persons had a probability of living in QJ at the level of 9.4% and in the Czech Republic we estimated this probability at the level of 10.2%. Unexpected was the finding that in Slovakia the people who lived in marriage were the least risky, while in the Czech Republic it was divorced people. In terms of urbanisation, intermediate populated areas were the most risky in Slovakia (11.4%) and densely populated areas in the Czech Republic (7.9%). The stated findings are relatively unexpected, because in terms of other dimensions of poverty and social exclusion (income poverty or material deprivation), sparsely populated areas are the most risky, which was confirmed by e.g. Weziak-Bialowolska (2016).

The above conclusions were obtained by contrasting analysis and subsequent estimation of the probability depending on individual factors, while the impact of other factors was fixed. In addition, we estimated the probabilities of QJ in the interaction of the three most important factors, namely economic activity, type of household and education, while the other relevant factors (age, marital status and degree of urbanization) were fixed. For a better overview, we focused only on the two most frequent statuses of economic activity, namely the status "unemployed or disabled" and the status "employed", while this is the most risky and the least risky status of economic activity in terms of (quasi-)joblessness. We estimated the probabilities of QJ for 54 groups of people in Slovakia, depending on the 2 statuses of economic activity, 9 household types and 3 categories of education ($2 \times 9 \times 3 = 54$) and we also estimated the probabilities of QJ for 54 groups of people in the Czech Republic. Our analyses showed that unemployed and disabled people clearly have a higher risk of living in a QJ household in all education levels and in all types of households. People from households with only 1 adult of productive age had a higher risk of QJ than people from other types of households, in all education levels, with more visible differences among unemployed or disabled people than in the case of employed people. For people who are from these most risky types of households, who are simultaneously unemployed or disabled and have the lowest level of education (ISCED 0-2), we estimated the probability of QJ above 75% in Slovakia and above 60% in the Czech Republic.

In conclusion, it should be said that the paper provides a limited view of the risk of living in a (quasi-) joblessness household due to the fact that we analyzed the impact of only selected factors. In addition, we focused on only one of the dimensions of poverty and social exclusion, which opens up space for further research and the expansion of our analyses by other aspects of social exclusion, such as the risk of poverty or material deprivation. Despite some limitations, the paper reveals the most risky groups of people in terms of very low work intensity in the Slovakia and Czech Republic. Social policy should focus primarily on the mentioned groups of people.

ACKNOWLEDGMENTS

This work was supported by the VEGA projects *The Impact of Inflation on Poverty and Social Exclusion in Slovakia and the EU* (No. 1/0285/24) and *The impact of the COVID-19 crisis on business demography and employment in the Slovak Republic and the EU* (No. 1/0561/21).

References

ALLISON, P. D. (2012). *Logistic Regression using SAS. Theory and Application.* 2nd Ed. Cary, NC: SAS Institute Inc.

ATKINSON, A. B., GUIO, A. C., MARLIER, E. (2017). (eds.) *Monitoring social inclusion in Europe.* Luxembourg: Publications Office of the European Union.

BATZER, E. (2020). *Contrasts and Linear Hypothesis Tests* [online]. <https://www.evanbatzer.com/posts/contrasts>.

BRADSHAW, J., WILLIAMS, J., LEVITAS, R., PANTAZIS, C., PATSIOS, D., TOWNSEND, P., MIDDLETON, S. (2000). The relationship between poverty and social exclusion in Britain. In: *26th General Conference of The International Association for Research in Income and Wealth*, Cracow, Poland.

BROWN, U., LONG, G. (2018). Poverty and welfare [online]. In: *Social Welfare*, Routledge, 19–34. <https://www.taylorfrancis.com/chapters/edit/10.4324/9781315202686-3/poverty-welfare-usha-brown-gil-long>.

DE GRAAF-ZIJL, M., NOLAN, B. (2011). Household joblessness and its impact on poverty and deprivation in Europe [online]. *Journal of European Social Policy*, 21(5): 413–431. <https://doi.org/10.1177/0958928711418854>.

DUDEK, H., LISICKA, I. (2013). Determinants of poverty – binary logit model with interaction terms approach. Econometrics [online]. *Ekonometria. Advances in Applied Data Analytics*, 3(41): 65–77. <https://www.researchgate.net/publication/270285045>.

DUDEK, H., SZCZESNY, W. (2021). Multidimensional material deprivation in Poland: a focus on changes in 2015–2017 [online]. *Quality & Quantity*, 741–763. <https://doi.org/10.1007/s11135-020-01024-3>.

DUIELLA, M., TURRINI, A. (2014). Poverty developments in the EU after the crisis: a look at main drivers [online]. *ECFIN Economic Brief*, 1–10. <https://data.europa.eu/doi/10.2765/72447>.

EUROSTAT. (2021). *Glossary: Persons living in households with low work intensity* [online]. <https://ec.europa.eu/eurostat/statisticsexplained/index.php?title=Glossary:Persons_living_in_ households_with_low_work_intensity>.

EUROSTAT. (2022a). *Risk-of-poverty in very low work intensity households at 64%* [online]. <https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20220804-1>.

EUROSTAT. (2022b). *Living conditions in Europe – poverty and social exclusion* [online]. <https://ec.europa.eu/eurostat/statisticsexplained/index.php?title=Living_conditions_in_Europe-_poverty_and_social_exclusion&oldid=575050>.

FILANDRI, M., STRUFFOLINO, E. (2019). Individual and household in-work poverty in Europe: understanding the role of labor market characteristics [online]. *European Societies*, 21(1): 130–157. <https://doi.org/10.1080/14616696.2018.1536800>.

GALLARDO, M. (2020). Measuring vulnerability to multidimensional poverty [online]. *Social Indicators Research*, 148(1): 67–103. <https://doi.org/10.1007/s11205-019-02192-y>.

GAO, M., LI, L., GAO, Y. (2022). *Statistics and Analysis of Targeted Poverty Alleviation Information Integrated with Big Data Mining Algorithm* [online]. Security and Communication Networks. <https://doi.org/10.1155/2022/1496170>.

GUIO, A. C., MARLIER, E., VANDENBROUCKE, F., VERBUNT, P. (2022). *Differences in Child Deprivation Across Europe: The Role of In-Cash and In-Kind Transfers* [online]. Child Indicators Research, 1–26. <https://doi.org/10.1007/s12187-022-09948-y>.

JOHNSTON, H., MCGAURAN, A. M. (2018). *Low Work Intensity Households and the Quality of Supportive Services: Detailed Research Report.* Dublin: National Economic and Social Council.

LAFUENTE, J. Á., MARCO, A., MONFORT, M., ORDÓÑEZ, J. (2020). Social exclusion and convergence in the EU: An assessment of the Europe 2020 strategy [online]. *Sustainability*, 12(5): 1843. <https://doi.org/10.3390/su12051843>.

LITTELL, R. C., STROUP, W. W., FREUND, R. J. (2010). *SAS for Linear Models.* 4th Ed. Cary, NC: SAS Institute Inc.

NOLAN, B., MARX, I. (2009). *Economic inequality, poverty, and social exclusion* [online]. The Oxford handbook of economic inequality, 315–341. <https://doi.org/10.1093/oxfordhb/ 9780199606061.013.0013>.

OECD. (2009). *Is work the best antidote to poverty?* [online]. Chapter 3, Employment Outlook, Paris: Organisation for Economic Co-operation and Development. <https://www.oecd.org/els/emp/ 45219514.pdf>.

RAVALLION, M. (2020). On measuring global poverty [online]. *Annual Review of Economics*, 12: 167–188. <https://doi.org/10.1146/annurev-economics-081919-022924>.

RUSNAK, Z. (2012). Logistic regression model in poverty analyses [online]. *Econometrics*, 35: 9–23. <https://www.ceeol.com/search/article-detail?id=119600>.

SACCONE, D., DEAGLIO, M. (2020). Poverty, emergence, boom and affluence: a new classification of economies [online]. *Economia Politica*, 37(1): 267–306. <https://doi.org/10.1007/s40888-019-00166-4>.

SAS. (2018). *SAS/STAT® 15.1 User's Guide. The GLM Procedure.* Cary, NC: SAS Institute Inc.

SCHAD, D. J., VASISHTH, S., HOHENSTEIN, S., KLIEGL, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial [online]. *Journal of Memory and Language*, 110: 104038. <https://doi.org/10.1016/j.jml.2019.104038>.

STANLEY, J., HENSHER, D. A., STANLEY, J., CURRIE, G., GREENE, W. H., VELLA-BRODRICK, D. (2011). Social exclusion and the value of mobility [online]. *Journal of Transport Economics and Policy (JTEP)*, 45(2): 197–222. <http://www.jstor.org/stable/23072175>.

SUZUKI, M., TANIGUCHI, T., FURIHATA, R., YOSHITA, K., ARAI, Y., YOSHIIKE, N., UCHIYAMA, M. (2019). Seasonal changes in sleep duration and sleep problems: A prospective study in Japanese community residents [online]. *PLoS One*, 14(4): e0215345. <https://doi.org/10.1371/journal.pone.0215345>.

ŠOLTÉS, E., HURBÁNKOVÁ, Ľ., KOTLEBOVÁ, E., ŠOLTÉSOVÁ, T., VOJTKOVÁ, M. (2018). *Chudoba a sociálne vylúčenie v EÚ a v SR: v kontexte stratégie Európa 2020.* Pardubice: Univerzita Pardubice, Fakulta ekonomicko-správní. ISBN 978-80-7560-186-5

ŠOLTÉS, E., KOMARA, S., ŠOLTÉSOVÁ, T. (2022). Exploration of poverty and social exclusion of Slovak population via contrast analysis associated with logit models [online]. *Quality & Quantity*, 1–27. <https://doi.org/10.1007/s11135-022-01573-9>.

ŠOLTÉS, E., VOJTKOVÁ, M. (2018). Work intensity in Slovakia in relationship with socio-economic characteristics of households [online]. *Entrepreneurial Business and Economics Review*, 6(1): 115–134. <https://doi.org/10.1007/s11135-022-01573-9>.

VERBUNT, P., GUIO, A.-C. (2019). Explaining differences within and between countries in the risk of income poverty and severe material deprivation: Comparing single and multilevel analyses [online]. *Social Indicators Research*, 144(2): 827–868. <https://doi.org/10.1007/s11205-018-2021-1>.

VLAČUHA, R., KOVÁČOVÁ, Y. (2021). *EU SILC 2020 Indikátory chudoby a sociálneho vylúčenia.* Bratislava: ŠÚ SR.

WANG, B., WU, P., KWAN, B., TU, M. X., FENG, C. (2018). Simpson's paradox: examples [online]. *Shanghai Archives of Psychiatry*, 30(2): 139. <http://dx.doi.org/10.11919/j.issn.1002-0829.218026>.

WEZIAK-BIALOWOLSKA, D. (2016). Spatial variation in EU poverty with respect to health, education and living standards [online]. *Social Indicators Research*, 125(2): 451–479. <https://doi.org/10.1007/s11205-014-0848-7>.

Subjective Well-Being in Czech and Slovak Cities

Katarína Plačková[1] | *Technical university of Košice, Košice, Slovakia*
Oto Hudec[2] | *Technical university of Košice, Košice, Slovakia*

Received 2.3.2023 (revisions received 5.5.2023, 2.7.2023), **Accepted** (reviewed) 29.7.2023, **Published** 15.12.2023

Abstract

Cities are home to a significant proportion of the population in the EU, providing access to job opportunities and public services and, subsequently, driving economic growth. However, cities also face social and environmental challenges such as poverty, prohibitively high housing costs, discrimination, crime, excessive noise and air pollution. This raises the issue of how residents in European cities perceive their lives and assess their overall well-being and satisfaction with the amenities in their city. A U-shaped relationship between life satisfaction and age is tested in a sample of European cities using data from the Quality of Life in European Cities survey, with higher levels of satisfaction expected among younger and older individuals. The results supported the hypothesis and provided evidence for the importance of considering age in the analysis of well-being in urban settings. Subjective well-being is not only influenced by personal factors such as age and individual experiences but also by the quality of the urban environment. The second part employs ordinal logistic regression to analyse individual and contextual factors of well-being in four Czech and Slovak cities, namely Prague, Ostrava, Bratislava, and Košice.

Keywords	***DOI***	***JEL code***
Well-being, European cities, U-curve, urban environment, ordinal regression	*https://doi.org/10.54694/stat.2023.9*	*I31*

INTRODUCTION

Statistical approaches conventionally describe economic growth using standard indicators such as GDP with aggregate data suggesting indirect effects on the quality of life of the population. Therefore, it is important to fill the statistical gaps and to design indicators that better monitor social and environmental progress (Eurostat, 2015). However, quality of life and subjective well-being are influenced by various factors including social relationships, the quality of the environment, freedom and security, accumulated wealth, income and free time, as well as risks and expectations for the future. Although goods and services produced for the market each year are used as a criterion of regional development, personal and social well-being should be the objective of our efforts. Both objective and subjective well-being can be measured through household income and consumption surveys. However, both methods

[1] Department of Regional Sciences and Management, Faculty of Economics, Technical University of Košice, Němcovej 32, 040 01 Košice, Slovakia, Slovakia. E-mail: katarina.plackova@tuke.sk, phone: (+421)948329993.
[2] Department of Regional Sciences and Management, Faculty of Economics, Technical University of Košice, Němcovej 32, 040 01 Košice, Slovakia, Slovakia. E-mail: oto.hudec@tuke.sk, phone: (+421)556023303.

have limitations such as high implementation costs and the inability to provide continuous welfare information to policy makers (Voukelatou et al., 2021).

Cities are often appreciated for their agglomeration advantages, dynamic nature, wealth, job opportunities, cultural and gastronomic amenities and potential for innovation and good governance. Hence, the multifaceted term of quality of life in a city can be determined by various factors including economic, social, cultural and environmental conditions. Liveability and habitability are similar concepts which both refer to the suitability of a city for living. Cities in Western Europe tend to rank highly in terms of overall quality of life. However, it is not always clear whether the opportunities and amenities offered by cities lead to personal well-being and fulfilment. Urban living can also be associated with competition, strained relationships, loneliness, anonymity, crime and a loss of generosity (Gajdoš and Hudec, 2020).

The relationship between local amenities and subjective well-being (SWB) has recently received much attention. SWB has been found to be lower in cities with poorer services and social conditions. Local amenities play an important role in quality of life and may also affect the location decisions of households (Colombo et al., 2014).

Overall subjective life satisfaction is therefore influenced by external circumstances such as the society in which people live as well as individual determinants (Diener et al., 1999). Likewise, subjective well-being is shaped by socio-demographic factors such as age, income or education which lead to different life situations and different expectations or preferences (Eurostat, 2015). This leads us to two hypotheses:

The first hypothesis suggests that people experience their happiest moments in their youth as well as later on in older age, and that the increase in subjective well-being in older people is associated with greater satisfaction in terms of material needs and interpersonal relationships (Álvarez, 2022; Easterlin and Plagnol, 2008). The aim is to verify the existence of this U-shaped subjective well-being in different parts of Europe, taking into account geographical differences. It is hypothesized that people in the northern and western parts of Europe (Eurostat, 2022) tend to be more satisfied with their lives than people in other parts of Europe due to various factors such as the health of residents, economic and social systems, attitudes towards older people, environmental conditions, or other public service factors (Okulicz-Kozaryn and Valente, 2019).

Subsequently, the research goes beyond examining age as a substantial factor in shaping subjective well-being and delves into the impact of a city or place's character on well-being. The second hypothesis is that the environment and other characteristics of a city play a significant role in determining life satisfaction. It is expected that urbanisation leads to higher wages, increased productivity, innovation and creativity as well as access to various public services and amenities. However, it is also expected that negative externalities associated with urbanization, such as the cost of living, pollution, and traffic congestion, may have a negative impact on subjective well-being (Lenzi and Perucca, 2020).

The use of subjective indicators related to urban life is a well-established method for evaluating quality of life and liveability in cities (Mouratidis, 2020). This paper embraces this approach and sets out to provide a deeper understanding of the impact of city characteristics on subjective well-being. Therefore, the relationship between the urban environment, individual personality factors and subjective well-being is analysed using data from the Quality of Life in European cities survey for four cities in Slovakia and the Czech Republic – Prague, Ostrava, Bratislava and Košice. This provides a motivation for developing a comprehensive ordinal logistic econometric model to explore the influence of both individual and contextual factors on subjective well-being.

1 THEORETICAL OVERVIEW

Well-being can be explained as a state of contentment, health or happiness. However, it is necessary to understand that human well-being is a much broader concept than just feeling happy at that moment. Rather, well-being involves the inherent nature and feelings of a person which can also be influenced by the urban environment, public services, economic opportunities and availability of housing and sense of community (Daskalopoulou et al., 2022). Therefore, the subjective well-being of a city can be understood as a reflection of the overall satisfaction and contentment of its inhabitants with their lives. Subjective well-being refers to how people perceive and evaluate their lives and specific areas and activities within their lives (Diener and Suh, 1997). From an economic point of view, an individual's satisfaction with their life is influenced by factors such as earnings, pensions, employment and the quality of their housing. From a social perspective, satisfaction can be linked to education, healthcare, trust and the level of crime in an individual's area of residence. In terms of psychology, subjective well-being involves both cognitive and affective evaluations of an individual's life (Diener et al., 2002).

The cognitive evaluation pertains to the overall satisfaction with one's life in addition to satisfaction with specific areas such as employment and relationships while affective appraisal refers to a person's emotions, moods and feelings. Positive affect is characterized by pleasant emotions while negative affect involves unpleasant emotions such as anger and sadness (Smith and Konik, 2022). These components of subjective well-being can differ in their stability and variability over time. The term subjective well-being is often used interchangeably with terms such as happiness, quality of life, and satisfaction with one's own life and its management.

Individual factors such as innate temperament, personality and one's attitude towards life form, mental resilience and the ability to cope with life's challenges can contribute to greater life satisfaction. But also external circumstances, such as financial resources, social connections, and the environment and society a person lives in, can impact a person's overall well-being (Diener et al., 1999). It is worth noting that everyone has different needs and experiences when it comes to happiness, and what brings happiness to one person may not necessarily bring happiness to another (Layous and Lyubomirsky, 2014). Hence, subjective well-being is influenced by both contextual circumstances and individual factors.

Numerous studies have addresses macroeconomic and social factors suggesting a positive relationship between subjective well-being and key factors such as GDP, income and property rights. Moreover, it has been documented that unemployment and poverty have negative effects on subjective well-being (Marton and Mojsejová, 2022). Negative impacts were found for variables such as corruption and working hours. Thus, subjective happiness may be potentially enhanced through sustained efforts to combat corruption and alleviate excessive workloads. Also unemployment and poverty affect subjective well-being in a negative way.

There is a debate over the use of subjective indicators to measure well-being. Indeed, it can be difficult to compare data across countries due to differences in survey methods and question order. The current study uses data from the Quality of Life in European cities survey to look at the relationship between age and happiness (Biermann et al., 2022). It examines whether well-being follows a U-shaped curve over the course of a person's life, with the highest levels of happiness occurring at younger and older ages. According to the U-shaped curve theory, well-being decreases during middle age (Álvarez, 2022; Toshkov, 2022). This can be a result of factors such as poorer health, negative life events and increased stress related to work and economic success (Blanchflower and Oswald, 2008). The lowest point of the curve is typically reached between the ages of 40 and 50 although there can be variations in this pattern among different social groups and countries. Older individuals may experience increased happiness due to greater satisfaction in material needs and interpersonal relationships, the anticipation of retirement and the possibility that happier people tend to live longer (Biermann et al., 2022).

It has been traditionally believed that happiness remains constant or slightly increases with age. However, new studies have suggested otherwise, pointing towards a convex U-shaped curve of subjective well-being throughout the life cycle with a low point in middle age (Blanchflower and Oswald, 2008). This U-curve shape can be observed even in large samples in different countries with the minimum point being on average around 48.3 years old (Blanchflower, 2021). However, empirical research cannot be expected to lead to an unequivocal confirmation of the U-curve. For instance, similar pattern of life satisfaction across both Britain and Germany could be divided into three stages. The U-shaped curve was a good fit for the first and second stages which showed decreasing and increasing well-being, respectively. However, in the third stage which begins in the late 60s, well-being was found to decline (Wunder et al., 2013).

The built environment can have a significant impact on life satisfaction including satisfaction with health, leisure activities and personal relationships. One important characteristic that has been identified as impacting subjective well-being is population density (Želinský et al., 2021). Research has suggested that public policies aimed at reducing population density in urban areas may improve subjective well-being (Li and Winters, 2017).

A geographical comparison revealed some interesting differences, higher levels of social and psychological well-being in countries with strong social support networks and high per capita public expenditure on health and public services (Hansen and Slagsvold, 2016). In addition, Western European countries tend to have higher levels of subjective well-being compared to Mediterranean and East European countries (Conde-Sala et al., 2017). This difference may be attributed to factors such as higher levels of loneliness, limited access to social support and poorer health (Hansen et al., 2016). Eastern Europeans tend to be less satisfied with life due to lower income, unemployment, perceived corruption, and weaker government performance (Djankov et al., 2016; Sarracino, 2010). Countries with wealthier societies tend to have better education and healthcare systems which can improve the quality of life for individuals (Rodríguez-Pose and Maslauskaite, 2012). Thus, it seems that examining subjective well-being from a geographical perspective is a logical approach. Therefore, it is of interest to explore the U-shaped curve relationship between age and well-being and in what way it is influenced by both economic and social factors. In particular, it may be useful to compare countries and their cities in order to understand the occurrence and causes of this pattern. This can help to determine whether the U-shaped curve is a valid representation of the relationship between age and well-being in a particular country or region, what is consistent with the first hypothesis.

While cities and places can have a significant impact on subjective well-being, the extent of this contextual influence is likely to vary depending on factors such as access to quality public services, cultural and social infrastructure, green spaces, and economic opportunities can contribute to overall subjective well-being, while contextual factors such as unemployment, pollution and high living costs may have a negative impact. Cities can offer access to better job opportunities, higher income, and improved consumption options, but they may also negatively impact well-being through increased living costs, pollution, and work pressure. The extent to which external environmental factors affect subjective well-being is an interesting question for further exploration (Loschiavo, 2021).

Well-designed urban public spaces and quality of public services appear to be important determinants in affecting people's well-being. Access to green spaces such as parks and gardens, as well as air pollution control contribute to overall wellbeing (Poortinga et al., 2021). Green urban spaces contribute to improving human health and well-being while mitigating the effects of climate change at the same time. There is evidence that green areas moderate air temperatures and consequently reduce energy consumption and urban forests act as carbon sinks and reduce air pollution and noise. The beneficial effects of urban green spaces in relation to human well-being have become a fundamental element for urban planners and policy makers (Giannico et al., 2021). Moreover, improving urban infrastructure may be a significant factor in increasing well-being and liveability (Chen, 2023). The relationship between satisfaction with municipal services (for example public education, transportation, safety or areas for sports and recreation) and subjective well-being is statistically significant, suggesting that quality public services affect well-being (Voukelatou et al., 2014).

Cities provide access to leading businesses, cultural institutions, and civic amenities, making them centres of production and activity in their regions. They also offer benefits in areas such as social care, health and business and the possibility of easy travel to other destinations (Pineo et al., 2018). From the urban development point of view, there are various concepts of how to make residents in cities happier. However, it is important to remember that a city is more than just its physical infrastructure (Glaeser, 2011) and that the primary feature is the people it is made up of. This means that any plans to increase "urban happiness" should prioritize the well-being and needs of people. One of the major challenges today is how to accommodate the increasing number of people moving into cities in a way that ensures their long-term health and well-being (Pineo et al., 2018).

The concentration of opportunities and services in urban areas can be very beneficial for the health and well-being of residents. However, the crowded living and working conditions of cities can also lead to the spread of pollution and diseases. Many rapidly growing cities struggle with various problems and health issues in particular. Moreover, cities in developed countries face new challenges such as health problems that can be partially attributed to the access to green spaces and poor air quality as well as a sedentary lifestyle leading to weight gain and obesity (Pineo et al., 2018).

The Nordic countries are examples of high happiness, and other countries such as Switzerland, the Netherlands, New Zealand, Canada and Australia show similar characteristics, which may point to a universal formula for a happy society that includes high-quality, non-corrupt government institutions that deliver on promises and provide sufficient support to citizens (Martela, 2020). Essential steps towards a happy society include building a well-functioning and trustworthy government, and maximising citizen participation in decision-making processes, fostering a sense of belonging, trust and social cohesion. Despite having a lower GDP than its neighbouring Nordic countries, Finland is considered the most stable, safest and best-governed country as well as being the least corrupt and very socially progressive. If people are the core of happy countries, they must also be at the centre of thinking about cities.

Therefore, the liveability of a city is determined by the quality of life it can provide for its residents. Cities around the world are increasingly shifting their political focus to promote well-being. Some of the biggest challenges for urban well-being include reshaping the built environment to provide accessible and safe space for human relationships and mutual interaction, movement and an active lifestyle. This focus on creating a clean, low-carbon and resilient urban environment can bring significant public health, social and economic benefits to citizens. Nowadays, more local governments have recognized the importance of well-being and making it a political priority. Life satisfaction provides a clearer direction to policy makers on where to focus new and effective policies in order to increase social well-being. The existing empirical research substantiates the second hypothesis, indicating a relationship between individual and contextual factors and wellbeing. An ordinal logistic regression model will be designed based on a questionnaire survey conducted in four cities across the Czech Republic and Slovakia.

2 DATA SOURCES AND METHODS

In 2004, the European Commission started a survey to look at citizens' perceptions of the quality of life in their home cities. This survey covers capital cities in addition to other large cities in the EU, EFTA, Great Britain, the Western Balkans and Turkey (European Union, 2013; Dijkstra et al., 2020). The survey asked citizens to express their opinions on various aspects of city life such as the quality of public healthcare, education, cultural or sports facilities and city amenities. People were also asked to rate their satisfaction with their housing, the life they lead and the financial situation in the household (European Commission, 2020).

These studies use datasets from the Quality of Life in European Cities Report of 2012, 2015, and 2019 to examine how citizens perceive life satisfaction in their respective cities (European Commission, 2015; European Commission, 2016; European Commission, 2020). The Flash Eurobarometer, launched by the European Commission, conducts surveys in all EU member states periodically, with a typical sample size of 500–1 000 respondents per city, contracted to TNS Opinion or Gallup. The interviews were conducted in the national languages via phone. Data collection for this study was conducted between 15th November to 7th December 2012 and between 21st May to 27th June 2015 (European Commission, 2015; European Commission, 2016). These two datasets were used in the research of the U-curve in this paper. Additionally, data from the 5th survey of European cities, carried out in 2019, was used to create a model for Czech and Slovak cities, with a target sample size of 700 complete interviews in each surveyed city, targeting individuals aged 15 and above (European Commission, (b)). The respondents were asked to rate their satisfaction with life on a scale of four ratings (very satisfied, fairly satisfied, not very satisfied, not at all satisfied).

2.1 U-curve of well-being by age

The U-curve thesis suggests that people's level of happiness and satisfaction with their lives follows a U-shaped pattern over time (Blanchflower and Oswald, 2008). The present study aims to find out whether the U-curve of well-being by age is valid by using available data from the Quality of Life in European cities survey in 2012 and 2015. Individuals are likely to experience an increased sense of satisfaction with their lives when they are young with their satisfaction decreasing as they get older. This reaches a tipping point before starting to rise again as they approach older age (Blanchflower and Oswald, 2008). The 2012 and 2015 Quality of Life Surveys provide valuable insights into the finer details of this complex phenomenon. The data enables a validation of satisfaction levels across the following age demographics:

1) Adolescents and young adults aged 15 to 24,
2) Adults aged 25 to 39,
3) Middle-aged adults aged 40 to 54,
4) Senior citizens aged 55 and over.

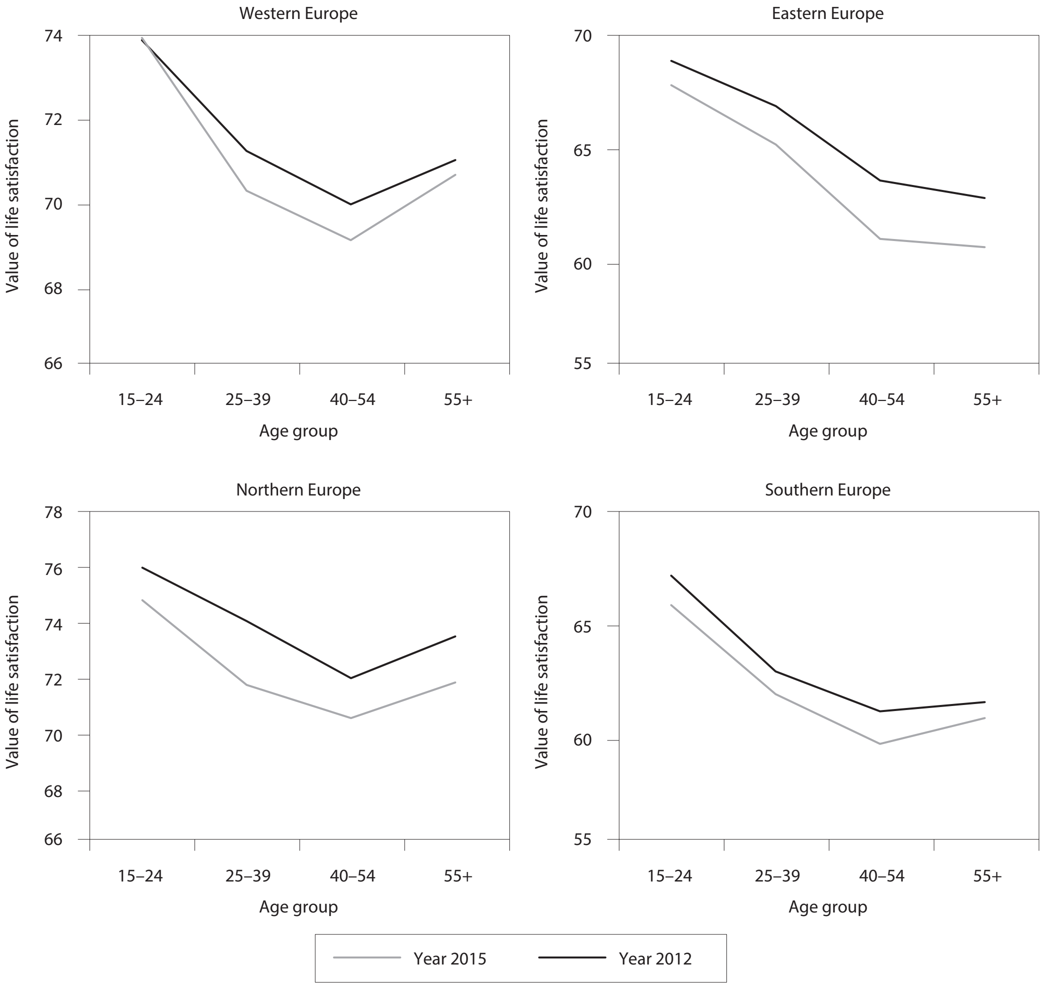
The objective of the study was to examine the nature of the curve of life satisfaction with regards to age as well as contrasting the findings from the years 2012 and 2015. A comprehensive analysis of data collected from various cities across Europe was used to identify any potential patterns in relation to the geographical location of the respondents and thereby providing a deeper understanding of the factors that may influence subjective well-being. Indeed, the "U-curve" hypothesis may not be consistent across different regions (Eurostat, 2022) as it may be influenced by geographical factors such as living standards, governance quality or culture. An examination of the life satisfaction values across different age groups within each geographical region will look at whether the pattern of the "U-curve" is consistent across different parts of Europe, and how it may be influenced by geographical factors. The study divides Europe into four distinct geographical regions – Northern, Southern, Eastern and Western (based on Nations Online)[3] in order to identify how living standards and governance quality may impact the subjective well-being of residents across different age groups.

The research process entailed the classification of the surveyed cities into four regions of Europe (Nations Online) – Southern Europe (24 cities), Northern Europe (19 cities), Eastern Europe (15 cities) and Western Europe (25 cities). Subsequently, the mean life satisfaction was calculated according to the respondents' affiliation to a region, for each region and age group based on data from both 2012 and 2015. The variations in the U-curve of life satisfaction became apparent when the cities were grouped by region and the mean values were calculated for each age group (Figure 1).

It can be concluded that the U-curve variation in subjective well-being is evident in Western Europe with a clear pattern of satisfaction declining in middle age, and then increasing again in older age.

[3] Nations Online – the regions of Europe are geographically divided. Eastern Europe is conventionally the geographical region east of Germany. Southern Europe is bounded by the Mediterranean Sea in the south and mainly refers to the subtropical southern region. Western Europe is bounded by the Atlantic Ocean to the west, the English Channel, the North Sea to the north and by the Alps in the south.

Figure 1 U-curve of well-being by age group in parts of Europe



Source: Own calculation

Similar patterns can be observed in Northern and Southern Europe. However, the data does not show a clear U-shaped curve in Eastern Europe. These findings align with the Continental Survey of Health, Ageing and Retirement in Europe (Blanchflower, 2021). Older adults in Northern Europe have better financial stability and health compared to those in Southern and Eastern Europe. Moreover, older adults in Eastern Europe suffer a higher risk of loneliness and limited access to social support both of which may contribute to their lower levels of subjective well-being (Leichsenring et al., 2021). However, the U-curve of subjective well-being has been criticised for only existing under certain socio-economic conditions (Deaton, 2008). The existence of a U-curve has been partially confirmed in the current research based on an extensive survey of 41 000 respondents in 78 cities. However, the results show that the U-curve is not a universal trend and is influenced by various factors such as economic and social conditions. Older people in Northern Europe often possess a superior level of financial security and overall well-being in comparison to their counterparts in Southern and Eastern Europe (Lee, 2021; Watson, 2005).

In Eastern Europe, the older demographic tends to report a less favourable state of subjective well-being, possibly owing to factors such as isolation, restricted access to social support and sub-optimal health status. This disparity is likely a result of the less developed social security systems in Eastern European countries in comparison to their Western European counterparts. Older Eastern European women are at significant risk of experiencing loneliness, due to the lower life expectancy of men (Hansen and Slagsvold, 2016). The negative portrayal of older people in the media also contributes to lower self-esteem and self-worth of older people (Robinson et al., 2008).

While research has shown that age is one of the demographic factors in determining an individual's level of subjective well-being, personal factors such as personality traits, coping styles and cognitive abilities can also play a role. This is in addition to environmental and contextual factors such as social support, income, access to healthcare and place-based factors such as the level of urbanisation and cultural values. Therefore, it is important to consider both individual and contextual factors with regard to subjective well-being.

2.2 Individual and contextual factors

While age is one of the significant factors with regards to subjective well-being (SWB), other personality characteristics can influence it including gender, personality traits, values, coping strategies, cognitive and emotional states. A person's personality and coping strategies may influence their ability to adapt to and cope with stress and challenges, subsequently affecting their SWB. At the same time, contextual factors including economic, social, environmental and cultural factors, as well as the physical environment and access to resources and opportunities can influence SWB. Therefore, SWB in European cities is expected to be affected by a variety of individual and contextual factors (Diener, 2023).

Income levels have a generally positive impact on SWB, with higher income individuals reporting higher levels of happiness and life satisfaction (Ferrer-i-Carbonell, 2005). Likewise, education levels are positively associated with SWB in urban areas, as individuals with higher levels of education tend to have better job prospects and greater access to resources and opportunities (Ruiu and Ruiu, 2019). On the other hand, unemployment rates are negatively correlated with SWB as people without a job may experience feelings of financial insecurity and social isolation. Environmental factors are also important with access to green spaces in urban areas being positively related to SWB (Giannico et al., 2021). Indeed, access to nature and outdoor activities can have a positive impact on mental and physical health. Contact with people as well as social support networks such as family and friends all provide emotional support and a sense of belonging. Neighbourhood safety might also be an important factor as well as political stability, both contributing to the feeling of security and confidence (Mouratidis, 2019). Conversely, poor air quality, noise, traffic congestion and high density are all negatively associated with SWB.

However, the question still remains as to how the attributes of a place affect overall SWB in a person. Indeed, it is of paramount importance that cities strive to promote a sense of liveability which encompasses the relationship between the individual and their surroundings. A well-built environment and quality public services should aim to meet the needs and expectations of residents. Despite these efforts, cities often face a plethora of issues including air and noise pollution, traffic congestion, poverty and overpopulation, all of which may have an impact on how residents perceive and evaluate their satisfaction with life in a city (Mouratidis, 2018).

Thus, the current research aims to understand the relationship between the urban environment, personality factors and subjective well-being. Given that there are significant cultural differences in the EU, this study is focusing on a particularly few factors that explain the effects of individual and contextual factors in Czech and Slovak cities. The cities explored have gone through similar processes of modernisation and urbanisation, shaping their urban environments in similar ways. Both countries are located in the same region of Europe and have been exposed to similar environmental and economic conditions.

This may also have influenced the way in which the cities have developed. As indicated, it makes sense to examine the effects of factors in a limited area which share common characteristics. In summary, the similarities between the cities in the Czech and Slovak Republics make them a valuable sample to study the relationship between urban environment, individual personality factors and subjective well-being, as the homogeneity of these cities allows for a clearer understanding of the specific factors that impact SWB.

The research utilises responses from residents living in four cities within the Czech Republic and Slovakia: Bratislava, Košice, Prague and Ostrava. The respondents evaluated their satisfaction with overall quality of life (SWB) as well as specific attributes of the cities they reside in such as public services, transportation and amenities. In addition, they evaluated their financial and job situation on a four-point scale (very satisfied, somewhat satisfied, rather unsatisfied, not at all satisfied). The items were reversed as needed to ensure that a higher score corresponded to a greater degree of satisfaction, agreement, or frequency before conducting the statistical analyses. We employed the original coding scheme of the Perception Survey on the Quality of life in European Cities 2019 – Codebook, as also in the recent studies using the same data by Lenzi and Perucca (2023) and Giannico et al. (2021) to ensure consistency across analyses. For the variable gender, number 1 was assigned to male and 2 to female.

In terms of the variable age, the following scale was used:

1. Age group from 15 to 24,
2. Age group from 25 to 44,
3. Age group from 45 to 64,
4. Age group 65 and above.

The perception of contextual factors is formulated in the same way, only the factor changes, here is an example of a question on cultural facilities: *"Generally speaking, please tell me if you are very satisfied, rather satisfied, rather unsatisfied or not at all satisfied with each of the following issues in your city or area - cultural facilities such as concert halls, theatres, museums and libraries."*

Variable values are 5 very satisfied/4 rather satisfied/3 I don't know/2 rather unsatisfied/1 not at all satisfied. Values increase with higher satisfaction.

The only exception is the phrasing of the question on corruption in public administration *"Please tell me whether you strongly agree, somewhat agree, ... – There is corruption in my local public administration."*. But again, values increase from negative to positive statements: 5 strongly agree/4 somewhat agree/3 don't know/2 somewhat disagree/1 strongly disagree. To maintain a consistent scale for city size, Ostrava and Košice are both assigned a value of 1, indicating that they are of similar size. The two country capitals, Bratislava, and Prague, are assigned values of 3 and 5, respectively, reflecting their larger sizes relative to Ostrava and Košice.

Therefore, the ordinal logistic model (alternative names ordered logistic regression or proportional odds logistic regression) is constructed by the following variables describing individual (1–5) and contextual (6–11) factors of SWB:

Table 1 Variables and codes Included in the model

	Variable	Codes
1.	Age	1 from 15 to 24 2 from 25 to 44 3 from 45 to 64 4 65 and above
2.	Gender	1 male 2 female

Table 1		(continuation)
	Variable	**Codes**
3.	Education	1 primary 2 secondary 3 bachelor or master 4 doctoral
4.	Your personal job situation	1 not at all satisfied 2 rather unsatisfied 3 I don't know 4 rather satisfied 5 very satisfied
5.	The financial situation of your household	
6.	Cultural facilities such as concert halls, theatres, museums, and libraries	
7.	Public spaces such as markets, squares, pedestrian areas	
8.	Living in a city: I'm satisfied to live in my city	1 strongly disagree 2 somewhat disagree 3 I don't know 4 somewhat agree 5 strongly agree
9.	Social capital: most people in my city can be trusted	
10.	Corruption: There is corruption in my local public administration	
11.	Population	1 Ostrava and Košice 3 Bratislava 5 Prague

Source: Perception Survey on the Quality of life in European Cities 2019: Codebook and author´s processing

The dependent ordinal variable *SWB refers to the level of satisfaction with the life* of respondent. The model takes into account contextual factors such as spaces and facilities, as well as individual factors such as age, gender and personal financial and household circumstances. The goal is to reveal the influence of these various factors on overall subjective well-being:

$$logit(P(SWB \leq y \mid X)) = \beta_0 + \beta_1 \cdot age + \beta_2 \cdot gender + \beta_3 \cdot education + \beta_4 \cdot job\_situation + \beta_5 \cdot financial\_situation + \beta_6 \cdot cultural\_facilities + \beta_7 \cdot public\_spaces + \beta_8 \cdot living\_in\_city + \beta_9 \cdot social\_capital + \beta_{10} \cdot corruption + + \beta_{11} \cdot population.$$

Logit(P(SWB ≤ y | X)) is the logarithm of the cumulative odds that the dependent variable SWB is less than or equal to y, given the values of the independent variables X. β_0 represents the intercepts for each level of the dependent variable, indicating the baseline odds for each category of SWB. The coefficients β_1, β_2, ..., β_{11} correspond to the independent variables X_1, X_2, ..., X_{11}, representing the effects of the predictors on the odds of SWB. In this formula, SWB represents the ordinal dependent variable, which can take on ordered categories 1, 2, 3, 4, 5. The independent variables X_1, X_2, ..., $X11$ are ordinal, only gender is binary. The goal of the ordinal logistic regression is to estimate the values of the coefficients β_1, β_2, ..., β_{11} to analyse the relationship between the independent variables and the cumulative odds of the dependent variable SWB. The cumulative odds represent the odds of the dependent variable being equal to or less than a particular level compared to all the categories above it. In other words, the cumulative chance for level 3 represents the combined chance that the dependent variable falls in level 3, 2 or 1, versus the chance that it is in one of the higher levels (4, 5). The model estimates coefficients for each independent variable that indicate the direction and strength of their relationship with the outcome variable SWB.

This formula shows how the logarithm of the chances of being placed in a higher SWB category depends on the values of the independent variables. The coefficients β_i indicate how much the log odds change for a unit increase in the independent variable *i* if all other variables remain constant. The numbers 2, 3, 4, 5 in the names (e.g., age2) are the levels of the factor. The p-values for these factors indicate the significance of the difference in terms SWB between each age group (25–44) and the reference group, which is the first age group (from 15 to 24). For example, the coefficient for factor (age)3 is –0.525, which means that the log odds in age group 3 is –0.525 compared to age group 1. To obtain the odds ratio, the coefficient needs to be exponentiated: exp(–0.525) = 0.592. As a result, in age group 3, compared to age group 1, the odds of a higher SWB category decrease by 40.8% (1 – 0.592).

The results including the regression coefficients and p-values are presented in Table 2.

Table 2 Regression values, intercepts and p-values of the variables from the model

Factor	Value	p-value	
factor(age)2	–0.22020	0.17110	
factor(age)3	–0.52543	0.00246	**
factor(age)4	–0.44478	0.00979	**
factor(gender)2	–0.05120	0.49289	
factor(job_situation)2	–0.23388	0.15315	
factor(job_situation)3	0.32883	0.02303	*
factor(job_situation)4	0.27499	0.02054	*
factor(job_situation)5	0.44143	0.00049	***
factor(financial_situation)2	–0.00679	0.96794	
factor(financial_situation)3	0.24730	0.43322	
factor(financial_situation)4	0.56759	0.00038	***
factor(financial_situation)5	0.73653	0.00004	***
factor(education)2	0.31177	0.03825	*
factor(education)3	0.40997	0.01161	*
factor(education)4	0.49579	0.00474	**
factor(education)5	0.36978	0.13432	
factor(education)9	0.15722	0.88207	
factor(live_incity)2	0.53133	0.08448	+
factor(live_incity)4	0.68065	0.01539	*
factor(live_incity)5	0.99748	0.00042	***
factor(social_capital)2	0.01501	0.90277	
factor(social_capital)4	0.11632	0.34452	
factor(social_capital)5	0.32838	0.02094	*
factor(corruption)2	0.18925	0.07701	+
factor(corruption)4	0.38634	0.00144	**
factor(corruption)5	0.28554	0.00446	**
factor(cultural_facilities)2	0.13266	0.58913	
factor(cultural_facilities)3	0.20434	0.43552	
factor(cultural_facilities)4	0.15304	0.47915	
factor(cultural_facilities)5	0.40596	0.06374	+
factor(public_spaces)2	0.63991	0.00252	**
factor(public_spaces)3	0.70827	0.01300	*

Table 2 (continuation)

Factor	Value	p-value	
factor(public_spaces)4	0.73109	0.00026	***
factor(public_spaces)5	0.94372	0.00001	***
factor(population)3	–0.18430	0.04872	*
factor(population)5	–0.18288	0.04720	*
1|2	–1.27125	0.00487	**
2|3	0.51132	0.24963	
3|4	0.60104	0.17600	
4|5	3.26443	0.00000	***

Note: With statistical significance: *** 0.001 ** 0.01 * 0.05 + 0.1.
Source: Own calculations using data from the Quality of Life in European Cities survey (2019)

The findings highlight the intricate relationship between various contextual and individual factors and overall well-being of residents in cities. The coefficients for the factor age are all negative, which means that higher age groups indicate lower SWB, for factor(age)3 and factor(age)4 are less than 0.05, which means that these effects are statistically significant. This confirms the results from the first part in the Czech and Slovak cities. It suggests that at middle age, SWB (subjective well-being) decreases, and it does not return to higher levels at the end of economic activity. The effect of gender on SWB is negative but not significant. However, a better job situation has a strong positive effect on SWB compared to the reference group of unemployed individuals. Similarly, a better financial situation in the household is significantly associated with higher SWB, compared to the reference group of very bad financial situations. Another significant individual factor is education. The coefficients suggest positive effects on SWB for individuals with secondary, bachelor, and master degrees, compared to those with primary education.

Subsequently, the coefficients for the contextual factors indicate how residents' evaluations of various aspects of the city affect their SWB, compared to the reference group (not at all satisfied or strongly disagree). The quality of space is important for SWB, the coefficients for levels 2, 4, and 5 are positive, which means that being more satisfied with living in the city is associated with higher log odds of being in a higher category of SWB. A very similar strong effect is in the case of satisfaction with the public spaces such as markets, squares, pedestrian areas. Perceived corruption in a city is significantly linked to lower levels of well-being among residents in the Czech and Slovak cities. While culture-led development has been associated with presumed positive impacts on SWB, the impact of cultural factors may not be as powerful as anticipated. The coefficient for factor social_capital5 indicates the effect of trusting most people in the city on SWB, compared to not trusting most people. The coefficient is positive, which means that trusting most people in the city is associated with higher log odds of being in a higher category of SWB. The p-value for factor social_capital5 is statistically significant, trusting most people (compared to not trusting most people) increases the log odds of being in a higher category of SWB by 0.328, and increases the odds of being in a higher category of SWB by 38.8%.

The coefficients for factor(population) estimate the effects of different city sizes on SWB, compared to the reference group (Ostrava and Košice). The coefficients are negative for levels 3 and 5, which means that living in a larger city (Bratislava or Prague) is associated with lower log odds of being in a higher category of SWB. This suggests that SWB decreases with the size of the city.

CONCLUSION

The paper makes a contribution to the understanding of the relationship between city context and well-being. The study highlights that a city's physical and social environment can impact people's emotional and psychological states, their sense of belonging, purpose, happiness and overall satisfaction with life. The results of the ordered logistic regression model support the hypothesis that city and place have a strong impact on well-being of residents. The contextual factors, such as satisfaction with living in the city, public spaces, trust, and cultural facilities and lower perceived corruption are statistically significant and contribute to an evaluation in a higher category of SWB. In particular, the analysis shows that larger cities tend to have decreasing levels of SWB. This finding corresponds with theoretical frameworks that suggest that complex stressors associated with urban living, such as increased competition, social fragmentation, and limited access to nature, can negatively affect individuals' overall well-being (Diener et al., 2018).

Residents' individual circumstances also contribute significantly to an understanding of their well-being. There was no evidence of an effect of gender, but there is a very strong effect of work and financial situation on SWB. Compared to the reference groups (unemployed individuals and individuals with a very poor financial situation), individuals with a better work situation and financial conditions were significantly more likely to have a higher SWB. Similarly, greater life satisfaction was associated with higher education.

The article contributes to the literature on the relationship between urban context and well-being and provides empirical evidence on the importance of urban planning and policy making for enhancing the well-being of urban residents.

The results inform policy-makers and city planners on the key components that can shape the experience of residents in a city such as urban design, public spaces, and community resources in their efforts to improve the well-being of residents. Urban design, public spaces and the availability of community resources are key components that can shape the experience of residents in a city (Olsen et al., 2019). Public spaces such as parks and plazas can offer residents a place to relax and connect with others, while green spaces can provide an escape from the built environment and help reduce stress levels (Fleming et al., 2016; Olsen et al., 2019).

Yet, not all residents are significantly influenced by the context of the city with more individual factors playing a role in shaping well-being. The negative relationship between age and wellbeing can be attributed to the predictors of loss of health, a partner, and friends (Blekesaune and Hansen, 2022), especially in the Central and Eastern European countries.

The results suggest that the relationship between well-being and the city context is complex and multi-faceted and should take into consideration both the city context and individual factors when developing initiatives aimed at improving well-being in cities. The study also sheds light on the relationship between age and well-being, which is consistent with the U-theory where well-being generally decreases in early adulthood, reaches a low point in middle age, and begins to increase again in old age (Blanchflower and Oswald, 2008). In Western, Northern and Southern Europe, the results of this study have confirmed the existence of this U-shaped curve although Eastern Europe has shown a different pattern. This may be due to differences in social support systems, health conditions and the perception of well-being of the older population. In Eastern Europe, factors such as loneliness and limited access to social support may contribute to a lower perception of well-being among the elderly (Leichsenring et al., 2021). For Eastern Europe, where retirees may not have as high levels of well-being, the results stress on the importance of addressing factors such as loneliness and limited access to social support in order to improve well-being of elderly.

This paper highlights the importance of taking both cultural and regional differences into account when understanding the relationship between age and well-being. Cities may not always have enough information about the different segments of their population, such as age, households, and gender, and how they relate to well-being. The paper's findings help shed light on these issues and highlight the importance of customised approach to understanding and improving well-being in cities. In particular, the paper's exploration of the U-curve of well-being and its relationship to age, as well as the importance of access to public services, such as public spaces and cultural amenities, can inform urban and regional policy aimed at improving well-being outcomes for residents.

ACKNOWLEDGMENTS

The work was supported by the Slovak Scientific Grant Agency (VEGA) Grant 1/0034/23.

References

ÁLVAREZ, B. (2022). The Best Years of Older Europeans' Lives [online]. *Social Indicators Research*, 160(1): 227–260. <https://doi.org/10.1007/s11205-021-02804-6>.

BIERMANN, P., BITZER, J., GÔREN, E. (2022). The relationship between age and subjective well-being: Estimating within and between effects simultaneously [online]. *The Journal of the Economics of Ageing*, 21: 100366. <https://doi.org/10.1016/j.jeoa.2021.100366>.

BLANCHFLOWER, T., OSWALD, A. J. (2008). Is well-being U-shaped over the life cycle? [online]. *Social science & medicine*, 66(8): 1733–1749. <https://doi.org/10.1016/j.socscimed.2008.01.030>.

BLANCHFLOWER, T. (2021). Is happiness U-shaped verywhere? Age and subjective well-being in 145 countries? [online]. *Journal of Population Economics*, 34(2): 575–624. <https://doi.org/10.1007/s00148-020-00797-z>.

BLEKESAUNE, M., HANSEN, T. (2022). The age and well-being "paradox": a longitudinal and multidimensional reconsideration [online]. *European Journal of Ageing*, 19(4): 1277–1286. <https://doi.org/10.1007/s10433-022-00709-y>.

COLOMBO, E., MICHELANGELI, A., STANCA, L. (2014). La Dolce Vita: Hedonic estimates of quality of life in Italian cities [online]. *Regional Studies*, 48(8): 1404–1418. <https://doi.org/10.1080/00343404.2012.712206>.

CONDE-SALA, J. L., PORTELLANO-ORTIZ, C., CALVÓ-PERXAS, L., GARRE-OLMO, J. (2017). Quality of life in people aged 65+ in Europe: associated factors and models of social welfare – analysis of data from the SHARE project (Wave 5) [online]. *Quality of life research*, 26(4): 1059–1070. <https://doi.org/10.1007/s11136-016-1436-x>.

DASKALOPOULOU, I., KARAKITSIOU, A., MALLIOU, C. (2022). A Multicriteria Analysis of Life Satisfaction: Assessing Trust and Distance Effects [online]. *Operations Research Forum*, 3: 59. <https://doi.org/10.1007/s43069-022-00170-8>.

DEATON, A. (2008). Income, health, & well-being around the world: evidence from the Gallup world poll [online]. *Journal of Economic perspectives*, 22(2): 53–72. <https://doi.org/10.1257/jep.22.2.53>.

DIENER, E. D., SUH, M. E. (1997). Subjective well-being and age: An international analysis. *Annual review of gerontology and geriatrics*, 17(1): 304–324.

DIENER, E., SUH, E. M., LUCAS, R. E., SMITH, H. L. (1999). Subjective well-being: Three decades of progress [online]. *Psychological bulletin*, 125(2): 276–302. <https://doi.org/10.1037/0033-2909.125.2.276>.

DIENER, E., LUCAS, R. E., OISHI, S. (2002). Subjective well-being: The science of happiness and life satisfaction. *Handbook of positive psychology*, 2: 63–73.

DIENER, E., OISHI, S., LUCAS, R. E. (2018). National accounts of subjective well-being [online]. *American Psychologist*, 73(5): 218–226. <https://doi.org/ 10.1037/a0038899>.

DIENER, E. (2023). Happiness: The science of subjective well-being. *Noba textbook series: Psychology*, Champaign, IL: DEF publishers.

DJANKOV, S., NIKOLOVA, E., ZILINKSY, J. (2016). The happiness gap in Eastern Europe [online]. *Journal of comparative economics*, 44(1): 108–124. <https://doi.org/10.1016/j.jce.2015.10.006>.

EASTERLIN, R. A., PLAGNOL, A. C. (2008). Aspirations, attainments, and satisfaction: Life cycle differences between American women and men [online]. *Journal of happiness studies*, 9: 601–619. <https://doi.org/10.1007/s10902-008-9106-5>.

DIJKSTRA, L., D'HOMBRES, B., MONTALTO, V., PONTAROLLO, N., BOLSI, P., CASTELLI, C., DE DOMINICIS, L., POELMAN, H. (2020). *Report on the Quality of Life in European Cities* [online]. <https://doi.org/10.2776/600407>.

EUROPEAN COMMISION. (2020). *Report on Quality of life in European Cities. October 7* [online]. Luxembourg. <https://doi.org/10.2776/600407>.

EUROPEAN COMMISION. (2015). *Flash Eurobarometer 366: Quality of Life in European Cities* [online]. Luxembourg. [cit. 1.5.2023]. <http://data.europa.eu/88u/dataset/S1035_366>.

EUROPEAN COMMISION. (2016). *Flash Eurobarometer 419: Quality of Life in European Cities* [online]. Luxembourg. [cit. 1.5.2023]. <http://data.europa.eu/88u/dataset/S2070_419_ENG>.

EUROPEAN UNION. (2013). *Quality of life in Cities. Perception survey in 79 European Cities* [online]. Luxembourg. [cit. 27.10.2022]. <https://ec.europa.eu/regional_policy/sources/studies/urban/survey2013_en.pdf>.

EUROSTAT. (2015). *Quality of life: facts and views* [online]. Luxembourg. [cit. 20.10.2022]. <https://ec.europa.eu/eurostat/web/products-statistical-books/-/KS-05-14-073>.

EUROSTAT. (2022). *Quality of life indicators – overall experience of life* [online]. Luxembourg. [cit. 18.12.2022]. <https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Quality_of_life_indicators_-_overall_experience_of_life>.

FERRER-I-CARBONELL, A. (2005). Income and well-being: an empirical analysis of the comparison income effect [online]. *Journal of public economics*, 89(5–6): 997–1019. <https://doi.org/10.1016/j.jpubeco.2004.06.003>.

FLEMING, C. M., MANNING, M., AMBREY, C. L. (2016). Crime, greenspace and life satisfaction: An evaluation of the New Zealand experience [online]. *Landscape and Urban Plannin*g, 149: 1–10. <https://doi.org/10.1016/j.landurbplan.2015.12.014>.

GIANNICO, V., SPANO, G., ELIA, M., D'ESTE, M., SANESI, G., LAFORTEZZA, R. (2021). Green spaces, quality of life, and citizen perception in European cities [online]. *Environmental research*, 196: 110922. <https://doi.org/10.1016/j.envres.2021.110922>.

GAJDOŠ, J., HUDEC, O. (2020). European cities: What matters for subjective well-being? *Theoretical & Empirical Researches in Urban Management*, 15(4): 5–20.

GLAESER, E. (2011). *Triumph of the city: How urban spaces make us human.* Pan Macmillan.

HANSEN, T., SLAGSVOLD, B. (2016). Late-life loneliness in 11 European countries: Results from the Generations and Gender Survey [online]. *Social Indicators Research*, 129(1): 445–464. <https://doi.org/10.1007/s11205-015-1111-6>.

CHEN, C. W. (2023). Can smart cities bring happiness to promote sustainable development? Contexts and clues of subjective well-being and urban liveability [online]. *Developments in the Built Environment*, 13: 100108. <https://doi.org/10.1016/j.dibe.2022.100108>.

LAYOUS, K. T. I. N., LYUBOMIRSKY, S. (2014). The how, why, what, when, and who of happiness. In: GRUBER, J., MOSKOWITZ, J. T. (eds.) *Positive emotion: Integrating the light sides and dark sides*, 473–495.

LEE, S. (2021). Social exclusion and subjective well-being among older adults in Europe: findings from the European social survey [online]. *The Journals of Gerontology*, Series B, 76(2): 425–434. <https://doi.org/10.1093/geronb/gbaa172>.

LEICHSENRING, K., SANDU, V., ZÓLYOMI, E. (2021). *Addressing loneliness and social isolation among older people in Europe* [online]. Policy Brief 2021/7, Vienna: European Centre. <https://www.euro.centre.org/publications/detail/4127>.

LENZI, C., PERUCCA, G. (2020). Urbanization and subjective well-being [online]. In: DELLA TORRE, S., CATTANEO, S., LENZI, C., ZANELLI, A. *Regeneration of the built environment from a circular economy perspective*. Springer Nature, 21–28. <https://doi:10.1007/978-3-030-33256-3_3>.

LENZI, C., PERUCCA, G. (2023). Economic inequalities and discontent in European cities [online]. *Urban Sustainability*, 3(1): 26. <https://doi.org/10.1038/s42949-023-00104-1>.

LI, Y., WINTERS, J. V. (2017). Urbanisation, natural amenities and subjective well-being: Evidence from US counties [online]. *Urban Studies*, 54(8): 1956–1973. <https://doi.org/10.1177/0042098016631918>.

LOSCHIAVO, D. (2021). Big-city life (dis)satisfaction? The effect of urban living on subjective well-being [online]. *Journal of Economic Behavior & Organization*, 192: 740–764. <https://doi.org/10.1016/j.jebo.2021.10.028>.

MARTELA, F., GREVE, B., ROTHSTEIN, B., SAARI, J. (2020). The Nordic exceptionalism: What explains why the Nordic countries are constantly among the happiest in the world [online]. *World happiness repor*t, 129–146. [cit. 4.8.2023]. <https://worldhappiness.report/ed/2020/the-nordic-exceptionalism-what-explains-why-the-nordic-countries-are-constantly-among-the-happiest-in-the-world/>.

MARTON, B., MOJSEJOVÁ, A. (2022). Macroeconomic Indicators and Subjective Well-Being: Evidence from the European Union [online]. *Statistika: Statistics and Economy Journal*, 102(4): 369–381. <https://doi.org/10.54694/stat.2022.19>.

MOURATIDIS, K. (2018). Rethinking how built environments influence subjective well-being: A new conceptual framework [online]. Journal of Urbanism: International Research on Placemaking and Urban Sustainability, 11(1): 24–40. <https://doi.org/10.1080/17549175.2017.1310749>.

MOURATIDIS, K. (2019). Compact city, urban sprawl, and subjective well-being [online]. *Cities*, 92: 261–272. < https://doi.org/10.1016/j.cities.2019.04.013>.

MOURATIDIS, K. (2020). Commute satisfaction, neighborhood satisfaction, and housing satisfaction as predictors of subjective well-being and indicators of urban livability [online]. *Travel Behaviour and Society*, 21: 265–278. <https://doi.org/10.1016/j.tbs.2020.07.006>.

NATIONS ONLINE PROJECT. *List of European Countries – Nations Online project* [online]. [cit. 17.10.2022]. <https://www.nationsonline.org/oneworld/europe.htm.>.

OKULICZ-KOZARYN, A., VALENTE, R. R. (2019). Livability and Subjective Well-Being Across European Cities [online]. *Applied Research Quality Life*, 14(1): 197–220. <https://doi.org/10.1007/s11482-017-9587-7>.

OLSEN, J. R., NICHOLLS, N., MITCHELL, R. (2019). Are urban landscapes associated with reported life satisfaction and inequalities in life satisfaction at the city level? A cross-sectional study of 66 European cities [online]. *Social Science & Medicine*, 226: 263–274. <https://doi.org/10.1016/j.socscimed.2019.03.009>.

PINEO, H., RYDIN, Y. (2018). *Cities, Health and Well-being* [online]. London: Royal Institution of Chartered Surveyors (RICS). [cit. 4.8.2022]. <https://discovery.ucl.ac.uk/id/eprint/10107484/>.
POORTINGA, W., BIRD, N., HALLINGBERG, B., PHILLIPS, R., WILLIAMS, D. (2021). The role of perceived public and private green space in subjective health and wellbeing during and after the first peak of the COVID-19 outbreak [online]. *Landscape and Urban Planning*, 211: 104092. <https://doi.org/10.1016/j.landurbplan.2021.104092>.
ROBINSON, T., GUSTAFSON, B., POPOVICH, M. (2008). Perceptions of negative stereotypes of older people in magazine advertisements: Comparing the perceptions of older adults and college students [online]. *Ageing & Society*, 28(2): 233–251. <https://doi.org/10.1017/S0144686X07006605>.
RODRÍGUEZ-POSE, A., MASLAUSKAITE, K. (2012). Can policy make us happier? Individual characteristics, socio-economic factors and life satisfaction in Central and Eastern Europe [online]. *Cambridge Journal of Regions, Economy and Society*, 5(1): 77–96. <https://doi.org/10.1093/cjres/rsr038>.
RUIU, G., RUYU, M. L. (2019). The Complex Relationship Between Education and Happiness: The Case of Highly Educated Individuals in Italy [online]. *Journal of Happiness Study*, 20: 2631–2653. <https://doi.org/10.1007/s10902-018-0062-4>.
SARRACINO, F. (2010). Determinants of subjective well-being in high and low income countries: Do happiness equations differ across countries? [online]. *The Journal of socio-economics*, 42(0): 51–66. <https://doi.org/10.1016/j.socec.2012.11.006>.
SMITH, C. A., KONIK, J. (2022). Who is satisfied with life? Personality, cognitive flexibility, and life satisfaction [online]. *Current Psychology*, 41(12): 9019–9026. <https://doi.org/10.1007/s12144-021-01359-6>.
TOSHKOV, D. (2022). The Relationship Between Age and Happiness Varies by Income [online]. *Journal of Happiness Studies*, 23(3): 1169–1188. <https://doi.org/10.1007/s10902-021-00445-7>.
VOUKELATOU, V., GABRIELLI, L., MILIOU, I., CRESCI, S., SHARMA, R., TESCONI, M., PAPPALARDO, L. (2021). Measuring objective and subjective well-being: dimensions and data sources [online]. *International Journal of Data Science and analytics*, 11(4): 279–309. <https://doi.org/10.1007/s41060-020-00224-2>.
WATSON, R. (2005). People over 50 in northern Europe feel healthier than their southern counterparts [online]. *British Medical Journal*, 330(7499): 1044. <https://doi.org/10.1136/bmj.330.7499.1044-b>.
WUNDER, C., WIENCIERZ, A., SCHWARZE, J., KÜCHENHOFF, H. (2013). Well-being over the life span: Semiparametric evidence from British and German longitudinal data [online]. *Review of Economics and Statistics*, 95(1): 154–167. <https://doi.org/10.1162/REST_a_00222>.
ŽELINSKÝ, T., HUDEC, O., MOJSEJOVÁ, A., HRICOVÁ, S. (2021). The effects of population density on subjective well-being: A case-study of Slovakia [online]. *Socio-Economic Planning Sciences*, 78: 101061. <https://doi.org/10.1016/j.seps.2021.101061>.

Life Expectancy Changes and Their Consequences for Pension System in Finland and the Czech Republic

Ondřej Šimpach [1] | *Prague University of Economics and Business, Prague, Czech Republic*

Received 1.5.2023 (revision received 25.7.2023), **Accepted** (reviewed) 31.8.2023, **Published** 15.12.2023

Abstract

Finland and the Czech Republic are among the countries where population ageing has been the most pronounced in the last decade. The aim of the paper is to describe future development of life expectancy in the context of pension system reforms that are currently prepared by the politicians in analysed countries.

One-year age-and-sex-specific mortality rates for population aged 0 to 100+ were taken from the Human Mortality Database for 1950–2021 and projected to 2050. Three stochastic models were calculated in R and compared. Suitable was Lee-Carter model modified by Li-Lee-Gerland (with rotation of b_x parameter) because of low infant mortality in both populations. Projected year-on-year change of life expectancy was comparable to the Eurostat, but absolute values were too optimistic in our case.

Values of temporary life expectancy between 60 and 70 years and indices of annual relative changes revealed relatively fast pace of increase in life expectancy in both populations which the pension systems should take into account.

Keywords

Ageing, Czech Republic, mortality projections, Finland, Li-Lee-Gerland model, temporary life expectancy

DOI

https://doi.org/10.54694/stat.2023.20

JEL code

J11, J18, C29

INTRODUCTION

Population ageing is a problem and challenge in many countries. The people are living longer and thus spending more time in retirement which is a challenge for the pension system and its sustainability. Permanently low natality and higher life expectancy are changing the shape of the age pyramid in the EU-27 towards much older population structure. (Eurostat, 2020). Finland and the Czech Republic are among the countries of the European Union where population ageing has accelerated in the last ten years. The share of people aged over 65 increased between 2009 and 2019 by 5.1 p.p. and 4.7 p.p.,

[1] Department of Statistics and Probability, Faculty of Informatics and Statistics, Prague University of Economics and Business, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic. E-mail: ondrej.simpach@vse.cz, phone: (+420)737665461. ORCID: <https://orcid.org/0000-0002-6978-4304>.

resp. In the decade of 2012 to 2022, Finland took second place after Poland as the share of population over 65 years increased by 5 p. p. The increase was lower (by 4.4 p. p.) in the Czech Republic.

Those countries do not belong to the states with the highest median age of population (that is in Italy and Portugal – see Eurostat, 2023a), but the pace of ageing is one of the fastest in the EU. There were 6.0% of people older than 80 years and 23.1% older than 65 years in Finland in 2022, which ranks it among the member states with the highest shares of the old people in the EU. The Czech Republic had 4.3% of people older than 80 years and 20.6% of older than 65 years. However, clear trend of ageing is visible here, too. "The largest increase in the median age in the EU between 2020 and 2022 can be observed in Croatia, Portugal, Greece, Spain, Czech Republic and Italy all rising by more than 0.5 year" (Eurostat, 2023c).

The share of older people in the total population is expected to increase in the coming years due to increased longevity. The life expectancy at birth has been rising until year 2019, i.e. before the start of the Covid-19 pandemic. Main reasons were the reduction in infant mortality, rising of living standards, improved lifestyles, better education and better healthcare and medicine. The increase in life expectancy is commonly attributed to better healthcare and increased living standards. However, Šimpach and Pechrová (2013) found only a weak dependence between life expectancy and living standards expressed by variables of household and municipality amenities and availability of health care.

Life expectancy at birth in the Czech Republic rose to 79.3 years in 2019, but decreased to only 77.2 years in 2022 which was below the level in 2010. The decrease was not that significant in Finland, from 82.1 in 2019 to 81.9 years in 2022. Life expectancy at birth was higher for females in both countries and was above 80 years (Eurostat, 2023d).

Higher share of older people leads to the "ageing at the top" of the population pyramid. On the other hand, the pyramid is shrinking on the bottom due to low fertility. This may lead to an increased burden on those at working age to provide for the social expenditure required by the ageing population for a range of related services (Eurostat, 2023b).

"The baby boomer generation, born between 1946 and 1949, will pose a specific challenge to the service system despite the fact that they are likely to age with better health and functional ability than previous generations" (FIHW, 2023). The Czech Republic will face similar process, but later, as baby boomer generation was born in the 1970s.

The increase of labour force of older age will be caused not only by the irregularities in the age structure, but also by the permanent increase of retirement age (Fiala and Langhamrová, 2014). This could bring consequences for the age structure of the employed persons. Fiala and Langhamrová (2014) expect that the average age of persons in productive age will increase in the Czech Republic, because the share of younger workers will decline and the share of employees over 50 years will increase relatively strongly.

According to the information from European Commission (EC, 2022a) the retirement age is between 63 to 65 years in Finland, depending on the year when the person was born (1954 or earlier up to year 1964). The retirement age of persons born in 1965 and later is adjusted with the life expectancy which will be determined at the age of 62 years. "Since January 2017, the retirement age has been raised by 3 months annually until to reach 65 years by 2027. Then, the retirement age will be linked to life expectancy" (EC, 2022a). The retirement age in the Czech Republic currently depends on the year of birth, sex and number of children of the person born until year 1971. The retirement age for all persons born after 1971 is 65 years in all categories (EC, 2022b).

Needed reform of the pension system has not yet been implemented in the Czech Republic. On the other hand, Finland, together with countries such as the Netherlands, Estonia and Denmark have implemented reforms that link retirement age to changes in life expectancy. Danish case was examined by Alvarez et al. (2021) and they found out that linking retirement age to life expectancy increases uncertainty about length of life after retirement and hence the financial costs of the pension system are more sensitive to changes in mortality. Besides "socio-economic disparities in lifespans persist regardless of the age at which individuals retire" (Alvarez et al., 2021). In addition, the later retirement can bring social tension between job applicants. There have been many stereotypes about alleged lower performance of older employees that leads to unwillingness to employ older people, which, in effect, may lead to a loss of the crucial knowledge possessed by them (Musilová and Režňáková, 2015).

Therefore, our article examines Finland and the Czech Republic and projects the future development of the life expectancy. The aim is to find out how the *life expectancy* will develop, which can serve a basis for deciding on pension system planning and reforms.

1 LITERATURE SURVEY

"One of the possibilities to take into account life expectancy in pension systems is to link some parameters of pension systems to the development of life expectancy, specifically automatically adjusting the retirement age to life expectancy" (Holub et al., 2020).

The life expectancy of people in Finland has been increasing continuously over the past decades and is above the EU average. "These gains in life expectancy were driven by steady reductions in deaths from cardiovascular diseases" (OECD, 2017). Besides, a reduction in mortality rates after the age of 65 has occurred since 2000. "Finland, healthy working life expectancy has increased irrespective of how health is measured but also working with health problems has become more prevalent" (Laaksonen et al., 2022).

Historical development of mortality is described in study of Kannisto et al. (1999). They pointed out that since 1950 the decline of mortality rates have suddenly accelerated due to the introduction of antibiotics. However, this development soon slowed down and was followed by a slowdown, for adult men even by a reversal. Around 1970 an unprecedented decline in the mortality of the elderly which raised life expectancies at younger ages as well (Kannisto et al., 1999).

The life expectancy in the Czech Republic was always comparable with that of the developed countries, even before the Second World War (Langhamrová, 2014). However, in 1945 was seen a sudden drop, mainly due to worsened hygienic conditions at the end of the war. Decreased infant (especially neonatal mortality) resulted in more than 50% increase in life expectancy at birth in 1950s. "Until the early sixties, the Czech Republic was one of the countries with the lowest infant mortality rates in the world" (Langhamrová, 2014). In the socialist era since the mid-sixties, the mortality rates started to worsen, and life expectancy stagnated. The improvement started after the revolution since 1990s (Langhamrová, 2014; Šimpach et al., 2014). Promising development, however, was broken due to the Covid-19 pandemic in years 2020 and 2021.

Impact of Covid-19 on the mortality rates in the Czech Republic and in Spain was examined by Šimpach and Šimpachová Pechrová (2021). They found out that Covid-19 pandemic affected the mortality rates in a way that they were higher and decreased at a slower pace than they would without taking 2020 into account in ages above 50 in Spain. The effect was less pronounced in the Czech Republic.

A methodology for analysis of development and changes in life expectancies was elaborated by Arriaga (1984). He presented a set of indices for interpreting change in life expectancies and also introduced technique for explaining change in life expectancy by change in mortality at each age group. Our article utilizes the indices to measure the changes in life expectancy in old-age groups in projection horizon.

2 METHODS

Mortality modelling can be based on deterministic or stochastic models. Deterministic modelling is based on prior set assumptions which are stated either by expert guess or by supportive statistical methods. The approach of the cohort-component method is the most often used (Leslie, 1945), which can be further enriched and developed with elements from probability theory. For example, Fiala and Langhamrová (2014) used deterministic component method to project the population of the Czech Republic in medium, low and high variant. They expected that life expectancy at birth will continuously increase, but at slower pace and that mortality continues to be relatively low.

On the other hand, stochastic projections are based on stochastic time series models of age-specific demographic measures that complement multivariate statistical methods. The models contain stochastic and error terms.

Syuhada and Hakim (2021) modelled the mortality rate by an Autoregressive (AR) model with a conditional heteroscedasticity effect that was accommodated by a stochastic model of Autoregressive Conditional Heteroscedastic (ARCH) and Stochastic Volatility Autoregressive (SVAR) model. They further forecasted Mortality-at-Risk (MaR) which is a risk measure that refers to the fall of mortality rates for given time period so their results could serve to insurance industry.

Šimpach and Šimpachová Pechrová (2021) used stochastic method, particularly original Lee-Carter model, to elaborate the projections of the mortality rates for the Czech Republic and Spain.

We model mortality rates also by stochastic method and consequently we calculate life expectancy using life tables algorithm. Mortality rates of x-year old in time t ($\mathbf{m}_{x,t}$) represents the ratio of the number of deaths of x-year old in time t ($\mathbf{D}_{x,t}$) and of the exposure to risk which is mid-year population of x-year old population ($\mathbf{E}_{x,t}$) – see Formula (1).

$$\mathbf{m}_{x,t} = \frac{\mathbf{D}_{x,t}}{\mathbf{E}_{x,t}} . \tag{1}$$

We chose stochastic approach, particularly Lee-Carter (LC) model which has the advantage of being comprehensible concept and giving historical values of mortality lower weights in the projections than to recent data. Lee and Carter (1992) elaborated a model of age-specific death rates with a time component $\mathbf{b}_{t,}$ a fixed relative age component $\mathbf{a}_x$, and a time series model (an autoregressive integrated moving average – ARIMA of the time component $\mathbf{k}_t$ (Booth et al., 2002). Original Lee-Carter model (1992) models the natural logarithms of age-specific death rates as (2).

$$m_x = e^{\mathbf{a}_x + \mathbf{b}_x \mathbf{k}_t + \boldsymbol{\varepsilon}_{x,t}} \text{ or in linearized form as } \ln \mathbf{m}_{x,t} = \mathbf{a}_x + \mathbf{b}_x \mathbf{k}_t + \boldsymbol{\varepsilon}_{x,t} , \tag{2}$$

where $\mathbf{a}_x$ are the age-specific profiles independent of time which represents the general mortality shape across age; vector of $\mathbf{b}_x$ are the additional age-specific components that determine the changes in each age group when $\mathbf{k}_t$ changes; $\mathbf{k}_t$ are the time-varying parameters of the level of mortality for all ages and $\boldsymbol{\varepsilon}_{x,t}$ is error term. Ages x take values from 0 to $\omega - 1$ (where ω is an age when any person from original population is not alive anymore). Time is represented by t and takes values from 1 to T (that is year 2021 in our case studies).

"The $\mathbf{b}_x$ profile tells us which rates decline rapidly and which rates decline slowly in response to changes in $\mathbf{k}_t$" (Lee and Carter, 1992). That means that it is derivation of logarithm of mortality rates $\mathbf{m}_{x,t}$ by the time (3).

$$\frac{d \ln \mathbf{m}_{x,t}}{dt} = \frac{d(\mathbf{a}_x + \mathbf{b}_x \mathbf{k}_t)}{dt} = \mathbf{b}_x \frac{d\mathbf{k}_t}{dt} . \tag{3}$$

The coefficient $\mathbf{b}_x$ can be negative for some ages, which indicates that mortality at those ages tends to rise when mortality in other ages is falling.

Error term $\boldsymbol{\varepsilon}_{x,t} \approx \mathrm{N}(0,\sigma_\varepsilon^2)$ captures age-specific historical influences not captured by the model and is supposed to be non-correlated and homoscedastic. However, "the observed pattern of the mortality rates shows a different variability at different ages, highlighting that the homoscedasticity hypothesis is quite unrealistic" (Russolillo, 2017). For modelling of the data at higher ages, certain models such as Kannisto can be used. For example, the Czech Statistical office (2020) uses model based on logistic curve which takes into account the slowing of the increase in mortality with age (Thatcher, Kannisto and Vaupel, 1998).

The model written in Formula (2) is underdetermined, therefore Lee and Carter (1992) normalized the $\mathbf{b}_x$ to sum to unity and the $\mathbf{k}_t$ to sum to 0, which implies that the $\mathbf{a}_x$ are simply the averages over time of the $\ln(\mathbf{m}_{x,t})$. Using these constrains $\sum_{x=1}^{N}\mathbf{b}_x = 1$ and $\sum_{t=1}^{T}\mathbf{k}_t = 0$, the least squares estimator for $\mathbf{a}_x$ can be obtained by (4), Danesi et al. (2015).

$$\hat{a}_x = \frac{\sum_{x=1}^{N} \ln m_{x,t}}{N}, \tag{4}$$

where N is the total number of used years. Under this normalization, $\mathbf{b}_x$ is the proportion of the change in overall logarithm of mortality attributable to age x. Indexes of the intensity of level of mortality $\mathbf{k}_t$ are next modelled as a time series (specifically, a random walk with drift) and forecasted as proposed by Lee and Carter (1992). The order of lags is determined based on recommendation by Akaike information criterion. For example, Russolillo (2017) used the ARIMA (0,1,0) model to forecast the index of mortality $\mathbf{k}_t$ for next 25 years. Šimpach and Dotlačilová (2015) applied ARIMA (1,1,0) with drift on the Czech population.

"The singular value decomposition (SVD) method can be used to find a least squares solution when applied to the matrix of the logarithms of the rates after the averages over time of the (log) age-specific rates have been subtracted" (Lee and Carter, 1992).

A model was further extended by Li and Lee (2005) to calculate coherent mortality forecasts for countries with low infant mortality. One of the disadvantages of the Lee-Carter model is namely that the age-specific set of the $\mathbf{b}_x$ parameter is estimated on the basis of historical data and does not develop in time. Li, Lee and Gerland (2013) examined low-mortality countries and observed that during long time forecast of mortality rates there was too high decrease of mortality in neonatal and low ages in comparison with decrease of mortality in older ages. Therefore, the rotation of $\mathbf{b}_x$ in time must be modelled in order to project long-term mortality changes. Because $\mathbf{b}_x$ is calculated as the first-order derivation of $\ln(\mathbf{m}_{x,t})$ at all ages (3), if we want to model change of $\mathbf{b}_x$ in time, we actually model the second-order differences of $\ln(\mathbf{m}_{x,t})$ at all ages. Li, Lee and Gerland (2013) modified the shape of the $\mathbf{b}_x$ curves that were estimated from historical data by standard LC model, by smoothing their values at adolescent and adult ages (15–65) to equal the average value for this age range, and then by reducing the values at infant and child ages (0–14) to this average. Coherent parameter $\mathbf{b}^c{}_x$ and ultimate parameters $\mathbf{b}^u{}_x$ are the same for males and females. Based on the transformation, a curve of the final age-specific change in mortality is thus smoothed.

When searching for suitable model to model mortality rates in Finland and the Czech Republic, we must look on the characteristics of the examined populations. Finland had the lowest infant mortality in the EU in 2021 (1.8 deaths per 1 000 live births). The Czech Republic was the fifth with 2.2 deaths per 1 000 live births which is the lowest value in recent years. This is an important feature that favours Li-Lee-Gerland model.

Finally, index $\mathbf{k}_t$ is projected to the future and the mortality rates are calculated from Formula (2) up to year 2050.

Fitted and forecasted mortality rates are then used to calculate life expectancy using life table algorithm (see e. g. Šimpach et al., 2013). First, the p_x – probability that an individual alive at time x will survive the interval $(x, x + 1)$ is calculated as (5).

$$p_x = \exp(-m_x) \cdot h_x, \tag{5}$$

where *exp()* is Euler constant, h_x is the length of the age interval (1 year in our case). Then the probability that an individual alive at time x will die in the interval $(x, x + 1)$, i. e. before the next birthday, is $q_x = 1 - p_x$. (see Gompertz, 1825; Chiang, 1961; Šimpach et al., 2013).

Consequently, the life tables are calculated. We start with hypothetical model cohort of 100 000 alive males and females at the age of 0 (so-called radix of the table) and calculate number of people left alive at age x (l_x) until the age $\omega - 1$, where ω is an age when no person from original population is alive as (6).

$$l_x = l_{x-1} \cdot p_{x-1} \text{ or } l_x = l_{x-1} \cdot (1 - q_{x-1}) \,. \tag{6}$$

We start with multiplying 100 000 (l_0) by the probability that a person aged 0 will die before reaching 1 year (q_0) to obtain the number of deaths at age 0 years (d_0) which is then subtracted from original population of 100 000 people. Hence, we gain the number of surviving people to age 1. The calculation continues to the age ω (100 in our case, because that is the strict limit of used R-packages).

For the calculation of life expectancy at birth, the projection of alive born children is needed, but we do not project fertility in our paper, so we do not have available the number of alive born children and the p_0 is not correctly calculated in our case. This does not pose any problem to further calculation of life expectancy, as we are mainly interested in life expectancy as older ages (e_{60}, e_{65}, e_{70}, e_{75}, e_{80}) rather than at birth (e_0).

For pension system policy is an important indicator the average number of years of life remaining at exact age x. Therefore, life expectancy e_x of the person at the exact age x years is calculated as (7). It is the ratio of the number of remaining years of life (T_x) and the number of survivors to exact age (l_x).

$$e_x = \frac{T_x}{l_x} \,, \tag{7}$$

where T_x (total number of years lived by alive population from age x to the end – i.e. remaining years to end age ω) is gained by summing number of years that x years old person lived until actual age (L_x). For ages above 1 year, it is assumed that deaths occur evenly over a year of age, the L_x is calculated as follows (8):

$$L_x = \frac{l_x + l_{x+1}}{2} \,. \tag{8}$$

The calculation of changes in life expectancy are based on Arriaga (1984) methodology. Particularly, we use temporary life expectancies (${}_ie_x$), i.e. life expectancies between two specific ages (x) and time (t), and indices that are calculated based on the comparison of temporary life expectancies. "The temporary life expectancy from age x to $x + i$ is the average number of years that a group of persons alive at exact age x will live from age x to $x + i$ years" (Arriaga, 1984). i is the difference between initial age x and last examined year $x + i$. Temporary life expectancy is calculated as (9).

$${}_ie_x = \frac{T_x - T_{x+i}}{l_x} \,. \tag{9}$$

When calculating the relative changes over time period of temporary life expectancy (${}_iRC_x^n$), the absolute difference between temporary life expectancy at particular age and time t (${}_ie_x^t$) and time $t + n$ (${}_ie_x^{t+n}$) is related to the difference between i and temporary life expectancy in time t (10).

$${}_iRC_x^n = \frac{{}_ie_x^{t+n} - {}_ie_x^t}{i - {}_ie_x^t} \,. \tag{10}$$

Because RC index does not enable comparison when there are different time-periods compared, an index of annual change of the RC index is calculated as (11).

$${}_iARC_x^n = \left[1 - (1 - {}_iRC_x^n)^{1/n}\right] \cdot 100 \,. \tag{11}$$

All measures are calculated for both populations, compared and interpreted in the context of pension reforms.

One-year age-and-sex-specific mortality rates for population aged 0 to 100+ for Finland and the Czech Republic were taken from the Human Mortality Database and they are available in a time series from 1950 to 2021.

Calculations were done in software R. There were 3 approaches compared. Approach (A) used packages *demography* (Hyndman et al., 2012, 2022) that estimates original Lee-Carter model and age-specific mortality rates (and also can estimate fertility rates and migration).

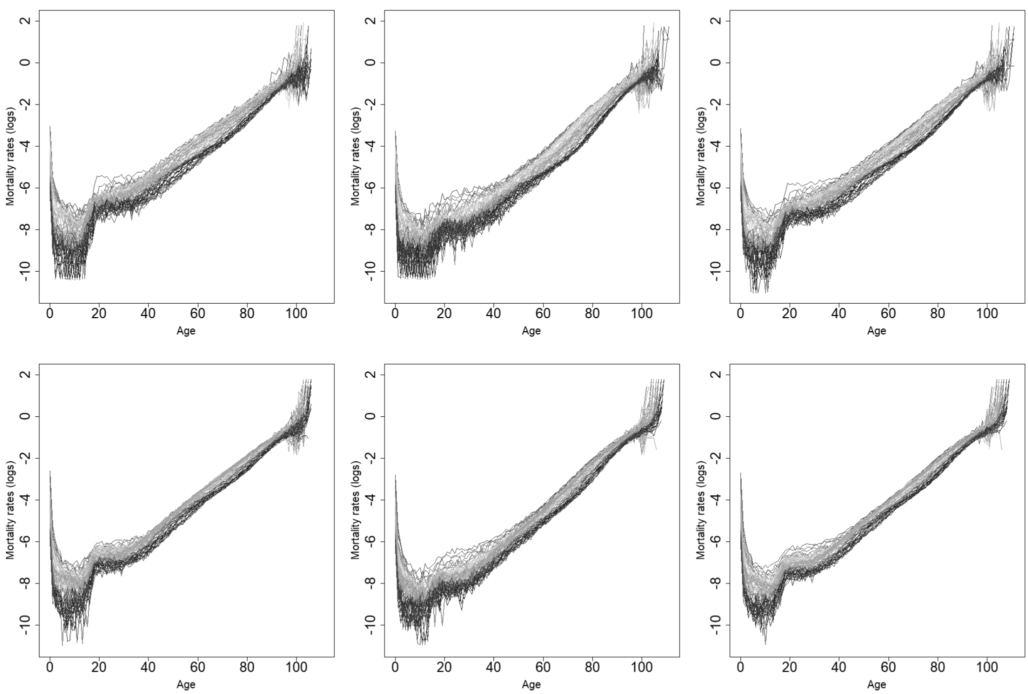
Approach (B) used package *MortCast* elaborated by Ševčíková et al. (2022) which estimates and projects age-specific mortality rates using augmented Lee-Carter method and related methods that are described in Ševčíková et al. (2016). An augmented Lee-Carter model is estimated here by function *leecarter.estimate*.

Approach (C) also used package *MortCast* by Ševčíková et al. (2022), but this time with extension by Li-Lee-Gerland (2013). Coherent parameter $\mathbf{b}^c_x$ and the ultimate $\mathbf{b}^u_x$ for rotation for male and female mortality rates are estimated here by a function *lileecarter.estimate*. Both parameters are then used for total population, too. Index $\mathbf{k}_t$ is projected by ARIMA model using *auto.arima* command of the package *forecast* by Hyndman (2012). This command selects optimal model based on the lowest value of Akaike Information Criterion (AIC) and corrected Akaike Information Criterion (AICc).

3 RESULTS

Figure 1 displays mortality rates in logarithms for population of males, females and total population in Finland and the Czech Republic. Light grey represents year 1950 and as the colour is darker, the

Figure 1 Mortality rates in logarithms for population of males (left), females (middle) and total population (right) in Finland (upper) and the Czech Republic (lower)



Source: Own elaboration based on the data from Human Mortality Database (HMD, 2023)

newer is the data. Black colour represents year 2021. It can be clearly seen that the mortality rates had declined over time in all populations. There was a high variability of mortality rates in population of males in higher ages, especially in Finland. It is natural that the largest changes of mortality rates are at the highest ages (approximately 80 years and above), where the mortality has got the different character than in lower ages as there are low numbers of deaths and also low numbers of living. Besides, this data is affected by systematic and random errors.

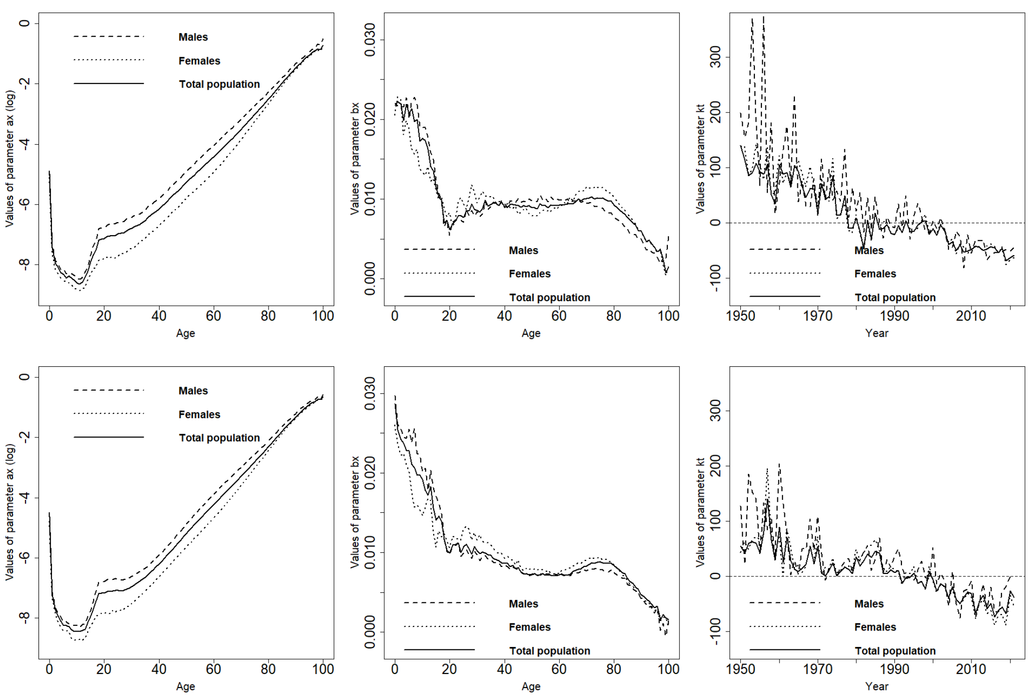
3.1 Mortality rates modelling

First, above stated mortality rates were modelled by three approaches: (A) Hyndman (2012, 2022), (B) Ševčíková et al. (2016) and (C) Ševčíková et al. (2016) with extension by Li-Lee-Gerland (2013).

Figure 2 displays the components of the approach A – Hyndman (2012, 2022) – parameter $\mathbf{a}_x$ (left), parameter $\mathbf{b}_x$ (middle) and $\mathbf{k}_t$ (right) for Finland (upper) and the Czech Republic (lower). The development of the parameter $\mathbf{a}_x$ had expected development according to Gompertz (1825). Parameter $\mathbf{b}_x$ has also correct development. It shows that mortality rates decline rapidly until the age of 20 in comparison with other ages. The parameter is almost constant until the ages from 70 to 80. Time-varying index of the level of mortality for all ages for males' population is very volatile at the beginning of the time series. The volatility decreased later, but the changes are still higher than in case of females and total population.

Regarding the Czech population, parameters $\mathbf{a}_x$, $\mathbf{b}_x$ and index $\mathbf{k}_t$ also develop according to the assumptions. Decrease of $\mathbf{b}_x$ parameters is again sharp until the age of 20, but unlike in Finnish

Figure 2 Development of parameters a_x, b_x and index k_t in approach A



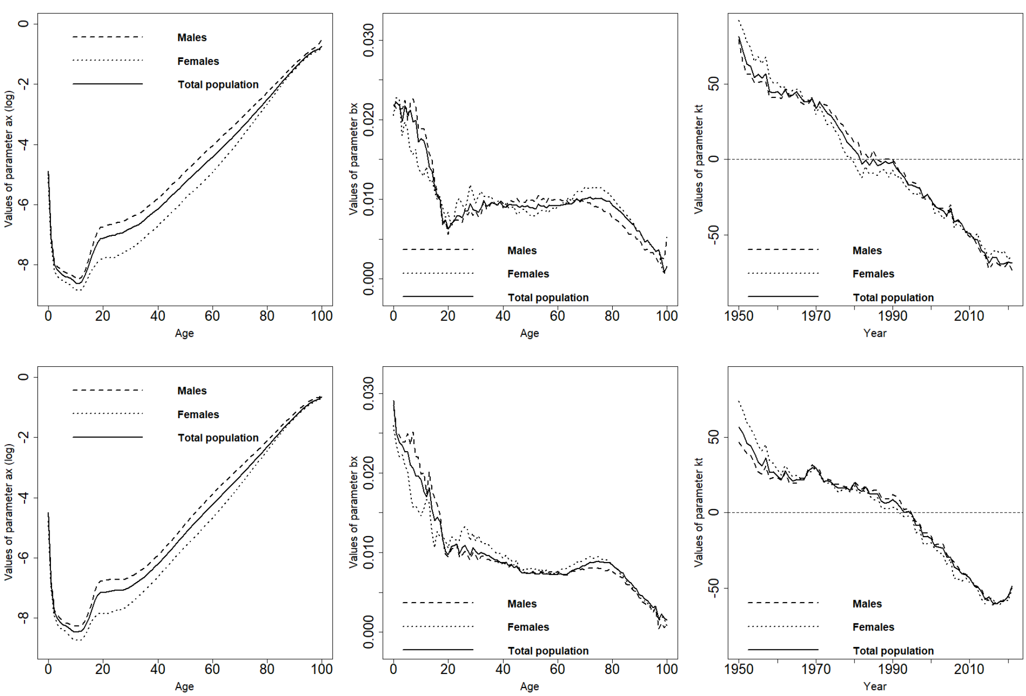
Source: Own elaboration based on the data from Human Mortality Database (HMD, 2023)

population it continuous until the age of 60 (despite than in slower pace). There is again high variability in the development of indices of total mortality $\mathbf{k}_t$ between 1950–1970, but this time for all populations. The decrease of $\mathbf{k}_t$ is slower than in Finnish population. We can expect that incorrect parameters will not be able to provide reliable forecasts.

Approach of Ševčíková et al. (2016) also uses Lee-Carter model, but with modification of parameter $\mathbf{a}_x$. From Figure 3 can be seen that approach B had almost similar parameters $\mathbf{a}_x$ and $\mathbf{b}_x$ as previous approach A. The highest difference is in index $\mathbf{k}_t$ probably because approaches A and B use different estimation methods. Hyndman (2012, 2022) uses Latent Class Analysis (LCA) and Ševčíková (2016) Singular Value Decomposition (SVD). Index for males' population is much smoother, the development of all populations is closer, and $\mathbf{k}_t$ is decreasing sharply throughout the whole period. The only exceptions are last two years (2020, 2021) due to Covid-19 pandemic. Time-varying index stagnated in Finland and increased in the Czech Republic. Here can be clearly seen the difference in handling the pandemic between the two states. In the Czech Republic, the effect of increased mortality was significantly manifested, and $\mathbf{k}_t$ not only stopped decreasing, but even increased.

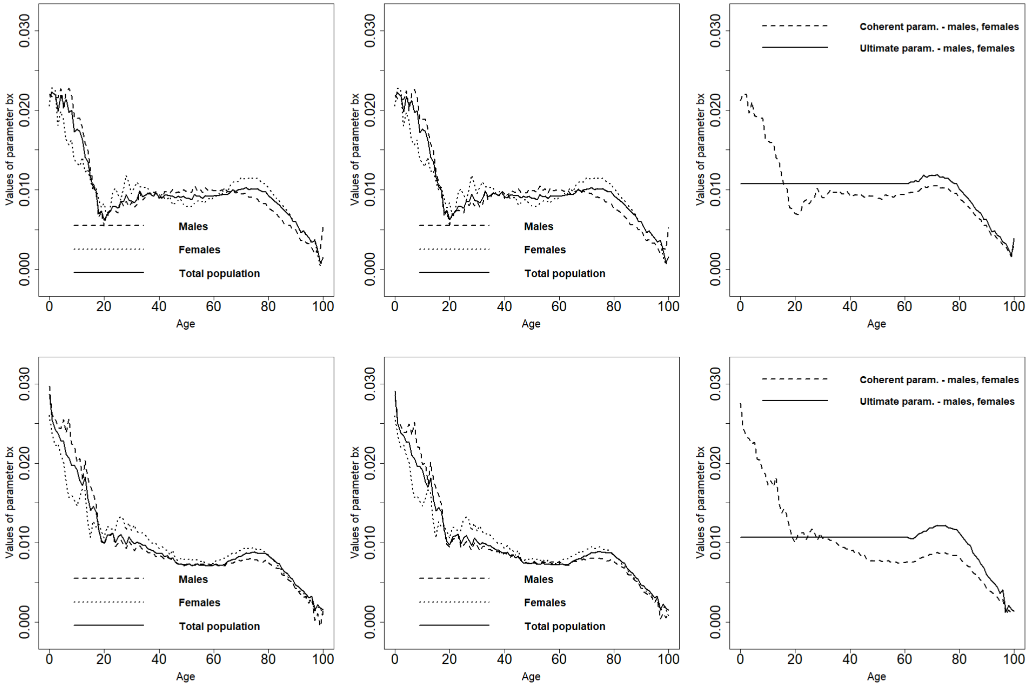
The approach C had similar parameter $\mathbf{a}_x$ and index $\mathbf{k}_t$ as approach B. Parameter $\mathbf{b}_x$ was rotated according to Li-Lee-Gerland (2013), because the decline of young people mortality is not that fast in developed populations as in developing populations, because the mortality is already very low there. Coherent parameter $\mathbf{b}^c_x$ and ultimate parameters $\mathbf{b}^u_x$ are the same for males and females, so they are used also for total population. $\mathbf{b}^u_x$ is constant up to age 60, $\mathbf{b}^c_x$ decreases up to age 60. Parameter $\mathbf{a}_x$ and index $\mathbf{k}_t$ are similar as in Ševčíková et al. (2016), so we present only comparison of parameters $\mathbf{b}_x$ of various approaches in Figure 4.

Figure 3 Development of parameters a_x, b_x and index k_t in approach B



Source: Own elaboration based on the data from Human Mortality Database (HMD, 2023)

Figure 4 Comparison of parameter b_x development in approaches A, B, C



Source: Own elaboration based on the data from Human Mortality Database (HMD, 2023)

3.2 Mortality rates projection

Mortality rates were projected in a way that index $\mathbf{k}_t$ was projected to the future and the mortality rates are calculated from the Formula (2). $\mathbf{k}_t$ was projected by ARIMA model using *auto.arima* command of the package *forecast* by Hyndman (2012), where optimal models were selected based on the lowest value of AIC and AICc criterion.

The results are displayed at Figure 5. It can be seen that confidence intervals are narrower in Finland than in the Czech Republic because of the impact of Covid-19 pandemic (volatility at the end

Figure 5 Projection of the k_t for Finland (upper) and the Czech Republic (lower)

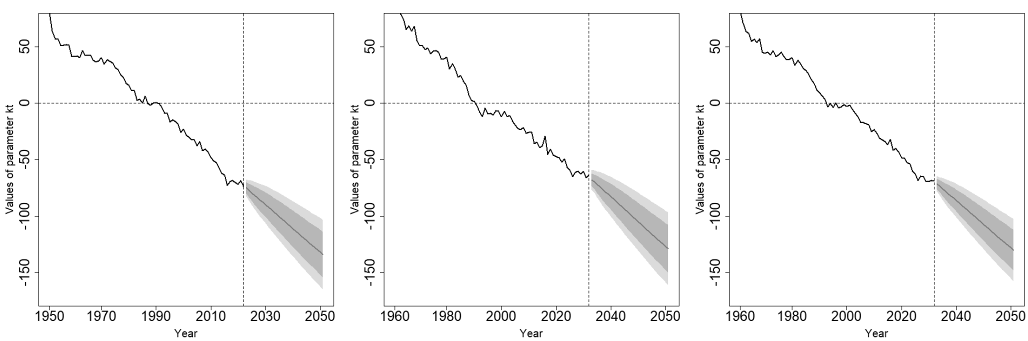
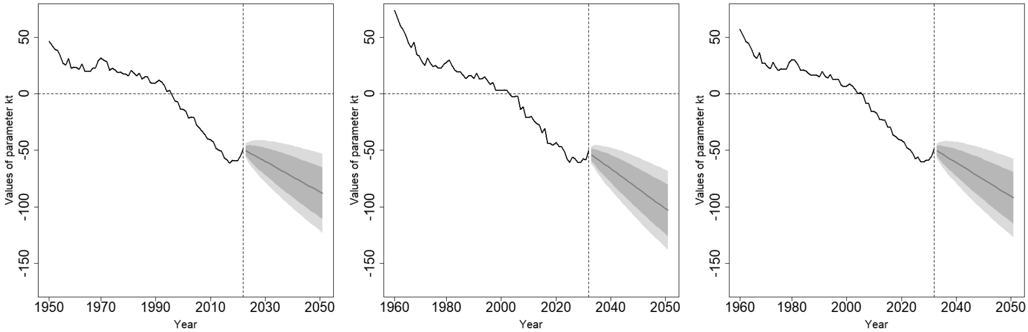
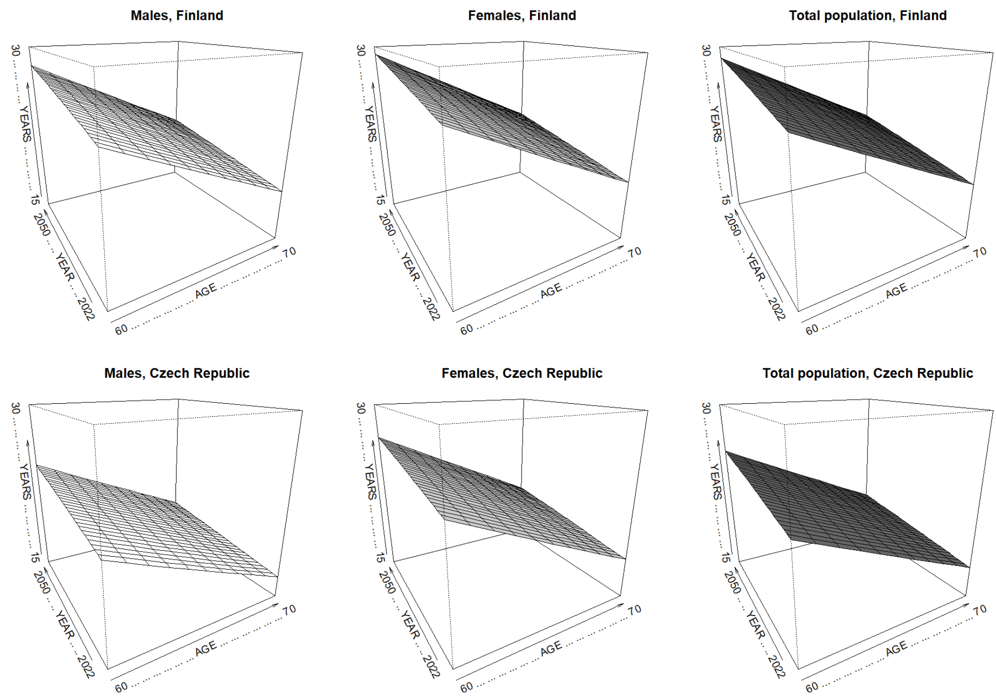


Figure 5 (continuation)



Source: Own elaboration based on the data from Human Mortality Database (HMD, 2023)

Figure 6 Life expectancy projection for Finland (upper) and the Czech Republic (lower) at ages from 60 to 70 years



Source: Own elaboration based on the data from Human Mortality Database (HMD, 2023)

of the time series increased). The best model for all Finnish populations was ARIMA(1,1,0) with drift. Optimal models for Czechs were ARIMA(0,1,0) with drift for males and total and ARIMA(1,1,0) with drift for females. The types of models are similar to Šimpach et al. (2014).

Mortality rates were projected by all three approaches. They were compared based on of the realism of the results and the fact that the mortality rates correspond to the assumptions and current developments.

The most feasible results provided approach of Ševčíková (2016) in modification by Li-Lee-Gerland (2013), so we present only mortality rates and life expectancy projections that are based on approach C. Mortality rates projections suggest continual decrease of mortality rates, that is expected in the future. Consequently, a life tables were calculated based on the empirical and projected data. Long increasing trend of life expectancy was interrupted by Covid-19 pandemic in 2020 and 2021. The effect is much higher for the case of the Czech Republic. However, the life expectancy will increase again in projected future as expected which can be seen in Figure 6.

The absolute values of the results were not optimal, because they were overrated by the model. Therefore, we focus on the pace of increase of life expectancy. Particularly we analyse first absolute differences and relative change of temporary life expectancies.

4 DISCUSSION

Because the paper focuses on the projection of life expectancies for the purpose of pension reform, we aim at the retirement age categories in this section. Time interval between 60 and 70 years is examined as the retirement age is within this range. Pace of increase of life expectancy of males and females is firstly analysed on the example of retirement age of 60 which tells us, how quickly the population will live longer, hence, how quickly the retirement age can/shall be postponed.

The results of our projections are compared with the Eurostat baseline projections (2023). This population projection is deterministic (what-if projection) and based on sets of assumptions for fertility, mortality and net migration. Besides, 5 sensitivity tests are done. "The assumptions formulated for mortality are based on the idea that in 2022 and 2023 the mortality rates have not completely aligned with levels observed before the epidemic, but that they will fully return to that level by 2024" (Eurostat, 2023f). Therefore, age-and-sex specific mortality rates for 2022 are derived from the averages of the years 2018–2021, for 2024 from 2018 and 2019, and 2023 are average of 2022 and 2024. "The age-specific mortality rates are smoothed using weighted regression B-splines, constrained to allow equal or increasing rates only after the age of 25" (Eurostat, 2023f).

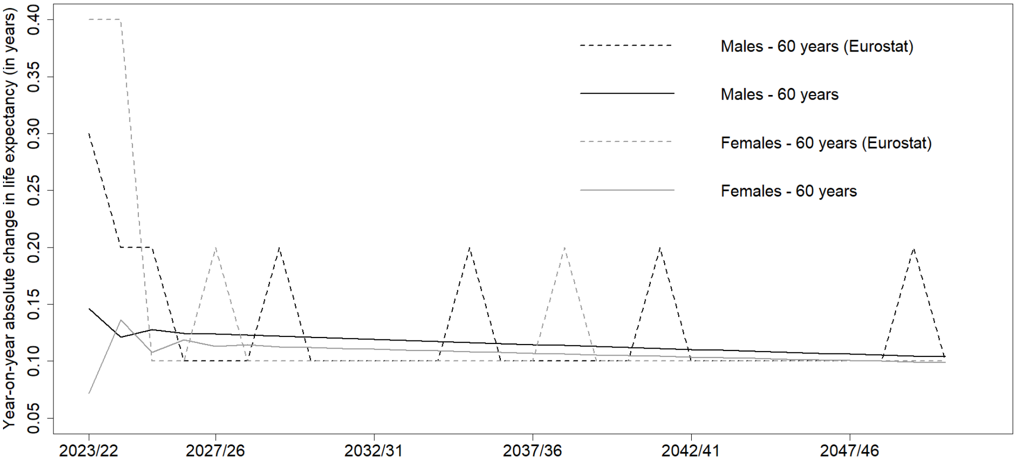
We utilized the data for projected life expectancy by age in completed years from baseline projection. Those data are derived from the period-cohort life table by applying an estimated age-specific gap between these two measures of life expectancy. "The projected life expectancy by age in completed years is provided for the convenience of the users, but it is not the outcome of a regular computation of a life table and it represents only an approximated measure of a life expectancy computed on age-period mortality data" (Eurostat, 2023e).

4.1 Year-on-year absolute changes in life expectancy

The life expectancy at the age of 60 will continue to increase in both examined countries. However, the pace of increase is slowing down. This can be seen from Figure 7 (for Finland) and Figure 8 (for the Czech Republic) where the year-on-year absolute change in life expectancy is displayed. Solid line represents the results of our models, while the dashed line illustrates the calculations of Eurostat (2023) which are rounded on 1 decimal place, so the differences are almost constant in many cases.

Year-on-year absolute change shows that life expectancy of 60 years old males will grow by 0.15 years (53 days) at the beginning, but only by 0.10 years (38 days) at the end of projection period. As the life expectancy of females is already high, the pace of growth will be slower (less than 50 days per year). It can be expected that life expectancy in absolute values will grow by 1 year after 8 years in case of males. It will take 9 years to life expectancy to grow by 1 year in case of females. Both applies to the beginning of the forecasted period. Later, it will take longer. This can have consequences for the pension reform considering that the retirement age is linked to the life expectancy in Finland. The retirement age will increase faster in case of males than females.

Figure 7 Year-on-year absolute change in projected life expectancy (in years) – Finland

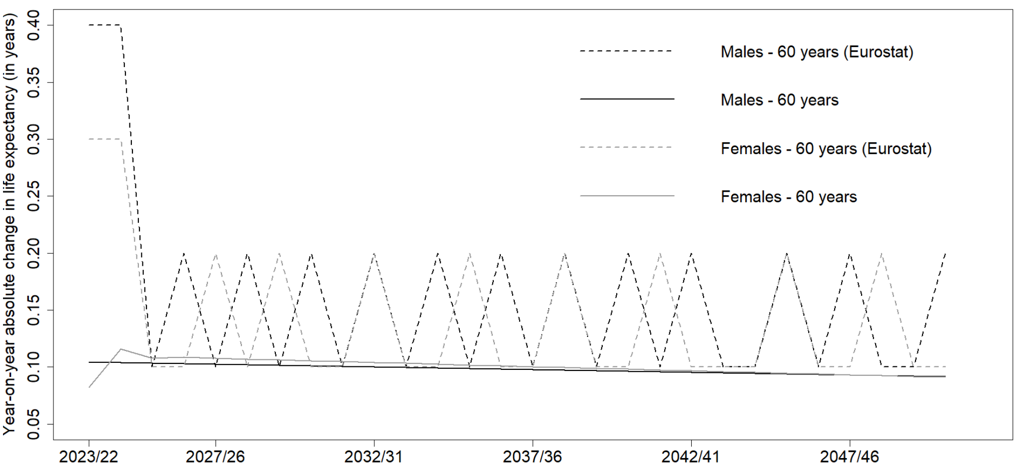


Source: Own elaboration based on the data from Human Mortality Database (HMD, 2023), Eurostat (2023)

Regarding the case of the Czech Republic, the pace of decrease is milder than in Finland. It will take 9.5 years, before the life expectancy of males will increase by 1 year. The pace is almost similar for females.

Comparison of our results with those of Eurostat shows the volatility at the beginning of the period and stability later in population of Finland. Year-on-year changes in Czech population are different every time in case of Eurostat projections.

Figure 8 Year-on-year absolute change in projected life expectancy (in years) – the Czech Republic



Source: Own elaboration based on the data from Human Mortality Database (HMD, 2023), Eurostat (2023)

4.2 Changes in life expectancy

Consequently, we look on the changes in life expectancy in detail. Temporary life expectancies are calculated between years 60 and 70 according to the Formula (9). Therefore, x is equal 60 and i is 10. Then the relative change in time is derived according to Formula (10). t is the starting year of projections 2022. Other years are 2025, 2030, 2035 etc., so n equals firstly to 3 and then to 5. The results for Finnish population are displayed at Table 1.

Temporary life expectancy between 60 and 70 years is according to the expectations higher for females in Finnish population and will increase in time. Every year, some days will be added to current life expectancy. For example, 0.0151 years represents 5.5 days that will live males aged between 60 and 70 longer in 2025 than in 2022. "The concept of relative change in temporary life expectancies refers to the years of life expectancy increase between two ages as a proportion of the maximum possible increase" (Arriaga, 1984). Hence, in relative terms this will mean an increase of life expectancy between 60 and 70 years by 6.88% of the maximum possible increase. Annual relative change suggests that the increase of life expectancy will be the highest between 2022 and 2025 (the index 2.3496 means that the increase will be by 134.96%). It implies that the highest decline in mortality will also occur during this period.

On the other hand, the relative change of life expectancy for females is the lowest in this first period between 2022 and 2025. In other periods, it overcomes the change for males. The highest increase of life expectancy and the highest decrease of mortality for females will occur between 2025 and 2030.

The results for the Czech Republic are displayed at Table 2. Notably, the temporary life expectancy is lower than in Finland. It is only 9.546 years for males in 2022, but according to the projection it will increase in time. Also, the pace of relative change and annual relative change is lower. The increase of life expectancy and decrease of mortality can be expected to be the highest for both sexes at the end of the projected period between 2045 and 2050.

Table 1 Change in life expectancy between 60 and 70 years in Finland

Year	Temp. life exp. ${}_ie_x$ ($i = 10, x = 60$)	Years added (absolute change)	Relative change ${}_iRC_x^n$ ($i = 10, x = 60, n = 3;5$)	Annual relative change ${}_iARC_x^n$ ($i = 10, x = 60, n = 3;5$)
		Males		
2022	9.780	0.0151	0.0688	2.3496
2025	9.796	0.0222	0.1084	2.2681
2030	9.818	0.0198	0.1086	2.2729
2035	9.837	0.0177	0.1087	2.2758
2040	9.855	0.0158	0.1088	2.2784
2045	9.871	0.0141	0.1089	2.2807
2050	9.885			
		Females		
2022	9.884	0.0074	0.0635	2.1622
2025	9.891	0.0125	0.1146	2.4057
2030	9.904	0.0110	0.1143	2.3974
2035	9.915	0.0097	0.1143	2.3991
2040	9.925	0.0086	0.1144	2.4005
2045	9.933	0.0076	0.1145	2.4017
2050	9.941			

Source: Own elaboration based on the data from Human Mortality Database (HMD, 2023)

This positive changes in increase of life expectancy and decrease of mortality shall or rather must be reflected in the reforms of pension systems.

Table 2 Change in life expectancy between 60 and 70 years in the Czech Republic

Year	Temp. life exp. $_ie_x$ (i = 10, x = 60)	Years added (absolute change)	Relative change $_iRC_x^n$ (i = 10, x = 60, n = 3;5)	Annual relative change $_iARC_x^n$ (i = 10, x = 60, n = 3;5)
		Males		
2022	9.546	0.0118	0.0415	1.4017
2025	9.564	0.0297	0.0683	1.4043
2030	9.594	0.0278	0.0684	1.4073
2035	9.622	0.0259	0.0685	1.4101
2040	9.648	0.0242	0.0687	1.4127
2045	9.672	0.0226	0.0688	1.4151
2050	9.695			
		Females		
2022	9.793	0.0106	0.0511	1.7347
2025	9.804	0.0176	0.0900	1.8689
2030	9.822	0.0161	0.0901	1.8703
2035	9.838	0.0146	0.0902	1.8723
2040	9.852	0.0133	0.0903	1.8741
2045	9.866	0.0121	0.0903	1.8757
2050	9.878			

Source: Own elaboration based on the data from Human Mortality Database (HMD, 2023)

4.3 Research limitations and challenges

Our research has limitation that comes from the limitation of the Lee-Carter model and optimal model selection. Age-specific profiles (parameter $\mathbf{a}_x$) are constant in time, which might not be true. Also, parameter $\mathbf{b}_x$ is constant in time.

An influence on results has also the length of the time series (Coale and Kisker, 1986; Šimpach et al., 2014). Also, Booth et al. (2006) who compared projections for 3 time periods and found out that "mean absolute error in log death rates is consistently grater for the long fitting period, while mean error in life expectancy is consistently smallest for the short fitting period."

Unlike in Šimpach et al. (2014) where Hyndman model was an optimal model, it failed in our case. An extension proposed by Li-Lee-Gerland had to be used, because Finland and the Czech Republic have low infant mortality. Besides, there is higher volatility in our data as we use up-to-date time series (Šimpach et al. (2014) used from 1920/1948 to 2012). Especially years 2020 and 2021 brought changes in the long-term development pattern of mortality due to Covid-19 pandemic (Šimpach and Šimpachová Pechrová, 2021).

Our model was probably not able to absorb the shock caused by the Covid-19 pandemic in 2020 and 2021, because after stabilization and the improvement of the situation in 2021, the resulting predictions of demographic indicators were significantly more optimistic in absolute terms than the predictions of Eurostat (Eurostat baseline projections, 2023); despite that year-on-year changes were comparable.

To eliminate the influence of Covid-19 pandemic, multidecrement life tables can be used for determining the change that life expectancy at birth for the theoretical situation when a particular cause of death did not occur. "These tables explain the impact of eradicating a cause of death on life expectancy without determining the effect of changing mortality on life expectancy at each age" (Arriaga, 1984).

We present only one stochastic type of method for life expectancy projection (through the projection of mortality rates and following calculation of life expectancy based on life tables algorithm). Besides, direct projection of empirical life expectancies is possible. For example, Šimpach et al. (2013) used ARIMA(1,0,0) without constant for males and ARIMA(0,2,1) without constant for females to model life expectancy at birth.

There are also deterministic methods based on expert guess, scenarios or regressions that can be used. Therefore, the challenge for future research is to project mortality rates by other methods and life expectancy and to compare the results with the projections done by stochastic Lee-Carter model.

CONCLUSION

The aim of the paper was to describe the future development of life expectancy in the context of pension system reforms that are currently prepared by the politicians in the countries where ageing is the fastest in the EU (Finland and the Czech Republic). One-year age-and-sex-specific mortality rates for both populations aged 0 to 100+ were taken from the Human Mortality Database and they were available in a time series from 1950 to 2021.

Unlike the projections of the statistical institutions such as Eurostat or Czech Statistical office, our paper does not use deterministic models, but rather stochastic ones that are able to capture the influence of the random component as they include error terms.

In contrast to the approach of Šimpach et al. (2013) who used ARIMA models to directly model and predict life expectancy, our stochastic modelling takes into account also mutual interaction among different ages. That means that the evolution of mortality (and life expectancy) does not cross over time. It can happen in case of ARIMA modelling (despite that this was not the case of Šimpach et al., 2013).

Three stochastic models were constructed, calculated in R software and compared. Consequently, mortality rates were projected, and life expectancy calculated up to year 2050. The results are compared to the deterministic projections of the Eurostat.

The most suitable model producing the most feasible results was Lee-Carter model modified by Li-Lee-Gerland because low infant mortality was typical for both populations, so the parameter $\mathbf{b}_x$ had to be rotated. The model projected decrease of mortality rates, which lead to increase of life expectancy. The pace of increase of life expectancy for females was slower in absolute terms, because it was currently already high, so the growth was not that obvious.

Life expectancy year-on-year increments projected by our model up to year 2050 are comparable to Eurostat projections of life expectancy for both populations – Finnish and Czech as we presented on the example of 60-years old population. However, in absolute terms, our stochastic projection gave more optimistic results than the projection of Eurostat.

Consequently, life expectancy between 60 and 70 years was calculated as same as its relative and annual relative changes according to the methodology by Arriaga (1984). According to the results, life expectancy will grow in the future, the pace will be faster in the Czech Republic than in Finland in absolute terms (years added), but the relative change and annual relative change is higher in Finland. Considering that Finnish retirement age is already linked to the life expectancy, it is necessary to re-consider it in the future, if the life expectancy increases too fast, then the retirement age could be too high soon. The Czech Republic retirement age shall also rise, but is not linked to the life expectancy yet. If the growth of it is too fast, then its link to the retirement age might not be optimal.

References

ALVAREZ, J.-A., KALLESTRUP-LAMB, M., KJÆRGAARD S. (2021). Linking retirement age to life expectancy does not lessen the demographic implications of unequal lifespans [online]. *Insurance: Mathematics and Economics*, 99: 363–375. <https://doi.org/10.1016/j.insmatheco.2021.04.010>.

ARRIAGA, E. E. (1984). Measuring and Explaining the Change in Life Expectancies [online]. *Demography*, 21(1): 83–96. <http://www.jstor.org/stable/2061029>.

BOOTH, H., HYNDMAN, R. J., TICKLE, L., DE JONG, P. (2006). Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions [online]. *Demographic Research*, (15): 289–310. <https://doi.org/10.4054/DemRes.2006.15.9>.

BOOTH, H., MAINDONALD, J., SMITH, L. (2002). Applying Lee-Carter under conditions of variable mortality decline [online]. *Population Studies*, 56: 325–336. <https://www.jstor.org/stable/3092985>.

COALE, A. J., KISKER, E. E. (1986). Mortality Crossovers: Reality or Bad Data? [online]. *Population Studies*, 40: 389–401. <https://doi.org/10.1080/0032472031000142316>.

CHIANG, CH. L. (1961). A Stochastic Study of the Life Table and Its Applications. III. The Follow-Up Study with the Consideration of Competing Risks [online]. *Biometrics*, 17(1): 57–78. <http://www.jstor.org/stable/2527496>.

CZECH STATISTICAL OFFICE. (2020). *Mortality tables – Methodology. Czech Statistical Office – Statistics* [online]. <https://www.czso.cz/csu/czso/umrtnostni-tabulky-metodika>.

DANESI, I. L., HABERMAN, S., MILLOSSOVICH, P. (2015). Forecasting mortality in subpopulations using Lee–Carter type models: A comparison [online]. *Insurance: Mathematics and Economics*, 62: 151–161. <https://doi.org/10.1016/j.insmatheco.2015.03.010>.

EC. (2022a). *Employment, Social Affairs & Inclusion: Finland – Old-age pension* [online]. European Commission. <https://ec.europa.eu/social/main.jsp?catId=1109&langId=en&intPageId=4524>.

EC. (2022b). *Employment, Social Affairs & Inclusion: Czech Republic – Old-age pension* [online]. European Commission. <https://ec.europa.eu/social/main.jsp?catId=1106&langId=en&intPageId=4478>.

EUROSTAT. (2020). *Struktura a stárnutí obyvatelstva* [online]. Eurostat Statistics Explained. <https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Population_structure_and_ageing/cs&oldid=510077>.

EUROSTAT. (2023a). *Median age of population, 2012 and 2022 (years)* [online]. Eurostat Statistics Explained. <https://ec.europa.eu/eurostat/statistics-explained/index.php?title=File:Median_age_of_population,_2012_and_2022_(years).png>.

EUROSTAT. (2023b). *Population structure and ageing* [online]. Eurostat Statistics Explained. <https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Population_structure_and_ageing>.

EUROSTAT. (2023c). *Population structure and ageing, Median age is highest in Italy and lowest in Cyprus* [online]. Eurostat Statistics Explained. <https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Population_structure_and_ageing#Median_age_is_highest_in_Italy_and_lowest_in_Cyprus>.

EUROSTAT. (2023d). *Mortality and life expectancy statistics* [online]. Eurostat Statistics Explained. <https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Mortality_and_life_expectancy_statistics&oldid=584067>.

EUROSTAT BASELINE PROJECTIONS. (2023). *Projected life expectancy by age (in completed years), sex and type of projection* [online]. Eurostat Data Browser. <https://ec.europa.eu/eurostat/databrowser/view/PROJ_23NALEXP__custom_5983317/default/table?lang=en>.

EUROSTAT. (2023e). *EUROPOP2023 – Population projections at national level (2022–2100)* [online]. Eurostat metadata. <https://ec.europa.eu/eurostat/cache/metadata/en/proj_23n_esms.htm>.

EUROSTAT. (2023f). *Population projections in the EU – methodology* [online]. Eurostat Statistics Explained. <https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Population_projections_in_the_EU_-_methodology>.

FIALA, T., LANGHAMROVÁ, J. (2014). Increase of Labor Force of Older Age – Challenge for the Czech Republic in Next Decades [online]. *Procedia Economics and Finance*, 12: 144–153. <https://doi.org/10.1016/S2212-5671(14)00330-X>.

FIHW. (2023). *Finnish Institute for Health and Welfare: Ageing policy* [online]. <https://thl.fi/en/web/ageing/ageing-policy>.

GOMPERTZ, B. (1825). On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies [online]. *Philosophical Transactions of the Royal Society of London*, 115: 513–585. <https://www.jstor.org/stable/107756>.

HMD. (2023). *Human Mortality Database* [online]. Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France). <www.mortality.org>.

HOLUB., M., LANGHAMROVÁ, J., VRABCOVÁ, J. (2020). Automatic adjustment of pension age to life expectancy. In: *RELIK 2020: Reproduction of the human capital – mutual links and connections*, 110–128.

HYNDMAN, R. J. (2012): *Demography: Forecasting mortality, fertility, migration and population data* [online]. R package. <https://cran.r-project.org/web/packages/demography/demography.pdf>.

HYNDMAN, R. J. (2022). *Demography: Forecasting mortality, fertility, migration and population data* [online]. <https://cran.r-project.org/web/packages/demography/demography.pdf>.

HYNDMAN, R. J., BOOTH, H., YASMEEN, F. (2012). Coherent mortality forecasting: the product ratio method with functional time series models [online]. *Demography*, 50: 261–283. <https://www.jstor.org/stable/23358841>.

KANNISTO, V., NIEMINEN, M., TURPEINEN, O. (1999). Finnish Life Tables since 1751 [online]. *Demographic research*, 1: 1–27. <https://doi.org/10.4054/DemRes.1999.1.1>.

LAAKSONEN, M., ELOVAINIO, M., KAINULAINEN, S., LEINONEN, T., JAASKELAINEN, T., RISSANEN, H., KOSKINEN, S. (2022). Changes in healthy and unhealthy working life expectancies among older working-age people in Finland, 2000–2017 [online]. *European journal of public health*, 32(5): 729–734. <https://doi.org/10.1093/eurpub/ckac119>.

LANGHAMROVÁ, J. (2014). The use of decomposition methods for the evaluation of the development of mortality in the Czech Republic in 1920–2012. In: *The 8th International Days of Statistics and Economics*, 807–816.

LEE, R. D., CARTER, L. R. (1992). Modeling and Forecasting U. S. Mortality [online]. *Journal of the American Statistical Association*, 87(419): 659–671. <https://doi.org/10.2307/2290201>.

LESLIE, P. H. (1945). On the Use of Matrices in Certain Population Mathematics [online]. *Biometrika*, 33(3): 183–212. <https://doi.org/10.2307/2332297>.

LI, N., LEE, R. D. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method [online]. *Demography*, 42: 575–594. <https://doi.org/10.1353%2Fdem.2005.0021>.

LI, N., LEE, R. D., GERLAND, P. (2013). Extending the Lee-Carter Method to Model the Rotation of Age Patterns of Mortality Decline for Long-Term Projections [online]. *Demography*, 50: 2037–2051. <https://doi.org/10.1007%2Fs13524-013-0232-2>.

MUSILOVÁ, H., REŽŇÁKOVÁ, M. (2015). Age Related Discrimination in the Context of Corporate Social Responsibility and Company Performance, the Case of the Czech Republic [online]. *Procedia Economics and Finance*, 23: 71–76. <https://doi.org/10.1016/S2212-5671(15)00340-8>.

OECD/EUROPEAN OBSERVATORY ON HEALTH SYSTEMS AND POLICIES. (2017). *Finland: Country Health Profile 2017* [online]. State of Health in the EU, Paris: OECD Publishing/Brussels: European Observatory on Health Systems and Policies. <https://doi.org/10.1787/9789264283367-en>.

RUSSOLILLO, M. (2017). Assessing Actuarial Projections Accuracy: Traditional vs. Experimental Strategy [online]. *Open Journal of Statistics*, 7: 608–620. <https://doi.org/10.4236/ojs.2017.74042>.

SYUHADA, K., HAKIM, A. (2021). Stochastic modelling of mortality rates and Mortality-at-Risk forecast by taking conditional heteroscedasticity effect into account [online]. *Heliyon*, 7(10): 1–7. <https://doi.org/10.1016/j.heliyon.2021.e08083>.

ŠEVČÍKOVÁ, H., LI, N., GERLAND, P. (2022). *MortCast: Estimation and Projection of Age-Specific Mortality Rates* [online]. <https://cran.r-project.org/web/packages/MortCast/MortCast.pdf>.

ŠEVČÍKOVÁ, H., LI, N., KANTOROVÁ, V., GERLAND, P., RAFTERY, A. E. (2016). Age-Specific Mortality and Fertility Rates for Probabilistic Population Projections [online]. *Dynamic Demographic Analysis*, 285–310. <https://doi.org/10.48550/arXiv.1503.05215>.

ŠIMPACH, O., DOTLAČILOVÁ, P. (2016). Age-Specific Death Rates Smoothed by the Gompertz–Makeham Function and Their Application in Projections by Lee–Carter Model [online]. In: ROJAS, I., POMARES, H. (eds.) *Time Series Analysis and Forecasting. Contributions to Statistics*, Springer, Cham. <https://doi.org/10.1007/978-3-319-28725-6_18>.

ŠIMPACH, O., DOTLAČILOVÁ, P., LANGHAMROVÁ, J. (2013). Logistic and ARIMA models in the Estimation of Life Expectancy in the Czech Republic. In: *31st International Conference on Mathematical Methods in Economics*, 915–920.

ŠIMPACH, O., DOTLAČILOVÁ, P., LANGHAMROVÁ, J. (2014). Effect of the Length and Stability of the Time Series on the Results of Stochastic Mortality Projection: An application of the Lee-Carter model. In: *1st International Work-Conference on Time Series (ITISE)*, 1375–1386.

ŠIMPACH, O., PECHROVÁ, M. (2013). Assessing the impact of standard of living on the life expectancy at birth using Vector Autoregressive Model. In: *31st International Conference on Mathematical Methods in Economics*, 921–926.

ŠIMPACH, O., ŠIMPACHOVÁ PECHROVÁ, M. (2021). Implications of the SARS-Cov-2 Pandemic for Mortality Forecasting: Case Study for the Czech Republic and Spain [online]. *Engineering Proceedings*, 5(1): 1–10. <https://doi.org/10.3390/engproc2021005058>.

THATCHER, R. A., KANNISTO, V., VAUPEL, J. W. (1998). *Force of Mortality at Ages 80 to 120.* Odense University Press, 104 p.

Payout Phase of Defined Contribution Systems: the Case of Slovakia

Igor Melicherčík[1] | *Comenius University, Bratislava, Slovakia*
Gábor Szűcs[2] | *Comenius University, Bratislava, Slovakia*

Received 8.3.2023 (revision received 30.6.2023), **Accepted** (reviewed) 4.7.2023, **Published** 15.12.2023

Abstract

The paper aims to assess various aspects concerning the payment phase of the old-age pension scheme, the so-called second pillar of the pension system in Slovakia. However, the conclusions may also be useful for other pension systems. Using the Lee-Carter model and standard actuarial methods, we conclude that the second pillar is advantageous for the high-income groups or in case of high performance of pension funds. We also address the issue of deferring the purchase of a lifetime annuity. Deferral can be beneficial when the yield of the pension fund exceeds a certain threshold value. This threshold usually raises with increasing age. We argue that the temporary pension is a disadvantageous product and its recent cancellation is correct. The main contribution of the paper subsists in a three-state model of long-term care insurance, using which we calculate corresponding replacement rates. Combined with a lifetime annuity, long-term care insurance can be beneficial.

Keywords	***DOI***	***JEL code***
Pension system in Slovakia, Svensson yield curve, Lee-Carter model, lifetime annuity, deferred annuity, long-term care insurance	*https://doi.org/10.54694/stat.2023.12*	*G22, H75, I13, J11, J26*

INTRODUCTION

Since January 2005, pensions in Slovakia are operated by a three-pillar system: the compulsory, Pay-As-You-Go (PAYG) first pillar, the second pillar in the form of old-age pension saving, and the third pillar as a voluntary supplementary pension saving. Only the first pillar is compulsory and future pensioners can redirect part of their contributions to the saving (second) pillar. In this case, the first pillar pension is reduced accordingly. This reduced pension can be supplemented by benefits from savings of the second pillar. Savings in the second pillar and the subsequent payment phase are governed by the Act of the National Council of the Slovak Republic No. 43/2004 Coll. on the old-age pension scheme, as amended ("Act 43/2004 Coll.").

Several publications addressed the question of whether participation in the second pillar is beneficial. The answer is not uniform and depends on a specific wage profile. For low-income groups, currently

[1] Faculty of Mathematics, Physics and Informatics, Comenius University, Mlynská dolina, 842 48 Bratislava, Slovakia. Corresponding author: e-mail: igor.melichercik@fmph.uniba.sk, phone: (+421)260295477.
[2] Faculty of Mathematics, Physics and Informatics, Comenius University, Mlynská dolina, 842 48 Bratislava, Slovakia.

valid Act 43/2004 Coll. only offers the possibility of paying benefits in the form of a lifetime annuity. According to Melicherčík et al. (2015), lifetime annuity benefits may not cover (even in the case of zero fees of insurance companies) the reduction of the first pillar pension. For higher-income savers, the probability of covering the reduction is higher due to the partial solidarity of the first pillar. In Gubalová et al. (2022), a calculation of the Global Pension Index for Slovakia is presented. The authors point to the problem that very low pensions for low-income groups negatively affect the value of the pension index of Slovakia. This is a serious issue, which is mainly caused by the inappropriate setting of saving strategies in the funded pillars of the Slovak pension system and the ways of receiving pension benefits from the second and third pillars. In Špirková et al. (2021), authors offer a critical view of pension savings in the Slovak pension system. Within several case studies, they analyze risk factors that may affect the amount of future pension benefits and discuss the importance of choosing a suitable payout product.

However, currently valid Act 43/2004 Coll. provides more possibilities for using pension savings. If the sum of pension benefits paid from other sources reaches a minimum reference amount (currently an average old-age pension), the pensioner may apply for a programmed withdrawal or a temporary pension. The programmed withdrawal also includes the possibility to withdraw the entire saved amount at once. This greatly expands the possibilities of using pension savings.

Recently, the first offers of life insurance companies to pay lifetime benefits from the second pillar have appeared. The level of benefits was low mainly due to the risk of longevity and low interest rates. On the other hand, many authors (see e.g., Milevsky, 1998; Šebo and Šebová, 2016) argued that the immediate purchase of a lifetime annuity is not an optimal use of pension savings. The Slovak government, therefore, came up with the amendment to Act 43/2004 Coll., where the payment phase is implemented in the form of a specific programmed withdrawal and a deferred purchase of a lifetime annuity. Temporary pensions and all other forms of programmed withdrawal are cancelled. One of the contributions of this paper is a discussion, of whether such a change is beneficial for pensioners.

Pension benefits should be set to meet the needs of pensioners. Due to health problems, long-term care is often necessary during retirement. The extent of care depends on the specific disability and can also be full-time care. Pensioners typically do not have the necessary resources to finance it. The main contribution of this paper is a model focused on using pension savings to purchase long-term care insurance. We have calculated the resulting replacement rates that can be achieved using realistic savings levels.

To summarize, the Slovak pension system after the reform in 2005 combines several principles. The PAYG first pillar is supplemented by a defined-contribution second pillar. The payment phase of the second pillar offers several options. Therefore, even though the article is based on the reality of the pension system in Slovakia, several issues that are also interesting for other pension schemes are discussed. The aim of the paper is to thoroughly evaluate the payment phase of the second pillar of the pension system in Slovakia. We will also discuss the appropriateness of the latest legislative changes in Act 43/2004. Based on the above, we set the following hypotheses:

H1: The benefits from the second pillar are not sufficient to compensate the corresponding reduction of the first pillar.

H2: Immediate purchase of a lifetime annuity is not an optimal use of pension savings.

H3: Temporary pension is an unsuitable pension product.

H4: Long-term care insurance is a reasonable pension product that could suitably expand the current options in the payout phase of the second pillar.

When discussing H1-H4, the Lee-Carter stochastic demographic mortality model (Lee and Carter, 1992) and the Svensson model (Svensson, 1994) are used. To convert savings into replacement rates, we use the approach of Špirková et al. (2019). Regarding deferred annuities, we follow the results of Milevsky (1998). However, we extend these calculations by considering the more complicated structure of the fees associated with annuities and the Svensson yield curve for discounting. Our three-state model is used for calculations related to long-term care insurance.

The paper is organized as follows. In the first Section, we assess the level of pension benefits when buying lifetime annuities. In the second Section, we first deal with programmed withdrawal and temporary pensions. Then, we discuss in detail the deferral of the purchase of lifetime annuities and the convenience of the latest changes in Act 43/2004 Coll. regarding the payment phase of the second pillar. The third Section deals with long-term care insurance. In the final part we conclude.

1 EXPECTED LEVEL OF SAVINGS AND LIFETIME ANNUITIES

The mandatory part of the pension system in Slovakia has two pillars: the public, compulsory, non-funded (Pay-As-You-Go) first pillar, and the private, fully funded second pillar. The contribution rate (for the old-age pensions) is currently set at 18% for the first pillar (in the case a pensioner decides to stay only in the public scheme only) or 12.5% for the first pillar and 5.5% for the second pillar (in the case a pensioner decides to participate in both pillars) with a future planned increase to 6%.

The adequacy of pension savings can be assessed in several ways. In Kilianová et al. (2006) authors introduced a retirement-years indicator (D_T). It was calculated as the ratio of the sum S_T saved at the time of retirement T and the last yearly wage W_T before retirement: $D_T = \frac{S_T}{W_T}$. This indicator can be easily recalculated to the replacement rate (the ratio of the first pension to W_T, cf. Melicherčík et al., 2015).

1.1 Survival and mortality probabilities

In order to recalculate the savings to the replacement rate, we have used an approach of Špirková et al. (2019). Let us assume that all ages are expressed in years. Pricing of annuity products has been based on the relevant survival probabilities ${}_tp_x$ representing the probability that an individual at age x years survives at least until the age of $x + t$ years, where $t \in \mathbb{N}$. Denote by q_x the probability that an individual being at age x years dies before age $x + 1$ years. Then ${}_tp_x = \prod_{h=0}^{t-1}(1 - q_{x+h})$. In our practical calculations we have used three sets of mortality rates (for $x \in \{62, 63, \dots, \omega\}$, where $\omega = 105$ years denotes the maximum age):

- $q_x^{(S)}$ representing the static mortality rates from the Mortality Tables of the SO SR (SO SR, 2022) for the total Slovak population in the year 2018,
- $q_x^{(H)}$ denoting the predicted mortality rates using the Lee-Carter longevity model for the future period from 2020 to 2063,
- $q_x^{(L)}$ representing the estimated mortality rates from the lower bound of the 99% prediction interval for the aforementioned Lee-Carter predictions; in our calculations we have applied them as components of the pessimistic (in terms of insurance) longevity model (see also Špirková et al., 2019).

For this mortality rates applies: $q_x^{(S)} \geq q_x^{(H)} \geq q_x^{(L)}$, $\forall x \in \{62, 63, \dots, 105\}$.

The Lee-Carter stochastic demographic mortality model is a statistical model that is widely used in the fields of demography, actuarial science, and old-age pension modelling to analyse and forecast mortality rates. The model was developed in 1992 by Ronald D. Lee and Lawrence R. Carter (Lee and Carter, 1992) and has since become one of the most popular and well-regarded mortality models. The Lee-Carter model is based on the idea that mortality rates can be expressed as a function of age and a time trend, which captures the overall pattern of mortality improvement over time. Specifically, the model assumes that the logarithm of the age-specific mortality rates follows a linear trend with age, with the slope of the trend varying over time according to a stochastic process. In our case, we have estimated the parameters of the Lee-Carter model based on data from the Human Mortality Database (HMD, 2022) for the total Slovak population from 1990 to 2019 using the 'demography' package (Hyndman, 2023) of the statistical software R (R Core Team, 2022). Subsequently, we have predicted the mortality rates for the future period of 2020–2063 for persons aged 62 to 105 years using the R package ‘forecast’ (Hyndman and Khandakar, 2008; Hyndman et al., 2023). The Lee-Carter longevity model, i.e., vector of mortality rates $q_x^{(H)}$ for $x \in \{62, 63, \ldots, 105\}$, has been constructed by selecting the diagonal elements of the matrix of predicted mortality rates. The pessimistic longevity model has been built similarly, but when constructing the vector $q_x^{(L)}$, $x \in \{62, 63, \ldots, 105\}$, we have used the diagonal elements of the matrix containing the lower bound values of 99% prediction intervals.

1.2 Svensson yield curve

In accordance with ECB (2022b), we have used the Svensson yield curve as a functional form for the spot interest rates depending on corresponding maturities. The Svensson yield curve is given by (Svensson, 1994):

$$R(t) = \beta_0 + \beta_1 \frac{1-\exp\left(-\frac{t}{\tau_1}\right)}{\frac{t}{\tau_1}} + \beta_2 \left[\frac{1-\exp\left(-\frac{t}{\tau_1}\right)}{\frac{t}{\tau_1}} - \exp\left(-\frac{t}{\tau_1}\right)\right] + \beta_3 \left[\frac{1-\exp\left(-\frac{t}{\tau_2}\right)}{\frac{t}{\tau_2}} - \exp\left(-\frac{t}{\tau_2}\right)\right], \tag{1}$$

where $R(t)$ is a yield from a bond investment with continuous compounding, t is a time to maturity, $t \in (0, T_{\max}]$, $T_{\max}$ is the maximum time to maturity, τ_1, τ_2, β_0, β_1, β_2, β_3 are parameters of the Svensson yield curve. The discounting factor corresponding to the maturity t is then given by $P(t) = e^{-R(t)t}$. The interpretation of the parameters is available e.g., in Aljinović et al. (2012): β_0 is the long-term asymptotic value of $R(t)$, β_1 is the spread between the long-term and short-term rates, i.e., $\beta_0 + \beta_1$ is the short-term rate (the rate corresponding to zero maturity). The parameters τ_1, τ_2, β_2, β_3 specify the positions, magnitudes and directions of two humps corresponding to the Svensson curve.

1.3 Pension annuity product

Consider a person with a retirement age x years having saved the amount S_T. The basic equivalence equation represents the expected present values of all cash-flows related to the yearly annuity payment P_x:

$$S_T = P_x a_x (1 + \beta) + P_x \alpha. \tag{2}$$

On the left-hand side stands the accumulated sum S_T representing a premium of the product. The value $P_x a_x$ is the expected present value of the whole life yearly paid annuity-immediate P_x, where:

$$a_x = \sum_{t=1}^{\omega-x} {}_t p_x \, P(t)$$

denotes the expected present value of a whole life 1 monetary unit (m. u.), paid at the end of each year under the condition that the person is alive and ω is the maximum age (regarding used life tables $\omega = 105$ years). According to the current version of Act 43/2004 Coll., the monthly benefits P_x are paid in the first seven years of the retirement period regardless of whether the beneficiary is alive. Therefore, we set ${}_t p_x = 1$ for $t = 1, 2, \ldots, 7$. Finally, α and β represent fees charged to the first and following annuity payments. Denote by

$$\tilde{a}_x = a_x(1 + \beta) + \alpha,$$

the value of the whole life 1 m. u. including fees. Dividing both sides of (2) by W_T and making some minor adjustments one has

$$\mathrm{RE}_x = \frac{P_x}{W_T} = \frac{D_T}{\tilde{a}_x}, \tag{3}$$

where RE_x is the replacement rate (the ratio of the yearly pension to the last yearly salary before retirement).

Table 1 contains replacement rates for different levels of savings and retirement ages calculated using the Svensson ECB all bonds curve (the parameters have been estimated in ECB (2022a) as of 9 June 2022) and fees $\alpha = 50\%$, $\beta = 8\%$. We have applied probabilities of death $q_x^{(H)}$ calculated using the Lee-Carter longevity model (see Section 1.1). In Table 2 the values of $\tilde{a}_x$ for various ages together with corresponding conditional life expectancies (LE_x) are presented. One can observe that the $\tilde{a}_{62}$ is slightly less than LE_{62}, which is a consequence of the appreciation of savings using interest rates. However, due to the 7-year warranty and transaction costs (especially, in the case of shorter LE_x, the charged α plays a more significant role), with increasing age, the price of $\tilde{a}_x$ becomes significantly higher than the expected lifetime LE_x. It is worth noting, that Act 43/2004 Coll. has been recently amended and the payout phase will (from 1 January 2024) have a different form. This will be in detail discussed in Section 2.

Table 1 Replacement rates of lifetime annuity payments for various levels of savings and initial ages of the pensioner (x in years) using mortality rates $q_x^{(H)}$ with 7-year guarantee applied

D_T / age x	62	65	70	75	80
1.5	0.0782	0.0830	0.0932	0.1063	0.1210
2.0	0.1043	0.1108	0.1243	0.1417	0.1613
2.5	0.1303	0.1385	0.1553	0.1772	0.2017
3.0	0.1564	0.1661	0.1864	0.2126	0.2420
3.5	0.1825	0.1938	0.2175	0.2480	0.2823
4.0	0.2085	0.2215	0.2485	0.2835	0.3226

Source: Own construction

Table 2 Values of $\tilde{a}_x$ for different ages x (in years) using mortality rates $q_x^{(H)}$ with 7-year guarantee applied together with corresponding conditional life expectancies at age x (LE_x in years) in the total Slovak population in 2019 according to HMD (2022)

x	62	65	70	75	80
$\tilde{a}_x$	19.18	18.06	16.09	14.11	12.40
LE_x	19.97	17.78	14.28	11.09	8.21

Source: Own construction according to HMD (2022)

To illustrate the calculated levels of replacement rates, let us consider a person contributing to the second pillar at 6% of the gross wage. Following recent legislative changes, the total contribution rate used for the reduction of the first pillar pension was increased from 18% to 22.75%.[3] Therefore, this saver will receive 16.75/22.75 of the pension from the first pillar designed for 50% replacement rate.[4] As a result, the required compensation from the savings of the second pillar is roughly $13.2\% \approx \frac{6}{22.75} \times 50\%$. Using the results from Table 1 one can conclude, that achieving such a replacement rate requires a level of savings of approximately 2.5 of yearly salaries. In Melicherčík et al. (2015) authors reported a realistic level

3 The first pillar pension is reduced by the part δ / 22.75 for the period of participation in the second pillar, where δ is the contribution rate (in percentage) to the second pillar. The original reduction ratio was δ / 18.

4 According to the amendment to Act of the National Council of the Slovak Republic No. 461/2003, the newly granted pensions will not be increased by the entire increase of wages (only by 95% of the increase of wages). The replacement rate of 50% will therefore be reduced in the future.

of savings D_T in the second pillar after 40 years of saving between 1.5 and 4 yearly salaries. Compensation for the reduction of the first pillar by savings from the second pillar is therefore questionable.

However, this calculation only applies to certain income groups. The amount of the starting pension is linear with respect to a quantity called the Average Personal Wage Point (APWP). This quantity is the average of the ratios of personal wages and average wages in the national economy for a working career.[5] To calculate the pension, however, the APWP is adjusted according to Table 3. The adjusted APWP is denoted by APWP*. The above calculation is valid for APWP ∈ [1, 1.25). For other values of APWP, the required compensation of the first pillar benefits can be estimated as:

$$\frac{6.00}{22.75} \times \frac{\text{APWP}^*}{\text{APWP}} \times 50\%. \tag{4}$$

Table 3 APWP and adjusted APWP (APWP*)

APWP	APWP*
APWP < 1	APWP + (1 – APWP) × 0.2
APWP ∈ [1, 1.25]	APWP
APWP > 1.25	min(1.25 + (APWP – 1.25) × 0.68 ,3)

Source: Own construction

The values of the required compensations according to (4) for selected values of the APWP are presented in Table 4. One can observe that the second pillar is beneficial for high-income groups, for which the required compensation rates are lower.

Table 4 Required compensations (in %) of the first pillar benefits

APWP	0.50	0.75	1.00	1.25	1.50	2.00	3.00	5.00	7.00
APWP*	0.60	0.80	1.00	1.25	1.42	1.76	2.44	3.00	3.00
Compensation	15.82	14.07	13.19	13.19	12.48	11.60	10.73	7.91	5.65

Source: Own construction

2 OTHER LEGAL POSSIBILITIES OF USING PENSION SAVINGS

2.1 Temporary pension and programmed withdrawal

According to current legislation, a saver is entitled to a programmed withdrawal or a temporary pension if the sum of pension benefits paid from other sources reaches a minimum reference amount (currently an average old-age pension). A temporary pension is an insurance product paid by insurance companies. The length of the contract can be 5, 7 or 10 years. Unlike the second pillar lifetime pension, the temporary pension does not include the insurer's obligation to pay 7 years of pension benefits. At the event of the beneficiary's death, the payment of benefits shall cease. In addition, a temporary pension does not insure longevity. The product can be valued using Formula (2) while omitting the condition ${}_tp_x = 1$ for $t = 1, 2, \ldots, 7$. To highlight the disadvantage of the temporary pension, we have calculated the results using fees $\beta = 8\%$ (the same as in the case of a lifetime annuity) and $\alpha = 0\%$ (50% for the lifetime annuity). Values $\tilde{a}_x$ used for the calculation of the replacement rates for temporary pensions can be found in Table 5. One can observe that the values $\tilde{a}_x$ are close to the lengths of the pensions. Therefore, it is questionable

[5] For details, see Act of the National Council of the Slovak Republic No. 461/2003, as amended (Articles 62 and 63).

(especially for lower ages and pension lengths) whether there is a big difference between the temporary pension and the withdrawal of the full amount (which is a legal form of the programmed withdrawal) with gradual spending without any institutional assistance. A motivational difference can be observed only for higher ages or lengths of temporary pensions. On the other hand, in this case, the risk of death before the end of the planned period is substantial. To conclude, a temporary pension is probably not a good alternative for using pension savings. The main reasons are low interest rates and low probabilities of death. The cancellation of this option in the current amendment to the law is therefore correct.

Table 5 Values of $\ddot{a}_x$ used for calculation of the replacement rates for temporary pensions using mortality rates $q_x^{(H)}$ for various ages of the pensioners (x in years) and durations of pensions

Age x / duration (in years)	5	7	10
62	4.9741	6.7022	9.0094
65	4.9392	6.6379	8.8837
70	4.8617	6.4923	8.5946
75	4.7361	6.2501	8.0914
80	4.4725	5.7547	7.1639

Source: Own construction

In contrast, the programmed withdrawal is not an insurance product. When using this method of payment, the savings remain in the pension management company (PMC), with which the pensioner concludes a retirement benefit plan. Under this agreement, the PMC will pay a pension from a personal account under pre-agreed terms. The beneficiary determines the monthly amount and the length of the retirement benefits. At the event of death, the remaining funds are subject to inheritance. It is worth noting that by using the programmed withdrawal, one can avoid the annuity fees. An interesting set of dynamic and static strategies of programmed withdrawal can be found in Šebo and Šebová (2016). The authors also calculated the expected value of the bequest corresponding to the respective strategies. Most of the strategies presented avoided ruin and can be considered a more effective alternative compared to lifetime annuities. However, according to the latest legislative changes, the programmed withdrawal is subject to income tax, which partially disadvantages this form of using pension savings.

2.2 Deferred purchase of an annuity

The legislation does not require the purchase of a lifetime annuity from the second pillar savings even in the case of receiving a pension from the first pillar. There is no reason to rush to buy an annuity when the pensioner's income is sufficient. Such a situation occurs, e.g., when the beneficiary is working after retirement age. In such a case, it may be advantageous to defer the purchase of an annuity, or not to buy an annuity at all, and to use the savings later for a more reasonable purpose or to leave them as a bequest. Paragraph 46i of Act 43/2004 Coll. gives a possibility to apply for the return on investment payment if the saver has reached the retirement age, while not being the recipient of the retirement pension or early retirement pension by programmed withdrawal. Note that the return on investment is not a retirement pension. Therefore, savings remain the property of the saver. It is worth noting that in some years the return on investment can be negative and therefore zero benefits may be paid.

A rational question is whether it is worth delaying the purchase of a lifetime annuity. In IFP (2022) authors discussed a combination of a programmed withdrawal and buying the lifetime annuity. They considered the programmed withdrawal with monthly payments one would receive when buying an annuity at retirement. While monthly benefits were paid, the remaining savings were invested. After 10 years, the lifetime annuity was purchased from the rest of the savings. The authors reported, that the resulting monthly lifetime benefit was with probability 93% higher than that resulting from the lifetime annuity purchased at retirement. A thorough analysis of strategies using deferred lifetime annuities considering stochastic interest rates and mortality rates can be found in Milevsky (1998).

Assume that a retiree at age x years has an amount $\tilde{a}_x$ to purchase a unit lifetime annuity. However, the purchase of an annuity may be deferred by one year and the funds $\tilde{a}_x$ may be invested instead. Assuming an annual return of r, such a strategy is advantageous if:

$$\tilde{a}_x(1 + r) \geq \tilde{a}_{x+1} + 1. \tag{5}$$

From (5), a minimum yield r_x^* can be derived to make this strategy beneficial:

$$r_x^* = \frac{\tilde{a}_{x+1}}{\tilde{a}_x} + \frac{1}{\tilde{a}_x} - 1. \tag{6}$$

Compared to the Milevsky (1998) approach, we consider in addition the transaction cost α associated with the first payment and different interest rates for different maturities according to the Svensson curve. For a deeper insight into Formula (6), let's write $\tilde{a}_x$ in a simpler form

$$\tilde{a}_x = (1 + \beta)\sum_{t=1}^{\omega-x} \frac{{}_tp_x}{(1 + y)^t} + \alpha,$$

where y is a yield to maturity. Supposing that the yield y is the same for x and $x + 1$ and using a natural actuarial identity:

$${}_tp_{x+1} = \frac{{}_{t+1}p_x}{{}_1p_x},$$

Formula (6) takes the form

$$r_x^* = \frac{1 + y}{{}_1p_x} - 1 - \frac{\beta}{\tilde{a}_x} - \frac{\alpha}{\tilde{a}_x}\left(\frac{1 + y}{{}_1p_x} - 1\right). \tag{7}$$

The first term on the right-hand side of (7) is increasing with respect to x. On the other hand, the terms with fees α and β are decreasing with respect to x. Thus, for zero fees, the minimum yield r_x^* increases with age and sooner or later the purchase of the lifetime annuity is advantageous. On the other hand, high fees can increase the optimal purchase age.

The minimum values of r_x^* calculated according to Formula (6) for selected ages are in Table 6. The results e.g., show that if the deferral of the lifetime annuity is to be beneficial for 10 years, a yield of around 4% is required. The calculations do not assume a 7-years payment guarantee as in Section 1.3.

Table 6 Minimum values of r_x^* for various ages (in years) using mortality rates $q_x^{(H)}$

x	r_x^*	x	r_x^*	x	r_x^*	x	r_x^*
62	0.0333	72	0.0399	82	0.0649	92	0.1452
64	0.0344	74	0.0421	84	0.0775	94	0.1751
66	0.0354	76	0.0449	86	0.0922	96	0.2188
68	0.0366	78	0.0490	88	0.1076	98	0.2805
70	0.0380	80	0.0553	90	0.1243	100	0.3621

Source: Own construction

Such a guarantee affects too much the price of the annuity and does not make sense for higher ages. Therefore, we omit it throughout this section. Results presented in Table 6 confirm that the minimum yield r_x^* is increasing with x.

The current amendment to Act 43/2004 Coll. assumes that it is advantageous to postpone the purchase of the lifetime annuity.[6] According to this change, the saved funds will be divided into two equal parts at the start of the payout phase. Define D_x as half of the median life expectancy common to men and women of the saver age. Half of the savings are then paid evenly on monthly basis during the period D_x (this is a specific form of programmed withdrawal, which is not taxed). The other half of the savings (invested in the meantime) will be used to buy a life annuity after the D_x period. Denote by M the amount saved at the retirement age. Suppose that one half of the savings were invested with a yield y. To continue with at least the same payments after period D_x, one has the inequality:

$$\frac{M}{2}\frac{(1+y)^{D_x}}{\tilde{a}_{x+D_x}} \geq \frac{M}{2D_x},$$

hence:

$$\tilde{a}_{x+D_x} \leq D_x(1+y)^{D_x}. \tag{8}$$

The values of D_x, $\tilde{a}_{x+D_x}$, and threshold yields for ages over 60 years are shown in Table 7. One can observe that e.g. for the beginning of using funds from the second pillar at the age of 64 years, a yield of 2.83% is required to continue with the same benefits after the period D_x. In recent financial market conditions, this is realistic, but not certain. Increasing y^* with increasing age x is consistent with the results of Table 6. Moreover, a programmed withdrawal during the D_x period will prevent loss of potential inheritance at the event of death. Finally, let's recall the fact that fixed transaction costs α have a greater impact on annuity prices at older ages. To summarize, the combination of a programmed withdrawal with the subsequent purchase of a lifetime annuity provides an interesting solution, balancing the reduction of the risk of losing funds at the event of earlier death with the potential problem of a reduction in benefits after switching to a lifetime annuity.

Table 7 The estimated values of D_x (in years) using the Mortality Tables of the SO SR (SO SR, 2022) for the total Slovak population in the year 2019, values of $\tilde{a}_{x+D_x}$ and threshold yields y^* for ages over 60 years, using mortality rates $q_x^{(H)}$ and no guarantee applied

Age x (in years)	60	62	64	66	68	70
D_x	11.25	10.40	9.57	8.75	7.97	7.19
$\tilde{a}_{x+D_x}$	13.59	13.06	12.50	11.93	11.34	10.74
y^*	0.01696	0.02211	0.02831	0.03610	0.04523	0.05739

Source: Own construction

3 LONG-TERM CARE INSURANCE

At present, the need for a long-term care in the case of dependency is a common problem. Most pensioners do not have the means to cover the associated costs. In this section, we present our model of long-term care insurance as well as the replacement rates it could provide. We have supposed that a person can be in one of the following three states:

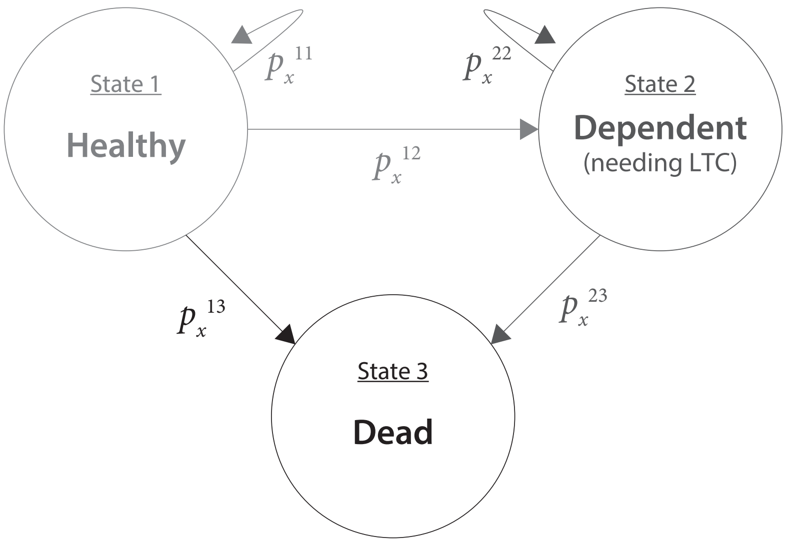
1. Healthy,

[6] It is in accord with already mentioned results of IFP (2022).

2. Dependent (needing a long-term care),
3. Dead.

Furthermore, we have assumed that a healthy person can become dependent or die and a dependent person cannot become healthy. A graphical representation of the model is shown in Figure 1.

Figure 1 Graphical representation of the long-term care (LTC) insurance model



Source: Own construction

Table 8 Shares requiring long-term care (LTC) in the population of Belgium (2007)

Age cohort (in years)	% of LTC	Age cohort (in years)	% of LTC
50–54	4.7465	70–74	15.1312
55–59	5.8612	75–79	20.5790
60–64	8.3343	80–84	36.6185
65–69	8.5914	85+	72.2630

Source: Willemé (2010)

Since realistic data of number of persons requiring long-term care in Slovakia are not available, we have used the data from Belgium (Willemé, 2010). The shares requiring long-term care in Belgium are in Table 8. In the first step, we have interpolated these data using weighted averages, with weights corresponding to population sizes at ages 65, 66, …, 105 years in the total Slovak population; see Mortality Tables of the SO SR for 2018 (SO SR, 2022). In the second step, we have fitted the interpolated values by a polynomial-exponential function. The resulting estimates of shares y_x requiring long-term care for various age cohorts x are available in Table 9.

Let us denote the probabilities of remaining in the corresponding states, respectively transitions between the states as follows:

- p_x^{ii} – the probability that an individual being at age x years in state $i \in \{1, 2\}$ remains in this state at least to age $x + 1$ years,

- ${}_{m}p_x^{ii}$ – the probability that an individual being at age x years in state $i \in \{1, 2\}$ remains in this state at least to age $x + m$ years, where $m \in \mathbb{N}$ is a multiple of year,
- p_x^{ij} – the probability that an individual being at age x years in state $i \in \{1, 2\}$ transits to state $j \in \{2, 3\}$, $j > i$ before age $x + 1$ years.

Table 9 Shares requiring long-term care (γ_x) for various ages x (in years) – our estimates

x	γ_x	x	γ_x	x	γ_x	x	γ_x
62	0.0833	73	0.1489	84	0.5005	95	0.8968
63	0.0840	74	0.1691	85	0.5549	96	0.9288
64	0.0846	75	0.1913	86	0.6049	97	0.9440
65	0.0852	76	0.2151	87	0.6477	98	0.9627
66	0.0859	77	0.2401	88	0.6890	99	0.9757
67	0.0865	78	0.2661	89	0.7239	100	0.9821
68	0.0888	79	0.2958	90	0.7623	101	0.9848
69	0.0944	80	0.3400	91	0.7911	102	0.9869
70	0.1034	81	0.3775	92	0.8204	103	0.9892
71	0.1157	82	0.4122	93	0.8492	104	0.9920
72	0.1310	83	0.4500	94	0.8693	105	0.9920

Source: Own construction

The following equations apply to the mentioned probabilities:

$$p_x^{11} + p_x^{12} + p_x^{13} = 1, \tag{9}$$

$$p_x^{22} + p_x^{23} = 1. \tag{10}$$

By shifting individuals of age x years and balancing the number of dependent ones we have:

$$(1 - \gamma_x)p_x^{12} - \gamma_x p_x^{23} = \gamma_{x+1}(1 - \bar{q}_x) - \gamma_x, \tag{11}$$

$$\bar{q}_x = (1 - \gamma_x)p_x^{13} + \gamma_x p_x^{23}. \tag{12}$$

Formulas (9)–(12) are not sufficient to determine all the necessary probabilities p_x^{ij}. Missing equations can be replaced by defining the relationship between the mortality probabilities p_x^{13} and p_x^{23}. One has more options for this definition, e.g.

1. $p_x^{13} = p_x^{23} = q_x^{(H)}$, the case of mortality rates using Lee-Carter longevity model,
2. $p_x^{13} = p_x^{23} = q_x^{(L)}$, the case of pessimistic longevity (in terms of insurance),
3. $p_x^{13} = q_x^{(L)}$ with additional equation $q_x^{(S)} = (1 - \gamma_x)\, p_x^{13} + \gamma_x p_x^{23}$ setting the whole-population mortality rate at the higher level; in this case the dependents (persons dependent on the long-term care) have lower life expectancy comparing to the healthy population,
4. $p_x^{13} = q_x^{(L)}$ and $p_x^{23} = \kappa \dfrac{q_x^{(S)} - (1 - \gamma_x)p_x^{13}}{\gamma_x}$, where κ is a constant factor such that the dependents have the half life expectancy comparing to the healthy population,

5. $p_x^{13} = q_x^{(S)}$ and $p_x^{23} = \lambda \dfrac{q_x^{(S)} - (1 - \gamma_x)p_x^{13}}{\gamma_x} = \lambda p_x^{13}$, where $\lambda > 1$ is a constant factor such that the dependents have the half life expectancy comparing to the healthy population.

Combining Formulas (9)–(12) with any of these options one can determine all the necessary probabilities p_x^{ij}, $i, j \in \{1, 2, 3\}$, $i \leq j$.

Denote by Z_x a random variable representing the time (in years) in which a healthy person of age x years switches to state 2 (long-term care dependency). It is obvious, that $\Pr(Z_x = 1) = p_x^{12}$. The probabilities for further times $k > 1$, where $k \in \mathbb{N}$, can be calculated as

$$\Pr(Z_x = k) = p_x^{11} \times p_{x+1}^{11} \times \ldots \times p_{x+k-2}^{11} \times p_{x+k-1}^{12}. \tag{13}$$

Suppose that a person of age x years becomes dependent after m years. The expected present value of long-term care lifetime benefits L_x is $L_x \times {}_{m|}a_x^L$, where:

$$_{m|}a_x^L = P(m+1) + \sum_{t=m+1}^{\omega-x-1} P(t+1) \times {}_{t-m}p_{x+m}^{22}. \tag{14}$$

The probabilities $_kp_y^{22}$ that the person of age y years remains in state 2 at least next $k \in \mathbb{N}$ years are calculated as follows:

$$_kp_y^{22} = p_y^{22} \times p_{y+1}^{22} \times p_{y+2}^{22} \times \ldots \times p_{y+k-1}^{22}. \tag{15}$$

The expected present value of unit long-term care benefits is then:

$$\mathrm{E}\left[{}_{Z_x|}a_x^L\right] = \sum_{m=1}^{\omega-x-1} {}_{m|}a_x^L \times \Pr(Z_x = m). \tag{16}$$

Denote by $\tilde{a}_x^L$ the value of unit long-term benefits including fees:

$$\tilde{a}_x^L = \mathrm{E}\left[{}_{Z_x|}a_x^L\right](1+\beta) + \alpha.$$

The replacement rate of long-term care benefits RL_x can be then calculated as:

$$\mathrm{RL}_x = \frac{L_x}{W_T} = \frac{D_T}{\tilde{a}_x^L}. \tag{17}$$

We have calculated replacement rates for three options of the relationship between the mortality probabilities p_x^{13} and p_x^{23} (see above). In our calculations, we have considered the fees $\alpha = 10\%$ and $\beta = 9\%$. Compared to the case of the lifetime annuity, the fee α is lower (50% for the lifetime annuity). This reflects the fact that the expected lifetime (EL) in state 2 is typically lower than the life expectancy when calculating a lifetime annuity and a high fee corresponding to the first pension could significantly lower the benefits. The decrease in the fee α is compensated with a higher fee β ($\beta = 8\%$ for the lifetime annuity). In our calculations, we have considered the level of savings $D_T = 3$ yearly salaries.

Option 2, where $p_x^{13} = p_x^{23} = q_x^{(L)}$, can be used to estimate the upper limit of the long-term care (LTC) insurance price. The replacement rates of the LTC benefits and the expected value of 1 m. u. lifetime benefits (paid in the case of necessary LTC) are presented in Table 10 (columns RR and $\mathrm{E}\left[{}_{Z_x|}a_x^L\right]$, respectively). Compared to the replacement rates for the lifetime annuities (Table 1) one can observe significantly higher values. For example, for the age of 62 years, the replacement rate for LTC benefits is more than 6 times higher than in the case of lifetime annuity. This value is influenced by two factors. As the age increases, the probability of transition to state 2 increases, causing the insurance price to rise.

On the other hand, life expectancy decreases with increasing age, which has the opposite effect on the insurance price. According to values $\mathrm{E}\left[{}_{Z_x|}a_x^L\right]$ in Table 10 for lower ages, the first factor prevails, for higher ages the decisive factor is the decrease of the life expectancy. In the last two columns of Table 10, we present the expected value of 1 m. u. lifetime benefits in the case of LTC benefits paid from the age according to the first column and expected lifetime in state 2 for a person with the age according to the first column (${}_{0|}a_x^L$ and EL in 2 respectively). Compared to $\mathrm{E}\left[{}_{Z_x|}a_x^L\right]$, the ${}_{0|}a_x^L$ values are significantly higher, which is related to the uncertainty of a healthy person's transition to state 2.

Option 5, in which applies $p_x^{13} = q_x^{(S)}$ and $p_x^{23} = \lambda p_x^{13}$, is appropriate to estimate the lower limit of the LTC insurance price. The corresponding values are in Table 11. Compared to Option 2, one can observe significantly higher replacement rates and lower expected lifetimes in years in state 2.

Option 3, where $p_x^{13} = q_x^{(L)}$ with additional equation $q_x^{(S)} = (1 - \gamma_x)\, p_x^{13} + \gamma_x p_x^{23}$, respects the fact, that people needing LTC have lower life expectancy comparing to healthy ones (see e.g., Murtaugh et al., 2001). Moreover, the mortality rate for the whole population is the realistic value $q_x^{(S)}$. The results corresponding to Option 3 are presented in Table 12. For example, the replacement rate corresponding to the age of 62 years is about 9 times higher compared to lifetime annuities (Table 1). The life expectancy of healthy people (it can be seen in the last column of Table 10) is significantly higher than that of people in the need of LTC (see the last column of Table 12).

Table 10 Replacement rates (RR) of long-term care benefits (D_T = 3) and expected lifetimes (EL) in years in the state 2 for different ages and mortality settings according to Option 2

Age *x* (in years)	RR	$\mathrm{E}\left({}_{Z_x|}a_x^L\right)$	${}_{0|}a_x^L$	EL in 2
62	0.9740	2.7340	17.5089	24.17
65	0.8696	3.0734	16.4735	22.03
70	0.7466	3.5947	14.5665	18.46
75	0.7567	3.5453	12.4312	14.91
80	0.8484	3.1523	10.1182	11.47

Source: Own construction

One can observe, that results in Tables 10–12 vary significantly, which shows a big impact of the p_x^{13} and p_x^{23} choices. The precise adjustment would require relevant data on the mortality of the dependents. Since we do not yet have these data, we present only three options, two of which represent the lower and upper limits of the insurance price.

Table 11 Replacement rates (RR) of long-term care benefits (D_T = 3) and expected lifetimes (EL) in years in the state 2 for different ages and mortality settings according to Option 5

Age *x* (in years)	RR	$\mathrm{E}\left({}_{Z_x|}a_x^L\right)$	${}_{0|}a_x^L$	EL in 2
62	2.3034	1.1031	8.8124	9.75
65	2.1393	1.1948	7.9958	8.64
70	1.9261	1.3372	6.6674	6.90
75	2.0513	1.2500	5.3240	5.25
80	2.4398	1.0363	4.0677	3.80

Source: Own construction

Table 12 Replacement rates (RR) of long-term care benefits (D_T = 3) and expected lifetimes (EL) in years in the state 2 for different ages and mortality settings according to Option 3

Age *x* (in years)	RR	$E\left({}_{Z_x|}a_x^L\right)$	${}_{0|}a_x^L$	EL in 2
62	1.4036	1.8691	10.0721	11.63
65	1.2726	2.0710	8.6030	9.66
70	1.1362	2.3307	7.4501	8.08
75	1.1460	2.3100	6.5696	6.89
80	1.2852	2.0498	5.6600	5.71

Source: Own construction

When deciding to purchase LTC insurance, it is useful to know the probability that it will be used sometime in the future. The probability p_x^U that a retiree at age *x* years will receive LTC benefits can be calculated as:

$$p_x^U = \sum_{m=1}^{\omega-x-1} \Pr(Z_x = m).$$

The probabilities p_x^U for selected ages can be found in Table 13. One can observe probabilities around 50% of receiving the LTC benefits. It is worth noting, that the probabilities of real using the insurance must be assessed together with the corresponding replacement rate, which is times higher compared to a classic lifetime pension.

Table 13 Probabilities that a retiree at age *x* will receive LTC benefits according to Options 1–5

Age *x* (in years) / Option	Option 1	Option 2	Option 3	Option 4	Option 5
62	0.4183	0.4829	0.5621	0.5816	0.4589
65	0.4345	0.5003	0.5806	0.5912	0.4666
70	0.4621	0.5278	0.6049	0.6181	0.4850
75	0.4653	0.5305	0.6049	0.6248	0.4774
80	0.4491	0.5125	0.5782	0.5990	0.4493

Source: Own construction

Buying the LTC insurance means giving up a large amount in favour of potential benefits. As in the case of a lifetime annuity, one may consider deferring the purchase of LTC insurance. Assume that a retiree at age *x* years has an amount $\tilde{a}_x^L$ to purchase a potential unit LTC benefit. If the individual decides to postpone the purchase for one year, the amount $\tilde{a}_x^L$ is invested with return *r* and three possibilities can arise:

- With probability p_x^{11} the retiree remains in a healthy state. After one year the retiree can buy the LTC insurance for the new price $\tilde{a}_{x+1}^L$.
- With probability p_x^{12} the retiree switches to the dependent state. After one year it is impossible to buy the LTC insurance and the retiree will lose the benefits with value ${}_{0|}a_{x+1}^L$.
- With probability p_x^{13} the retiree dies. In such a case, the amount of $\tilde{a}_x^L(1 + r)$ remains as an inheritance.

The expected profit from the one-year deferral of the purchase of the LTC insurance can then be calculated as:

$$\text{Edprof}_x = p_x^{11}\left(\tilde{a}_x^L(1 + r) - a_{x+1}^L\right) + p_x^{12}\left(\tilde{a}_x^L(1 + r) - {}_{0|}a_{x+1}^L\right) + p_x^{13}\tilde{a}_x^L(1 + r).$$

For Edprof$_x$ to be positive, the yield must be greater than the threshold value:

$$r_x^{L*} = \frac{p_x^{11}\tilde{a}_{x+1}^L + p_x^{12}{}_{0|}a_{x+1}^L}{\tilde{a}_x^L} - 1.$$

The minimum values of r_x^{L*} for selected ages are presented in Table 14. The threshold yields r_x^{L*} are positive for ages above 80 years. On the other hand, according to Tables 10–12, the values of ${}_{0|}a_{x+1}^L$ are significantly higher compared to $\tilde{a}_x^L$ and the pensioner faces the risk of losing the LTC benefits when he/she decides to defer the buying the LTC insurance. One can observe an increase in threshold yields with respect to age (except for minor anomalies at ages 62 and 65 years). This is a consequence of increasing (with respect to age) p_x^{12} probabilities and the resulting risk of falling into the state of dependence.

Table 14 The minimum values of yearly returns in case of one-year deferral of the LTC insurance purchase for selected ages according to Options 1–5

Age (in years) / Option	Option 1	Option 2	Option 3	Option 4	Option 5
62	–0.0522	–0.0508	–0.0417	–0.0408	–0.0577
65	–0.0556	–0.0530	–0.0378	–0.0367	–0.0642
70	–0.0256	–0.0256	–0.0184	–0.0179	–0.0395
75	–0.0098	–0.0109	–0.0049	–0.0036	–0.0203
80	0.0129	0.0097	0.0170	0.0232	0.0134

Source: Own construction

It is probably unrealistic to renounce all savings in favour of purchasing the LTC insurance. The authors in Murtaugh et al. (2001) presented an interesting idea of purchasing a combined product of a lifetime annuity and LTC insurance. They argued that a cohort buying such a combination had a higher mortality rate than a cohort buying a lifetime annuity only. Therefore, setting the mortality rates according to Option 3 is appropriate for calculating the price of the combined product. In Table 15, lifetime annuity replacement rates calculated with mortality $q_x^{(S)}$ are displayed. Compared to Table 1, where replacement rates are calculated with mortality rates $q_x^{(H)}$, the values from Table 15 are higher. The saved amount D_T (in our calculations we use $D_T = 3$ and age $x = 62$ years) can be divided between the purchase of the lifetime annuity and the LTC insurance. For example, a 1:1 split of the saved amount offers a replacement rate for the lifetime annuity of 8.37% and an additional 140.36%/2 = 70.18% if needed the LTC. Other options can be calculated using the linearity of replacement rates with respect to D_T.

Table 15 Replacement rates of whole-life annuity payments for various levels of savings and current initial ages of the pensioner using mortality rates $q_x^{(S)}$ with 7-year guarantee applied

D_T / age x (in years)	62	65	70	75	80
1.5	0.0837	0.0899	0.1027	0.1185	0.1365
2.0	0.1116	0.1199	0.1370	0.1581	0.1820
2.5	0.1396	0.1499	0.1712	0.1976	0.2275
3.0	0.1675	0.1799	0.2054	0.2370	0.2730
3.5	0.1954	0.2099	0.2397	0.2766	0.3185
4.0	0.2233	0.2399	0.2739	0.3161	0.3640

Source: Own construction

CONCLUSIONS

The introduction of the second pension pillar created a variety of possible types of pension benefits in Slovakia. We discussed several alternatives for using the second pillar savings.

When recalculating savings from the second pillar to the replacement rate, we have used an approach from Špirková et al. (2019). Lifetime annuities may have a problem to compensate the shortening of the first pillar. We found that the validity of hypothesis H1 depends on the level of income and the development of financial markets. For low-income cohorts, compensating for the reduction in the first pillar would require unrealistically high returns on assets. For high-income cohorts, participation in the second pillar is highly likely to be beneficial.

In the analysis of deferred annuities, we modified the approach of Milevsky (1998), considering a more general fee structure and discounting using the Svensson yield curve. The later purchase of an annuity can be an attractive solution for savers who do not need a pension benefit from the second pillar immediately after retirement. However, the advantage of such a strategy requires an adequate yield for invested savings. In accordance with Milevsky (1998), the threshold yield typically increases with the saver's age. This also applies to the new way of paying out savings from the second pillar, which is a combination of a programmed withdrawal and the later purchase of a lifetime annuity. On the other hand, however, unpaid money is still the property of the pensioner. Thus, the validity of hypothesis H2 depends on age and the level of market returns. In general, we can recommend deferring the purchase of an annuity only at the early stages of retirement.

An interesting possibility of using savings is the investment return withdrawal. This option can also be combined with a later purchase of the annuity. Its advantage is that unpaid savings still belong to the property of the saver. The drawback is that in some years zero benefits might be paid.

An early purchase of a temporary pension seems disadvantageous. There is little difference between a temporary pension benefit and gradually spending the saved amount. More promising benefits can only be obtained for higher ages or lengths of temporary pensions. On the other hand, this carries a high risk of death before the contract expiration. Thus, our calculations presented in Section 2.1 prove the validity of hypothesis H3.

The main contribution of the paper is the valuation of long-term care insurance. We used a three-state model, where it was necessary to differentiate the mortality of healthy people and those in need of long-term care. Since we did not have exact data, we formulated five options of relationships between the mentioned mortality rates. Options with the lowest and the highest mortality were used for the upper and the lower estimate of the insurance price, respectively. We consider Option 3 from Section 3 as a realistic setting. In this setup, the population in general has high mortality probabilities $q_x^{(S)}$, while the healthy have low mortality rates $q_x^{(L)}$. The resulting pension benefits can guarantee a reasonable level of long-term care. Also in the case of the LTC insurance, an increase in the threshold returns for investment can be observed for the advantageous deferral of the purchase. Deferred purchase of the LTC insurance is suitable for lower ages. Later, the risk of falling into a state of dependence and losing potential benefits increases.

The LTC insurance product implies renouncing of savings in favour of purchasing insurance. A tempting alternative is to purchase a combined product consisting of the life annuity and the LTC insurance. For a cohort that opts for this choice, the cost of the lifetime annuity may be lower (Murtaugh et al., 2001). This is consistent with the Option 3 mortality setting. Along with (Murtaugh et al., 2001), we believe that the combined product is realistic and that hypothesis H4 about the meaningfulness of the LTC insurance is valid.

In the next research, it would be useful to obtain more accurate data on the number of people in the need of long-term care in Slovakia. Based on them, the proportions of dependents γ_x could be refined. An interesting separate research task is to investigate the relationship between the mortality rates of healthy p_x^{13} and dependents p_x^{23}. For this, it is also necessary to obtain relevant data.

ACKNOWLEDGEMENTS

This work was supported by the Slovak Research and Development Agency under the project APVV-20-0311 and Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic under the project VEGA 1/0760/22.

References

ALJINOVIĆ, Z., KATALINIĆ, K., POKLEPOVIĆ, T. (2012). Best fit model for yield curve estimation. *Croatian Operational Research Review*, 3: 28–40.

ECB. (2022a). *Euro area yield curves* [online]. European Central Bank. [cit. 15.6.2022]. <https://www.ecb.europa.eu/stats/financial_markets_and_interest_rates/euro_area_yield_curves/html/index.en.html>.

ECB. (2022b). *The Yield Curve Technical Notes of ECB* [online]. European Central Bank. [cit: 15.6.2022]. <https://www.ecb.europa.eu/stats/financial_markets_and_interest_rates/euro_area_yield_curves/html/technical_notes.pdf>.

GUBALOVÁ, J., MEDVEĎOVÁ, P., ŠPIRKOVÁ, J. (2022). The Global Pension Index of Slovakia [online]. *Statistika: Statistics and Economy Journal*, 102(2): 117–137. <https://doi.org/10.54694/stat.2021.38>.

HMD. (2022). *Human Mortality Database* [online]. Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France). [cit. 17.6.2022]. <https://www.mortality.org>.

HYNDMAN, R. (2023). *Demography: Forecasting Mortality, Fertility, Migration and Population Data* [online]. R package version 2. [cit. 13.2.2023]. <https://CRAN.R-project.org/package=demography>.

HYNDMAN, R., ATHANASOPOULOS, G., BERGMEIR, C., CACERES, G., CHHAY, L., O'HARA-WILD, M., PETROPOULOS, F., RAZBASH, S., WANG, E., YASMEEN, F. (2023). *Forecast: Forecasting functions for time series and linear models* [online]. R package version 8.21. [cit. 13.2.2023]. <https://pkg.robjhyndman.com/forecast>.

HYNDMAN, R. J., KHANDAKAR, Y. (2008). Automatic time series forecasting: the forecast package for R [online]. *Journal of Statistical Software*, 26(3): 1–22. <https://doi.org/10.18637/jss.v027.i03>.

IFP. (2022). *Default strategy in pension saving: The case of Slovakia* [online]. Institute of Financial Policy, Ministry of Finance of the Slovak Republic. [cit. 17.8.2022]. <https://www.mfsr.sk/sk/financie/institut-financnej-politiky/publikacie-ifp/ekonomicke-analyzy/51-default-strategy-pension-savings-case-slovakia.html>.

KILIANOVÁ, S., MELICHERČÍK, I., ŠEVČOVIČ, D. (2006). Dynamic accumulation model for the second pillar of the Slovak pension system. *Finance a uver – Czech Journal of Economics and Finance*, 56(11–12): 506–521.

LEE, R. D., CARTER, L. R. (1992). Modeling and Forecasting U. S. Mortality [online]. *Journal of the American Statistical Association*, 87(419): 659–671. <https://doi.org/10.2307/2290201>.

MILEVSKY, M. A. (1998). Optimal asset allocation towards the end of the life cycle: To annuitize or not to annuitize? [online]. *Journal of Risk and Insurance*, 65(3): 401–426. <https://doi.org/10.2307/253657>.

MELICHERČÍK, I., SZŰCS, G., VILČEK, I. (2015). Investment strategies in defined-contribution pension schemes. *AMUC*, 84(2): 191–204.

MURTAUGH, C. M., SPILLMAN, B. C., WARSHAWSKY, M. J. (2001). In sickness and in health: an annuity approach to financing long-term care and retirement income. *Journal of Risk and Insurance*, 68(2): 225–253.

R CORE TEAM. (2022). *R: A language and environment for statistical computing* [online]. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org>.

ŠEBO, J., ŠEBOVÁ, Ľ. (2016). Suboptimality of immediate annuitization in private pension schemes. *Ekonomický časopis*, 64(2): 169–190.

SO SR. (2022). *Mortality Tables of the SO SR 2018* [online]. Statistical Office of the Slovak Republic. [cit. 15.6.2022]. <https://slovak.statistics.sk>.

ŠPIRKOVÁ, J., DOBROVIČ, M., VINCZEOVÁ, M. (2021). A Critical View on Pension Savings in Slovakia [online]. *Statistika: Statistics and Economy Journal*, 101(4): 370–380. <https://doi.org/10.54694/stat.2021.11>.

ŠPIRKOVÁ, J., SZŰCS, G., KOLLÁR, I. (2019). Detailed view of a payout product of the old-age pension saving scheme in Slovakia. *Ekonomický časopis*, 67(7): 761–777.

SVENSSON, L. E. O. (1994). Estimation and Interpreting Forward Interest Rates: Sweden 1992–1994. *NBER Working Paper Series*, 4871.

WILLEMÉ, P. (2010). The Long-Term Care System for the Elderly in Belgium [online]. *ENEPRI Research Report*, 70. <http://dx.doi.org/10.2139/ssrn.2033672>.

Efficiency Evaluation of Water Sector in the Czech Republic: Two-Stage Network Dea

Emília Zimková[1] | *Matej Bel University, Banská Bystrica, Slovakia*
Petra Gurčíková[2] | *Matej Bel University, Banská Bystrica, Slovakia*
Michaela Vidiečanová[3] | *Matej Bel University, Banská Bystrica, Slovakia*
Ľubomír Pintér[4] | *Matej Bel University, Banská Bystrica, Slovakia*
Colin Lawson[5] | *University of Bath, Bath, United Kingdom*

Received 15.6.2023, **Accepted** (reviewed) 21.8.2023, **Published** 15.12.2023

Abstract

The water and wastewater services, usually provided in a monopoly regime, do not offer the operators natural incentives toward efficiency and innovation. Therefore the main aim of the regulatory institutions is to stimulate a competitive environment. The contribution measures technical efficiency of 21 water and waste water companies in the Czech Repulic. For the period 2018–2020, the two-stage slacked-based model (SBM) by Kaoru Tone and Miki Tsutsui (2009) was applied. The results of this study are heterogeneous. Only one company out of 21 can be identified as an overall technically efficient unit during all three analyzed years. It is Vodohospodárská společnost Olomouc (VSO) followed by Pražské vodovody a kanalizace (PVK) that is very close to full technical efficiency. Our results therefore reveal a strong potential for the decrease of ineficiency of the water sector in the Czech Republic. Another important outcome is the fact that regulation of the water industry in the Czech Republic is highly fragmented.

Keywords	***DOI***	***JEL code***
Performance measurement, optimization, regulation, water industry, data envelopment analysis, Network SBM, Covid-19	*https://doi.org/10.54694/stat.2023.26*	*C01, D22, D42, L51*

1 Department of Finance and Accounting, Faculty of Economics, Matej Bel University, Tajovského 10, 975 90 Banská Bystrica, Slovakia. E-mail: emilia.zimkova@umb.sk, phone: (+421)484466322. ORCID: <https://orcid.org/0000-0003-1648-8103>.
2 Department of Quantitative Methods and Information Systems, Faculty of Economics, Matej Bel University, Tajovského 10, 975 90 Banská Bystrica, Slovakia.
3 Department of Finance and Accounting, Faculty of Economics, Matej Bel University, Tajovského 10, 975 90 Banská Bystrica, Slovakia.
4 Department of Finance and Accounting, Faculty of Economics, Matej Bel University, Tajovského 10, 975 90 Banská Bystrica, Slovakia. E-mail: lubomir.pinter@umb.sk, phone: (+421)484466325.
5 Department of Economics, Faculty of Economics, University of Bath, Claverton Down, Bath BA2 7AY, UK. E-mail: hsscwl@bath.ac.uk, phone: (+44)1225388388.

INTRODUCTION

For more than 20 years, national regulators of network industries in many countries have been using benchmarking tools. These instruments are not used by national regulators in Central European countries yet. An issue of network industries regulation became heavily discussed topic in the professional as well as the wider public. The intensity of interest has increased over the past years, under the influence of the financial crisis and the Covid-19 pandemic. The importance of network industries was reinforced by Russia's aggression in Ukraine, but also by the increasingly intense climate crisis.

Regulated network industry entities usually offer electricity, gas, heat and water. These are commodities that significantly affect not only the consumer basket of households, but also all economic units. In our contribution, we will focus on research in the water segment, which, despite its vital importance, is absent in the academic literature in Central Europe. In addition, the ongoing climate change clearly confirms the exceptionality and rarity of this commodity.

Fresh water is, together with the sun, vital and its value will increase every year. Its share in the total volume of water on Earth is only 2.5%, up to 96.5% of the total volume of water on Earth is in form of oceans. The rest is in form of a saline ground-water (0.39%) and saline lakes (0.07%) (Gleick, 1993). Water research from an economic point of view is one of the areas that are extremely important, yet knowledge in this area is not sufficient.

From the economic theory point of view in the area of drinking water production, distribution and supply and wastewater collection and treatment, water companies, as well as other operators of public water supply and public sewerage systems, are natural local monopolies. In a given locality, there is always one supplier defined by its operating territory, or by operated water supply or sewerage systems, without the possibility to choose by consumers.

Thus, the water and wastewater services (WWS), usually provided in a monopoly regime, do not offer the operators natural incentives toward efficiency and innovation in opposition to competitive markets. The fostering of a competitive environment in the WWS is one of the main aims of the regulatory institutions. Regulators more and more employ benchmarking as a way to create markets and, therefore to encourage the WWS to be more productive. Benchmarking can be briefly defined as a process of seeking excellence through the systematic comparison of performance measures with reference standards (Marques, 2011). In our contribution the Data envelopment analysis is used to benchmark water providers operating in the Czech Republic in years 2018–2020.

Among main advantages of benchmaring use in the WWS, the following can be pointed out: [i] strong incentives are provided to operators to be efficient and innovative mitigating the costs of operation and capital expenses; [ii] on-going pressure is put on the water utilities to improve service quality; [iii] a fairer recovery of costs and of the capital investments is assured, and [iv] an increase of transparency and sharing of information minimizing its asymmetry between different stakeholders (especially between the regulator and the operator) (Marques, 2006).

The supervision of the water management in the Czech Republic is surprisingly fragmented. It is mainly carried out by the Ministry of Finance and the Ministry of Agriculture, but the Ministry of the Environment, the Ministry of Health, the Ministry of Transport and the Ministry of Defense also have partial competences. Every year, the Ministry of Finance oversees the creation of the prices of drinking and waste water. The Ministry of Agriculture ensures the regulation of water companies, supervises compliance with the binding rules agreed by the Ministry of Finance, collects data on the costs of water and wastewater companies and carries out inspections of WWS. It ensures that customers know who to contact with complaints about water companies when cooperation is needed.

The Department of Water Protection at the Ministry of the Environment covers the protection of the quantity and quality of surface and underground water, protection against floods, planning in the field of water at the national and international level, international cooperation in the field of water protection, economic, financial and administrative instruments for water protection, the creation of legislation and standards in the field of water protection.

After privatization in 1993, 11 state water companies in the Czech Republic were divided into 40 regional water companies and more than 1 200 small water intermediaries. In 2020, there were 7 729 owners and 3 041 operators of water companies in the Czech Republic. The number of owners increased by 249 compared to 2019, and the number of operators increased by 49 (Ministry of Agriculture, 2022).

Due to the high number of owners and operators of water infrastructure, several associations dealing with water management issues were created. The most important position is held by the organization SOVAK, which brings together physical and legal entities whose activities include the supply, removal and purification of waste water. The company includes the owners, managers, operators of water supply and sewerage systems for the public. The main goal of SOVAK is to formulate and defend the common interests of all members, to ensure the coordination of activities and services according to the needs and interests of members, to cooperate with professional organizations (EurEau – European Union of National Associations of Water Suppliers and Wastewater Service Providers) and to publish a professional magazine. The Association for Water of the Czech Republic and the Association of Owners of Water Infrastructure also aim to bring together water companies.

The regulation of water management in the Czech Republic is extremely fragmented compared to other countries, as it is ensured by six ministries. In the Czech Republic. In the Czech Republic, there is no independent professional regulatory body for water management, and the regulation of network sectors has long been perceived in the Czech Republic as the regulation of electricity, gas and heat.

The production process of water management companies in the past period was significantly affected by the Covid-19 pandemic and the measures taken by governments in an attempt to mitigate the spread of the pandemic. In the Czech Republic, among the most significant impacts of the Covid-19 pandemic on the activities of water companies were: restrictions on tourism, including the closure of accommodation and restaurant facilities, restrictions on the operation of schools and offices, restrictions on contact with customers, the expansion of work from home, and the adoption of adequate hygiene measures. Absences of employees due to the disease Covid-19, or Quarantines in case of infection of a close person caused a difficult organization of work, which was manifested by reduced reconstruction of property, with the exception of the emergency situations elimination.

The research in this paper focuses on the technical efficiency of 21 water companies operating in the territory of the Czech Republic, which provide abstraction, treatment, delivery, cleaning of water for more than two thirds of economic subjects in the Czech Republic. We apply two-stage slacked-based model (SBM) approach of data envelopment analysis (DEA) and evaluate the production process in water companies operating in the territory of the Czech Republic in 2018–2020.

The paper is structured as follows. The first part contains a brief literature overview, the second part provides description of data and model specifications, the third part reports main research findings that are discussed and in the final one concludes.

1 LITERATURE REVIEW

In the following part of the paper, we will summarize empirical studies dealing in recent years with the issue of efficiency evaluation based on the application of the DEA method in water management.

The paper of Thanassoulis (2000a) provides an introduction to the basic DEA models for assessing efficiency under constant and variable returns to scale. It also outlined the use of DEA by OFWAT, the regulator of water companies in England and Wales. The best of the effciency rating offered by DEA or OLS regression were allocated to the water utilities. The efficiency ratings had an impact on the cap placed on the company´s charges. Another paper writen by Thanassoulis (2000b) details the use of

DEA to estimate potential savings in the specific context of water distribution and discussed the use of the results obtained.

Liu and Fukushige (2019) investigated the efficiency of water utilities in Japan using a DEA model and regression analysis. The authors implement a two-stage analysis method, which involves measuring the relative efficiency of water supply and sewerage services using DEA in the first stage, and then, based on regression models, the authors examine the relationships between prices and estimated efficiency scores.

In their study, the authors Molinos-Senante and Maziotis (2021) examine the efficiency of water and sewerage companies in Chile during the years 2010–2018. This study estimates the cost-effectiveness of the water sector using the stochastic non-parametric data envelopment (StoNED) method, which combines the advantages of data envelopment analysis (DEA) and stochastic frontier analysis (SFA). The results from this study also indicate the need to consider the quality of service for the set water tariffs. The study also showed that public water companies showed higher price efficiency, followed by fully private and finally concession water companies. On the other hand, when the authors analyzed the trend of cost efficiency, the results showed that full private water companies showed an upward trend in their efficiency, while public water companies deteriorated their efficiency.

Another group of authors, Lombardi et al. (2019) conducted an empirical analysis using DEA in the Italian water sector. The study focuses on a selected sample of 68 Italian water companies between 2011 and 2013. The authors found that public water companies have the highest efficiency, purchasing and employing inputs more efficiently, compared to mixed or private firms. Considering water loss in research, the study shows that companies located in the north of Italy are more efficient than those operating in the center and south. Also, small companies showed better net efficiency results, followed by large and extra large companies.

The efficiency of Mexican water companies was analyzed by Salazar-Adams (2021) using double bootstrap DEA. In the first stage, he calculated the efficiency of each company based on a set of inputs and outputs, and in the second stage, the efficiency score is regressed against a set of explanatory variables that affect the efficiency of water companies. In his study, the author concludes that decentralization from the state to the municipalities in Mexico has not significantly increased the efficiency of water utilities because municipal utilities are as efficient as those run by the state. However, the water reform paved the way for new organizational schemes such as inter-municipal enterprises and privately managed enterprises. Intermunicipal enterprises in the sample are on average more efficient than state and municipal enterprises. Several privately managed companies in the sample are, on average, the most efficient type of water utility in Mexico. A possible explanation is local regulation, which compensates for the lack of institutional capacity that often occurs in developing countries.

A study by Liang et al. (2021) developed an improved two-stage network DEA analysis model assuming variable returns to scale in terms of both weights and solution methods, which determined the weights of each stage with the share of input resources. In particular, they measured the overall efficiency of water resource systems, water use efficiency, and wastewater treatment efficiency of 11 provinces in western China from 2008–2017. A panel Tobit regression model was used to further analyze the influencing factors of total efficiency, water use efficiency and wastewater treatment efficiency.

Another group of authors Zhou et al. (2018) investigated the efficiency of water management in China based on a non-radial non-oriented DEA model. The Slacks-Based Measure (SBM) model can simultaneously optimize required inputs and required outputs and project each unit to the "farthest" point on the efficient frontier.

Our contribution is the first empirical study on the technical efficiency of the water and waste water companies in the Czech Republic. To assess the technical efficiendy of water companies, an advanced Network SBM (slack-based network DEA model, variable return to scale) is used.

2 DATA AND MODEL SPECIFICATION

2.1 Data specification

As part of the analysis of technical efficiency, we analyze 21 private water companies operating in the territory of the Czech Republic: Pražské vodovody a kanalizace (PVK), Severočeské vodovody a kanalizace (SVK), Severomoravské vodovody a kanalizace Ostrava (SVKO), Brněnské vodárny a kanalizace (BVK), ČEVAK (CEV), Vodárenská akciová společnost, a.s. Brno (VAS), Ostravské vodárny a kanalizace (OVK), Stredočeské vodárny (SCV), Vodárna Plzeň (VPL), Vodárny a kanalizace Karlovy Vary (VKKV), Královéhradecká provozní (KHP), Vodovody a kanalizace Hodonín (VKH), Vodovody a kanalizace Mladá Boleslav (VKMB), Vodovody a kanalizace Břeclav (VKB), Vodovody a kanalizace Přerov (VKP), Vodovody a kanalizace Vsetín (VKV), Vodovody a kanalizace Kroměříž (VKK), Vodohospodářská společnost Rokycany (VSR), Vodohospodářská a obchodní společnost (VOS), Vodohospodářská společnost Benešov (VSB), Vodohospodářská společnost Olomouc (VSO).

To perform the efficiency analysis, we use two different types of variables, financial and physical. We drew data on financial variables from the available financial statements of companies, primarily from balance sheets and profit and loss statements and data regarding physical variables were gained from the annual reports of individual companies. Table 1 characterizes the selected variables of the model, which represent individual inputs, outputs and intermediate products. The selection of variables was based on the empirical studies listed above.

Table 1 Description of variables

Variable	Notation	Description
Operating costs	OPEX	Costs associated with economic activity, in thousands CZK
Investment	I	Investments into water infrastructure, in thousands CZK
The length of the water supply network	LN	The length of the water supply network without connections, in km
Volume of invoiced drinking water	W	Volume of invoiced drinking water intended for implementation, in m^3
Number of customers	CUS	Number of customers supplied with drinking water

Source: Authors

Table 2 presents the descriptive statistics of the variables of the water companies for the year 2020, and Tables 3 and 4 show the descriptive statistics for the years 2019 and 2018.

In 2020, the operating cost variable reaches an average value of approximately 1 410 649 thousand CZK. There are considerable differences in the minimum and maximum values for all monitored variables. As for the average value of investments, in 2020 they represented 131 308 thousand CZK. The average

Table 2 Descriptive statistics of variables in 2020

2020	OPEX	I	LN	W	CUS
Mean	1 410 649.71	131 308.71	2 218.42	16 441.71	325 807.71
Median	766 684	96 440	1 133	8 369	169 482
Standard Deviation	1 815 278.27	157 739.16	2 345.14	21 345.89	356 914.17
Minimum	48 567	2 319	288.95	1 094	22 500
Maximum	7 621 119	693 962	9 724	91 239	1 330 000
Count	21	21	21	21	21

Source: Authors

length of the water supply network in 2020 was 2 218 km. The average value of supplied drinking water during 2020 is 16 441 m^3. Compared to 2019, the observed value decreased slightly. The average number of customers who are supplied with drinking water is 325 808 in 2020. We can observe higher standard deviations of all variables, which are caused by the different sizes of water companies.

Table 3 Descriptive statistics of variables in 2019

2019	OPEX	I	LN	W	CUS
Mean	1 351 443.38	125 721.05	2 220.32	16 921.19	324 980.33
Median	739 950	85 833	1 130	7 265	168 340
Standard Deviation	1 708 483.89	140 866.20	2 336.98	22 608.34	354 653.28
Minimum	47 969	2 667	290.63	1 076	22 670
Maximum	7 039 511	665 260	9 705	97 190	1 317 000
Count	21	21	21	21	21

Source: Authors

Table 4 Descriptive statistics of variables in 2018

2018	OPEX	I	LN	W	CUS
Mean	1 251 179.05	115 179.24	2 205.12	16 950.67	323 162.86
Median	705 075	91 457	1 127	7 005	167 119
Standard Deviation	1 499 535.03	120 493.84	2 322.72	22 593.11	351 508.42
Minimum	44 483	3 401	290.69	1 174	22 583
Maximum	5 838 638	539 701	9 670	97 746	1 300 000
Count	21	21	21	21	21

Source: Authors

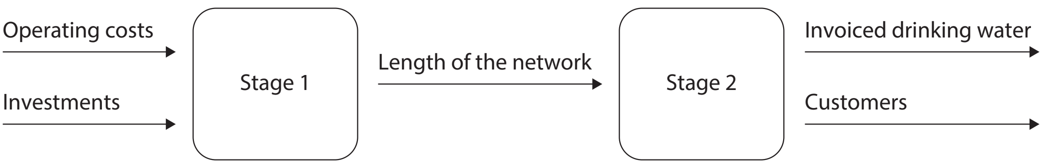
The selection of inputs and outputs was influenced by availability of the data and by empirical studies of other authors. Empirical studies most often consider operational costs, employee costs and capital costs, or investments as inputs. Among the traditional outputs, the authors include the volume of supplied drinking water and the number of customers connected to the public water supply. Some authors consider the length of the water supply network as input, others as output. In our research, we apply the network DEA model, which refers to the length of the network as an intermediate product.

To analyze the efficiency of water companies in the Czech Republic, we use a two-level network structure. The first phase represents the costs of maintenance and operation of the infrastructure, and the second phase is the use of the network for the supply of drinking water to customers. In the first phase, the water company must cover the costs of operation and maintenance. The costs include expenditure on employee wages, energy consumption, capital employed and are the inputs of the first stage. The mentioned expenses enable the infrastructure, which in our case is the length of the water supply network, to be functional. Thus, the infrastructure represents the output of the first phase. From a cost perspective, a water company is efficient if it is successful in operating and maintenance activities using minimum costs. In the second phase, the water company uses the infrastructure for water supply. Using the network DEA model, we can therefore calculate the cost efficiency of the enterprise, the efficiency of water service provision and the overall efficiency of the process. The length of the network in this model represents the connection between the mentioned two processes. In the first phase, the length of the network enters the process as an output, and in the second process, the length of the network is represented as an input. In the research, the operating costs (OPEX) and investments (I) of the water companies appear as the inputs of the first phase. The volume of invoiced drinking water (W) and the number of customers connected to the public water supply (CUS) are used as outputs in the second phase. The length of the network (LN) is considered an intermediate product, which is an output in the first phase and an input in the second phase.

The first stage shows how the company maintains its infrastructure by spending the minimum cost (cost efficiency), and the second stage shows how it utilizes its infrastructure to deliver water and wastewater services to its customers.

Figure 1 represents the given network model.

Figure 1 Two stage network DEA model



Source: Authors

When using the network SBM model, it is important to assign weights to individual divisions. We consider both phases to be equally important, so we decided to give them the same weight of 0.5 and 0.5, since the cost management process of the company and the process of providing water services to its customers are the basic tasks of water companies. An equally important issue is determining the connection between individual divisions. According Tone and Tsutsui (2009), there are 4 types of links: a free link, in which connected activities are under the control of the company while maintaining continuity between input and output. Another type is a fixed link, where the linked activities are unchanged and the intermediate product is not under the control of the company. There are two more types of less used links, so called good link and bad link. In our analysis, we use a free link, as each water company can guide the length of the water network.

2.2 Model specification

Jablonský and Dlouhý (2015) cluster the stages of the production process into serial, parallel or their combination. The serial model assumes a multi-stage production process in which a certain output represents the input to the next stage.

In our study, a two-stage network DEA model is used. The efficiency of the 1st and 2nd level is defined as follows:

$$\max \theta_0^{\,1} = \frac{\vec{w}'\,\vec{z}_0'}{\vec{v}'\,\vec{x}_0'} \qquad \max \theta_0^{\,2} = \frac{\vec{u}'\,\vec{y}_0}{\vec{w}'\,\vec{z}_0}, \tag{1}$$

under conditions:

$$\vec{w}'\vec{z} - \vec{v}'x \le \vec{0}' \qquad \vec{u}'\vec{y}_0' - \vec{w}'\vec{z} \le \vec{0}'$$
$$\vec{v}' \ge \vec{0}', \vec{w}' \ge \vec{0}'; \vec{u}' \ge \vec{0}'$$
$$\vec{w}'\vec{z} - \vec{v}'x \le \vec{0}' \qquad \vec{u}'\vec{y}_0 - \vec{w}'\vec{z} \le \vec{0}'$$
$$\vec{v}' \ge \vec{0}', \vec{w}' \ge \vec{0}'; \vec{u}' \ge \vec{0}'.$$

The overall efficiency can be expressed as:

$$\max \theta_0^{\ 0} = \frac{\vec{u}'\,\vec{y}_0}{\vec{v}'\,\vec{x}_0'}\,, \tag{2}$$

under conditions:

$$\vec{u}'Y - \vec{v}'X \leq \vec{0}'$$
$$\vec{v}' \geq \vec{0},\ \vec{u}' \geq \vec{0}.$$

Assuming that the output of the 1st stage is also the input of the 2nd stage, the overall efficiency is expressed by the following equation:

$$\theta_0^{\ 0} = \frac{\vec{w}'\,\vec{z}_0'}{\vec{v}'\,\vec{x}_0} \times \frac{\vec{u}'\,\vec{y}_0'}{\vec{v}'\,\vec{x}_0} = \frac{\vec{u}'\,\vec{y}_0'}{\vec{v}'\,\vec{x}_0}\,, \tag{3}$$

under conditions:

$$\vec{u}'Y - \vec{w}'Z \leq \vec{0}'$$
$$\vec{w}'Z - \vec{v}'X \geq \vec{0}'$$
$$\vec{v}' \geq \vec{0},\ \vec{u}' \geq \vec{0},\ \vec{w}' \geq \vec{0}\,.$$

In the case of writing the model as a linear programming problem, we get the following form:

$$\max \theta_0^{\ 0} = \vec{u}'\,y_0\,, \tag{4}$$

under conditions:

$$\vec{v}'\vec{x}_0 = 1$$
$$\vec{u}'Y - \vec{w}'Z \leq \vec{0}'$$
$$\vec{w}'Z - \vec{v}'X \geq \vec{0}'$$
$$\vec{v}' \geq \vec{0},\ \vec{u}' \geq \vec{0}.$$

If $\vec{v}'^{*}$, $\vec{u}'^{*}$,$\vec{w}'^{*}$, represent an optimal solution, we express the overall efficiency of the 1st and 2nd stages as follows:

$$\theta_0^{\ 1} = \frac{\vec{w}'^{*}\,\vec{z}_0}{\vec{v}'^{*}\,\vec{x}_0}$$
$$\theta_0^{\ 2} = \frac{\vec{u}'^{*}\,\vec{y}_0}{\vec{w}'^{*}\,\vec{z}_0} \tag{5}$$
$$\theta_0^{\ 0} = \frac{\vec{u}'^{*}\,\vec{y}_0}{\vec{v}'^{*}\,\vec{x}_0}\,,$$

where: X = input matrix, Y = output matrix, Z = intermediate product matrix, $\vec{x}_0$ = input´s vector of DMU, $\vec{y}_0$ = output´s vector of DMU, $\vec{z}_0$ = intermediate product vector of DMU, $\vec{v}$ = input weight vector, $\vec{u}$ = output weight vector, $\vec{w}$ = intermediate product weight vector, $\theta_0^{\ 0}$ total efficiency of DMU, $\theta_0^{\ 1}$ efficiency of the first stage of DMU, $\theta_0^{\ 2}$ efficiency of the second stage of DMU.

While the traditional DEA model cannot systematically analyze the relationship between the overall efficiency and the efficiency of each stage, the Network SBM model considers the interactions between different stages and can integrate the efficiency evaluation of each network node in the system with the overall efficiency evaluation of the system. Tone and Tsutsui (2009) points out that it is one of the methods applicable to the comprehensive assessment of structural efficiency within the DMU.

3 RESULTS AND DISCUSSION

Table 5 shows the technical efficiency of water companies measured by the Network SBM model (VRS) for the years from 2018 to 2020. The first part of the table represents the efficiency of the 1st phase, which represents the operational management process in the area of costs. In the 1st phase, the inputs were operating costs and investments. These expenses serve to ensure the functionality of the network. So we used the length of the network as the output of the stage 1. The results shown in Table 5 indicate that in Phase 1, several water companies did not perform their operations and maintenance activities by minimum costs. Most of the water companies were identified as cost-inefficient under Phase 1. Based on the results, we observe that in 2020 only 1 out of 21 production units was technically efficient, namely Vodohospodárská společnost Olomouc (VSO). It is followed by Pražské vodovody a kanalizace (PVK) (99.9%), Severomoravské vodovody a kanalizace Ostrava (SVKO) (99.9%), and Severočeské vodovody a kanalizace (SVK) (97.65%), which are very close to full technical efficiency. Other entities achieved a technical efficiency of 76.63% and less. This result suggests that most water companies should better allocate their resources by improving day-to-day operations, maintaining infrastructure and catching up with the most efficient companies.

In the second phase, companies use the water supply network to supply water, and therefore it represents the input of the 2nd phase. The volume of supplied drinking water and the number of customers connected to the public water supply are the outputs of the second phase. In this phase, we record up to 15 technically efficient companies during the year 2020. The Severomoravské vodovody a kanalizace Ostrava (SVKO), which was almost technically efficient production unit in the first phase, achieved surprisingly unfavorable result, and in the second phase it reached the efficiency of 30.79%.

Table 5 Technical efficiency of water companies in 2018–2020 assessed by two-stage DEA model

DMU 2020	Efficiency of the 1. phase	Efficiency of the 2. phase	Overall efficiency	Order
PVK	0.99998	1	0.99999	2
SVK	0.97654	1	0.98813	3
SVKO	0.99998	0.30799	0.65399	12
BVK	0.76633	1	0.86771	4
CEV	0.64068	1	0.78099	7
VAS	0.64832	1	0.78664	6
OVK	0.72385	1	0.83981	5
SCV	0.53739	1	0.69909	8
VPL	0.52660	1	0.68990	9
VKKV	0.50215	1	0.66857	11
KHP	0.41580	1	0.58737	13
VKH	0.51091	1	0.67629	10
VKMB	0.42342	0.46947	0.43712	17
VKB	0.40970	1	0.58126	14
VKP	0.43615	0.58729	0.48205	16
VKV	0.36030	0.22282	0.30312	21
VKK	0.35344	1	0.52229	15
VSR	0.27156	1	0.42712	18
VOS	0.26686	0.59209	0.33536	20

Table 5 (continuation)

DMU 2020	Efficiency of the 1. phase	Efficiency of the 2. phase	Overall efficiency	Order
VSB	0.24268	0.69961	0.36036	19
VSO	1	1	1	1
DMU 2019	**Efficiency of the 1. phase**	**Efficiency of the 2. phase**	**Overall efficiency**	**Order**
PVK	0.99998	1	0.99999	2
SVK	0.89861	1	0.94660	3
SVKO	1	0.2262	0.61310	21
BVK	0.72219	1	0.83869	7
CEV	0.61802	1	0.76392	10
VAS	0.64932	0.67555	0.65965	19
OVK	0.86103	1	0.92533	4
SCV	0.73847	1	0.84956	6
VPL	0.62442	1	0.76879	9
VKKV	0.68778	0.64752	0.67137	15
KHP	0.57315	1	0.72866	11
VKH	0.81036	0.42014	0.63569	20
VKMB	0.72888	0.56885	0.66141	18
VKB	0.76461	0.57250	0.68137	14
VKP	0.86160	0.44153	0.66718	16
VKV	0.66200	0.78743	0.71196	12
VKK	0.78047	0.82343	0.79930	8
VSR	0.54215	0.89340	0.66563	17
VOS	0.74675	1	0.85501	5
VSB	0.52121	1	0.68526	13
VSO	1	1	1	1
DMU 2018	**Efficiency of the 1. phase**	**Efficiency of the 2. phase**	**Overall efficiency**	**Order**
PVK	0.99999	1	1	1
SVK	0.79583	1	0.88631	4
SVKO	1	0.15721	0.5786	20
BVK	0.76699	1	0.86813	5
CEV	0.644	1	0.78346	8
VAS	0.67288	0.21533	0.48884	21
OVK	0.91392	1	0.95502	3
SCV	0.74777	1	0.85568	6
VPL	0.64733	1	0.78591	7
VKKV	0.75115	0.73261	0.7432	10
KHP	0.5805	1	0.73458	11
VKH	0.8535	0.53582	0.70721	14

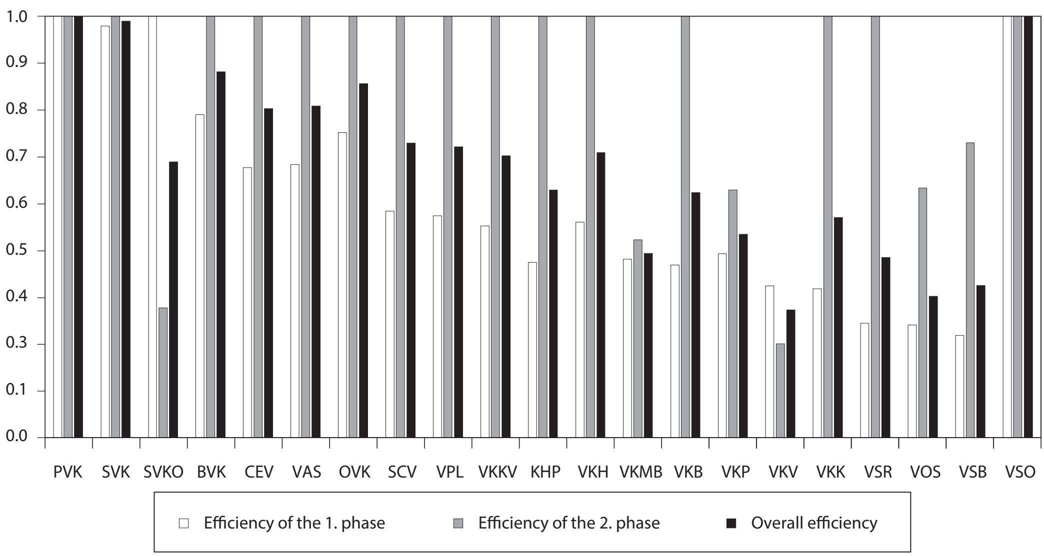
Table 5 (continuation)

DMU 2018	Efficiency of the 1. phase	Efficiency of the 2. phase	Overall efficiency	Order
VKMB	0.74387	0.5408	0.65725	16
VKB	0.81354	0.45802	0.65406	17
VKP	0.88052	0.34152	0.62814	19
VKV	0.6928	0.5783	0.64594	18
VKK	0.77986	0.64208	0.71949	13
VSR	0.58015	1	0.7343	12
VOS	0.77982	0.71869	0.75304	9
VSB	0.49577	1	0.66289	15
VSO	1	1	1	1

Note: Pražské vodovody a kanalizace (PVK), Severočeské vodovody a kanalizace (SVK), Severomoravské vodovody a kanalizace Ostrava (SVKO), Brněnské vodárny a kanalizace (BVK), ČEVAK (CEV), Vodárenská akciová společnost in Brno (VAS), Ostravské vodárny a kanalizace (OVK), Stredočeské vodárny (SCV), Vodárna Plzeň (VPL), Vodárny a kanalizace Karlovy Vary (VKKV), Královéhradecká provozní (KHP), Vodovody a kanalizace Hodonín (VKH), Vodovody a kanalizace Mladá Boleslav (VKMB), Vodovody a kanalizace Břeclav (VKB), Vodovody a kanalizace Přerov (VKP), Vodovody a kanalizace Vsetín (VKV), Vodovody a kanalizace Kroměříž (VKK), Vodohospodářská společnost Rokycany (VSR), Vodohospodářská a obchodní společnost (VOS), Vodohospodářská společnost Benešov (VSB), Vodohospodářská společnost Olomouc (VSO).
Source: Authors

As for the overall efficiency, we can also see the results in Table 5 together with the order of the analyzed subjects. Only one entity can be identified as an overall technically efficient production unit during all three analyzed years, namely Vodohospodárská společnost Olomouc (VSO). Pražské vodovody a kanalizace (PVK) is very close to full technical efficiency (99.9% in 2020, 2019, almost 100% in 2018), which ranks 2nd overall. In 2020 and 2019, Severočeské vodovody a kanalizace (SVK) was in third place (98.8% in 2020, 94.66% in 2019). In 2028, the third position belonged to Ostravské vodárny a kanalizace (OVK).

Figure 2 Technical efficiency of the water companies in Czech Republic in 2020

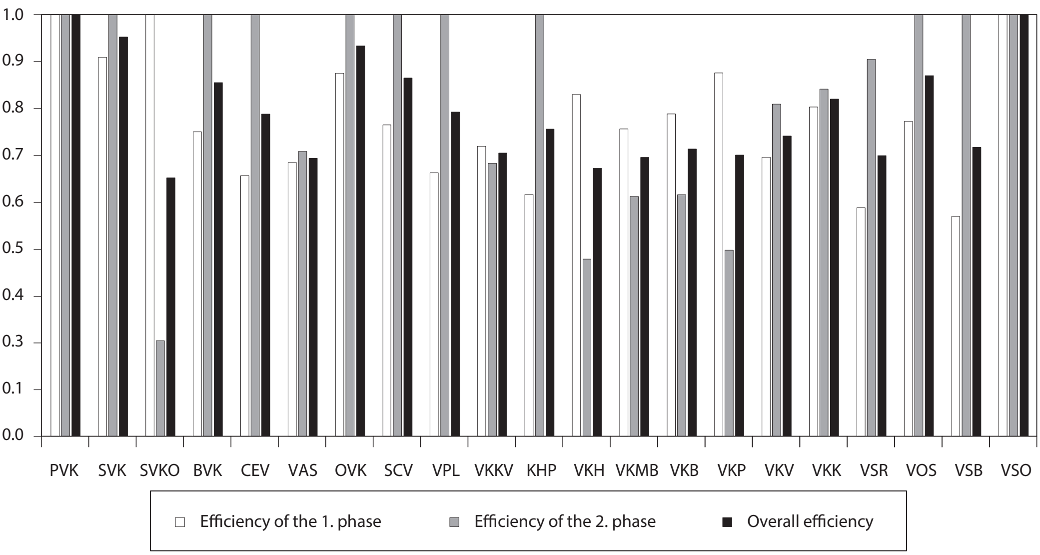


Source: Authors

Vodárenská akciová společnost (VAS), based in Brno, which provides services in the south of Moravia, Vysočina and part of the Pardubice region, recorded a significant increase in overall technical efficiency. While in 2018 it operated with the lowest technical efficiency of the assessed companies and ranked 21st (48.88%), in 2019 it ranked 19th (65.96%) and in 2020 it ranked 6th (78.66%).

Technical efficiency was calculated by each individual year separately, no data pooling was used. Figure 2 illustrates the ranking of water companies in the Czech Republic in 2020. The best result was achieved by the company Vodohospodárská společnost Olomouc (VSO), which was an efficient production unit in both phases of the process, and therefore also in overall efficiency. The VSO belongs to smaller water companies. It is worth to notice that the overall efficiency has almost been achieved by Pražské vodovody a kanalizace (PVK), i.e. the largest water company in the Czech Republic. Severočeské vodovody a kanalizace (SVK) is approaching the level of full technical efficiency as it is the third best production unit. The company Vodovody a kanalizace Vsetín (VKV) achieved the lowest efficiency score.

Figure 3 Technical efficiency of the water companies in Czech Republic in 2019

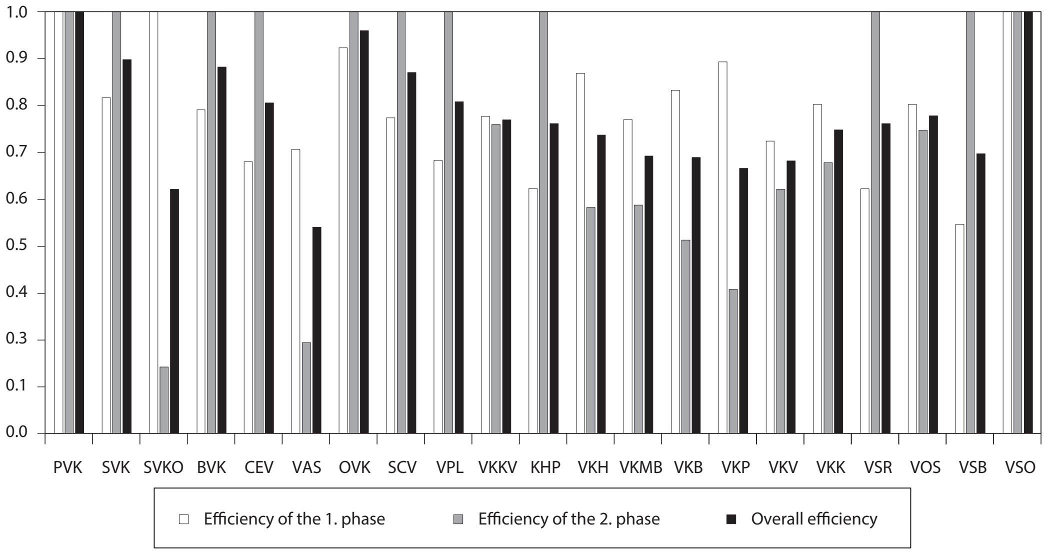


Source: Authors

Figure 3 shows the ranking of water companies from in the Czech Republic in 2019. The overall ranking of companies in terms of efficiency did not change significantly. Again, the Vodohospodářská společnost Olomouc (VSO) is the overall efficient unit, and we consider Pražské vodovody a kanalizace (PVK) and Severočeské vodovody a kanalizace (SVK) to be nearly efficient. The Severomoravské vodovody a kanalizace Ostrava (SVKO) is the least efficient production unit.

Figure 4 illustrates the ranking of water companies from in the Czech Republic in 2018. The best result was achieved by the company Vodohospodářská společnost Olomouc (VSO), which was an efficient production unit in both phases of the process, and therefore also in overall efficiency in all three analyzed years. Pražské vodovody a kanalizace (PVK) almost achieved overall efficiency, and Ostravské vodárny a kanalizace (OVK) took third place. The lowest efficiency score was achieved by Vodárenská akciová společnost (VAS).

Figure 4 Technical efficiency of the water companies in Czech Republic in 2018



Source: Authors

CONCLUSION

In conclusion, the importance of efficiency in the water industry cannot be overstated. It not only ensures the responsible use of a precious resource but also has wide-reaching implications for the environment, public health, economics, and overall sustainability. Therefore, continuous efforts to improve efficiency in the water sector are crucial for addressing the challenges of the 21st century.

The two-stage Network DEA model provides a thorough overview of the efficiency of individual process phases, which we consider important in complex production systems such as the water management system. Managers can thus observe the efficiency of individual phases as well as the overall efficiency within water management. Our findings show that water utilities should improve their cost performance by allocating their costs more efficiently and creating new practices that could reduce water supply costs and increase the volume of water supplied and the number of customers.

The results of our analysis can be interesting for policy makers in the water sector as well as for the management of individual investigated water companies operating in the Czech Republic. The results can also be an important benefit for regulatory authorities. In the Czech Republic, there is a clear absence of an independent central body that would have the competences of regulators in developed countries. The Authority could iniciate wide policies that are necessary in today´s economic and climate conditions. Is should introduce incentives that could lead to improved cost-effectiveness and efficiency in the provision of water management services. Findings at the level of individual stages of the production process can help the regulatory body to design policies based on incentives in the form of financial rewards.

We recommend to use the outputs of such an analysis in the determination of price tariffs for customers.

The information obtained as part of the analysis can lead to the improvement of the efficiency of water management in the Czech Republic, but also to the motivation of strategic planning aimed at increasing sustainability and resilience. Our future research will include the results of the technical efficiency of water management enterprises in the Czech Republic and Slovak Republic together.

ACKNOWLEDGEMENT

The research done in this paper is associated with the grant scheme VEGA 1/0442/22 New DEA models applicable for regulatory purposes in financial and network industries.

References

EAGRI. (2022). *eAGRI Portal* [online]. Prague: Ministry of Agriculture of the Czech Republic.

GLEICK, P. H. (1993). *Water in crisis.* Oxford University Press, New York. ISBN: 9780195076288

JABLONSKÝ, J., DLOUHÝ, M. (2015). *Modely hodnocení efektivnosti a alokace zdroju.* Prague: Profesional publishing. ISBN: 9788074311550

LIANG, X., LI, J., GUO, G,. LI, S., GONG, Q. (2001). Evaluation for water resource system efficiency and influencing factors in western China: A two-stage network DEA-Tobit model [online]. *Journal of cleaner production,* 328: 129674. <https://doi.org/10.1016/j.jclepro.2021.129674>.

LIU, J., FUKUSHIGE, M. (2019). Efficiency and pricing of water supply and sewerage services in Japan [online]. *Utilities policy*, 62: 957–1787. <https://doi.org/10.1016/j.jup.2019.100984>.

MOLINOS-SENANTE, M., MAZIOTIS, A. (2021). Benchmarking the efficiency of water and sewerage companies: Application of the stochastic non-parametric envelopment of data (stoned) method [online]. *Expert Systems With Applications,* 186: 957–4174. <https://doi.org/10.1016/j.eswa.2021.115711>.

SALAZAR-ADAMS, A. (2021). The efficiency of post-reform water utilitties in Mexico [online]. *Utilities policy*, 68: 101153. <https://doi.org/10.1016/j.jup.2020.101153>.

THANASSOULIS, E. (2000a). DEA and its use in the regulation of water companies [online]. *European Journal of Operational Research*, 127(1): 1–13. <https://doi.org/10.1016/S0377-2217(99)00436-1>.

THANASSOULIS, E. (2000b). The use of data envelopment analysis in the regulation of UK water utilities: Water distribution [online]. *European Journal of Operational Research*, 126(2): 436–453. <https://doi.org/10.1016/S0377-2217(99)00303-3>.

TONE, K., TSUTSUI, M. (2009). Network DEA: A slack-based measure approach [online]. *European journal of operational research*, 197(1): 243–252. <https://doi.org/10.1016/j.ejor.2008.05.027>. ISSN 0377-2217

ZHOU, X., LUO, R., YAO, L., CAO, S., WANG, S., LEV, B. (2018). Assessing integrated water use and wastewater treatment systems in China: A mixed network structure two-stage SBM DEA model [online]. *Journal of Cleaner Production*, 185: 533–546. <https://doi.org/10.1016/j.jclepro.2018.02.278>.

Investment into Low-Carbon Economy in the CEE NUTS-2 Regions: Are EU Funds Used where Needed?

Ľubica Štiblárová[1] | *Technical University of Košice, Košice, Slovakia*

Received 15.3.2023 (revision received 19.5.2023), **Accepted** (reviewed) 31.5.2023, **Published** 15.12.2023

Abstract

This paper explores the spending decisions about the EU Cohesion Policy 2014–20 investment in the low-carbon economy in the NUTS 2 regions of Central and Eastern European countries with regard to their climate need, proxied by the carbon emissions. By estimating non-spatial and spatial econometric models, which take into account the spatial scope of the Cohesion Policy, we do not confirm a statistically significant positive relationship between climate need proxied by carbon emissions in 2013 and the EU funds to a low-carbon economy in the programming period 2014–20. Our results, therefore, suggest that the EU funds with the low-carbon thematic objective have not been primarily spent in the regions with the highest carbon emissions prior to the examined programming period, calling for increasing awareness and necessary technical assistance for the beneficiaries, along the place-based strategies in the implementation of the Cohesion Policy in the next programming periods.

Keywords	DOI	JEL code
European structural and investment funds, allocation, low-carbon economy, spatial dependence, carbon emissions	*https://doi.org/10.54694/stat.2023.13*	*C21, E22, F36*

INTRODUCTION

The European Union (EU) is considered as a leader in focusing on the topic of climate change and mitigating its effects (Schreurs and Tiberghien, 2010; Siddi, 2020; Antimiani, Costantini and Paglialunga, 2023). Not only the recent Fit for 55 package proposed by the European Commission in 2021 aims to decrease net greenhouse emissions by at least 55% by 2030 (compared to the levels in 1990), but the EU considers carbon neutrality as its long-term goal in climate and energy policy (Schreurs and Tiberghien, 2010; Panarello and Gatto, 2023). The European Green Deal (EGD) should serve that purpose; the EGD has the ambition of zero net greenhouse emissions by 2050, which would ensure the status of the EU

[1] Faculty of Economics, Technical University of Košice, Němcovej 32, 040 01 Košice, Slovakia. E-mail: lubica.stiblarova@tuke.sk, phone: (+421)556022665.

as the first carbon-neutral continent (Cassetti et al., 2023; European Commission, 2023a). Some authors consider it as a *"roadmap of key policies for the EU's climate agenda"* (Siddi, 2020: 4).

This climate action, later translated into sustainable economic growth and decarbonization (Kedward and Ryan-Collins, 2022; Panarello and Gatto, 2023), should be financed by the NextGenerationEU Recovery Plan and the recent EU's common budget 2021–27 through the European Investment and Structural funds. For objective Greener Europe, more than 127 billion EUR is planned to be invested in lowering carbon emissions, circular economy, and mitigation of climate change within the period 2021–27 (European Commission, 2023b).

The EU Cohesion policy has already been supportive in promoting the EU's transition to a low-carbon economy in previous programming periods via European Regional Development Fund (ERDF). In the programming period 2014–20, more than 60 billion EUR have been planned to be invested in the low-carbon economy, while there has also been an obligatory condition to allocate a particular minimum share of the ERDF payments to the low-carbon economy thematic objective (European Commission, 2023c).[2] The member states exceeded these requirements and scheduled double the amount spent in the previous programming period 2007–13.

To provide desired effects, i.e., a reduction of carbon dioxide emissions, there should also be effective allocation mechanisms serving that purpose. Were investments into a low-carbon economy spent in the climatically disadvantaged regions? While there has been provided an excessive empirical evidence on the effects of EU funds in terms of reducing regional disparities and promoting economic growth (see, e.g., Ederveen, De Groot and Nahuis, 2006; Eggert et al., 2007; Becker, Egger and Von Ehrlich, 2010; Mohl and Hagen, 2010), the spending strategies of the member states' authorities themselves, i.e., whether financial resources are primarily implemented in the locations where needed, are not in the main area of interest, although, should be considered as crucial driving forces of the EU economic effects as well (Medve-Balint, 2018).

The aim of this paper is to answer this research question on the sample of Central and Eastern European (CEE) NUTS2 regions during the programming period 2014–20. We choose the CEE countries as the EU recipients who richly draw on the Cohesion Policy, but also due to the fact that compared to the Western European member states, these countries lag behind in the fulfillment of the EU's climate and energy goals (see, e.g., CORDIS, 2022). It is therefore more than appropriate to invest in the regions where most needed.

The rest of this paper is organized as follows. The first section provides a literature review on the effects of the Cohesion Policy in the EU. The second section describes a methodology and used data on examining the relationship between the low-carbon investment on the NUTS2 level and *"climate need"* measured as CO_2 emissions in the CEE countries. We provide results and discussion in the next section, while the last section concludes our comments, with policy recommendations on this matter.

1 LITERATURE REVIEW

The Cohesion Policy plays an important role in the sustainable transition of the EU member states; this task does not only consist of financing low-carbon investment but also in supporting the transformation of socio-economic and technical conditions. This includes, for instance, the development of infrastructure, technologies, or building capacities, which should be later translated into a systematic change in unsustainable production and consumption systems (European Commission, 2020b).

In this regard, the main branch of the empirical research is, therefore, devoted to the effects of the Cohesion Policy on convergence, economic growth, and/or employment. Here, the authors

[2] At least 20% of the payments from ERDF had to be allocated to low-carbon economy objective in more developed regions, 15% in transition regions, and 12% in less developed regions.

mostly rely on the β-convergence models from neoclassical growth theory (Baumol, 1986; Barro and Sala-i-Martin, 1992).

A majority of studies confirm a positive contribution of the Cohesion Policy to the recipient member states (see, e.g., Cappelen et al., 2003; Ederveen, De Groot and Nahuis, 2006; Ramajo et al., 2008; Mohl and Hagen, 2010). For instance, Cappelen et al. (2003) find a positive effect of the EU funds payment on regional economic growth performance, stating that historical changes in funds functioning helped to build an even more effective Cohesion Policy. Counter-factual evaluation of the Cohesion policy is provided by Becker, Egger, and Von Ehrlich (2013), or Pellegrini et al. (2013), who with the use of regression discontinuity design find a positive impact of the Cohesion Policy in the EU regions during 1994–2006. In a similar way, the positive causal effects of the EU funds on employment are observed in the Italian Objective 1 regions by Giua (2017).

Analogously, Mohl and Hagen (2010) show that Objective 1 payments do have a positive, statistically significant effect on regional GDP in EU countries during the programming period 2000–06.[3] However, Mohl and Hagen (2010) also state that the total amount of EU funds for Objectives 1, 2, and 3 do not seem to have a statistically significant positive effect on regional economic growth. Becker, Egger and Von Ehrlich (2010) confirm a positive effect of the EU funds for Objective 1 on EU regional GDP per capita growth, but they do not find employment growth effects at all. Similar evidence is brought by other studies which do not find significant effects of the Cohesion Policy (see, e.g., Boldrin and Canova, 2001).

Compared to the aggregate view on the EU funds, a lower granularity can take into account unobserved heterogeneity regarding sectors in the economy, objectives of the Cohesion Policy, and regions (see, e.g., Mohl and Hagen, 2010; Scotti, Flori and Pammolli, 2022). In this vein, Scotti, Flori and Pammolli (2022) provide detailed evidence of the effects of the EU funds across different sectors in the programming period 2007–14.[4] The authors find that the investment promising instantaneous and long-term growth has been confirmed for the energy, R&D, and transport sectors. On the contrary, environmental investment does not seem to provide an immediate stimulus to regional development.

In particular, the issue of environmental investments within the Cohesion Policy and their impact has been addressed, for example, by Ptak (2016), Streimikiene (2016), Kozera et al. (2022), or Gouveia, Henriques and Amaro (2023). Streimikiene (2016) investigates the role of EU funds in supporting sustainable energy development, with the main objective of energy efficiency allowing, among other things, a reduction of carbon emissions. The author focuses on the Baltic states in the programming periods 2007–13 and 2014–20 and states that the EU payments helped the Baltic countries to increase their energy productivity, especially Lithuania. The share of renewables has increased as well, where Estonia dominated in this area.

Kozera et al. (2022) examine low-carbon investment within the Cohesion Policy 2014–20 implemented in Polish municipalities. The authors focus on regional EU payments and find differences across regions, but also supported areas. For instance, the greatest portion of the EU funds has been used to promote infrastructure for clean transport and its improvement regarding energy efficiency.

Ptak (2016) discusses the expected effects of EU funds in the programming period 2014–20 with regard to the EU climate and energy targets in 9 EU member countries.[5] While the highest effect of the EU funds in the area of a low-carbon economy is expected in Poland, Ptak (2016), with a few exceptions, considers the achievement of set goals as a challenge. At the same time, the author states that selected countries differ widely in terms of meeting environmental targets.

[3] Mohl and Hagen (2010) consider the following EU countries: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Netherlands, Portugal, Spain, Sweden, and the United Kingdom.

[4] The authors focus on particular sectors: energy, environment, human resources, IT infrastructure, research and development, rural development, social infrastructure, tourism, and transportation.

[5] Austria, Czech Republic, Estonia, Germany, Latvia, Lithuania, Poland, Slovakia, and the United Kingdom.

Gouveia, Henriques and Amaro (2023) examine the efficiency of ERDF expenditure intended to support a low-carbon economy in 23 EU member states.[6] Using a Value-Based Data Envelopment Analysis (VBDEA), the authors only find less than half of examined countries as efficient recipients; while Spain seems to be the leading country in robust efficiency, the Czech Republic is the worst in the list of inefficient member states. Due to the discovered inefficiency, Gouveia, Henriques and Amaro (2023) suggest providing enhanced EU financing mechanisms and technical expertise regarding low-carbon economy for the recipient countries.

Dilba et al. (2016) present even more unfavorable evidence in their report. The authors focus on the spending plans of the EU funds and projects regarding climate in nine CEE countries.[7] In spite of the large available financial resources, Dilba et al. (2016) state that the full potential for the clean energy transition of the CEE countries has been unexploited and the climate requirements under the EU law have been implemented superficially. For instance, the EU funds in Poland have been rather spent on sustaining than transforming the recent coal-based economy.

Liobikiene and Mandravickaite (2016) investigate the link between the Cohesion Policy and greenhouse gas emissions in the programming period 2007–13 from the production-based perspective. The authors state that despite technological progress with regard to emissions, it is not possible to compensate for the effect of production scale in the Baltic countries, Bulgaria, or Poland. Also, Liobikiene and Mandravickaite (2016) declare that economic growth generated by the Cohesion Policy did not significantly contribute to structural changes to decrease greenhouse gas emissions.

Ionescu, Zlati and Antohi (2021) focus on Cohesion Policy from the point of view of the sustainability goals of the 2030 Agenda. Based on a cluster analysis in the monitored period 2006–20, the authors emphasize the need to reduce disparities between EU regions from the point of view of sustainable development. Ionescu, Zlati and Antohi (2021) also call for the improvement of sustainable development through the promotion of regional financial autonomy and administrative capacity.

For the emissions, Naqvi (2021) finds similar heterogeneous evidence, both spatial and temporal in EU NUTS 2 regions. Why are the effects of the Cohesion Policy noticeable in some countries/regions and negligible in others? There might be various reasons behind this matter – both on the aggregate and regional/sectoral levels. In this regard, the previous studies aim attention to the recipients' absorption capacity, institutional quality, or mismatch between the development needs and spending strategies of EU funds.

The first mentioned, the recipients' absorption capacity, has been proven to be one of the factors determining the effects of the EU funds. The European Commission (2020a) claims that it is vital to promote absorption capacity through country-specific suggestions, such as effective institutions and tax systems, anti-corruption frameworks, etc., which can be later translated into ease of investment. A similar scenario in terms of speeding up processes is recommended at the EU level in order to protect the EU's financial interests. In accordance with this, Arbolino and Di Caro (2021) find that an increased absorption rate and allocation of EU funds are associated with a reduction of regional employment losses during recessions, calling for a prompt and accurately managed Cohesion Policy. However, the recessionary periods tend to be associated with the decreased member state's ability to spend EU funds, i.e., the absorption paradox (Tatulescu and Patruti, 2014).

For institutional quality, it is expected that high-quality institutions are allied with higher GDP per capita, and employment, for which they present one of the prerequisites of long-term economic and social convergence (see, e.g., Mascherini and Mizsei, 2022). Montresor, Pecci and Pontarollo (2020)

---

[6] The sample consists of Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Denmark, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Poland, Portugal, Romania, Slovakia, Spain, Sweden, the United Kingdom.

[7] Croatia, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, and Slovakia.

examine this topic in 180 European NUTS-2 regions during the time period 1989–2006 using a spatial econometric approach. The authors find the conditional effect of regional institutional quality. Based on the results, Montresor, Pecci and Pontarollo (2020) confirm that an increase in regional institutional quality is associated with higher effectiveness of Objective 1 within the Cohesion Policy, because of which they recommend supporting regions' institutional capacity in order to improve the efficiency of the EU funds on productivity growths.

Similar suggestions are provided by Mendez and Bachtler (2022) who investigate the relationship between the quality of government and administrative performance for ERDF of 173 European regions in the time period 2007–13. The authors state that the quality of government seems to be a crucial driver of administrative performance and a high government quality may boost the Cohesion Policy implementation.

Jager (2022) confirms the importance of the institutions for the EU innovation projects in the Italian regions during 2007–20 only in the short-term; as the author claims, a high institutional quality allows for strengthening the effectiveness of the EU funds on patents.

While the empirical literature offers extensive results about the shortcomings of the Cohesion Policy with respect to the absorption capacity and institutional quality, there is a lack of systematic evidence on the possible mismatch between the development needs of the member states and their spending strategies of EU funds. A rare exception presents a study by Medve-Balint (2018) which examines Southern and Eastern EU member states' national spending strategies in two programming periods of the Cohesion Policy (2007–13 and 2014–20) with respect to the recipients' domestic development needs. The author examines five spending strategies – physical infrastructure, R&D and info-communication technology, business support, human capital, and institution building. With the use of Kendall's tau-b correlation, Medve-Balint (2018) reports that the spending strategies do not correspond to the domestic development needs of the recipient countries, indicating a possible misallocation problem.[8] In particular, recipients tend to prioritize physical infrastructure projects over R&D or human capital investment, which could bring long-term benefits.

This paper investigates a similar issue, with emphasis placed on the low-carbon theme in the Cohesion Policy, such as Kozera et al. (2022). However, our approach differs from the seminal work of Medve-Balint (2018) in several ways, by which we contribute to the recent empirical knowledge on this matter. Firstly, we focus on the regional level of the EU funds and the related climate needs of the regions. The regional examination can not only complement country-level analysis but may reveal detailed evidence, i.e., capture the within-country heterogeneity (Rodriguez-Pose and Ganau, 2022) on the actual EU spending in the CEE regions with respect to their climate needs.

The rationale for looking at the regional level is quite straightforward; EU funds are not distributed to specific regions, but at the national level. It must be said, however, that data on the socio-economic development of NUTS-2 regions are used to calculate the amount of national EU funding, after which each member state submits the proposal on how the total allocation will be handled through national and/or regional programmes (for more, see, European Commission, 2021).[9] In other words, the European Commission does not allocate the EU funds directly to the regions, but to programmes and categories of regions due to the fact that the EU Cohesion Policy is coordinated, not a common policy. This gives the opportunity for discrepancies between the actual expenditure and the needs of given regions.

Secondly, we do not rely on correlation coefficients, but estimated econometric model where we also allow for spatial dependence. The previous literature on the effects of the EU funds has highlighted the spatial scope of the Cohesion Policy since the spillover effects from the EU funds may arise

8 In Medve-Balint's (2018) paper, misallocation can be understood as a situation when development needs in selected spending categories do not correspond to the share of funds allocated to them.

9 In the programming period 2014–20, data on the socio-economic development of NUTS-2 regions have been used to divide regions into three categories: "less developed", "transitional", and "more developed".

in the spirit of the new economic geography (see, e.g., Krugman and Venables, 1995; Mohl and Hagen, 2010; Smit et al., 2015; Antunes et al., 2020; Scotti, Flori and Pammolli, 2022). Moreover, Scotti, Flori and Pammolli (2022) highlight the role of spatial spillovers in the EU funds allocation with regard to the granularity of data; the authors observe spillover effects that differ across sectors and geographical levels. For instance, the highest indirect spillovers are detected in the transport sector. Omitting the spatial aspect of the EU funds could thus lead to biased estimates.

2 METHODOLOGY AND DATA

To examine the relationship between the regional expenditure on EU funds to a low-carbon economy and the carbon dioxide emissions in the CEE regions, we estimate the baseline model in the following form:

$$EU\ funds\_carbon_i = \beta_0 + \beta_1 \log CO_{2i} + \sum_{c=1}^{C} \delta_c \log CV_{ci} + \mu_j + \varepsilon_i\,, \tag{1}$$

where $EUfunds\_carbon_i$ presents the total eligible expenditure spent on the EU projects with the thematic objective low-carbon economy, CO_{2i} stands for carbon dioxide emissions which presents our proxy for *"climate need"*, and $CVci$ stands for control variables (human capital in R&D, population, area, unemployment, and GDP per capita, namely) for the regions i within the programming period 2014–20. We also add a dummy variable (μ_j) denoting recipient country j to control for country-specific effects, and ε_i stands for the error term.

The model is estimated for cross-sectional data on the NUTS 2 level in the CEE countries.[10] Since the goal is to examine whether the EU funds' expenditure of the programming period 2014–20 has been spent in regions with high *climate need* (i.e., with a high level of carbon dioxide emissions), our main variable of interest, CO_{2i}, measures emissions prior to the start of the programming period, i.e., in the year 2013.[11] The rationale behind the decision of a lagged CO_2 variable in our model specification lies behind the fact that the European Parliament and the European Council decide on the total budget for the recipient countries based on national governments' proposals before the start of the programming period. Since the EC does not allocate expenditure directly to the regions which need it, national governments, in accordance with the EU allocation rules, have freedom in selecting their spending priorities (see, e.g., Medve-Balint, 2018). This can possibly lead to the situation when final expenditures are spent in regions which do not have such an urgent climate need from the point of view of carbon emissions). The sign of the estimated coefficient can, therefore, reveal whether EU funds have corresponded to the actual climate need (positive estimate) or on the contrary, have mismatched the climate need measured as regional carbon emissions (negative estimate), which may signal the possible misallocation of the EU funds on the low-carbon economy.[12]

Unlike evidence on the misallocation brought by Medve-Balint (2018), we extend the model by considering spatial dependence in the EU funds implementation in the spirit of new economic geography (Krugman and Venables, 1995). The previous empirical literature suggests that not only the spatial patterns but also spillover effects may appear in the context of the Cohesion Policy (see, e.g., Mohl and Hagen, 2010; Scotti, Flori and Pammolli, 2022). For this reason, we test spatial dependence in EU funds expenditure using Moran's I test, and then, to account for these possible effects, we estimate spatial econometric models. Here, we follow Elhorst (2014) and apply a specific-to-general approach (Elhorst,

[10] All NUTS 2 regions of the CEE countries are included in the analysis, regardless of their status (more developed/transition/less developed regions) since all regions have been supported by EU funds within the low-carbon economy thematic objective under "Greener, carbon-free Europe" objective (PO02).

[11] The remaining (control) variables in the model are measured as the average values for the programming period 2014–20 to take into account the socio-economic conditions of the given period.

[12] While referring to misallocation, we follow the terminology used in Medve-Balint (2018) – for more, see Footnote 8.

2014) by first estimating the non-spatial linear model and then assessing whether the baseline model should be extended by interaction spatial effects.

As a result, we choose a spatially lagged explanatory variables (SLX) model which includes exogenous interaction effects ($\boldsymbol{WX}$):

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

and spatial Durbin error (SDEM) model which extends SLX model (Formula 2) by considering the spatial dependence among the observation in the error term ($\boldsymbol{u} = \boldsymbol{\lambda Wu} + \boldsymbol{\varepsilon}$).[13] The estimation of the SLX and SDEM models allows us to examine direct effects of explanatory variables (through vector of estimated regression coefficients $\boldsymbol{\beta}$ related to the matrix of explanatory variables $\boldsymbol{X}$), but also indirect effects through the coefficients related to spatially lagged explanatory variables ($\boldsymbol{\theta}$).

The complete variables description with data sources is available in Table A1 in the Annex. The dependent variable, *EUfunds_carbon*, has been calculated based on individual location data published by the European Commission (DG REGIO) as a sum of total eligible expenditure for implemented projects with the thematic objective low-carbon economy, under *"Greener, carbon-free Europe"* objective (PO02), mapped to NUTS 2 regions (as % GDP). The proxy for climate need, the CO_{2i} variable, has been also individually calculated for the NUTS 2 regions from yearly emission gridmaps published in the EDGAR database (Emissions Database for Global Atmospheric Research). Here, we joined spatial point data of carbon dioxide emissions based on longitude and latitude coordinates to NUTS 2 polygons. The administrative NUTS 2 boundaries are retrieved from the GISCO statistical unit dataset provided by Eurostat.

The control variables have been selected in accordance with the previous empirical studies dealing with regional convergence and effects of EU funds (see, e.g., Cappelen et al., 2003; Kozera et al., 2022; Scotti, Flori and Pammolli, 2022). We control for human capital in R&D activities ($R\&D_i$), total regional area ($Area_i$), population (Pop_i) in examined NUTS 2 regions, unemployment ($Unemp_i$), and regional GDP per capita ($GDPpc_i$). The inclusion of these variables allows us to examine whether projects with low-carbon theme have been implemented in regions with higher or lower share of human capital in R&D ($R\&D_i$), in larger rural areas or rather smaller, but more dense cities ($Area_i$, Pop_i). For the remaining variables,

Table 1 Descriptive statistics

	N	Min	Mean	Median	S.D.	Max
EUfunds_carbon	61	0.165	2.656	2.627	1.638	8.324
CO_2	61	9 070.088	102 421.761	48 713.267	149 122.978	815 499.350
R&D	58	0.112	0.839	0.557	0.813	3.712
Pop	58	642 679.286	1 726 750.333	1 446 866.357	798 574.704	4 511 471.143
Area	61	496.000	18 595.276	16 263.000	12 478.170	64 586.000
Unemp	61	13.386	46.121	40.371	23.222	101.200
GDPpc	61	9 285.714	20 440.984	17 771.429	9 758.162	58 385.714

Source: Own elaboration based on data from EDGAR, the European Commission, and Eurostat

[13] The SLX and SDEM models have been selected based on information criteria (AIC, BIC), the log-likelihood value, and the likelihood ratio (LR) test. The SDEM model has been selected by AIC and the log-likelihood value, while the SLX model has been recommended by BIC and LR test, for which we report the results of both model estimations.

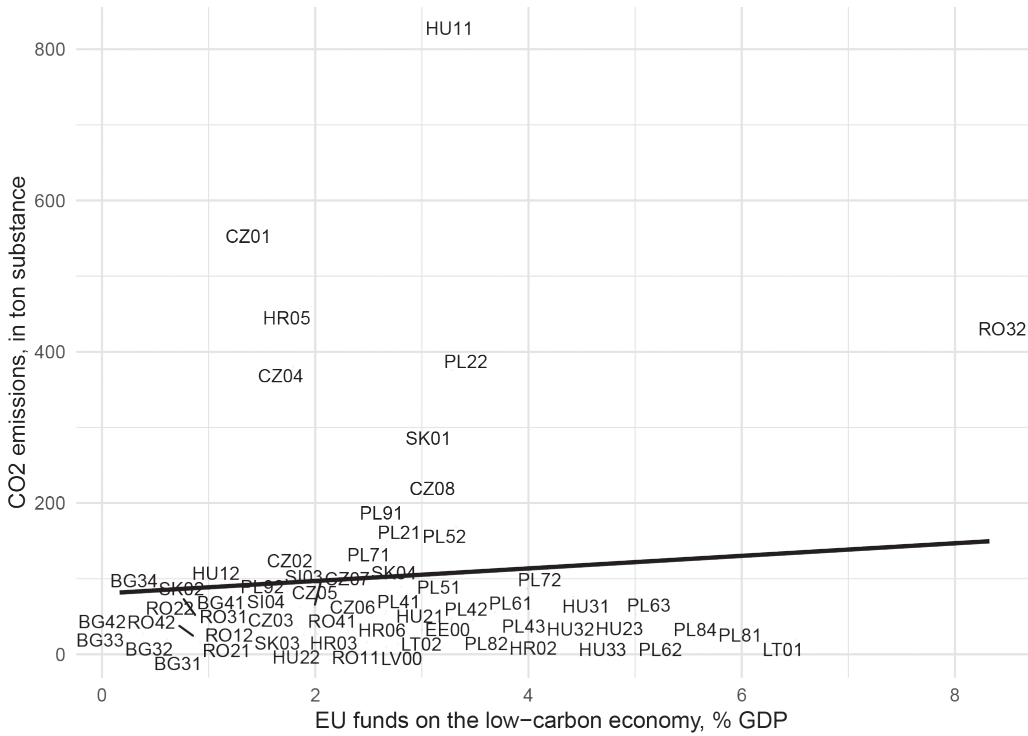
Unemp$_i$ and *GDPpc*$_i$, we expect that EU funds are primarily implemented in less developed regions in terms of lower GDP per capita and/or higher unemployment.

Our dataset consists of NUTS 2 regions in the Central and Eastern European countries. We provide a list of names of considered NUTS 2 regions in Table A2 in the Annex, while descriptive statistics is available in Table 1.

3 RESULTS AND DISCUSSION

In this section, we first depict the overall relationship between CO_2 emissions and EU expenditure on low-carbon within the NUTS 2 CEE regions (see Figure 1) and discuss the spatial distribution of considered variables, focusing on regional differences. Then, based on the examination of spatial dependence, we provide estimation results for non-spatial and spatial models and draw conclusions from the analysis.

Figure 1 Relationship between CO_2 emissions and EU low-carbon investment in NUTS 2 CEE regions

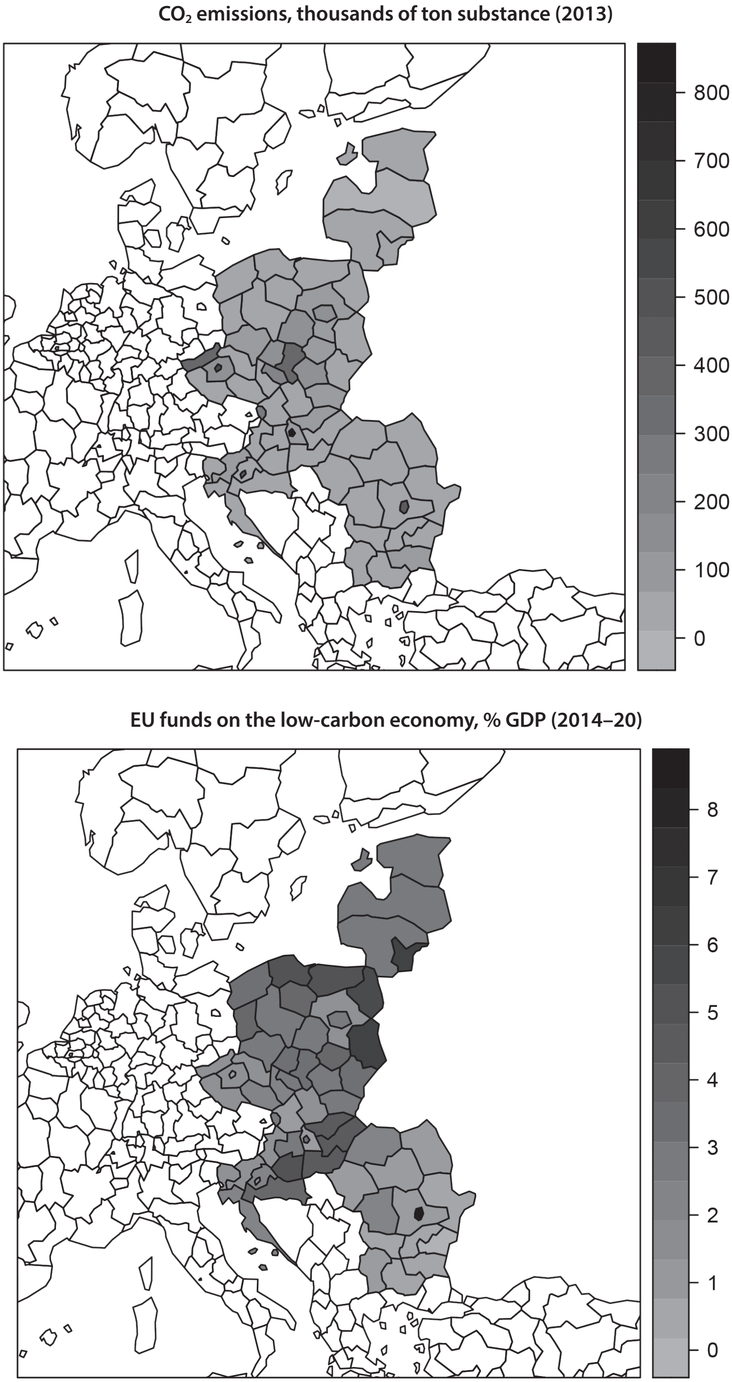


Source: Own elaboration based on data from EDGAR and the European Commission

From Figure 1, we cannot easily say that the EU funds for the low-carbon economy have been spent in the regions with higher carbon emissions. The trend line has a slight positive slope which is rather caused by unusual observations – by excluding them, the line is almost flat. The unusual observations mostly present capital cities, e.g., Budapest (HU11), Prague (CZ01), or Zagreb (HR05).

These metropolises show significantly higher emissions (see Figure 2, left) compared to spent EU funds in a low-carbon economy. We can assume that such results are related to the fact that regions in which the capital cities are located often show higher growth compared to the others. Hence, with gaining

Figure 2 Spatial distribution of CO_2 emissions and EU low-carbon investment in NUTS 2 CEE regions



Note: © EuroGeographics for the administrative boundaries.
Source: Own elaboration based on data from EDGAR and the European Commission

a status of transition or more developed regions, there is a general decrease in EU expenditure which is in line with e.g., Prota, Viesti and Bux, (2020).

On the other hand, such conclusions cannot be drawn for the Romanian capital, Bucharest, which receives a huge portion of EU funds for the low-carbon economy (% of GDP) and at the same time, peaks in pollution mainly because of traffic jams and motor vehicles in Romania (see also Figure 2, left).

Bucharest belongs to the most polluted cities in Europe and even the European Commission reproached the Romanian government for systematically exceeding the PM10 concentration limits in this region since 2007, failing to improve this situation (for more, see, European Commission, 2018). Overall, poor environmental conditions in Romania can be attributed to the transport and energy sectors, with the country's dominating long-standing orientation on the oil and gas industry (see, e.g., Fernandez et al., 2022).

Along the mentioned capital cities, the remaining regions of our sample do not show large discrepancies at the given scale, except for the Polish region Silesia (PL22) which experiences significantly more carbon emissions compared to other Polish areas. Silesia is one of the most urbanized and industry-oriented Polish regions where coal mining and related sectors played a crucial role in determining regional economic growth for a long time. Also for this reason, a smooth transition to a low-carbon economy in this region should be helped through EU funds.

While comparing the spatial distribution of carbon emissions and EU low-carbon investment (Figure 2), we can again state that there is no clear similar pattern, i.e., it does not seem that high shares of the EU funds are allocated to regions with high levels of carbon emissions, as in Figure 1 (except for Bucharest). This evidence might be related to the built-in internal mechanisms and criteria for the distribution of funds, i.e. the concentration principle, which concerns the territorial agenda. A high portion of the EU funds (% GDP) on the low-carbon economy is going to Vilnius County (LT01), Polish, and Hungarian regions where neighboring regions also experience similar levels of investment, which can suggest spatial autocorrelation in the EU funds expenditure.

Table 2 Spatial autocorrelation (Moran's I) – EU funds on low-carbon economy

	Spatial weights matrix W	
	Contiguity-based	**Distance-based (6-nearest neighbors)**
EU funds on the low-carbon economy	0.311*** (<0.001)	0.177** (0.002)

Note: We report Moran I statistic (p-values in parentheses). *p < 0.1, **p< 0.05, ***p< 0.01.
Source: Own elaboration based on data from the European Commission

To test this, we apply Moran's I test for which the results are provided in Table 2. The Moran's I test confirms that the spatial distribution of the EU funds is positively autocorrelated – this result is statistically significant and robust using either contiguity-based or distance-based spatial weights matrix W. Our evidence of spatial dependence is, therefore, in line with previous studies of e.g., Mohl and Hagen (2010), Scotti, Flori and Pammolli (2022) and omitting this could lead to biased estimates.

In order to avoid such a problem, the non-spatial model where we investigate the effect of carbon need on the EU funds expenditure on the low-carbon economy is followed by spatial model estimation. The results are provided in Table 3. In all model specifications (columns (I)–(VIII)), we do not confirm a statistically significant relationship between a climate need proxied by carbon emissions in 2013 and the EU funds to a low-carbon economy in the programming period 2014–20 for NUTS 2 CEE regions. Our evidence, therefore, validates the previous results of Medve-Balint (2018) about the possible mismatch between the domestic development needs and spending strategies of the EU funds. While Medve-Balint (2018) claims that EU funds spent on the R&D and human capital projects did not reflect the domestic development needs of recipient countries in 2007–20, our model estimations provide similar evidence,

Table 3 Estimation results

	Non-spatial models						Spatial models	
	OLS						SLX	SDEM
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)
Constant	−1.166 (1.127)	−0.473 (1.034)	4.595* (2.333)	3.956 (2.537)	11.158** (4.362)	9.889* (5.134)	7.467 (10.970)	10.573 (7.509)
CO_2	0.029 (0.104)	−0.024 (0.094)	−0.010 (0.092)	0.098 (0.128)	0.062 (0.116)	0.049 (0.142)	0.211 (0.178)	0.177 (0.113)
R&D		0.216* (0.128)	0.249* (0.127)	0.351** (0.168)	0.459** (0.170)	0.415** (0.189)	0.534** (0.234)	0.468*** (0.151)
Pop			−0.374** (0.185)	−0.537** (0.249)	−1.160*** (0.383)	−1.227*** (0.395)	−1.316*** (0.462)	−1.506*** (0.297)
Area				0.189 (0.210)	0.150 (0.194)	0.173 (0.207)	0.352 (0.266)	0.357** (0.166)
Unemp					0.633** (0.247)	0.695** (0.297)	0.547 (0.357)	0.727*** (0.232)
GDPpc						0.197 (0.491)	0.158 (0.552)	0.376 (0.352)
Lag.CO_2							0.064 (0.312)	0.038 (0.217)
Lag.R&D							−0.144 (0.352)	−0.078 (0.216)
Lag.Pop							−1.151 (0.986)	−1.498** (0.668)
Lag.Area							0.504 (0.545)	0.493 (0.349)
Lag.Unemp							0.218 (0.678)	0.635 (0.478)
Lag.GDPpc							1.124 (1.011)	1.208* (0.621)
Country-specific effects	YES	YES	YES	YES	YES	YES	YES	YES
R^2	0.660	0.694	0.716	0.722	0.758	0.759	0.865	0.871
N	61	58	58	58	58	58	58	58
AIC	99.900	94.395	92.037	92.933	86.755	88.528	86.841	86.114
BIC	127.341	123.241	122.944	125.900	121.783	125.616	156.896	158.230
LogLik							−9.421	−8.057
LR test							−2.727 (0.099)	

Note: Robust standard errors are reported in parentheses; *p < 0.1, **p < 0.05, ***p < 0.01. For LR test, we report LR ratio (p-value in parentheses). The spatial models are estimated using a contiguity-based spatial weights matrix (the model estimations using a distance-based spatial weights matrix are qualitatively similar, and thus, not included for the sake of brevity).
Source: Own elaboration based on data from EDGAR, the European Commission, and Eurostat

but for the EU low-carbon investment and on a lower granularity of data, at the regional level for CEE countries. For the country-specific evidence, we can only refer to Kozera et al. (2022) who calculate the correlation between the Polish municipalities' EU investment in a low-carbon economy and air pollutant emissions. While the authors find a correlation of 0.64 between the value of projects per 100 km^2 and air pollutant emissions, there is almost zero correlation between the value of projects per 1 000 inhabitants and air pollutant emissions.

The reasons behind this situation may also lie in the readiness and abilities of the actors to prepare their projects at this level. As mentioned by Vironen, McMaster and den Hoed (2019) the recipients within this thematic objective may present small organizations with limited capacities and a lack of technical expertise, which can hinder the preparation of project proposals of sufficient quality. A similar problem could also arise on the side of the implementing authorities, where recipients could have problems obtaining the necessary information due to the newness of the theme. From the perspective of a low-carbon economy, in addition to enhancing administration capacity and skills at the local level which has been proposed by CORDIS (2022), more targeted allocation at the regional level could also possibly help the CEE countries which lag behind Western fulfillment of the EU's climate and energy goals.

For the remaining (control) variables, we observe a statistically significant positive relationship between human capital in R&D, unemployment, and the EU funds expenditure. The EU funds seem to be implemented in areas with higher levels of human capital in R&D which can suggest that human capital is a prerequisite for investment in the low-carbon economy. In line with Cappelen et al. (2003), our assumption about the EU funds to less developed regions in terms of higher unemployment (columns (V), (VI), and (VIII)) has been confirmed as well. On the other hand, a decrease in the average population in 2014–20 (columns (III)–(VIII)) has been associated with an increase of EU funds to a low-carbon economy, i.e., more populated regions do not necessarily receive a higher portion of the EU funds.

While looking at spatial model estimations, the results regarding the direct effects of the EU funds' determinants remain qualitatively similar and thus, robust. The direction of the indirect (spillover) effects, i.e., effects related to the lagged explanatory variables is the same for the SLX model and SDEM model estimation (columns (VII)–(VIII)), however, a statistical significance of the indirect effect of population and GDP per capita has been confirmed only for SDEM, for which this result should be taken with caution. This might be related to examined spatial scale; as Smit et al. (2015) state, the reduced ability to identify the spillover effects might be related to the relatively high NUTS 2 level (which has been selected due to the data availability) since the spillover effects can be formed within their boundaries, at a lower scale. Regardless of that, the spatial model estimations confirm a non-statistically significant relationship between carbon emissions and EU funds (columns (VII)–(VIII)) to the low-carbon economy in the NUTS2 CEE regions.

CONCLUSION

Over the past decades, the EU takes steps towards mitigating the effects of climate change, a reduction of carbon emissions, with a recent target of being the first carbon-neutral continent. The EU Cohesion Policy plays a crucial role in achieving this goal by providing funding for the low-carbon economy and related climate actions.

Since the European Commission does not allocate EU funds directly to the regions with high climate needs, the aim of this paper was to investigate whether discrepancies between the regions' needs and funding did not appear in the sample of NUTS 2 regions in Central and Eastern Europe within the programming period 2014–20. We contribute to the empirical strand of this literature by providing regional evidence capturing the within-country heterogeneity in the CEE countries, but also considering spatial aspects of the EU funds through estimation of the spatial econometric models. To the best of our knowledge, such an approach has not yet been used in this context.

While we detect spatial dependence in the EU funding on the low-carbon economy in 2014–20, which is in line with the existing empirical literature (see, e.g., Smit et al., 2015; Antunes et al., 2020; Scotti, Flori and Pammolli, 2022), the non-spatial and spatial model estimations do not confirm a statistically significant positive relationship between a climate need proxied by carbon emissions in 2013 and the EU funds to a low-carbon economy in the programming period 2014–20. Our results, therefore, suggest that the EU funds with the low-carbon thematic objective have not been primarily spent in the regions with the highest carbon emissions prior to the examined programming period. This evidence can be considered as a part of the criticism of the Cohesion Policy in the environmental field as in Liobikiene and Mandravickaite (2016), or Dilba et al. (2016).

Since the mismatch between the spending strategies and development needs may be also linked with the expected effects of the EU funds, it would be appropriate to avoid this scenario from the point of view of effectiveness. At the same time, the presented analysis does not indicate misallocation as the cause of the overall inefficiency of the Cohesion Policy, rather it raises questions as to whether the given situation could not be prevented in the following programming periods. One of the possibilities may lie in the increasing awareness and necessary technical assistance for the beneficiaries, which could indirectly boost the demand and competition for such targeted EU support. This is important, especially in the CEE region which lags behind Western Europe in fulfilling the EU's climate and energy goals and is still dependent on energy-intensive industries.

As we do not detect robust spillover effects while considering spatial models, a lower granularity of data which could reveal broader spillover effects might be used in further work. Expansion to multiple thematic objectives and regions within the EU might be subject to our future research as well.

ACKNOWLEDGMENT

This work was supported by the Scientific Grant Agency VEGA under Grant No. 1/0394/21.

References

ANTIMIANI, A., COSTANTINI, V., PAGLIALUNGA, E. (2023). Fossil Fuels Subsidy Removal and the EU Carbon Neutrality Policy [online]. *Energy Economics*, 119: 106524. <https://doi.org/10.1016/j.eneco.2023.106524>.

ANTUNES, M. et al. (2020). The Impact of Structural Funds on Regional Growth: A Panel Data Spatial Analysis [online]. *Intereconomics*, 55: 312–319. <https://doi.org/10.1007/s10272-020-0921-1>.

ARBOLINO, R., DI CARO, P. (2021). Can the EU Funds Promote Regional Resilience at Time of Covid-19? Insights from the Great Recession [online]. *Journal of Policy Modeling*, 43(1): 109–126. <https://doi.org/10.1016/j.jpolmod.2020.10.001>.

BARRO, R. J., SALA-I MARTIN, X. (1992). Convergence [online]. *Journal of Political Economy*, 100(2): 223–251. <https://doi.org/10.1086/261816>.

BAUMOL, W. J. (1986). Productivity Growth, Convergence and Welfare: What the Long Run Data Show [online]. *American Economic Review*, 76(5): 1072–1085.

BECKER, S. O., EGGER, P. H., VON EHRLICH, M. (2010). Going Nuts: The Effect of EU Structural Funds on Regional Performance [online]. *Journal of Public Economics*, 94(9–10): 578–590. <https://doi.org/10.1016/j.jpubeco.2010.06.006>.

BECKER, S. O., EGGER, P. H., VON EHRLICH, M. (2013). Absorptive Capacity and the Growth and Investment Effects of Regional Transfers: A Regression Discontinuity Design with Heterogeneous Treatment Effects [online]. *American Economic Journal: Economic Policy*, 5(4): 29–77. <https://doi.org/10.1257/pol.5.4.29>.

BOLDRIN, M., CANOVA, F. (2001). Inequality and Convergence in Europe's Regions: Reconsidering European Regional Policies [online]. *Economic Policy*, 16(32): 206–253. <https://doi.org/10.1111/1468-0327.00074>.

CAPPELEN, A. et al. (2003). The Impact of EU Regional Support on Growth and Convergence in the European Union [online]. *Journal of Common Market Studies*, 41(4): 621–644. <https://doi.org/10.1111/1468-5965.00438>.

CASSETTI, G. et al. (2023). The Interplay among Covid-19 Economic Recovery, Behavioural Changes, and the European Green Deal: An Energy-Economic Modelling Perspective [online]. *Energy*, 263(C): 125798. <https://doi.org/10.1016/j.energy.2022.125798>.

CORDIS. (2022). *Empowering Central and Eastern Europe to Meet Energy and Climate Commitments* [online]. <https://cordis.europa.eu/article/id/442128-empowering-central-and-eastern-europe-to-meet-energy-and-climate-commitments>.

DILBA, J. et al. (2016). *Climate's Enfants Terribles: How New Member States' Misguided Use of EU Funds Is Holding Back Europe's Clean Energy Transition* [online]. <https://c2e2.unepccc.org/kms_object/climates-enfants-terribles-how-new-member-states-misguided-use-of-eu-funds-is-holding-back-europes-clean-energy-transition/>.

EDERVEEN, S., DE GROOT, H. L. F., NAHUIS, R. (2006). Fertile Soil for Structural Funds? A Panel Data Analysis of the Conditional Effectiveness of European Cohesion Policy [online]. *Kyklos*, 59(1): 17–42. <https://doi.org/10.1111/j.1467-6435.2006.00318.x>.

EGGERT, W. et al. (2007). Konvergenz- und Wach-stumseffekte der Europäischen Regionalpolitik in Deutschland [online]. *Perspektiven der Wirtschaftspolitik*, 8(2): 130–146. <https://doi.org/10.1111/j.1468-2516.2007.00237.x>.

ELHORST, J. P. (2014). *Spatial Econometrics from Cross-Sectional Data to Spatial Panels*. Springer.

EUROPEAN COMMISSION. (2018). *Action brought on 12 October 2018 – European Commission v Romania* [online]. <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1557176797270&uri=CELEX:62018CN0638>.

EUROPEAN COMMISSION. (2020a). *Annual Sustainable Growth Strategy 2021* [online]. <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020DC0575&from=en>.

EUROPEAN COMMISSION. (2020b). *Supporting Sustainability Transitions under the European Green Deal with Cohesion Policy* [online]. <https://eurocid.mne.gov.pt/sites/default/files/repository/paragraph/documents/9193/reportsusttransiten.pdf>.

EUROPEAN COMMISSION. (2021). *Profile: EU Cohesion Spending by NUTS Regions* [online]. <https://cohesiondata.ec.europa.eu/stories/s/EU-spending-by-NUTS-regions/ym2g-jhtp/>.

EUROPEAN COMMISSION. (2023a). *A European Green Deal. Striving to Be the First Climate-Neutral Continent* [online]. <https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en>.

EUROPEAN COMMISSION. (2023b). *Policy Objective: Greener Europe* [online]. <https://cohesiondata.ec.europa.eu/themes/2/21-27>.

EUROPEAN COMMISSION. (2023c). *Low-Carbon Economy* [online]. <https://ec.europa.eu/regional_policy/policy/themes/low-carbon-economy_en>.

FERNANDEZ, J. M. et al. (2022). Street-Level Methane Emissions of Bucharest, Romania and the Dominance of Urban Wastewater. *Atmospheric Environment: X*, 13: 100153. <https://doi.org/10.1016/j.aeaoa.2022.100153>.

GOUVEIA, M., HENRIQUES, C., AMARO, A. (2023). Are ERDF Efficient in Strengthening the Switch to a Low-Carbon Economy? Some Insights with Value-Based Data Envelopment Analysis. In: HENRIQUES, C., VISEU, C. (eds.) *EU Cohesion Policy Implementation – Evaluation Challenges and Opportunities*, Springer, Cham.

GIUA, M. (2017). Spatial Discontinuity for the Impact Assessment of the EU Regional Policy: The Case of Italian Objective 1 Regions [online]. *Journal of Regional Science*, 57(1): 109–131. <https://doi.org/10.1111/jors.12300>.

IONESCU, R.-V., ZLATI, M. L., ANTOHI, V. M. (2021). European Union's Regions between Cohesion and Sustainability [online]. *Technological Forecasting and Social Change*, 169(120862). <https://doi.org/10.1016/j.techfore.2021.120862>.

JAGER, T. (2022). *Institutional Quality and the Effectiveness of EU Funds for Innovation in Italy* [online]. <https://diposit.ub.edu/dspace/bitstream/2445/187947/1/TFM-ECO_Jager_2022.pdf>.

KEDWARD, K., RYAN-COLLINS, J. (2022). A Green New Deal: Opportunities and Constraints. In: ARESTIS, P., SAWYER, M. (eds.) *Economic Policies for Sustainability and Resilience*, Palgrave Macmillan, Cham.

KOZERA, A. et al. (2022). Regional Diversity of Low-Carbon Investment Support from EU Funds in the 2014–2020 Financial Perspective Based on the Example of Polish Municipalities [online]. *Renewable and Sustainable Energy Reviews*, 168: 112863. <https://doi.org/10.1016/j.rser.2022.112863>.

KRUGMAN, P., VENABLES, A. (1995). Globalization and the Inequality of Nations [online]. *Quarterly Journal of Economics*, 110(1995): 857–880. <https://doi.org/10.2307/2946642>.

LIOBIKIENE, G., MANDRAVICKAITE, J. (2016). The EU Cohesion Policy Implications to GHG Emissions from Production-Based Perspective [online]. *Environmental Science & Policy,* 55(1): 178–185. <https://doi.org/10.1016/j.envsci.2015.10.007>.

MASCHERINI, M., MIZSEI, B. (2022). *Does Europe Lead the Way in Institutional Quality?* [online]. Luxembourg: Publications Office of the European Union. <https://www.eurofound.europa.eu/sites/default/files/ef_publication/field_ef_document/ef22026en.pdf>.

MEDVE-BALINT, G. (2018). The Cohesion Policy on the EU's Eastern and Southern Periphery: Misallocated Funds? [online]. *Studies in Comparative International Development*, 53: 218–238. <https://doi.org/10.1007/s12116-018-9265-2>.

MENDEZ, C., BACHTLER, J. (2022). The Quality of Government and Administrative Performance: Explaining Cohesion Policy Compliance, Absorption and Achievements across EU Regions [online]. *Regional Studies*. <https://doi.org/10.1080/00343404.2022.2083593>.

MOHL, P., HAGEN, T. (2010). Do EU Structural Funds Promote Regional Growth? New Evidence from Various Panel Data Approaches [online]. *Regional Science and Urban Economics*, 40(5): 353–365. <https://doi.org/10.1016/j.regsciurbeco.2010.03.005>.

MONTRESOR, E., PECCI, F., PONTAROLLO, N. (2020). Structural Funds, Institutional Quality and Regional Economic Convergence in EU: A Spatial Econometric Approach. In: THILL, J. C. (eds.) *Innovations in Urban and Regional Systems*, Springer, Cham.

NAQVI, A. (2021). Decoupling Trends of Emissions across EU Regions and the Role of Environmental Policies [online]. *Journal of Cleaner Production*, 323(129130). <https://doi.org/10.1016/j.jclepro.2021.129130>.

PANARELLO, D., GATTO, A. (2023). Decarbonising Europe – EU citizens' Perception of Renewable Energy Transition amidst the European Green Deal [online]. *Energy Policy*, 172: 113272. <https://doi.org/10.1016/j.enpol.2022.113272>.

PELLEGRINI, G. et al. (2013). Measuring the Effects of European Regional Policy on Economic growth: A Regression Discontinuity Approach [online]. *Papers in Regional Science*, 92(1): 217–233. <https://doi.org/10.1111/j.1435-5957.2012.00459.x>.

PROTA, F., VIESTI, G., BUX, M. (2020). The Contribution of European Cohesion Policy to Public Investment. In: CERNIGLIA, F., SARACENO, F. (eds.) *A European Public Investment Outlook*.
PTAK, M. (2016). Shift Towards a Low-Carbon Economy. An Impact of Support Under Operational Programmes on Climate and Energy Targets [online]. *Economic and Environmental Studies*, 16(1): 85–102.
RAMAJO, J. et al. (2008). Spatial Heterogeneity and Interregional Spillovers in the European Union: Do Cohesion Policies Encourage Convergence across Regions? [online]. *European Economic Review*, 52(3): 551–567. <https://doi.org/10.1016/j.euroecorev.2007.05.006>.
RODRIGUEZ-POSE, A., GANAU, R. (2022). Institutions and the Productivity Challenge for European Regions [online]. *Journal of Economic Geography*, 22(1): 1–25. <https://doi.org/10.1093/jeg/lbab003>.
SCHREURS, M. A., TIBERGHIEN, Y. (2010). European Union Leadership in Climate Change: Mitigation through Multilevel Reinforcement. In: HARRISON, K., MCINTOSH, SUNDSTROM, L. (eds.) *Global Commons, Domestic Decisions: The Comparative Politics of Climate Change*, MIT Press.
SCOTTI, F., FLORI, A., PAMMOLLI, F. (2022). The Economic Impact of Structural and Cohesion Funds across Sectors: Immediate, Medium-To-Long Term Effects and Spillovers [online]. *Economic Modelling*, 111(105833). <https://doi.org/10.1016/j.econmod.2022.105833>.
SIDDI, M. (2020). *The European Green Deal: Assessing its Current State and Future Implementation* [online]. FIIA Working Paper, May 2020/4. <https://iris.unica.it/retrieve/e2f56ed9-fe47-3eaf-e053-3a05fe0a5d97/WP114_European%20Green%20Deal.pdf>.
SMIT, M. J. et al. (2015). Rural Development Funding and Agricultural Labour Productivity: A Spatial Analysis of the European Union at the Nuts2 Level [online]. *Ecological Indicators*, 59: 6–18. <https://doi.org/10.1016/j.ecolind.2015.05.061>.
STREIMIKIENE, D. (2016). Review of Financial Support from EU Structural Funds to Sustainable Energy in Baltic States [online]. *Renewable and Sustainable Energy Reviews*, 58: 1027–1038. <https://doi.org/10.1016/j.rser.2015.12.306>.
TATULESCU, A., PATRUTI, A. (2014). Structural Funds and Economic Crises: Romania's Absorption Paradox [online]. *Procedia Economics and Finance*, 16: 64–72. <https://doi.org/10.1016/s2212-5671(14)00775-8>.
VIRONEN, H., MCMASTER, I., DEN HOED, W. (2019). *Transition to a Low-Carbon Europe: The Supporting Role of Cohesion Policy* [online]. <https://eprc-strath.org/wp-content/uploads/2021/10/IQ-Net-Thematic-Paper-2019S.pdf>.

ANNEX

Table A1 Variables description and data sources

Variable	Description	Source
EUfunds_carbon	Location data on a sum of total eligible expenditure for projects with the thematic objective low-carbon economy mapped to NUTS 2 regions, %GDP	European Commission
CO_2	Carbon dioxide emissions spatial point data (longitude and latitude coordinates) joined to NUTS 2 polygons, in ton substance	EDGAR, Eurostat
R&D	R&D personnel and researchers, the percentage of total employment (numerator in full-time equivalent)	Eurostat
Pop	Total population on 1 January, total sex, total age class, number	Eurostat
Area	Total area, square kilometer	Eurostat
Unemp	unemployment, from 15 to 74 years, thousand persons	Eurostat
GDPpc	GDP in Purchasing power standard (PPS, EU27 from 2020), per inhabitant	Eurostat

Source: Own elaboration based on data from EDGAR, the European Commission and Eurostat

Table A2 List of considered NUTS 2 regions

Code	Name	Country	Code	Name	Country
BG31	Severozapaden	Bulgaria	PL22	Slaskie	Poland
BG32	Severen tsentralen	Bulgaria	PL41	Wielkopolskie	Poland
BG33	Severoiztochen	Bulgaria	PL42	Zachodniopomorskie	Poland

Table A2 (continuation)

Code	Name	Country	Code	Name	Country
BG34	Yugoiztochen	Bulgaria	PL43	Lubuskie	Poland
BG41	Yugozapaden	Bulgaria	PL51	Dolnoslaskie	Poland
BG42	Yuzhen tsentralen	Bulgaria	PL52	Opolskie	Poland
CZ01	Praha	Czechia	PL61	Kujawsko-Pomorskie	Poland
CZ02	Stredni Cechy	Czechia	PL62	Warminsko-Mazurskie	Poland
CZ03	Jihozápad	Czechia	PL63	Pomorskie	Poland
CZ04	Severozápad	Czechia	PL71	Lódzkie	Poland
CZ05	Severovýchod	Czechia	PL72	Swietokrzyskie	Poland
CZ06	Jihovýchod	Czechia	PL81	Lubelskie	Poland
CZ07	Stredni Morava	Czechia	PL82	Podkarpackie	Poland
CZ08	Moravskoslezsko	Czechia	PL84	Podlaskie	Poland
EE00	Eesti	Estonia	PL91	Warszawski stoleczny	Poland
HR02	Panonska Hrvatska	Croatia	PL92	Mazowiecki regionalny	Poland
HR03	Jadranska Hrvatska	Croatia	RO11	Nord-Vest	Romania
HR05	Grad Zagreb	Croatia	RO12	Centru	Romania
HR06	Sjeverna Hrvatska	Croatia	RO21	Nord-Est	Romania
LV00	Latvija	Latvia	RO22	Sud-Est	Romania
LT01	Sostines regionas	Lithuania	RO31	Sud - Muntenia	Romania
LT02	Vidurio ir vakaru Lietuvos regionas	Lithuania	RO32	Bucuresti - Ilfov	Romania
HU11	Budapest	Hungary	RO41	Sud-Vest Oltenia	Romania
HU12	Pest	Hungary	RO42	Vest	Romania
HU21	Közép-Dunántúl	Hungary	SI03	Vzhodna Slovenija	Slovenia
HU22	Nyugat-Dunántúl	Hungary	SI04	Zahodna Slovenija	Slovenia
HU23	Dél-Dunántúl	Hungary	SK01	Bratislavský kraj	Slovakia
HU31	Észak-Magyarország	Hungary	SK02	Západné Slovensko	Slovakia
HU32	Észak-Alföld	Hungary	SK03	Stredné Slovensko	Slovakia
HU33	Dél-Alföld	Hungary	SK04	Východné Slovensko	Slovakia
PL21	Malopolskie	Poland			

Source: Own elaboration

Multicriteria Evaluation of Randomized Response Techniques for Population Mean

Ondřej Vozár[1] | *Prague University of Economics and Business, Prague, Czech Republic*
Luboš Marek[2] | *Prague University of Economics and Business, Prague, Czech Republic*

Received 24.7.2023, **Accepted** (reviewed) 3.8.2023, **Published** 15.12.2023

Abstract

Increasing demand in surveys focused on quantitative characteristics (population mean among others) of controversial issues like corruption, tax evasion, drugs consumption of sensitive variables like spending on drugs or illegal sources of income which lead to lively research, are of randomized response techniques for quantitative variables. Therefore, we propose complex multicriteria evaluation methodology for randomized response techniques for population mean. Based on extensive review in the literature, following ranges of criteria were proposed: statistical properties of estimator, implementation and parameter choice, respondent burden and credibility and confidentiality protection of respondents' data. Finally, we evaluate in this setting standard techniques using scramble variables and recently proposed techniques of dichotomous question.

Keywords	***DOI***	***JEL code***
Randomized response techniques, scramble variable, multicriteria evaluation, survey sampling, Horvitz-Thompson estimator, population mean	*https://doi.org/10.54694/stat.2023.32*	*C83, C10*

INTRODUCTION

The field randomized response techniques for quantitative variables have been experiencing rapid development both in theory and practice (e.g. Christofides and Chaudhuri, 2013; or Chaudhuri et al., 2016; doctoral thesis of Cobo Rodrigues, 2018) since the first proposal fifty years ago (Eriksson, 1973). There are three main approaches in this field:

- Methods using scramble variables (Eriksson, 1973; Eichorn and Hayre, 1983), where instead of true values respondent provides linearly transformed values of sensitive variables depending on results of a random experiment.

[1] Department of Statistics and Probability, Faculty of Informatics and Statistics, Prague University of Economics and Business, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic. E-mail: vozo01@vse.cz.
[2] Department of Statistics and Probability, Faculty of Informatics and Statistics, Prague University of Economics and Business, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic. E-mail: marek@vse.cz. ORCID: <https://orcid.org/0000-0003-4761-1936>.

- Methods using scramble variables with use of auxiliary variable known for the whole population strongly correlated to surveyed sensitive variable (Diana and Perri, 2013; Cobo Rodriguez, 2018).
- Methods using dichotomous response (Antoch et al., 2022), where respondent provides only dichotomous response ("Yes/No") instead of any numerical value related to value of surveyed sensitive variable.

Recently, research has also been focused on comparison of techniques to find optimal techniques both in theoretical criteria and practical performance of competing estimators. For example, Azeem and Ali (2023) study vast number of estimators using scramble variables. To our best knowledge, all studies only focus on statistical criteria like unbiasedness of estimators, their variance and performance of asymptotic confidence interval. There is a need for methodology incorporating also relevant issues like design of randomization of responses (choice of scramble variables), comfort and credibility for respondents, risk of disclosure of respondent data.

Main purpose of the paper is proposing such methodology and evaluate standard techniques using scramble variables and techniques using dichotomous response. Techniques using auxiliary variable are excluded at this first stage to keep the clarity of our evaluation. Moreover, the application of methods using auxiliary variable is not feasible in the most of the real life populations. First, there is no such auxiliary variables nature of studied population (convicts, members of small community). The second, more important objection is, that if auxiliary variable is strongly correlated with the sensitive variable, the auxiliary variable is also sensitive. Therefore, such auxiliary variable would be also available in very poor quality.

The rest of the paper is organized as follows. In Section 1, basic notions of estimation of population mean of a finite population by randomized response techniques and notation is introduced. Section 2 reviews evaluated randomized response techniques, both standard ones using scramble variables and recently proposed ones with use of dichotomous response. In Section 3, ranges of quality criteria and evaluation methodology are proposed, and methods reviewed in the previous section are evaluated. The main findings and conclusions of the paper are summarized in the last section.

1 BASIC NOTIONS FOR ESTIMATION OF POPULATION MEAN

The purpose of survey sampling is to estimate characteristics of a finite population $U = \{1, 2, \ldots, N\}$ of N unambiguously identified objects. For a quantitative variable Y the most common objective is to estimate its population total $t_Y = \sum_{i\in U} Y_i$ or population mean $\overline{t}_Y = t_Y/N$. To achieve that, a random sample s of fixed sample size n is selected with probability $p(s)$. Using probabilities π_i, ($\pi_i = \sum_{s\ni i} p(s)$) of selection of i^{th} unit of the population U, population mean is then estimated by unbiased Horvitz-Thompson estimator

$$\overline{t}_{Y,HT} = \frac{1}{N} \sum_{i\in s} \frac{Y_i}{\pi_i}. \tag{1}$$

For statistical properties of estimators and theorethical proofs consult Horvitz and Thompson (1952) and Section 2.8 in Tillé (2006) for details.

Because the surveyed variable is sensitive, respondents frequently refuse to answer or provide fabricated answers. Therefore, survey statisticians try to obtain at least randomized variable Z correlated to variable of interest Y. At the second stage, randomization of responses is always carried out independently for each unit selected in sample s on the sampling procedure $p(s)$. Randomized response Z is further transformed to random variable R, which follows standard model of randomized responses proposed by Arnab (1994):

$$E_q(R_i) = Y_i,\ Var_q(R_i) = \phi_i \textit{ for all } i \in U,\ Cov_q(R_i,R_j) = 0,\ \textit{if } i \neq j\ i, j \in U\,, \tag{2}$$

where E_q, Var_q and Cov_q denote mean, variance and covariance with respect to probability distribution $q(r|s)$ of randomization of response of a selected sample s. Finally, population mean is estimated by unbiased Horvitz-Thompson estimator using transformed randomized responses R_i instead of values of sensitive variable Y_i

$$\bar{t}_Y^R = \frac{1}{N} \Sigma_{i \in s} \frac{R_i}{\pi_i}, \tag{3}$$

where upper subscript R denotes the used randomized response technique.

2 EVALUATED METHODS

This section presents both standard methods using scramble variable for population mean without use of auxiliary variable and methods using dichotomous response by Antoch et al. (2022). Only these methods are evaluated. In total, five methods are evaluated in the paper.

2.1 Standard methods using scramble variables

Idea of scramble variables comes from the seminal paper of Eriksson (1973). Ten years later, Eichhorn and Hayre (1983) generalized this concept and laid theoretical foundation of scramble variable. Idea of scramble variables in setting of Erikson (1973) is as follows. Respondent generates value of random scramble variable S, unknown to an interviewer. Then respondent provides transformed response $Z = SY$. The distribution of the S must be chosen to mask the sensitive variable Y. This setup was generalized by Arcos et al. (2015), who defined a model describing all cases for methods using scramble variables including general formula for population mean estimator of Horvitz-Thompson type and its variance. This model is widely used both in theoretical research for comparison of methods and applications.

2.1.1 Method of Eriksson (1973)

Eriksson (1973) proposed that each respondent randomly selects a card from a package of L cards with numbers $b_1, b_2, \ldots, b_L$. The value in card selected is unknown to an interviewer. The i^{th} respondent provides transformed value $b_i y_i$ instead of original value y_i. Randomized response of i^{th} respondent is then defined as

$$Z_{i,E} = y_i \, S_1, \tag{4}$$

where S_1 is a scramble variable with non-zero mean μ_1 and positive variance σ_1^2.

Transformed randomized response is then given as

$$R_{i,E} = \frac{Z_{i,A}}{\mu_1}. \tag{5}$$

Unbiased population mean estimator of Horvitz-Thompson type is then

$$\bar{t}_{Y,E}^{HT,R} = \frac{1}{N} \sum_{i \in s} \frac{R_{i,E}}{\pi_i}. \tag{6}$$

2.1.2 Method of Chaudhuri (1987)

Chaudhuri (1987) modified method of Eriksson (1973) as follows. Each respondent randomly selects one card from two packages. The first package consists of L cards with numbers $b_1, b_2, \ldots, b_L$; the second one consists of K cards with numbers $c_1, c_2, \ldots, c_K$. Both selected cards are unknown to an interviewer. The ith respondent provides transformed value $b_i y_i + c_i$ instead of original value y_i. Randomized response of ith respondent is then defined as

$$Z_{i,CH} = y_i\, S_1 + S_2\,, \tag{7}$$

where S_2 S_1 are scramble variables with non-zero means μ_1, μ_2 and positive variances σ_1^2, σ_2^2.

Transformed randomized response is then given as

$$R_{i,CH} = \frac{Z_{i,CH} - \mu_2}{\mu_1}\,. \tag{8}$$

Unbiased population mean estimator of Horvitz-Thompson type is then

$$\bar{t}_{Y,CH}^{HT,R} = \frac{1}{N}\sum_{i\in s}\frac{R_{i,CH}}{\pi_i}\,. \tag{9}$$

2.1.3 Method of Bar-Lev et al. (2004)

Bar-Lev et al. (2004) modified proposal of Eriksson (1973) in following way. With fixed probability p unknown both to respondent and interviewer, each respondent reports its true value of sensitive variable y_i. With probability $1 - p$ each respondent randomly selects a card from a package of L cards with numbers $b_1, b_2, \ldots, b_L$. Then, the ith respondent provides transformed value $b_i\, y_i$ instead of original value y_i. Randomized response of ith respondent is then defined as

$$Z_{i,B} = \begin{cases} y_1 & \text{with probability } p\,, \\ y_i S_1 & \text{with probability } 1 - p\,, \end{cases} \tag{10}$$

where S_1 is a scramble variable with non-zero mean μ_1 and positive variance σ_1^2, and p fixed probability $(0 < p < 1)$, that respondent provides the true value of sensitive variable y_i.

If $p + (1 - p)\mu_1 \neq 0$, transformed randomized response is then given as

$$R_{i,B} = \frac{Z_{i,B}}{p + (1 - p)\mu_1}\,. \tag{11}$$

Unbiased population mean estimator of Horvitz-Thompson type is then

$$\bar{t}_{Y,B}^{HT,R} = \frac{1}{N}\sum_{i\in s}\frac{R_{i,B}}{\pi_i}\,. \tag{12}$$

The main concern is usually respondent privacy. Low probabilities like $p = 0.1$ or $p = 0.2$ are chosen, because respondent privacy is the main concern of randomized response techniques.

2.2 Methods using dichotomous responses

Methods using scramble variables have several drawbacks. The first drawback is missing practical guidelines for designing scramble variable. In the literature it is recommended to rely on the survey statistician experience (Chaudhuri, 1987). The second drawback is, that the calculations can be too demanding for respondents, and they can lead to severe errors or refusal. The third drawback is that this method can be less trustworthy for respondents, because they can feel that interviewer can guess somewhat the sensitive value. Moreover, if the interviewer knows the values of scramble variable, he can calculate the true value of sensitive variable.

To resolve these issues, Antoch et al. (2022) proposed completely different approach. They assume that the surveyed sensitive variable Y is both non-negative and bounded from above, i.e., $0 < m \leq Y \leq M$. They assume the both bounds m, M of the variable Y are known. Each respondent generates, independently of the others, a pseudorandom number U from the uniform distribution on interval (m, M), while the interviewer does not know this value. The respondent then answers a simple question: "Is the value of Y greater than U?". For example: "Is your monthly income greater than U?"

Note, that even if an interviewer knows the value of random number U, he cannot guess the true value of U accurately (unless $Y = U = M$). Therefore, they proposed more accurate estimator using the values of random numbers U. Unbiased variance estimators using plug-in technique of Arnab (1994) and Arnab (1995) were derived by Vozár (2023). However, the variance estimators for both methods require knowledge of random numbers U.

2.2.1 Original method of Antoch et al. (2022)

Randomized response of i^{th} respondent follows alternative distribution with parameter $\frac{y_i - M}{M - m}$

$$Z_{i,(m,M)} = \begin{cases} 1 \text{ with probability } \dfrac{y_i - M}{M - m}, \text{ if } U_i < y_i . \\ 0 \text{ with probability } 1 - \dfrac{y_i - M}{M - m}, \text{ if } U_i \geq y_i . \end{cases} \tag{13}$$

Transformed randomized response is then given as

$$R_{i,(m,M)} = m + (M - m)\, Z_{i,(m,M)} . \tag{14}$$

Unbiased population mean estimator of Horvitz-Thompson type is then

$$\bar{t}^{\,HT,R}_{Y,(m,M)} = \frac{1}{N} \sum_{i \in s} \frac{R_{i,(m,M)}}{\pi_i} . \tag{15}$$

2.2.2 Method of Antoch et al. (2022) using values of random numbers

Randomized response of i^{th} respondent incorporates information on random number in the following manner

$$Z_{i,\alpha,(m,M)} = \begin{cases} 1 - \alpha + 2\alpha \dfrac{U_i - m}{M - m} \text{ with probability } \dfrac{y_i - m}{M - m}, \text{ if } U_i < y_i , \\ -\alpha + 2\alpha \dfrac{U_i - m}{M - m} \text{ with probability } 1 - \dfrac{y_i - M}{M - m}, \text{ if } U_i \geq y_i , \end{cases} \tag{16}$$

where α is a tuning parameter. Its value is a priori set by the interviewer, is fixed and unknown to the respondent. Antoch et al. (2022) derived its optimal value minimizing variance for case of sample with constant selection probabilities π_i.

Transformed randomized response is then given as

$$R_{i,\alpha,(m,M)} = (M - m)\, Z_{i,\alpha,(m,M)} + m \tag{17}$$

Unbiased population mean estimator of Horvitz-Thompson type is then

$$\bar{t}^{HT,R}_{Y,\alpha,(m,M)} = \frac{1}{N}\sum_{i\in s}\frac{R_{i,\alpha,(m,M)}}{\pi_i}. \qquad (18)$$

3 EVALUATIONS OF METHODS

In the first subsection, range of evaluation criteria randomized response techniques for population mean relevant for design of real-life survey are discussed and multicriteria evaluation methodology is proposed. In the second part, methods presented in Section 2 are evaluated. As explained before, only methods without use of auxiliary variables are evaluated.

3.1 Quality criteria of randomized response techniques

When choosing a technique and the values of its parameters for a given sample survey, in addition to purely statistical considerations (unbiasedness and consistency of estimates), also other criteria must also be considered (see Chaudhuri and Mukerjee, 1988; Chaudhuri and Christofides, 2013; the paper of Blair et al., 2015; the doctoral thesis of Cobo Rodriguez, 2018).

The first range of quality criteria for estimates based on the randomized response technique represents statistical criteria, which are a necessary condition for all newly proposed estimates and statistical methods in general. Namely, they are unbiasedness (at least asymptotic one), consistency of estimators, and consistent estimates of variances allowing the construction of asymptotic confidence intervals. The next three ranges of estimation quality criteria are already specific only to the techniques of randomized response.

The remaining ranges of the quality of estimates are related to the randomized response technique used. We sum up these issues in these three ranges of criteria:

- technical implementation of a random experiment of randomized response,
- the optimal choice of the parameters of the randomized response technique and the existence of theoretical results for the choice of these parameters, when it is always necessary to balance between the accuracy of the estimates and the protection of respondents' privacy ("accuracy-privacy trade-off", Chaudhuri and Mukerjee, 1988),
- clarity and credibility of the randomized response technique for the respondent.

The implementation of a randomized response is important for the practical feasibility of the survey, and more non-statistical aspects must be considered, such as education, cultural habits, the relationship, and trust of the investigated population towards the organizers and users of this survey (Chaudhuri and Christofides, 2013; Blair et al. 2015). Most applications of randomized response for a quantitative variable have focused on a discrete random variable taking on a relatively small number of values. Therefore, the generation of the masking variable was realized in the past by simple tools such as drawing from a deck of cards, fate, and wheel of fortune. For practically continuous sensitive variables acquiring many values, these methods are impractical, so it is necessary to use random number generators. These procedures are used more and more often due to the use of web, tablets, or mobile applications when polling.

When applying the randomized response technique, the choice of the parameters of the chosen method is crucial, because there is an inverse relationship between the accuracy of the estimates on the one hand and the risk of a sufficiently accurate estimate or even the disclosure of the respondent's sensitive data on the other ("accuracy-privacy trade-off", Chaudhuri and Mukerjee, 1988). It is observed (Blair et al., 2015; Warner, 1965) that respondents perceive these risks of revealing their confidential data sensitively. For qualitative variables (specifically, the relative frequency of occurrence of a given character), theoretical results and practical guidelines were obtained by Blair et al. (2015). For quantitative variables, we do not know such results, many authors (e.g. Chaudhuri, 1987) refer to the experience of the statistician designing the given statistical survey.

Critical to the success of the survey is clarity and user comfort for the respondent. Several studies (see discussion Blair et al., 2015) show that even relatively simple methods such as Warner's mirror question method are incomprehensible to a non-negligible part of respondents (e.g. 20% or more). Also, complicated techniques or techniques requiring complex arithmetic operations (e.g., Chaudhuri's method, 1987) may lead to refusal to answer, gross errors, or cause respondents to mistrust that the interviewer is trying to trick them into revealing a sensitive answer.

The criteria for the evaluation of methods of randomized response are summarized in Table 1, including the weights for the evaluation of individual techniques of randomized response in the next part of the chapter. The same weighting was chosen for all four criteria ranges, each criterion also has the same weight within each criteria range. Each criterion is evaluated on a four-point scale ranging from 0 to 3 with the following meaning:

- 0: the method completely violates the given criterion (completely unsatisfactory),
- 1: the method partially fulfills the given criterion, but insufficiently (unsatisfactory),
- 2: the method fulfills the given criterion to a sufficient extent (satisfactory),
- 3: the method meets the given criterion in full (excellent, optimal value).

We assigned equal weight to each range of criteria. Also, each criterion has the same weight within the given range of criteria. This is because we consider all ranges of criteria and criteria to be equally important.

Table 1 Evaluation criteria of randomized response techniques for population mean

Range	Criterion	Weight
1. Statistical properties of estimator	1.1 Unbiasedness and consistency of estimator	1/12
	1.2 Unbiasedness and consistency of variance estimator	1/12
	1.3 Functional asymptotic confidence intervals	1/12
2. Implementation and parameter choice	2.1 Difficulty of technical implementation of the method	1/8
	2.2 Rules for the optimal choice of method parameters	1/8
3. Respondent burden and credibility	3.1 Difficulty and clarity of the method for the respondent	1/8
	3.2 Credibility of the method for the respondent	1/8
4. Confidentiality protection of respondents' data	4.1 The degree of risk of disclosure of sensitive information	1/8
	4.2 Leakage of sensitive information when knowing the result of a randomized response	1/8

Source: Own construction

3.2 Evaluation of randomized response techniques for population mean

In the multicriteria evaluation, we compare standard methods for estimating the population mean without using an auxiliary variable with newly proposed estimates using a dichotomous response. A total of five estimates using the respective methods are evaluated:

1) estimator $\bar{t}_{Y,E}^{HT,R}$ using Eriksson technique (1973),
2) estimator $\bar{t}_{Y,CH}^{HT,R}$ using Chaudhuri technique (1987),
3) estimator $\bar{t}_{Y,B}^{HT,R}$ using technique of Bar-Lev et al. (2004),
4) estimator $\bar{t}_{Y,(m,M)}^{HT,R}$ using dichotomous response (Antoch et al., 2022),
5) estimator $\bar{t}_{Y,\alpha,(m,M)}^{HT,R}$ using dichotomous response with knowledge random number (Antoch et al., 2022).

In terms of the statistical properties of the estimation, the methods are completely comparable. The estimators have all the desired statistical properties (Eriksson, 1973; Chaudhuri, 1987; Bar-Lev et al., 2004; Antoch et al., 2022), the respective Horvitz-Thompson-type estimators are unbiased and consistent. Variance estimates of using the Arnab (1994) technique (see Vozár, 2023 for estimators using dichotomous response) are also unbiased and consistent. Therefore, all methods fully meet criteria 1.1 and 1.2 and they are rated with four points. The results of simulation studies (Vozár, 2023) show that the variance estimates are strongly influenced either by outliers of transformed weighted responses (Eriksson technique, 1973; Chaudhuri, 1987; Bar-Lev et al., 2004) or by negative values in Horvitz-Thompson estimates and variance estimates for methods using dichotomous responses (Antoch et al., 2022). All methods are therefore evaluated with two points in criterion 1.3.

The evaluation of the methods, including its justification, according to the second range of criteria dealing with the implementation and choice of method parameters is summarized in Table 2. The evaluation of the methods, including its justification, according to the third range of criteria dealing with the respondent's burden and credibility is summarized in Table 3.

Table 2 Evaluation of implementation and parameter choice

Method	Criterion	Evaluation rationale	Score
$\overline{t}_{Y,E}^{HT,R}$	2.1 Difficulty of technical implementation of the method	Fairly undemanding, standard techniques (fate, cards, random number generator)	3
	2.2 Rules for the optimal choice of method parameters	Arbitrary, relying on the experience of the statistician designing the survey	1
$\overline{t}_{Y,CH}^{HT,R}$	2.1 Difficulty of technical implementation of the method	Fairly undemanding, standard techniques (fate, cards, random number generator)	3
	2.2 Rules for the optimal choice of method parameters	Arbitrary, relying on the experience of the statistician designing the survey	1
$\overline{t}_{Y,B}^{HT,R}$	2.1 Difficulty of technical implementation of the method	Fairly undemanding, standard techniques (fate, cards, random number generator)	3
	2.2 Rules for the optimal choice of method parameters	Arbitrary, relying on the experience of the statistician designing the survey	1
$\overline{t}_{Y,(m,M)}^{HT,R}$	2.1 Difficulty of technical implementation of the method	Fairly undemanding, standard techniques (fate, cards, random number generator)	3
	2.2 Rules for the optimal choice of method parameters	Rules of thumb for choice of an interval for generating random numbers (using a priori information about the range and quantiles of the sensitive variable)	2
$\overline{t}_{Y,\alpha,(m,M)}^{HT,R}$	2.1 Difficulty of technical implementation of the method	Fairly undemanding, standard techniques (fate, cards, random number generator)	3
	2.2 Rules for the optimal choice of method parameters	Rules of thumb for choice of an interval for generating random numbers (using a priori information about the range and quantiles of the sensitive variable)	2

Source: Own construction

Table 3 Evaluation of respondent burden and credibility

Method	Criterion	Evaluation rationale	Score
$\overline{t}_{Y,E}^{HT,R}$	3.1 Difficulty and clarity of the method for the respondent	The method is quite understandable, risk of numerical error	2
	3.2 Credibility of the method for the respondent	The method can be seen as tricky, because the interviewer knows the masking variables and recalculates the sensitive value	1
$\overline{t}_{Y,CH}^{HT,R}$	3.1 Difficulty and clarity of the method for the respondent	The method may be perceived as less comprehensible due to the two card moves and more complex arithmetic operations, higher risk of numerical error	1
	3.2 Credibility of the method for the respondent	The method can be seen as tricky, because the interviewer knows the masking variables and recalculates the sensitive value	1

Table 3 (continuation)

Method	Criterion	Evaluation rationale	Score
$\bar{t}_{Y,B}^{HT,R}$	3.1 Difficulty and clarity of the method for the respondent	The method is quite understandable, risk of numerical error	2
	3.2 Credibility of the method for the respondent	The method can be seen as tricky, because the interviewer knows the masking variables and recalculates the sensitive value. Reliability is lower because the respondent may be asked to provide the true value of a sensitive variable	1
$\bar{t}_{Y,(m,M)}^{HT,R}$	3.1 Difficulty and clarity of the method for the respondent	Easy method[1)], "Yes/No" answer to a simple question	3
	3.2 Credibility of the method for the respondent	The method is relatively credible1), it depends on how the respondent perceives the risk associated with knowing the value of the random number in the question	2
$\bar{t}_{Y,\alpha,(m,M)}^{HT,R}$	3.1 Difficulty and clarity of the method for the respondent	Easy method[1)], "Yes/No" answer to a simple question	3
	3.2 Credibility of the method for the respondent	The method is relatively credible,[1)] it depends on how the respondent perceives the risk associated with knowing the value of the random number in the question	2

Note: [1)] The difficulty and credibility of the method depends on the technical way of implementing random number generation.
Source: Own construction

The evaluation of the methods, including its justification, according to the fourth range of criteria dealing with the protection of the confidentiality of the respondent's data is summarized in Table 4.

Table 4 Evaluation of confidentiality protection of respondents' data

Method	Criterion	Evaluation rationale	Score
$\bar{t}_{Y,E}^{HT,R}$	4.1 The degree of risk of disclosure of sensitive information	Fairly good	2
	4.2 Leakage of sensitive information when knowing the result of a randomized response	One hundred percent, given the knowledge of the masking variable	0
$\bar{t}_{Y,CH}^{HT,R}$	4.1 The degree of risk of disclosure of sensitive information	Fairly good	2
	4.2 Leakage of sensitive information when knowing the result of a randomized response	One hundred percent, given the knowledge of the masking variable	0
$\bar{t}_{Y,B}^{HT,R}$	4.1 The degree of risk of disclosure of sensitive information	Fairly good	2
	4.2 Leakage of sensitive information when knowing the result of a randomized response	One hundred percent, given the knowledge of the masking variable	0
$\bar{t}_{Y,(m,M)}^{HT,R}$	4.1 The degree of risk of disclosure of sensitive information	Good, except for extremely low or high values	2
	4.2 Leakage of sensitive information when knowing the result of a randomized response	The risk is very low except for very low or high values, leakage of the exact value only if $y_i = M$	2
$\bar{t}_{Y,\alpha,(m,M)}^{HT,R}$	4.1 The degree of risk of disclosure of sensitive information	Good, except for extremely low or high values	2
	4.2 Leakage of sensitive information when knowing the result of a randomized response	The risk is very low except for very low or high values, leakage of the exact value only if $y_i = M$	2

Source: Own construction

The evaluation of the methods in the individual ranges of quality criteria for the individual randomized response techniques, including their order, is summarized in Table 5. From the point of view of statistical criteria (unbiasedness and consistency of estimates), all compared techniques are equivalent.

The evaluation results showed that the newly proposed methods compared to the standard methods bring improvements in the areas of respondent burden, method credibility, and respondent confidentiality protection. Moreover, for methods using dichotomous response there are rules for choice of an interval for random numbers based on very rough a priori information about the interval of values of the sensitive variable. Since knowledge of random numbers is needed to estimate the variance to construct asymptotic confidence intervals, we are inclined to use the more precise estimate $\bar{t}^{HT,R}_{Y,\alpha,(m,M)}$ in practice. Therefore, methods using dichotomous response can be implemented as dynamic questionnaires in computer or web assisted surveys. The survey tool would use its own random number generator and the generated random number U would be used to create question "Is your monthly income greater than U?". Value of random number U is stored to compute estimate $\bar{t}^{HT,R}_{Y,\alpha,(m,M)}$ and its variance estimator. This approach would require very careful training of intervievers and drafting survey information for respondents'.

Table 5 Evaluation of randomized response technique for population mean

Criterion	Scoring (weighted) and method ranking				
	$\bar{t}^{HT,R}_{Y,E}$	$\bar{t}^{HT,R}_{Y,CH}$	$\bar{t}^{HT,R}_{Y,B}$	$\bar{t}^{HT,R}_{Y,(m,M)}$	$\bar{t}^{HT,R}_{Y,\alpha,(m,M)}$
Cumulative point rating (total)	1.853 **(3.-4.)**	1.728 (5.)	1.853 (3.-4.)	2.383 (1.-2.)	2.383 (1.-2.)
1. Statistical properties of estimator	0.633 (1.-5.)	0.633 (1.-5.)	0.633 (1.-5.)	0.633 (1.-5.)	0.633 (1.-5.)
1.1 Unbiasedness and consistency of estimator	0.250 (1.-5.)	0.250 (1.-5.)	0.250 (1.-5.)	0.250 (1.-5.)	0.2501 (1.-5.)
1.2 Unbiasedness and consistency of variance estimator	0.250 (1.-5.)	0.250 (1.-5.)	0.250 (1.-5.)	0.250 (1.-5.)	0.250 (1.-5.)
1.3 Functional asymptotic confidence intervals	0.133 (1.-5.)	0.133 (1.-5.)	0.133 (1.-5.)	0.133 (1.-5.)	0.133 (1.-5.)
2. Implementation and parameter choice	0.500 (3.-5.)	0.500 (3.-5.)	0.500 (3.-5.)	0.625 (1.-2.)	0.625 (1.-2.)
2.1 Difficulty of technical implementation of the method	0.375 (1.-5.)	0.375 (1.-5.)	0.375 (1.-5.)	0.375 (1.-5.)	0.375 (1.-5.)
2.2 Rules for the optimal choice of method parameters	0.125 (3.-5.)	0.125 (3.-5.)	0.125 (3.-5.)	0.250 (1.-2.)	0.250 (1.-2.)
3. Respondent burden and credibility	0.375 (3.-4.)	0. 250 (5.)	0.375 (3.-4.)	0.625 (1.-2.)	0.625 (1.-2.)
3.1 Difficulty and clarity of the method for the respondent	0.250 (3.-4.)	0.125 (4.-5.)	0.250 (3.-4.)	0.375 (1.-2.)	0.375 (1.-2.)
3.2 Credibility of the method for the respondent	0.125 (3.-5.)	0.125 (3.-5.)	0.125 (3.-5.)	0.250 (1.-2.)	0.250 (1.-2.)
4. Confidentiality protection of respondents' data	0.375 (3.-5.)	0.375 (3.-5.)	0.375 (3.-5.)	0.500 (1.-2.)	0.500 (1.-2.)
4.1 The degree of risk of disclosure of sensitive information	0.250 (1.-5.)	0.250 (1.-5.)	0.250 (1.-5.)	0.250 (1.-5.)	0.250 (1.-5.)
4.2 Leakage of sensitive information when knowing the result of a randomized response	0.125 (3.-5.)	0.125 (3.-5.)	0.125 (3.-5.)	0.250 (1.-2.)	0.250 (1.-2.)

Source: Own construction

CONCLUSION

We proposed methodology of multicriteria evaluation of randomized response techniques for population mean, which incorporate all, both statistical and non-statistical issues relevant for implementation in real-life survey using randomized response. Based on extensive review in the literature, we identified four ranges of criteria relevant to randomized response techniques: statistical properties of estimator, implementation and parameter choice, respondent burden and credibility and confidentiality protection of respondents' data. This methodology was applied to compare standard techniques based on scramble variables and recently proposed techniques using dichotomous response. The techniques using dichotomous response are superior, because of its ease of implementation, higher comfort of respondents (answer "Yes/No" instead of calculations) and better protection of sensitive data. Even if interviewer knows the results of randomized response, he cannot guess the true sensitive values. Because both past and recent studies (Azeem and Ali, 2023 among others) focus mostly on statistical properties of the estimator, we proposed for the future research methodology and set of model populations with different shapes of distribution to compare estimators more objectively. This methodology must study the influence of the following factors affecting the performance of the studied techniques: shape, location, and variability of studied sensitive variable, population and sample size, choice of parameters (i.e. scramble variables). The assessment of extended scope of methods from this paper using the set of these model populations is the topic of our paper under preparation.

There are several areas for the future research. The first area is to assess performance methods and variance estimators (covering all methods in Azeem and Ali, 2023) depending on the size of sample, population and the shape of sensitive variables. The second one is to apply these techniques for the real data – wage data of the Czech and Slovak Republic. The paper under preparation focuses also on important application in official statistics – estimation of year-on-year growth of mean of sensitive variable. Applying the simulation study with the use of distributions of sensitive variables with different shape, we also resolved parameter choice of interval of random number and tuning parameter α. The use of robust techniques for estimates applying dichotomous variables, construction of variance estimators using of robust methods and bootstrap techniques represent open problems.

ACKNOWLEDGMENT
This work was supported by Institutional Support to Long-Term Conceptual Development of Research Organization, the Faculty of Informatics and Statistics of the Prague University of Economics and Business, Prague. The authors are grateful to the unknown reviewers for their valuable comments that considerably improved the contents of this paper.

References

ANTOCH, J., MOLA, F., VOZÁR, O. (2022). New Randomized Response Technique for Estimating the Population Total of a Quantitative Variable [online]. *Statistika: Statistics and Economy Journal*, 102(2): 205–227. <https://doi.org/10.54694/stat.2022.11>.

ARCOS, A., RUEDA, M., SINGH, S. (2015). A generalized approach to randomized response to quantitative data. *Quantity and Quality*, 49: 1239–1256.

ARNAB, R. (1994). Nonnegative variance estimation in randomized response surveys [online]. *Communication in Statistics – Theory and Methods*, 23(6): 1743–1752. <https://doi.org/abs/10.1080/03610929408831351>.

ARNAB, R. (1995). Optimal estimation of a finite population total under randomized response surveys. *Statistics*, 27: 175–180.

AZEEM, M., ALI S. (2023) A neutral comparative analysis of additive, multiplicative, and mixed quantitative randomized response models [online]. *PLoS ONE*. 18(4): e0284995. <https://doi.org/10.1371/journal.pone.0284995>.

BAR-LEV, S., K., BOBOVITCH, E., B, BOUKAI (2004). A note on randomized response models for quantitative data [online]. *Metrika*, 60: 255–260. <https://doi.org/10.1007/s001840300308>.

BLAIR, G., KOSUKE, I., ZHOU, Y. (2015). Design and analysis of randomized response technique. *Journal of American Statistical Association*, 101: 1304–1319.

CHAUDHURI, A. (1987). Randomized response surveys of a finite population: A unified approach with quantitative data [online]. J*ournal of Statistical Planning and Inference*, 15: 157–165. <https://doi.org/10.1016/0378-3758(86)90094-7>.

CHAUDHURI, A., MUKERJEE, R. (1988). *Randomized Response, Theory and Techniques.* New York: Marcel Dekker.

CHAUDHURI, A., CHRISTOFIDES, T. (2013). *Indirect Questioning in Sample Surveys.* Heidelberg: Springer. ISBN 978-3642-36275-0

CHAUDHURI, A., CHRISTOFIDES, T., RAO, C. (2016). *Handbook of Statistics 34. Data Gathering, Analysis and Protection of Privacy Through Randomized Response Techniques.* Amsterdam: Elsevier. ISBN 978-0444-63570-9

COBO RODRIGUEZ, B. (2018). *Using auxiliary information in indirect questioning techniques.* Doctoral thesis, University of Granada.

DIANA, G., PERRI, P. F. (2013). Scrambled response models based on auxiliary variables. In: TORELLI, N., PESARI, F., BAR-HEN, A. (eds.) *Advances in Theoretical and Applied Statistics*, Berlin: Springer-Verlag, 281–291.

EICHHORN, B., HAYRE, L. S. (1983). Scramble randomized response methods for obtaining sensitive data. *Journal of Statistical Planning and Inference*, 7: 307–316.

ERIKSSON, S. (1973). A new model for randomized response [online]. *International Statistical Review*, 41: 101–113. <https://doi.org/10.2307/1402791>.

HORVITZ, D., THOMPSON, D. (1952). A generalization of sampling without replacement from a finite universe [online]. *Journal of American Statistical Association*, 47: 663–685. <https://doi.org/10.1080/01621459.1952.10483446>.

TILLÉ, Y. (2006). *Sampling Algorithms.* New York: Springer. ISBN 978-0387-30814-2

VOZÁR, O. (2023). Unbiased variance estimator of the randomised response techniques for population mean [online]. *Statistika: Statistics and Economy Journal*, 103(1): 113–120. <https://doi.org/10.54694/stat.2022.38>.

WARNER, S. (1965). Randomized response: A survey technique for eliminating evasive answer bias [online]. *Journal of American Statistical Association*, 60: 63–69. <https://doi.org/10.2307/2283137>.

Determinants of Total Factor Productivity in India: an Econometric Analysis

Aamir Ahmad Teeli[1] | *Central University of Tamil Nadu, Thiruvarur, India*
Suadat Hussain Wani[2] | *University of Kashmir, Srinagar, India*
Adil Hussain Reshi[3] | *Baba Ghulam Shah Badshah University, Rajouri, India*

Received 9.4.2023 (revision received 16.7.2023), **Accepted** (reviewed) 2.10.2023, **Published** 15.12.2023

Abstract

Productivity rise as a source of growth has high potential to uplift millions out of poverty. Mitigating the productivity gap can play vital role in helping developing countries to offset with developed countries. In this regard, present study attempts to scrutinize the relationship and significance of various determinants of total factor productivity (TFP) in India for the period 1990–1991 to 2019–2020. To this end, autoregressive distributed lag bounds testing approach has been utilized. The results pinpoint that trade openness and unemployment have a positive impact on total factor productivity. On the other hand, inflation has a negative but insignificant relationship with total factor productivity, both in the short and long run. The results imply that exports need to be diversified, and import of capital goods should be encouraged to enhance TFP in the long run. Further, foreign-owned business entities need to be encouraged to hire domestic workers, which would result in enhanced TFP in the long run.

Keywords	***DOI***	***JEL code***
Economic growth, total factor productivity, ARDL, trade openness	*https://doi.org/10.54694/stat.2023.15*	*O4, O40, F43*

INTRODUCTION

Economic growth is of the first magnitude for the prosperity, plenteousness and development of a country in terms of its economy. It is the difference in economic growth that divides the world into developed, developing and underdeveloped economies. For developing countries like India and China, it is substantial to utilize all their available resources. Apart from the efficient employment of labor and capital, attention requires to be given to factor productivity. The contribution of productivity rise in gearing up the pace of economic growth is well documented in growth literature. It has been noted that the Asian Tigers (Hong Kong, Singapore, South Korea, and Taiwan) and economic performance of Japan have recently geared up because of consistent productivity improvements (Simbeye, 1992; World Bank, 1993). Nourishing factor

[1] Department of Economics, Central University of Tamil Nadu, Thiruvarur 610 005, Tamil Nadu, India. E-mail: imamireconomics@gmail.com.
[2] Department of Economics, University of Kashmir, Hazartbal 190006, Srinagar, India. E-mail: suadat.scholar@kashmiruniversity.net.
[3] Department of Economics, Baba Ghulam Shah Badshah University, Rajouri 185234, Jammu and Kashmir, India.

productivity provides an opportunity for countries to increase the welfare of their people and achieve higher levels of economic growth and development.

In neoclassical growth accounting framework, economic growth arises from two sources: factor accumulation and productivity growth. However, neoclassical paradigm sees efficiency and technological progress as exogenous processes. But later on, during the mid-1980s, with emergence of new growth theorists, the standard neoclassical growth models (Solow, 1956; Ramsey, 1928; Samuelson, 1958) were challenged. New growth theorists formally incorporated technical progress to take into account what may be called "non-traditional" determinants of growth. Thus, endogenous growth models have brought into focus the role played by endogenous policy changes in affecting the efficiency of factors of production and, ultimately their productivity. In fact, this means that many of the growth determinants involved can only indirectly affect production. Rather, these factors affect the efficiency of real inputs, including capital, labour and possibly human capital. Therefore, these determinants of production growth directly affect total factor productivity (TFP). So any economic theory has to make a choice which factors to identify as a source of economic progress. In economic literature, performance is most frequently measured by levels of profitability, liquidity, stability and productivity indicators (Latruffe et al., 2016). The latest named indicator, i.e., productivity, which is issued to be discussed in this paper, is not easily measured and there are multiple approaches to its operationalization (Capalbo et al., 2015). While the widely used partial factor productivity may be defined as the ratio of output over a specific input (for instance, labor productivity is often measured as total sales over the number of employees; see Machek and Hnilica, 2012), the approach of total factor productivity (TFP) allows for aggregating the set of outputs and inputs (Sheng et al., 2017). At the same time, using TFP as a productivity measurement primarily reflects a multi-input production system (Gaitán-Cremaschi et al., 2016). Unlike the measures of simple profitability and partial productivity, TFP measure allows for the observation of how efficiently producers make use of multiple market inputs to produce the total output (Sheng et al., 2020). Hence, to serve the goal of this study, the TFP approach has been used.

For developing countries like India, maintaining persistently high economic growth is necessary in order to cope with the challenges they are facing. Economic growth can only be maintained when all the sources of growth are taken care of. Productivity increase as a source of growth has the high potential to get millions out of poverty for the simple reason that labor is the only asset that belongs to the lowest sections of the country. Filling the productivity gap can play a key role in helping developing countries catch up with developed countries. Aiyar et al. (2013) argued that growth in China and India can be explained by total factor productivity growth. Talking about the TFP growth in India since reforms, it ranges from –2.8 to 3.3 during the period from 1991 to 2019 (KLEMS data 2022). The total productivity growth rate was 0.66 percent between 1980 and 1993; it increased to 0.97 percent during 1994–2002, and it further increased to 1.84 percent annually during 2003–2014, showing acceleration over time. The average annual growth rate of labor productivity increased from 3.4% during 1980–2002 to 6.4% during 2003–2014, showing that labor has become more productive over time. In comparison, between 1980 and 2002, capital productivity growth remained positive at 0.06%. But it decelerated during 2003–2014 to –0.65%. Over the period of 1980–2014, the Indian economy experienced a small negative growth in capital productivity (Golder et al. 2017).

There are many macroeconomic variables having the potential to affect total factor productivity of an economy. Openness, human capital investments, exchange rate, macroeconomic stability, unemployment rate, fiscal policies and many more. In recent years, many studies in the literature have attempted to analyse how various macroeconomic policies affect TFP growth. However, the impact of most of these macroeconomic factors on total productivity remains an open question with no clear conclusions. In this regard, the present study seeks to analyse the effect of macroeconomic factors on TFP in India. This paper is organized as follows: Section 1 presents review of existing literature regarding the potential determinants of TFP. Section 2 explains the data and methodology used in the study. Section 3 presents empirical analysis including results and discussion and final section provides the conclusion with some policy implications and scope for future research in this field.

1 REVIEW OF LITERATURE

As mentioned earlier, according to new growth theorists, there are two main sources of growth: factor accumulation and productivity growth, and this productivity growth is determined endogenously by various factors. There are many macroeconomic variables that have the potential to affect the total factor productivity of an economy. Openness, human capital investments, macroeconomic stability, unemployment rate, fiscal policies, and many more. Some potential factors which are examined in the study are listed below, along with their theoretical explanations.

1.1 Openness and TFP

Trade has the possible capacity to affect total factor productivity in the economy in several ways. Grossman and Helpman (1991), Romer (1992), and Barro and Sala-I-Martin (1995) were among others who argued that more open countries have a greater ability to absorb technological advances generated in the leading nations. However, Coe and Helpman (1995) argued that the transfer of technology and knowledge spillovers from advanced countries to developing countries through trade routes would be more effective in economies with better and more advanced education.

The routes via which openness can have potential impacts on productivity levels include making use of economies of scale, competition exposure, and comparative advantage. However, endogenous growth theory has expanded on the idea of scale economies, implying that trade may increase the generation and diffusion of knowledge through mechanisms like "learning by doing." According to the economic literature, FDI is essential for the expansion of TFP. Technology transfer and job creation are both benefits of FDI. It creates positive externalities through connections with regional suppliers and clients, expertise gained from adjacent international companies, staff training programmes, and knowledge spillovers to the domestic economy. However, there is also some literature that supports the opposite evidence, Grossman and Helpman (1991) argue that trade could hurt productivity, when competition and access obstacles are increased, negative externalities may result. Additionally, productivity may suffer in cases where FDI results in significant reverse flows, such as repatriation of earnings and dividends.

Myriad researchers have strived to investigate the openness–TFP growth nexus using both micro- and time-series cross-analysis. Some scholars support a positive linkage between the two (Dollar, 1992; Coe and Helpman, 1995; Sachs and Warner, 1995; Edwards, 1998; Cameron, Proudman and Redding, 1999; Miller and Upadhyay, 2000). Coe and Helpman (1995) studied the impact of openness on TFP in the Organisation for Economic Cooperation and Development (OECD) countries, the results encouraged the view that trade openness enhances technology transfer and thus causes the TFP growth. Edwards (1998) tries to solve measurement and endogeneity problems associated with some previous studies on openness and the TFP issue by using nine indices of trade policy and using instrument variable regression. He found a positive correlation between openness and productivity growth that was robust to the use of an openness indicator, an estimation technique, a time period, and a functional form. Rodriguez and Rodrik (2001) found that trade openness is positively correlated with total factor productivity growth. Dollar and Kraay (2003) findings indicated that greater trade openness is associated with higher total factor productivity. Xie and Zhang (2021) research demonstrated a positive relationship between openness and total factor productivity.

1.2 Inflation and TFP

Another set of variables that has the potential to play a role in determining the level of productivity is macroeconomic stability. A stable macroeconomic environment can influence the quality and efficient allocation of factors of production and their rate of utilization. Inflation, inflation variability, and exchange rate stability are some of the parameters of macroeconomic stability. These variables affect confidence, which in turn affects the efficient utilization of resources. However, theoretical and empirical works on this subject have not specifically focused on TFP. Therefore, it is unclear how most of these macroeconomic environmental factors impact TFP. Economic uncertainty and price variability may induce excess capacity and, hence, cause underutilization of factors (Fischer, 1993). Additionally, an inefficient combination of input resources may result from inflation. It may lead to an incorrect understanding of the relative pricing level, which might result in ineffective investment plans that have a negative impact on productivity growth (Clark, 1982). Positive productivity growth effects can arise from improvements in the innovation process due to inflation, as suggested by Mansfield (1980). Inflation causes uncertainty about the relative price levels and hence leads to inefficient investment plans, which thereby affect productivity inversely (Clark, 1982). Inflation disrupts investment plans by imposing a higher rate of tax on corporate profits (Feldstein, 1983), and these higher effective tax rates on corporate income affect productivity. Inflation distorts price signals and reduces the ability of economic agents to operate efficiently, affecting their productivity (Smyth, 1995).

Kose and Terrones (2020) found a negative relationship between inflation and total factor productivity growth. Månsson and Olsson (2021) examined the impact of inflation on total factor productivity growth from a New Keynesian perspective. The results revealed a negative relationship between inflation and productivity growth. Jarrett and Selody (1982), using a VAR model on the inflation rate and productivity growth, found that for every additional 1 percent of inflation, labor productivity growth in Canada falls by 0.3 percent. Selody extended earlier work by Clark (1982), using VAR methods, and revised this estimate to 0.2 percent for Canada and 0.1 percent for the U.S. (Ram, 1984). Applying similar methodologies to those of Jarrett and Selody (1982) to U.S. inflation and productivity growth data found Granger causality running from inflation to productivity, not the other way around, and found productivity growth decreased by 0.9 percent with every 1 percent increase in the rate of inflation. Grimes (1990) using time series data from 21 industrial countries for the period of 1962–1987, the results concluded that 1 percent extra inflation would reduce the rate of productivity growth by 0.1 percent on average. While some studies, including those by Berlemann and Christmann (2016), Sbordone and Kuttner (1994), Cameron, Hum, and Simpson (1996), Freeman and Yerger (1997), and Hondroyiannis and Papapetrou (1998), argue that the negative connection between inflation and productivity growth is spurious and is due to cyclical co-movements in the two variables.

1.3 Unemployment and TFP

The theoretical relationship between unemployment and productivity is yet to be resolved in the empirical economic literature. Apart from representing a colossal waste of a country's manpower resources, unemployment leads to lower income and well-being and a low level of skill acquisition, thereby impacting productivity levels (Akinboyo, 1987; Raheem, 1993). While some researchers argue that there is a positive relationship between the two (Diachavbre, 1991; Krugman, 1994). The relationship between the unemployment rate and productivity is such that when there is a high rate of unemployment in an economy, the opportunity for job losses increases, which encourages the workers to work efficiently without shrinking. This would imply that when there is a high unemployment rate, the efficiency of current workers goes up. Thus, the threat of losing a job during times of high unemployment may force workers to work to their full capacity, which ultimately can increase productivity. However, in some situations, it can work the opposite way as well. While observing an ample supply of laborers, foreign companies can set in and hire more people, which will increase employment and can contribute to economic growth and overall productivity. By creating more job opportunities, foreign companies can help reduce unemployment rates and provide income and stability to individuals, which can, in turn, positively impact productivity at the macro level.

The unemployment-TFP growth nexus has remained highly controversial and inconclusive among economists. There has been a surge in empirical studies concerning the unemployment-productivity growth nexus in developed as well as developing countries, as opined by Blanchflower and Oswarld (1994), Blanchflower and Katz (1999), and Bell et al. (2002). Hence, there is ample justification why analysing the productivity and unemployment nexus within the labor markets has received a relevant amount of attention in the economic literature. Tobin (1993) argued that when there is a short-run technology shock, it may induce a negative effect on employment and a positive effect on unemployment. McConnell and Brue (1986) argued that TFP growth is impaired by unemployment. Basu et al. (2006) find evidence for a negative relationship between the two using the VAR methodology. Contrary to this, Uhlig (2006) pointed out that the correlation between productivity growth and unemployment is positive, less volatile, and more persistent, such that this correlation varies with the span of time under consideration. Gordon (1997) also pointed out that there is a link between productivity and unemployment, and this link implies a time frame, especially when looked at from the long-run perspective. Amassoma and Nwosa (2013) show an insignificant effect of unemployment on productivity. Grubb, Jackman, and Layard (1982), and Braun (1984) put forth an explanation advocating that the link between unemployment and productivity depends upon what he describes as "wage aspirations", which adjust slowly to shifts in productivity growth.

It has been suggested in the literature that the type and magnitude of the effect of these macroeconomic factors on TFP growth is influenced by the economic and social climate or the environmental quality. The quality of the environment is related to the degree of technical development, political stability, and the quality of institutions. It should be pointed out that no consensus has emerged on the nature of the impact of these factors on TFP. While some studies found a positive impact of a few factors on TFP, others have reported a negative impact. Yet, a few others have reported no impact. Although there are some research works that discuss the issue of total factor productivity and unemployment. However, it is crucial to evaluate the credibility and relevance of these sources to ensure their applicability to the specific research context.

2 METHODOLOGY

Model specification

Following the KLEMS approach, the present study estimates Formula (1):

$$logTFP_t = \propto_0 + \propto_1 logTOP_t + \propto_2 logINF_t + \propto_3 logUEMP_t + \varepsilon_t \quad (1)$$

In Formula (1), TFP represents Total factor productivity at time t, TOP shows trade openness, INF inflation, UEMP unemployment rate, and ε is the error term.

2.1 Autoregressive distributed lag approach of cointegration

ARDL is a statistical technique that is used in modeling the variables are stationary of I(0) or I(1) or a combination of both. In such a situation applying simple cointegration becomes impossible. Pesaran and Shin (1995) and Pesaran et al. (1996) proposed the Autoregressive Distributed Lag (ARDL) approach of cointegration or bound procedure for analysing the long-run relationship between the variables which are either level stationary or stationary of first difference or combination of both. Here autoregressive refers to gaps in the dependent variable, and distributed lags refer to lags in the independent variable. In economic analysis, Autoregressive distributed lag model (ARDL) – Pesaran and Shin (1999) and Pesaran et al. (2001), Granger (1981), Engle and Granger (1987), and Johansen and Juselius (1990), cointegration techniques have popularly become the methods for estimating long-run relationships between the variables that are non-stationary as it is simple to apply and easy to interpret and involves just a single equation. ARDL cointegration technique does not require pre-testing for unit roots, unlike other Techniques (Emeka and Aham, 2016). However, it is necessary to carry unit root tests in order to avoid the model crash in case if any of the variables come out to be I(2). So ARDL model is used to determine the relationship between the variables having different order of integration (Pesaran and Shin, 1999, and Pesaran et al., 2001), while the issue of variables having different lags is also sorted out in the ARDL technique. The issue of variables having different optimal lags can be dealt with easily while using ARDL, which is impossible to take into account in case of cointegration, which makes ARDL superior to its alternatives (Duasa, 2007). Before applying this statistical method, one must make sure that none of the variables is I(2). Finally, a dynamic error correction model (ECM) can be derived from ARDL through a simple linear transformation. The technique of cointegration (ARDL) is specified as follows:

$$\Delta TFP_t = \propto_0 + \sum_{i=0}^{n} \propto_{1i} LnTFP_{t-i} + \sum_{i=0}^{n} \propto_{2i} LnTOP_{t-i} + \sum_{i=0}^{n} \propto_{3i} INF_{t-i} + \sum_{i=0}^{n} \propto_{4i} LnUNEMP_{t-i} + \beta_1 LnTFP_{t-1} + \beta_2 LnTOP_{t-1} + \beta_3 INF_{t-1} + \beta_4 LnUNEMP_{t-1} + \epsilon_t \,. \quad (2)$$

The parameters ($\propto_0 - \propto_4$) measure short-run relationships, and ($\beta_1 - \beta_4$) capture long-run relationships among variables. Formula (2) would be estimated through the ARDL framework. After checking for cointegration, the next step is to estimate the "error correction model" mainly for two purposes. First, it helps to examine short-run dynamics. Second, to provide information about the speed of adjustment in the model. Thus given these benefits, the following ECM models have been specified:

$$\Delta TFP_t = \propto_0 + \sum_{i=0}^{n} \propto_{1i} LnTFP_{t-i} + \sum_{i=0}^{n} \propto_{2i} LnTOP_{t-i} + \sum_{i=0}^{n} \propto_{3i} INF_{t-i} + \sum_{i=0}^{n} \propto_{4i} LnUNEMP_{t-i} + \lambda_1 ECM_{t-1} + \epsilon_t \,. \quad (3)$$

In Formula (3), the term ECM denotes the error correction term, which shows the speed of adjustment, whereas other variables have already been defined.

2.2 Data source and measurement of variables

The present study uses annual time series data from 1990–91 to 2019–20. The data for the dependent variable, total factor productivity, has been collected from the KLEMS database of RBI. There are different approaches of measuring TFP. However, there is no consensus on which of the measure is best for all purposes (Mehadevan, 2003). For the present study, the data on Total factor productivity estimated by KLEMS is directly used without making any changes. In the KLEMS approach, total factor productivity is measured by applying growth accounting. For measuring the total factor productivity of individual industries, the gross production function framework is used, taking capital, labor, energy, materials, and services as five inputs. Then, aggregation of these individual measurements is done in order to reach the total factor productivity. Data on openness (Trade as a percentage of GDP, where trade includes the total volume of both imports as well as exports) is used. Unemployment rate (usual principal status approach) as a percentage of the total labor force is applied. Annual Inflation at consumer prices (CPI, base year 2011–12) extracted from the World Bank database has been used in the present study.

3 EMPIRICAL RESULTS AND DISCUSSION

Before going for any econometric analysis, it is mandatory to check the properties of data variables. Table A1 in the Annex presents descriptive nature of the variables used in the study. In addition to mean, median and standard deviation, results of kurtosis and skewness are also presented to get a better idea about the variables used in the study. Mean and standard deviation are presented for determining range and coverage of data set. Besides, kurtosis and skewness measure the peakedness or flatness and degree of asymmetry of the series respectively. Similarly, results of Jarque-Bera test confirm that all the variables are normally distributed and suitable for further econometric analysis. The mean value of variables is found to be greater than standard deviation, which indicates stability of variables over the study period.

3.1 Unit root results

The second step is to test stationary properties of selected variables. According to Hill et al. (2001), regression should not be applied to variables that are non-stationary as it would lead to spurious regression. Thus, it is necessary to check stationary nature of data set. In this study, ADF and Phillips-Perron (PP) have been used to check stationary nature of variables. The results of both these tests are presented in Table 1. TFP is stationary at level with intercept in both tests, and for trend and intercepts, it is stationary only in the case of ADF. Openness and inflation are non-stationary at level for both tests. The unemployment rate is stationary at level with intercept and trend for both the tests; while when considering only intercept, it is stationary only in Phillips-Perron test. At first difference, all the included variables are found to be stationary.

Table 1 Unit root test results

	Intercept		Intercept & trend	
Variables	**ADF test-stat.**	**PP test-stat.**	**ADF test-stat.**	**PP test-stat.**
Level				
Tfp	−3.75(0.008)	−3.35(0.021)	−3.64(0.042)	−3.18(0.107)
Opn	−1.48(0.528)	−1.48(0.529)	−0.52(0.976)	−0.70(0.962)
Inf	−2.87(0.063)	−2.37(0.157)	−2.96(0.160)	−2.77(0.216)
Ur	−0.05(0.945)	−2.81(0.068)	−8.18(0.000)	−15.45(0.000)
First diff.				
ΔTfp	−6.95(0.000)	−7.25(0.000)	−7.04(0.000)	−7.31(0.000)
ΔOpn	−4.55(0.001)	−4.58(0.001)	−4.58(0.001)	−4.79(0.003)
ΔInf	−7.38(0.000)	−7.70(0.000)	−7.23(0.000)	−7.54(0.000)
ΔUr	−6.39(0.000)	−11.95(0.000)	−6.43(0.000)	−12.52(0.000)

Note: The corresponding p-values are given in parenthesis, ADF and PP are determined by Mackinnon's (1996) formula.
Source: Author's calculations

Table 2 ARDL bound test estimates

K	**F-statistic**	**Significance (%)**	**I(0)**	**I(1)**
		10%	2.37	3.20
3	8.411	5%	2.79	3.67
		1%	3.65	4.66

Source: Author's calculations

From unit root test, we found that the variables of the study were a combination of I(0) and I(1). Thus, given the stationary nature of data, ARDL modelling and Bounds testing are found to be optimal choice for estimation. It is important to mention here that bound test is applied to examine whether a long-run relationship exists or not. In bound testing approach, null hypothesis of no cointegration among variables is tested against alternative hypothesis of cointegration. Results of bounds test in Table 2 provide sufficient evidence for existence of long-run relationship, as F-Statistic was found to be greater than upper bond values at 5% level of significance. It implies rejection of null hypothesis and acceptance of alternative hypothesis. Therefore, bound test confirms long-run relationship among the variables under consideration. Next, we proceed to estimate long-run and short-run effects of selected variables on TFP in India.

3.2 Long-run results

The long-run results are presented in Table 3. The results reveal that most of the variables are statistically significant and have expected signs. Trade openness has a positive relationship with TFP, where 1% increase in openness increases TFP by 0.115%. The results align with findings of Dollar (1992), Coe and Helpman (1995), Lee (2004), Nachega and Fontanie (2006), Thomson (2006), and Bandara and Karunaratne (2013). Nachega and Fontaine (2006) have shown positive impact of trade openness on productivity in case of Niger though he has not provided detailed explanation for his findings. However, in case of India, Malik et al. (2021) have found negative impact of exports on total factor productivity and positive impact of imports and capital formation. The possible reason for negative impact of exports is basket of commodities exported by the country. In case of positive impact of imports, Kim et al. (2009) found that imports promote total factor productivity through increase in competition among domestic firms owing to import of consumer goods and technology diffusion embodied in imports from developed countries. In case of Sri Lanka, Bandara and Karunaratne (2013) found positive impact of export oriented industralization on growth performance of the country as compared to inward looking import substitution industralization. Their study concluded that the failure of the country to embark on the path of trade openness at right time deprived it of the opportunity to be one of the high performing miracle economies of Asia. The main source from which a developing country can benefit through openness is to establish political stabilty to attract foreign investment and tecnnology which enhances productivity and growth in long run. Sanjoy Saha (2012) using three proxies as Trade GDP, Export-GDP, and Import-GDP ratio, found evidence that all three proxies have a positive and significant impact on TFP.

Table 3 Long-run estimates

Variables	Coefficient	Std. error	t-statistic	Prob.
Opn	0.115	0.049	2.34	0.028
Inf	–0.169	0.109	–1.54	0.135
Ur	8.580	5.747	3.05	0.006

Source: Author's calculations

The findings of present study are in line with the above discussed literature. Trade openness in the present study includes both exports and imports and in case of India, there is huge trade deficit indicating high volume of imports as compared to exports. The findings of present study indicate a positive impact of imports through increased competition in domestic market and technological diffusion.

Inflation is negatively related to TFP though it is statistically insignificant. The results are in line with prior expectations (Akinlo and Adejumo, 2016; Gomez, 2020; and Johnson et al., 2021). There are a number of reasons which explain the negative relation between inflation and factor productivity. First, continuous and high rises in prices create uncertainty in markets regarding profitability and future investment. Second, rising prices discourage investors from investing in projects which leads to an increase in factor productivity in the long run. Finally, rising prices reduce demand for real money balance and thus lead to a decline in the growth rate of factor productivity.

Finally, the unemployment rate shows a positive relationship with TFP, with a coefficient of 8.58, which aligns with the findings of Lach and Tsiddon (1992), Uhlig (2006), Bassanini and Duval (2006). This could be due to factors such as labor reallocation, threat of losing jobs which leads to increased efforts on the part of workers, and increased competition for jobs leading to higher productivity standards, all this can lead to an increased effort and productivity, or it can be attributed to the effects of technological advancements replacing certain jobs with more productive alternatives. Here it's crucial to remember that this relationship might change based on the situation and specific factors at play.

3.3 Short-run results

Table 4 presents the result of the short-run analysis. The results show that trade openness has a negative impact on TFP, with 1% change in openness leads to 0.17 % decline in factor productivity. However, taking lagged value into account, it shows that trade in the previous period has a positive impact on factor productivity, which is in line with the theoretical background. These results align with findings of Malik et al. (2021) who have found the same sort of relation between openness and productivity. The coefficient of inflation shows negative relation with TFP but again it is statistically insignificant. Similarly, the results show that unemployment has a positive relationship with TFP with a coefficient of 11.8, which means a 1% increase in the unemployment rate increases TFP by 11.8 % in the short run.

Table 4 ECM of ARDL model (short-run estimates)

Variables	Coefficient	Std. error	t-statistic	Prob.
Opn	–0.173	0.065	–2.62	0.015
Opn_{-1}	0.077	0.026	2.91	0.007
Inf	–0.114	0.070	–1.47	0.154
Ur	11.81	2.060	4.44	0.000
Coint eq.(–1)	–0.672	0.090	–7.02	0.000
R-square		Adjusted R-square		
0.69		0.67		

Source: Author's calculations

The results show that sign of lagged error correction representation ($ECMt-1$) is statistically significant at 1% level with a negative sign. It shows the existence of a relationship between the variables used in the study. The coefficient indicates the speed of adjustment from short run disequilibrium to long run equilibrium. The value of the coefficient is –0.672, which indicates that from short run to long run, adjustments are corrected by 0.672% every year. It should be noted that the larger the error term, the faster the rate at which economies return to equilibrium. The value of adjusted R-square is 0.67, which shows that 0.67% of changes in dependent variables are explained by independent variables. Thus, the results show that the model used in the present study is robust and best fitted.

4 SENSITIVITY ESTIMATES

Table 5 shows that there is no serial correlation in the model, confirmed by Breusch–Godfrey serial correlation L.M. test. Besides, to check for autoregressive conditional heteroscedasticity, Breusch–Pagan–Godfrey and Glejser were run. The results presented that there is no sign of heteroscedasticity. Table 5 shows values of F-statistics along with probability value which is greater than a 5% significance level. Thus, these findings represent that our model is the best fit and perfect.

Table 5 Diagnostic tests

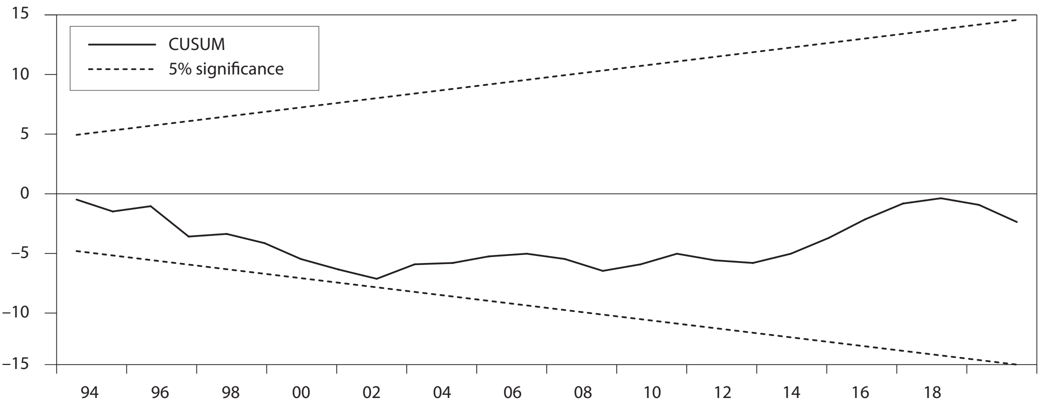
	Obs*R-square	F-statistic	Prob.
Breusch–Godfrey serial correlation LM test	0.505	0.186	0.77
Heteroskedasticity test Breusch–Pagan–Godfrey	5.719	1.301	0.33
Heteroskedasticity test Glejser	3.651	1.140	0.34

Source: Author's calculations

4.1 Stability estimates

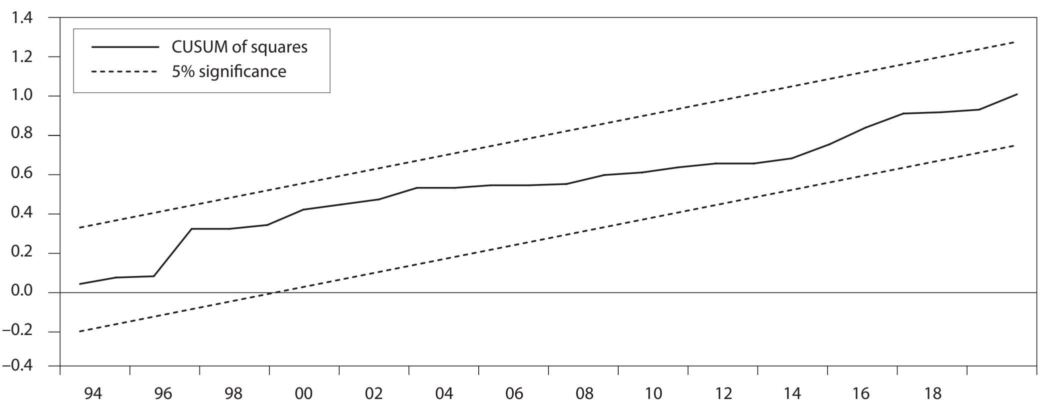
To test stability, CUSUM and CUSUMQ were used. There is a rule of thumb; if the blue line is within red dashed lines, it indicates that the model is stable. Two red lines represent a 5% level

Figure 1 CUSUM results



Source: Author's estimation

Figure 2 CUSUM results



Source: Author's estimation

of significance. The results of both presented in Figure 1 and Figure 2, respectively, validates the stability of our model.

CONCLUSION

The present study attempted to examine the impact of openness, inflation and unemployment on TPF in India in post-reform period. The results of this study provide valuable insights about short-run and long-run relationship between trade openness, inflation, unemployment and TFP. The results indicate that openness and unemployment rate impact TFP positively, while inflation does not have any significant impact. Based on the results, it can be concluded that due to limited job opportunities, workers do their work efficiently, which results in an increase in their productivity. From the findings about trade openness, it can be concluded that increase in competition and technology transfer are drivers of TPF and hence need to be encouraged. This highlights the necessity to promote diversified exports and import such goods that would lead to transfer of technology into the nation. Further, foreign-owned business entities need to be encouraged to hire domestic workers, which would result in enhancement of TFP in long run. Openness need to be increased by enhancing trade and foreign investment. Thus, it can be concluded that initiatives like Make in India, export promotion and reduction in trade barriers have high potential to increase TFP in the country.

In future, determinants of sectoral productivity can be an interesting area of research as certain industries or sectors may have more significant impact on overall productivity of a country due to their size, technological advancements or labour-intensive nature. Therefore, it will be very interesting to focus on the specific industries that have a substantial influence on overall productivity. Also, instead of using overall unemployment, impact of changes in existing levels of employment on TFP can be analysed by future researchers in order to identify the sectors with higher levels of economies of scale.

References

AIYAR, M. S., DUVAL, M. R. A., PUY, M. D., WU, M. Y., ZHANG, M. L. (2013). *Growth slowdowns and the middle-income trap.* International Monetary Fund.

AKINBOYO, T. (1987). Unemployment and productivity: A study of the Nigerian economy [online]. *Journal of Development Economics*, 25(2): 245–259. <http://dx.doi.org/10.22004/ag.econ.307335>.

AMASSOMA, D., NWOSA, P.I. (2013). The impact of unemployment rate on productivity growth in Nigeria: An error correction modeling approach. *Journal of Economics and Sustainable Development*, 4(9): 90–103.

BASSANINI, A., DUVAL, R. (2006). Employment Patterns in OECD Countries: Reassessing the Role of Policies and Institutions [online]. *OECD Economic Studies*, 42(1): 61–96. <http://www.oecd.org/publications>.

BANDARA, Y., KARUNARATNE, N. D. (2013). Globalisation, policy reforms and productivity growth in developing countries: Evidence from Sri Lanka [online]. *Global Business Review*, 14(3): 429–451. <https://doi.org/10.1177/0972150913496810>.

BASU, S., FERNALD, J. G., OULTON, N., SRINIVASAN, S. (2006). The case of the missing productivity growth: Or, does information technology explain why productivity accelerated in the United States but not the United Kingdom? *NBER Macroeconomics Annual*, MIT Press, 20: 9–82.

BERLEMANN, M., CHRISTMANN, R. (2016). Macroeconomic determinants of European total factor productivity: The role of applied research. *Economic Modelling*, 52: 931–942.

BLANCHFLOWER, D. G., KATZ, L. F. (1999). Unemployment and productivity. In: *Handbook of Labor Economics, Elsevier*, 3: 2567–2627.

CAMERON, L. A., HUM, D., SIMPSON, W. (1996). The relationship between inflation and productivity growth: Evidence from a panel of OECD countries. *European Economic Review*, 40(3–5): 1671–1681.

CAMERON, N. D., HUM, D., SIMPSON, W. (1996). Stylized facts and stylized illusions: Inflation and productivity revisited [online]. *Canadian Journal of Economics*, 29(1): 152–162. <https://doi.org/10.2307/136156>.

CAMERON, G., PROUDMAN, J., REDDING, S. (1999). *Technology transfer, R&D, trade and productivity growth.* Working Paper, London, UK: London School of Economics.

CAPALBO, S. M., ANTLE, J. M. (2015). *Agricultural Productivity: Measurement and Explanation*. New York: Routledge.

CLARK, K. P. (1982). Inflation and the productivity decline. *American Economic Review, Papers and Proceedings*, 72(2): 149–154.

COE, D. T., HEPLMAN, E. (1995). International R&D spillovers. *European Economic Review*, 39(5): 859–887.

DIACHAVBRE, J. C. (1991). Unemployment and productivity: Evidence from cross-country panel data. *Journal of Economic Studies*, 18(4): 15–26.

DOLLAR, D. (1992). Outward-oriented developing economies do grow more rapidly: Evidence from 95 LDCs, 1976–1985 [online]. *Economic Development and Cultural Change*, 40(3): 523–544. <https://doi.org/10.1086/451959>.

EDWARDS, S. (1998). Openness, productivity and growth: What do we really know? [online]. *Economic Journa*l, 108(447): 383–398. <https://doi.org/10.1111/1468-0297.00293>.

ENGLANDER, S., GURNEY, A. (1994). Medium-term determinants of OECD productivity. *OECD Economic Studies*, 22: 49–109.

FELDSTEIN, M. (1983). Inflation, income taxes, and the rate of interest: A theoretical analysis. In: *Inflation, tax rules, and capital formation.* University of Chicago Press, 28–43.

FISCHER, S. (1993). The role of macroeconomic factors in growth [online]. *Journal of Monetary Economics*, 32(3): 485–512. <https://doi.org/10.1016/0304-3932(93)90027-D>.

FREEMAN, D. G., YERGER, D. B. (1997). Inflation and productivity growth: A panel study of the 1960s and 1970s. *Journal of Money, Credit and Banking*, 29(2): 233–247.

FRIED, H., LOVELL, C. A. K., SCHMIT, S. (1993). *The measurement of productive efficiency.* Oxford, UK: Oxford University Press.

GRIFFITH, R., REDDING, S., SIMPSON, H. (2003). *Productivity convergence and foreign ownership at the establishment level.* Discussion Paper No. 572, London, UK: Centre for Economic Performance.

GRIMES, A. (1990). The impact of inflation on productivity growth in 21 industrial countries: A time series analysis (1962–1987). *Journal of Economic Research*, 25(2): 45–62.

GROSSMAN, G. M., HELPMAN, E. (1991). Quality Ladders and Product Cycles [online]. *Quarterly Journal of Economics*, 106. <https://doi.org/10.2307/2937947>.

HANSON, G. (2001). *Should countries promote foreign direct investment?* G-24 Discussion Paper 9, Geneva, Switzerland: UNCTAD.

HASAN, R. (2002). The impact of imported and domestic technologies on the productivity of firms: Panel data evidence from Indian manufacturing firms [online]. *Journal of Development Economics*, 69(1): 23–49. <https://doi.org/10.1016/S0304-3878(02)00051-2>.

HAKEL, J. E., PEREIRA, S. C., SLAUGHTER, M. J. (2002). *Does inward foreign direct investment boost the productivity of domestic firms?* [online]. NBER Working Paper No. 8724, Cambridge, MA: National Bureau of Economic Research. <https://doi.org/10.1162/rest.89.3.482>.

HERCOWITZ, Z., LAVI, Y., MELNICK, R. (1999). The impact of macroeconomic factors on productivity in Israel, 1960–96. *Bank of Israel Economic Review*, 52: 103–124.

HONDROIANNIS, G., PAPAPETROU, E. (1998). Temporal causality and the inflation-productivity relationship: Evidence from eight low inflation OECD countries [online]. *International Review of Economics and Finance*, 7(1): 117–135. <https://doi.org/10.1016/S1059-0560(99)80020-3>.

JARRET, J. P., SELODY, J. G. (1982). The productivity–inflation nexus in Canada. *The Review of Economics and Statistics*, 64(3): 361–367.

JOFNSON, C., WILSON, D., THOMPSON, E. (2021). Inflation and total factor productivity: A cross-industry analysis. *Economic Inquiry*, 49(2): 657–676.

KELLER, W., YEAPLE, S. R. (2003). *Multinational enterprises, international trade, and productivity growth: Firm level evidence from United States* [online]. NBER Working Paper No. 9504, Cambridge, MA: National Bureau of Economic Research. <https://doi.org/10.1162/rest.91.4.821>.

KHAN, S. U. (2006). Macro determinants of total factor productivity in Pakistan. *SBP Research Bulletin*, 2(2): 383–401.

KOSE, M. A., TERRONES, M. E. (2020). *How does inflation affect productivity growth?* International Monetary Fund.

KRUGMAN, P. (1994). Unemployment and productivity: Evidence from industry panel data. *The Quarterly Journal of Economics*, 109(4): 835–864.

LATRUFFE, L., et al. (2016). Measurement of sustainability in agriculture: a review of indicators [online]. *Studies in Agricultural Economics*, 118(3): 123–130. <http://repo.aki.gov.hu/id/eprint/2092>.

LACH, S., TSIDDON, D. (1992). The Behavior of Total Factor Productivity: Evidence from Israeli Industry. *The Review of Economics and Statistics*, 74(4): 617–624.

MACHEK, O. (2014). The relationship between financial performance and total factor productivity: evidence from the Czech agricultural sector [online]. *International Journal of Social Sciences and Humanity Studies*, 6(2): 57–67. <https://dergipark.org.tr/en/pub/ijsshs/issue/26216/276012>.

MCCONNELL, C. R., BRUE, S. L. (1986). *Macroeconomics: Principles, problems, and policies.* McGraw-Hill.

MAHADEVAN, R. (2003). To measure or not to measure total factor productivity growth? [online]. *Oxford Development Studies*, 31(3): 365–378. <https://doi.org/10.1080/1360081032000111742>.

MILLER, S. M., UPADHYAY, M. P. (2000). The effects of openness, trade orientation, and human capital on total factor productivity [online]. *Journal of Development Economics*, 63(2): 399–423. <https://doi.org/10.1016/S0304-3878(00)00112-7>.

NACHEGA, J. C., FONTAINE, T. (2006). *Economic growth and total factor productivity in Niger* [online]. Working Paper WP/06/208, Washington, DC: International Monetary Fund. <https://ssrn.com/abstract=941270>.
NELSON, R. (1981). Research on productivity growth and productivity differences: Dead ends and new departures [online]. *Journal of Economic Literature*, 19(3): 1029–1064. <https://www.jstor.org/stable/2724327>.
REA, T., SHESTALOVA, V. (2006). *Alternative measures of total factor productivity growth* [online]. [cit. 24.6.2014]. <http://papers.ssrn.com/so13/papers.cfr.pdf>.
RAM, R. (1984). Causal ordering across inflation and productivity growth in the post-war United States [online]. *The Review of Economics and Statistics*, 64(3): 472–477. <https://doi.org/10.2307/1925004>.
RAMSEY, F. P. (1928). A mathematical theory of saving [online]. *The economic journal*, 38(152): 543–559. <https://doi.org/10.2307/2224098>.
SALGADO, R. (2002). *Impact of structural reforms on productivity growth in industrial countries* [online]. Working PaperWP/02/10, Washington, DC: International Monetary Fund. <https://ssrn.com/abstract=879330>.
SHARMA, C. (2016). R&D, technology transfer and productivity in the Indian pharmaceutical industry [online]. *International Journal of Innovation Management*, 20(01): 1650010. <https://doi.org/10.1142/S1363919616500109>.
SHENG, Y., BALL, V. E., ERICKSON, K., MESONADA, C. S. J. (2022). Cross-country agricultural TFP convergence and capital deepening: Evidence for induced innovation from 17 OECD countries. *Journal of Productivity Analysis*, 58(2–3): 185–202.
SIMBYE, A. (1992). *International Cooperation and Productivity: The Case of Nigeria. Productivity: Key to Economic Development.* Lagos: National Productivity Centre, 22–30.
SMYTH, D. J. (1995a). Inflation and total factor productivity in Germany [online]. *Weltwirtschfliches Archiv*, 131(2): 403–405. <https://doi.org/10.1007/BF02707442>.
UHLIG, H. (2006). Explaining the Phillips curve: Time-series versus cross-sectional information. *American Economic Review*, 96(1): 62–85.
WONG, S. T., SENG, B. S. (1997). *Total factor productivity growth in Singapore: Methodology and trends.* Paper presented at the Capital Stock Conference.
XIE, Y., ZHANG, J. (2021). Does trade openness promote productivity growth in China? A case study of different ownership sectors [online]. *Economic Modelling*, 98: 360–371. <https://www.frontiersin.org/articles/10.3389/fenvs.2021.809339/full>.

ANNEX

Table A1 Descriptive statistics

Variables	*Tfp*	*Opn*	*Inf*	*Ur*
Mean	0.583	35.347	7.286	5.536
Median	0.850	38.751	6.550	5.565
Maximum	3.300	55.793	13.900	5.750
Minimum	–2.800	16.000	3.300	5.200
Std. deviation	1.603	12.969	3.224	0.147
Skewness	–0.347	0.043	0.438	–0.393
Kurtosis	2.215	1.593	1.934	2.342
Jarque-Bera	1.373	2.482	2.380	1.315
Probability	0.503	0.289	0.304	0.518
Sum	17.500	1 060.420	218.60	166.090
Sum sq. dev.	74.581	4 878.060	301.434	0.629
Observations	31	31	31	31

Source: Author's calculations

25th International Scientific Conference *Applications of Mathematics and Statistics in Economics (AMSE 2023)*

Stanislava Hronová[1] | *Prague University of Economics and Business, Prague, Czech Republic*

The 25th International Scientific Conference *Applications of Mathematics and Statistics in Economics* was held from 30 August to 3 September 2023 in Rajecké Teplice, in the foothills of Malá Fatra.[2] The organiser of this year's conference was the Department of Quantitative Methods and Information Systems of the Matej Bel University in Banská Bystrica. The conference was attended by more than 40 experts from the Czech Republic, Slovakia, and Poland representing the Matej Bel University in Banská Bystrica, Prague University of Economics and Business, Wroclaw University of Economics and Business, University of Wroclaw, University of Pardubice, Masaryk University, Czech Technical University in Prague, University of Economics in Bratislava, Comenius University in Bratislava, the Czech Statistical Office, and the Statistical Office of the Slovak Republic.

The opening session of the conference was devoted to looking back upon the 25 years of the conference, which has been held regularly (with the exception of the Covid year 2020) since 1997 with alternating hosting among Slovakia, Poland, and the Czech Republic. Symbolic certificates of fidelity were awarded to four participants who had not missed any of the events (Rudolf Zimka and Peter Laco from the Matej Bel University in Banská Bystrica and Richard Hindls and Stanislava Hronová from the Prague University of Economics and Business). The festive nature of this opening was also added by a very kind letter of greetings from the Rector of the Matej Bel University in Banská Bystrica Vladimír Hiadlovský.

The expert programme of the conference was opened by the President of the Czech Statistical Office Marek Rojíček with an invited lecture entitled *Communication strategy and new tools for dissemination in the Czech Statistical Office*, in which he stated that challenges for official statistics can be in the simplest form characterised as the answers to the three basic questions: "What to measure?", "What are the data sources?", and finally "How to communicate?" data and related stories and how to find the most efficient way of transferring the information to individual users. The Czech Statistical Office systematically cooperates with journalists, especially in the public media. Another way of direct communication with users is through social networks (X, Instagram, and LinkedIn).

The second invited lecture on *Using of innovations and their impact on the selected conclusions of the Population and Housing Census 2021* was given by Ľudmila Ivančíková, Director of the Social Statistics and Demography Directorate of the Statistical Office of the Slovak Republic. Her contribution focused on innovations implemented in the 2021 Census of Population, Houses and Flats in Slovakia that had

1 Department of Economic Statistics, Faculty of Informatics and Statistics, University of Economics and Business, Prague, W. Churchilla Sq. 4, 130 67 Prague 3, Czech Republic. E-mail: hronova@vse.cz.

2 More information at: <*https://www.amse-conference.eu*>.

a direct impact on timeliness, completeness, data quality, and the detail of the outputs – the use of the sign of life method with its impact on completeness of coverage, fully electronic collection and item non-response rates, or data integration and the associated shorter time for processing and dissemination of the main results.

Other conference sessions were held in 6 sections: Efficiency Studies and Economic Analyses, Economic Issues and Official Statistics, Economic Demographics and Demographic Economics, Statistical Models and Techniques, Applications of Statistics and Pension Systems, Macroeconomic Models and History of Statistics. It is very difficult to highlight the most interesting contributions; therefore, I would like to highlight only some of the contributions that I consider to be of high quality, interesting, and methodologically innovative.

In the *Efficiency Studies and Economic Analyses* section, the most attention was caught by a contribution called *Comparison of the efficiency of national health systems of the European Union countries before and during the COVID-19 pandemic.* The author (M. Biernacki) tried to compare activities of national health systems aimed at ensuring the quality of health of citizens. The comparison was made in external praxeological areas, i.e. in the area of citizens' well-being and in the economic area.

Papers presented in the *Economic Issues and Official Statistics* section focused on current topics related to price developments, household budget survey, tourism taxation, etc. A contribution called *Regional price levels in the Czech Republic: recent update* by P. Musil, J. Fischerová, and J. Kahoun was devoted to the possibilities of monitoring consumer price developments by Region. Currently, the update of regional price levels seems to be very beneficial, nevertheless COVID-19 and related government measures pose further challenges. The structure of household consumption habits changed substantially, e.g. e-commerce grew up rapidly. With respect to data sources available, the updated regional price levels will be estimated for the year 2020, temporary changes in the economy will be taken out. Users will have updated estimates comparable with previous results.

The papers in the section *Economic Demographics and Demographic Economics* were mainly devoted to problems of the so-called silver generation, which Polish colleagues have been dealing with for a long time. Of these, I consider it appropriate to highlight the contribution of B. Zmyslona and A. Marciniuk entitled *Analysis of differences in the living conditions of the "silver" generation, including pension and healthcare systems.* The authors compared the living conditions of the silver generation in Poland, Slovakia, and the Czech Republic. It is clear that changes in the age structure of the population are increasing the burden on the pension and health care systems. From the point of view of social policy, this means an increase in the demand of this population group for health care and long-term care services. The burden on the health care system is assessed on the basis of the structure of expenditure on prevention and treatment incurred both by the state and by households of the elderly. The pension system is influenced by, for instance, the expected life in retirement and in the future, the current state of public spending and public debt, real economic growth, and the labour force of the elderly.

In the *Statistical Models and Techniques* section, a theoretical paper by O. Vozár and L. Marek on *Real life application of randomized response techniques to estimation of population means: a case of Slovak and Czech wage data in 2016–2019* attracted a lot of attention and discussion. The authors proceeded from the well-known fact that direct questioning of a sensitive variable leads to a high rate non-response or reporting by highly distorted values. The authors proposed how to estimate in real conditions the population mean of a quantitative variable with broad span of values in short time series to also estimate reliable year-on-year growth rates.

The section *Applications of Statistics and Pension Systems* brought papers on various topics with the dominance of the two areas mentioned in the section title. In the context of the population ageing and the need to reform pension systems, the benefit of the paper called the *Amendment to the Act No. 43/2004 Coll. on old-age pension savings and risks associated with the payment phase* should be highlighted.

The authors (J. Špirková and S. Zelinová) discussed the impact of the amendment to the Act on old-age pension savings that established a precisely defined default investment strategy, which should help savers achieve the highest possible rate of appreciation of their assets in pension companies, and thus higher future pensions. The conditions for the payment of pensions in 2023 remain unchanged, but from the beginning of 2024, the payment phase of the second pillar will fundamentally change. At retirement, the (old age) pensioner will be able to withdraw half of the amount saved; the other half will be further appreciated and paid out in the form of a lifetime pension. The wording of the amendment to the Act leads to a significant decrease in the amount of the pension payment when switching from a programme withdrawal to a lifetime pension by almost 50%, provided that the saver's savings are divided in half. The scenarios offered by authors solve and reduce the risk of the transition between programme selection and a lifetime pension.

In the last section *Macroeconomic Models and History of Statistics*, two theoretical papers were presented (*The impact of multi-dimensional model's critical parameters on the equilibrium's stability and the existence of limit cycles and On some qualitative properties of a liquidity-growth model*). The authors (R. Zimka, M. Demetrian and E. Zimková) of the latter paper presented a two-dimensional nonlinear dynamical model describing the development of liquidity and profit of a firm around its equilibrium. They analysed the qualitative properties of liquidity and profit in a neighbourhood of their boundary values which separate the domain of prosperity from the domain of bankruptcy of a firm and they showed that the method of Bautin bifurcation can describe the possible reality in relations between a firm and its bank more suitably than Hopf bifurcation method. The last paper (by P. Závodský and O. Šimpach) was traditionally devoted to the history of statistics (*230 years of statistical theory in Bohemia*).

Hiking is a traditional part of the AMSE conference, this time in the Súľovské Rocks area. A group of the fittest set out on a challenging route through Roháč saddle to the Súľov Castle and back to Lietavská Zavadka. However, most of the participants chose the medium difficulty option, i.e. a visit to the Wedding Palace in Bytča and a climb to the Súľov Castle. History lovers were not left out with a visit to the Wedding Palace, the village of Súľov, and other important historical buildings in the area.

A full programme of AMSE 2023 including abstracts of the papers presented, can be found at: <*http://www.amse-conference.eu*>. You can also find there information about the history of AMSE and links to previous years of the international conference.[3]

The tradition of alternating hosts (Slovakia – Poland – Czech Republic) continues, and the 26th AMSE conference, which will be hosted by colleagues from the Department of Statistics at Wroclaw University of Economics and Business, will be held in Pawlowice, a suburb of Wroclaw, Poland, at the turn of August and September 2024.

[3] In this report on the Conference, texts of the Book of abstracts <*www.amse-conference.eu*> were used.

17th Year of the *International Days of Statistics and Economics (MSED 2023)*

Tomáš Löster[1] | *Prague University of Economics and Business, Prague, Czech Republic*

From 7th to 9th September 2023, a worldwide conference of the *International Days of Statistics and Economics (MSED)* took place at the Prague University of Economics and Business.[2] The conference belongs to traditional professional events; this year, the sixteenth year of this event was held. Prague University of Economics and Business (the Department of Statistics and Probability and the Department of Managerial Economics) was the main organizer, as usual; and was helped by the Faculty of Economics, the Technical University of Košice, and Ton Duc Thang University, as co-organizers. The conference ranks among important statistical and economic conferences, which can be proved by the fact that Online Conference Proceedings were included in the Conference Proceedings Citation Index (CPCI), which has been integrated within the Web of Science, Clarivate Analytics since 2011.

The traditional goal of this international scientific conference was a presentation of the contributions of individual authors and a discussion of current issues in the field of statistics, demography, economics, and management and their interconnection.

This year's conference and presentation were again in a hybrid form (online and real presentation at the university), which caused the participation of foreign nationals to be active.

The online implementation of the conference took place in individual channels of the conference teams in MS Teams (according to partial sections). The number of registered conference participants was a total of 112, of which 51 were foreign, e.g. Turkey (13), Slovakia (11), etc. Among the conference participants were 16 doctoral students.

The received papers were first evaluated in terms of scientific content and suitability of the topic concerning the focus of the conference. After the exclusion of unsatisfactory abstracts, a double independent anonymous review procedure took place in the spring of this year.

We would also like to invite researchers, doctoral students, and the wide professional public to the next *International Days of Statistics and Economics*, which will take place at the Prague University of Economics and Business traditionally in early September 2024.

1 Faculty of Informatics and Statistics, Prague University of Economics and Business, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic. E-mail: tomas.loster@vse.cz.

2 More at: <*http://msed.vse.cz*>.

41st *Mathematical Methods in Economics (MME 2023)* International Conference

Petra Zýková[1] | *Prague University of Economics and Business, Prague, Czech Republic*
Josef Jablonský[2] | *Prague University of Economics and Business, Prague, Czech Republic*

The Mathematical Methods in Economics (MME) conference has a very long history and tradition. It is one of the most important scientific events organised in the Czech Republic in the fields of operational research, econometrics, mathematical economics, and related research areas. In 2023, the 41st International Conference on *Mathematical Methods in Economics* was held in the city of Prague from 13 to 15 September.[3] In addition to the local organiser (the Department of Econometrics, Faculty of Informatics and Statistics, Prague University of Economics and Business), leading organisers of the MME conference were the Czech Society for Operations Research (CSOR) and the Czech Econometric Society.

The total number of participants in this year's MME conference was more than 110. Participants came from the Czech Republic, Slovakia, Lithuania, Belgium, Spain, Turkey, Greece and the Great Britain. The programme started with an opening ceremony, where the Chair of the Organising Committee, Josef Jablonský, introduced the main programme and all the facilities. Then, the plenary session started with two exciting lectures. The first one, titled *Network Data Envelopment Analysis: The Prevalent Methodological Approaches and Some Recent Developments*, was presented by Professor Dimitris K. Despotis from the University of Piraeus, Greece. Associate Professor Miloš Kopa from Charles University, Faculty of Mathematics and Physics, Prague, delivered the second plenary speech about *Decision Making in Finance via Stochastic Dominance*. After the plenary session, the conference was divided into four parallel sessions. The total number of presentations was almost 90. All accepted papers are published in the *Proceedings of the MME 2023*. As in previous years, they have been submitted for indexing in the Web of Science and Scopus database. All accepted abstracts are published in the *Book of Abstracts of the MME 2023*.

It has been a long tradition for PhD students to compete for the best paper during the MME conferences. The competition is organised and honoured by the CSOR. All submitted papers were peer-reviewed, and the programme committee further evaluated the papers with positive referee reports. Ten best-selected papers were presented at the conference in two special sessions, and the evaluation committee decided on the winners. The six best papers were awarded after a conference dinner in the restaurant Červený Jelen. Jana Junová (Charles University, Prague, Czech Republic), with her paper 'Estimation of the General Measure of Stochastic Non-dominance', was the winner of the competition. Monika Matoušková (Charles University, Prague, Czech Republic), with her paper 'Distributionally Robust Fixed Interval Scheduling with Heterogeneous Machines under Uncertain Finishing Times', ranked second. Dominik Kavřík

1 Department of Econometrics, Faculty of Informatics and Statistics, Prague University of Economics and Business, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic. E-mail: petra.zykova@vse.cz.
2 Department of Econometrics, Faculty of Informatics and Statistics, Prague University of Economics and Business, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic. E-mail: jablon@vse.cz.
3 More information at: <*https://mme2023.vse.cz*>.

(Prague University of Economics and Business, Prague, Czech Republic) ranked third with his paper 'Filtering Methods for Output Gap Estimation and the Empirical Taylor Curve: A Comparative Study'. Monika Kaľatová (Charles University, Prague, Czech Republic), Jan Rejthar (Prague University of Economics and Business, Prague, Czech Republic) and Mira Mauleshova (Czech University of Life Science, Prague, Czech Republic) won the remaining three awards.

The conference was organised at high level. All sessions took place in the New Building of the Prague University of Economics and Business. The welcome evening took place in the Academic Club Restaurant which is located on the top floor of the Paradise Building of the Prague University of Economics and Business.

An essential part of all conferences is a social programme that offers many opportunities to discuss various problems in an informal environment. The organisers have prepared three trips: Most participants took part in a guided tour of Konopiště Castle. The second trip was a hiking excursion to Tetín, and the third trip was a Puzzle hunt in Prague. The conference dinner took place at restaurant Červený Jelen on Hybernská Street.

This year's annual meeting of the CSOR decided that the 42nd MME conference will be held in the city of Ústí na Labem Jan Evangelista Purkyně University, the Faculty of Science, the Department of Informatics on 11–13 September 2024. The conference website is available at: <*https://mme2024.ujep.cz*>.

31st International Conference *Interdisciplinary Information Management Talks (IDIMT 2023)*

Petr Doucek[1] | *Prague University of Economics and Business, Prague, Czech Republic*
Lea Nedomová[2] | *Prague University of Economics and Business, Prague, Czech Republic*

The *Interdisciplinary Information Management Talks (IDIMT)*[3] conference is traditionally organized by the Department of Systems Analysis of the Faculty of Informatics and Statistics at the Prague University of Economics and Business, in co-operation with Johannes Kepler University Linz.

This year's 31st IDIMT conference took place from 6 to 8 September 2023 in the heart of Hradec Králové in the historic New Adalbertinum building, which was built according to the project of the architect Paul Ignatius Bayer between 1671 and 1710 as the seat of the Jesuit Order and as its dormitory for the Hradec Králové region. The main theme of this year's edition was "New Challenges for ICT and Management". The conference attracted papers from a total of 128 authors, with 42 submitted papers being accepted together with twelve invited papers. The authors come from ten different countries: Austria, Croatia, Czech Republic, Germany, Greece, Hungary, Palestine, Slovakia, Slovenia and Sweden. Two years of the COVID-Pandemic have taught us to handle and to overcome much of its impact and negative effects on all aspects of our lives. Digitalization, technology in general, and redirected scientific research played an important role in this recovery of science, research, society, business and economy, as demonstrated in this volume.

The conference was opened for the first time by Michael Sonntag from Johannes Kepler University in Linz. During the plenary session, the chairs of the main nine sessions then met to discuss a current issue of our time, namely artificial intelligence, large language models (like ChatGPT) and their possible applications and implications for the future of scientific conferences. It was agreed e.g. that some uses of AI in a research paper might be perfectly acceptable and not even require disclosure (like spell checkers), some might need to be made obvious to reviewers as well as readers (e.g. a summary created automatically, language "polishing", data evaluation – whether marked individually like a reference or globally like a disclaimer at the end remained unclear), while some are at least currently unacceptable (writing complete sections of the paper – here again drafts, material collection etc. might have to be handled differently). Another aspect discussed was whether/how such "artificial" contributions might be recognized and if unenforceable (because of e.g. undetectability) prohibitions are a good strategy.

We have chosen the following nine topics for 2023 IDIMT sessions:

- Crisis Management and ICT (Georg Neubauer, Karin Rainer),
- Cyber Security (Michael Sonntag),

---

[1] Faculty of Informatics and Statistics, Prague University of Economics and Business, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic. E-mail: doucek@vse.cz.

[2] Faculty of Informatics and Statistics, Prague University of Economics and Business, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic. E-mail: nedomova@vse.cz.

[3] More detail information about the conference at: <*https://idimt.org*>.

- Virtual Collaboration, Teaching & Learning (Anne Jantos),
- Autonomous Vehicles and Smart Environments (Erwin Schoitsch),
- Management of ICT Systems (Petr Doucek),
- Social Media (Antonín Pavlíček),
- Digital Transformation of Supply Chain Management (Radoslav Delina),
- Ethical Aspects of Working with Data (Anton Lisnik),
- Special session: Early Career & Student Showcase (Michael Sonntag).

The nine sessions of the conference were divided into two parallel streams, which started after the plenary session. The topics of the sessions covered a wide range of issues with which the discipline of information management is associated today – the main theme being the challenges for ICT in the post-Covid era and the associated expectations. The plenary session was immediately followed by the PhD session, where PhD students presented the contents of their intended dissertations and discussed the ideas and intentions of these with the conference participants. The most attended session was the one on "Digital Transformation of Supply Chain Management". The topic is relevant nowadays, as it reflects the European Union's efforts to transfer the processes of both business and state and public administration into cyberspace. The reflection is evident both in the EU 2030 strategy and its elaboration on the conditions of the Czech Republic and other Member States. Among the traditionally well-attended sessions was the section "Cyber Security". Its relevance is permanent, but currently the main areas of discussion are two EU directives - "NIS2" and "DORA" and their impact on business companies and public and state administration. The general economic and political situation has fed the topic "Crisis Management and ICT", where presenters addressed issues related to crises and scenarios for their course and resolution. The role of information systems and the different aspects of their management and integration into current affairs. Issues related to digital transformation, security, the impact of new international standards to the deployment of chatbots in the teaching process were the focus of the session "Management of ICT Systems". Friday's conference programme was significantly enlivened by two topics, namely the session "Ethical Aspects of Working with Data", which reminded participants that ICT is not just zeroes and ones, technical devices and networks, but also the ways we work on and with them. However, to work with ICT we need educated people and preferably young people. The topic of pedagogy and the deployment of ICT for education was the subject of the session "Virtual Collaboration, Teaching & Learning". The session, although short, provided interesting insights.

Overall, the conference was an important opportunity for ICT and management professionals to best share their knowledge, develop new practices and also plan future collaborations. This year's meeting was particularly challenging in the current era of expected digitalization and a sharp increase in work demanded. And this means, above all, increased pressure on the requirements of IT as well as its staff and employees. We will therefore look forward again to the next IDIMT conference to see what new things it will bring. The event will traditionally take place next year in September 2024. For more information about the conference and previous editions, please visit the website: <*https://idimt.org*>.

As a result of the conference, besides the presented results of the scientific work, the collaboration between Prague University of Economics and Business, Johannes Kepler University Linz and other universities, which were represented in the wide plenary of participants, was deepened.

This conference was partially co-funded through project IGA 409033 of the Faculty of Informatics and Statistics, Prague University of Economics and Business, and Johannes Kepler University Linz, Austria.

Papers
We publish articles focused at theoretical and applied statistics, mathematical and statistical methods, conception of official (state) statistics, statistical education, applied economics and econometrics, economic, social and environmental analyses, economic indicators, social and environmental issues in terms of statistics or economics, and regional development issues.

The journal of Statistika has the following **sections**:

The ***Analyses*** section publishes complex and advanced analyses based on the official statistics data focused on economic, environmental, social and other topics. Papers shall have up to 12 000 words or up to 20 1.5-spaced pages.

Discussion brings the opportunity to openly discuss the current or more general statistical or economic issues, in short what the authors would like to contribute to the scientific debate. Contribution shall have up to 6 000 words or up to 10 1.5-spaced pages.

In the ***Methodology*** section we publish articles dealing with possible approaches and methods of researching and exploring social, economic, environmental and other phenomena or indicators. Articles shall have up to 12 000 words or up to 20 1.5-spaced pages.

Consultation contains papers focused primarily on new perspectives or innovative approaches in statistics or economics about which the authors would like to inform the professional public. Consultation shall have up to 6 000 words or up to 10 1.5-spaced pages.

Book Review evaluates selected titles of recent books from the official statistics field (published in the Czech Republic or abroad). Reviews shall have up to 600 words or 1–2 1.5-spaced pages.

The ***Information*** section includes informative (descriptive) texts, information on latest publications (issued not only by the Czech Statistical Office), or recent and upcoming scientific conferences. Recommended range of information is 6 000 words or up to 10 1.5-spaced pages.

Language
The submission language is English only. Authors are expected to refer to a native language speaker in case they are not sure of language quality of their papers.

Recommended paper structure
Title — Authors and Contacts — Abstract (max. 160 words) — Keywords (max. 6 words / phrases) — Introduction — 1 Literature survey — 2 Methods — 3 Results — 4 Discussion — Conclusion — Acknowledgments — References — Annex (Appendix) — Tables and Figures (for print at the end of the paper; for the review process shall be placed in the text).

Authors and contacts
Rudolf Novak,1 Institution Name, City, Country
Jonathan Davis, Institution Name, City, Country
1 Address. Corresponding author: e-mail: rudolf.novak @domainname.cz, phone: (+420)111222333. ORCID (URL).

Main text format
Times 12 (main text), 1.5 spacing between lines. Page numbers in the lower right-hand corner. *Italics* can be used in the text if necessary. *Do not* use **bold** or <u>underline</u> in the text. Paper parts numbering: 1, 1.1, 1.2, etc.

Headings
1 FIRST-LEVEL HEADING (Times New Roman 12, bold)
1.1 Second-level heading (Times New Roman 12, bold)
1.1.1 Third-level heading (Times New Roman 12, bold italic)

Footnotes
Footnotes should be used sparingly. Do not use endnotes. Do not use footnotes for citing references.

References in the text
Place references in the text enclosing authors' names and the year of the reference, e.g. „…White (2009) points out that…", „…recent literature (Atkinson and Black, 2010a, 2010b, 2011; Chase et al., 2011: 12–14) conclude…". Note the use of alphabetical order. Between the names of two authors please insert „and", for more authors we recommend to put „et al.". Include page numbers if appropriate.

List of references
Arrange list of references alphabetically. Use the following reference styles: [book] HICKS, J. (1939). *Value and Capital: An Inquiry into Some Fundamental Principles of Economic Theory*. 1st Ed. Oxford: Clarendon Press. [chapter in an edited book] DASGUPTA, P. et al. (1999). Intergenerational Equity, Social Discount Rates and Global Warming. In: PORTNEY, P., WEYANT, J. (eds.) *Discounting and Intergenerational Equity*. Washington, D.C.: Resources for the Future. [on-line source] CZECH COAL. (2008). *Annual Report and Financial Statement 2007* [online]. Prague: Czech Coal. [cit. 20.9.2008]. <http://www.czechcoal.cz/cs/ur/zprava/ur2007cz.pdf>. [article in a journal] HRONOVÁ, S., HINDLS, R., ČABLA, A. (2011). Conjunctural Evolution of the Czech Economy. *Statistika: Statistics and Economy Journal*, 91(3): 4–17. [article in a journal with DOI]: Stewart, M. B. (2004). The Employment Effects of the National Minimum Wage [online]. *The Economic Journal*, 114(494): 110–116. <http://doi.org/10.1111/j.0013-0133.2003.0020.x>.
Please **add DOI numbers** to all articles where appropriate (prescribed format = link, see above).

Tables
Provide each table on a separate page. Indicate position of the table by placing in the text "<u>insert Table 1 about here</u>". Number tables in the order of appearance Table 1, Table 2, etc. Each table should be titled (e.g. Table 1 Self-explanatory title). Refer to tables using their numbers (e.g. see Table 1, Table A1 in the Annex). Try to break one large table into several smaller tables, whenever possible. Separate thousands with a space (e.g. 1 528 000) and decimal points with a dot (e.g. 1.0). Specify the data source below the tables.

Figures
Figure is any graphical object other than table. Attach each figure as a separate file. Indicate position of the figure by placing in the text "<u>insert Figure 1 about here</u>". Number figures in the order of appearance Figure 1, Figure 2, etc. Each figure should be titled (e.g. Figure 1 Self-explanatory title). Refer to figures using their numbers (e.g. see Figure 1, Figure A1 in the Annex).

Figures should be accompanied by the *.xls, *.xlsx table with the source data. Please provide cartograms in the vector format. Other graphic objects should be provided in *.tif, *.jpg, *.eps formats. Do not supply low-resolution files optimized for the screen use. Specify the source below the figures.

Formulas
Formulas should be prepared in formula editor in the same text format (Times 12) as the main text and numbered.

Paper submission
Please email your papers in *.doc, *.docx or *.pdf formats to statistika.journal@czso.cz. All papers are subject to double-blind peer review procedure. Articles for the review process are accepted continuously and may contain tables and figures in the text (for final graphical typesetting must be supplied separately as specified in the instructions above). Please be informed about our Publication Ethics rules (i.e. authors responsibilities) published at: <*http://www.czso.cz/statistika_journal*>.

Managing Editor: Jiří Novotný
Phone: (+420) 274 054 299
E-mail: statistika.journal@czso.cz | **web:** www.czso.cz/statistika_journal
Address: Czech Statistical Office | Na padesátém 81 | 100 82 Prague 10 | Czech Republic

Subscription price (4 issues yearly)
CZK 66 per copy + postage.
Printed copies can be bought at the Publications Shop of the Czech Statistical Office (CZK 66 per copy).
Address: Na padesátém 81 | 100 82 Prague 10 | Czech Republic

Subscriptions and orders
Czech Statistical Office | Information Services
Na padesátém 81 | 100 82 Prague 10 | Czech Republic
Phone: (+420) 274 052 733, (+420) 274 052 783
E-mail: objednavky@czso.cz

Design: Toman Design
Layout: Ondřej Pazdera
Typesetting: Družstvo TISKOGRAF, David Hošek
Print: Czech Statistical Office

103rd year of the series of professional statistics and economy journals of the State Statistical Service in the Czech Republic: ***Statistika*** (since 1964), ***Statistika a kontrola*** (1962–1963), ***Statistický obzor*** (1931–1961) and ***Československý statistický věstník*** (1920–1930).

Published by the Czech Statistical Office
ISSN 1804-8765 (Online)
ISSN 0322-788X (Print)
Reg. MK CR E 4684