STATISTIKA

STATISTICS AND ECONOMY JOURNAL

VOL. **105** (1) 2025

CZECH STATISTICAL OFFICE

CONTENTS

ANALYSES

5 Ondřej Vozár, Luboš Marek
Application of Randomized Response Techniques Using Dichotomous Response for Mean Wage in Czechia and Slovakia

18 Martin Mandel, Jan Vejmělek
Shaping Inflation Expectations in the Czech Economy: a Case of Financial Analysts and Corporate Managers

33 Anna Drahozalová, Václav Rybáček
Making Sense of Exchange Rate Pressures in Macroeconomic Statistics

49 Jolana Gubalová, Petra Medveďová, Jana Špirková
The Sustainability Pension Sub-Index in the Context of the Indicators Weights

65 Aldin Brajić, Saliha Brajić
Structural Modeling of Health Services' Quality Level as a Determinant of User Satisfaction of the Tertiary Health Care

87 Zoltán Bánhidi, Imre Dobos, Noémi Kalló, Ariella Janka Tarjáni
Digital Country Rankings for the Visegrád Group Countries with DEA and TOPSIS

102 Samapriya Trivedi, Shambhavi Mishra
A Quantile Regression Modelling Approach to Study Gender Wage Gap in India

METHODOLOGY

123 Jiří Vopravil, Táňa Dvornáková, Barbora Linhartová Jiřičková
Household Finance and Consumption Survey (HFCS) in Czechia

BOOK REVIEW

131 Ondřej Šíma
A Must Read for Those Who Wish to Learn More about Economic Analysis (Not Only) – Book Review

INFORMATION

135 Michal Čigáš
Impact of EU Membership on Statistical Business Register

138 Lucie Legnerová, Pavlína Mildnerová
Monitoring Sustainable Development through Indicators

144 Conferences

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**), in the international citation database of peer-reviewed literature **Scopus** (since 2015), and in 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 | Czechia
E-mail: statistika.journal@csu.gov.cz | www.csu.gov.cz/statistika_journal

Application of Randomized Response Techniques Using Dichotomous Response for Mean Wage in Czechia and Slovakia

Ondřej Vozár[1] | *Prague University of Economics and Business, Prague, Czechia*
Luboš Marek[2] | *Prague University of Economics and Business, Prague, Czechia*

Received 18.8.2024, **Accepted** (reviewed) 9.9.2024, **Published** 14.3.2025

Abstract

Research of controversial topics (drug consumption, corruption) requires reliable estimates of population means of sensitive variables (spending on drugs, illegal sources of income). To avoid non-response and fabricated responses surveys using randomized response techniques (RRT) for quantitative variables are conducted. The paper focuses on surveys conducted regularly in time to study evolution of population mean of a sensitive variable. This topic has not been explored for RRT yet. In applications of RRT is critical a choice of its parameters. The goal is to find rules of thumb if mean of a sensitive variable change in time. We focus only on methods using dichotomous variable (Antoch et al., 2022), because they were designed for variables with many values (Vozár and Marek, 2023). The different scenarios are applied on the Czech and Slovak wage data from Average Earnings Information System in years 2017–2019 using prior information. The scenarios evaluated in extensive simulation study focusing both on mean wage and year-on-year growth.[3]

Keywords	***DOI***	***JEL code***
Randomized response techniques, dichotomous response, wage distribution, survey sampling, comparability over time, population mean	*https://doi.org/10.54694/stat.2024.42*	*C83, C10, E24*

[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, Czechia. 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, Czechia. E-mail: marek@vse.cz. ORCID: <https://orcid.org/0000-0003-4761-1936>.
[3] Article based on the *AMSE 2023* Conference contribution.

INTRODUCTION

The field randomized response techniques (RRT) for quantitative variables have been rapidly developing both in theory and application (e.g. Chaudhuri and Christfides, 2013; or Chaudhuri et al., 2016) since the first pioneering paper fifty years ago by Eriksson (1973). Vozár and Marek (2023) summarized three main approaches:

- Methods using scramble variables (Eriksson, 1973; Eichorn and Hayre, 1983), where instead of true values respondents provide linearly transformed values of sensitive variables depending on results of a random experiment.
- Methods using scramble variables using auxiliary variables known for the whole population strongly correlated to estimated sensitive variables (Diana and Perri, 2013).
- 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 studied sensitive variables.

They argument, that methods using dichotomous response (Antoch et al., 2022) are more fit for sensitive quantitative variables with broad range and numbers of its values (i.e. variables in financial units). They also provide more comfort to respondents (no need of numerical calculations like in methods using scramble variables) and protection of confidentiality of respondents' data. We also exclude techniques using auxiliary variables, because their use is not feasible in most of the real-life populations. The main objection is, that if auxiliary variable is strongly correlated with the sensitive variable, the auxiliary variable must be also sensitive. Therefore, such an auxiliary variable would be unavailable or available in very poor quality.

The rest of the paper is organized as follows. Section 1 summarizes basic notions of survey sampling of a finite population, principles of randomized response techniques, estimation of population mean in this setting, notations and methods using dichotomous responses by Antoch et al. (2022). Section 2 deals with presentation of wage data, including Average Earning Information Systems of Czechia and Slovakia, evolution of wage distribution in studied period of 2016–2019, evolution of wage distribution and choice of wage distribution model. In Section 3, different scenarios for parameter choice of RRT are proposed, setting of simulation study and its numerical results are discussed. The last section summarizes the main findings and conclusions of the paper.

1 EVALUATED METHODS OF DICHOTOMOUS RESPONSES

This section presents basic notion of survey sampling and randomized response techniques, brief review of RRT for population mean and studied methods using dichotomous responses by Antoch et al. (2022).

1.1 Basic notions of survey sampling and randomized response techniques

The goal of survey sampling is to estimate characteristics of a finite population $U = \{1,2,\dots,N\}$ of N unique objects (units). For a quantitative variable Y it is its population total $t_Y = \sum_{i \in U} Y_i$ or population mean $\overline{t_Y} = t_Y / N$ mostly. To achieve that, a random sample s of fixed sample size n is selected with probability $p(s)$. Using probabilities $\pi_i, \left(\pi_i = \sum_{s \ni i} p(s)\right)$ of selection of i^{th} unit of the population U, unbiased Horvitz-Thompson estimator then estimates population mean:

$$\overline{t}_Y^{HT} = \frac{1}{N} \sum_{i \in s} \frac{y_i}{\pi_i}, \tag{1}$$

where subscript HT refers to type for the estimator (1). Statistical properties of estimators and proofs are presented in Section 2.8 in Tillé (2006). If the surveyed variable is sensitive, respondents often refuse to answer or provide fabricated answers. Instead, interviewers try to obtain randomized variable Z correlated to variable of interest Y. Randomization of responses is always carried out independently for each unit selected in sample s with probability $p(s)$. Randomized responses Z are then transformed to random variables R following standard model by Arnab (1994):

$$E_q(R_i) = Y_i, Var_q(R_i) = \phi_i \text{ for all } i \in U, Cov_q(R_i, R_j) = 0, \text{ if } i \neq j\ i, j \in U, \tag{2}$$

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:

$$\overline{t}_Y^{HT,R} = \frac{1}{N}\sum_{i \in s}\frac{R_i}{\pi_i}, \tag{3}$$

where upper subscript R denotes the used randomized response technique.

1.2 Methods using dichotomous responses

Standard methods using scramble variables have several drawbacks (Antoch et al., 2022; Vozár and Marek, 2023):

- Missing practical guidelines for designing scramble variable.
- Calculations required from respondents can be too demanding or misleading for respondents (they can lead to severe errors or non-response).
- Method can be less trustworthy for respondents, because they can feel that interviewer can guess somewhat the sensitive value. In addition, if with knowledge of the values of scramble variable, true value of sensitive variable can be directly calculated.

To resolve the drawback, Antoch et al. (2022) proposed completely different approach. Assuming that the surveyed sensitive variable Y is non-negative and bounded from above ($0 < m \leq Y \leq M$) and both bounds m, M of the variable Y are known. Each respondent draws (independently of the others), a random number U from the uniform distribution on interval (m,M). The interviewer does not know this value. Finally, the respondent answers only 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. Vozár (2023) derived unbiased variance estimators using plug-in technique of Arnab (1994) by assuming knowledge of random numbers U. Without use of random numbers U approximate confidence intervals can estimated by using computer-intensive methods like bootstrap.

1.2.1 Original method of Antoch et al. (2022)

Randomized response of i^{th} respondent follows alternative distribution with parameter $\frac{y_i - m}{M - m}$:

$$Z_{i,\alpha,(m,M)} = \begin{cases} 1 - \alpha + 2\alpha\dfrac{U_i - m}{M - m} & \text{with probability } \quad \dfrac{y_i - m}{M - m}, \text{ if } U_i < y_i, \\ -\alpha + 2\alpha\dfrac{U_i - m}{M - m} & \text{with probability } 1 - \dfrac{y_i - m}{M - m}, \text{ if } U_i \geq y_i, \end{cases} \tag{4}$$

Transformed randomized response is then given as:

$$R_{i,(m,M)} = m + (M - m)Z_{i,(m,M)}. \tag{5}$$

Unbiased population mean estimator of Horvitz-Thompson type is then:

$$\overline{t}_{Y,(m,M)}^{HT,R} = \frac{1}{N}\sum_{i \in s}\frac{R_{i,(m,M)}}{\pi_i}. \tag{6}$$

1.2.2 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:

$$Z_{i,\alpha,(m,M)} = \begin{cases} 1-\alpha+2\alpha\dfrac{U_i-m}{M-m} & \text{with probability } \quad \dfrac{y_i-m}{M-m}, \text{ if } U_i < y_i, \\ -\alpha+2\alpha\dfrac{U_i-m}{M-m} & \text{with probability } 1-\dfrac{y_i-m}{M-m}, \text{ if } U_i \geq y_i, \end{cases} \tag{7}$$

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. Vozár (2024) found out in extensive simulation study that values $\alpha = 0.5$ or $\alpha = 0.7$ provided good results for broad class of distributions of sensitive variables.

Transformed randomized response is then equal to:

$$R_{i,\alpha,(m,M)} = (M-m)\,Z_{i,\alpha,(m,M)} + m\,. \tag{8}$$

Unbiased population mean estimator of Horvitz-Thompson type is then:

$$\overline{t}_{Y,\alpha,(mM)}^{HT,R} = \frac{1}{N}\sum_{i \in s}\frac{R_{i,\alpha,(m,M)}}{\pi_i}. \tag{9}$$

2 CZECH AND SLOVAK WAGE DATA AND ITS STATISTICAL DISTRIBUTION

In this section the studied Czech and Slovak Wage data and statistical model of wage distribution are presented. Since no anonymized microdata on wages are available, we simulate the corresponding populations from the estimated wage distribution. First, statistical surveys Average Earnings Information Systems (ISPV) in both countries are shortly presented. Then, the evolution of Czech and Slovak wage distributions in years 2016–2019 is summarized. The last subsection discusses the methods of modelling Czech and Slovak wage data from Average Earnings Information Systems and provides estimated wage distributions for studied data. To achieve comparability, Slovak wage data were converted to Czech crowns.

2.1 Average Earnings Information System (ISPV) in Czechia and Slovakia

The wage data used come from corporate statistical surveys conducted within the Czech and Slovak statistical services, always on behalf the national Ministry of Labour and Social Affairs. Our study will focus to the so-called wage sphere, i.e. mainly wages in the private sector, while the data cover the population of employers with ten or more employees.

The sampling plans in both countries aim to cover the largest possible volume of wages paid with the lowest possible range of company selections and, at the same time, to provide representative results for the size groups of companies in terms of the number of employees. Data for size groups are important for economic policymaking and for other stakeholders:

- Census of large enterprisers with 250 and more employees.
- Survey sampling for the size group of medium-sized enterprises with 50 to 249 employees and small enterprises with 10 to 49 employees.

- Once every two years, a supplementary survey is carried out on the smallest enterprises with 1 to 9 employees, however, data for these enterprises are not included in the data from which the wage distributions studied in this chapter are estimated.

Although a small number of employers are selected, the ISPV data cover a substantial part of employees in the wage sphere of the Czech and Slovak economies (over 2.20 million of employees in Czechia, over 1.05 million of employees in Slovakia). In this study, we will deal with the average gross monthly wage in the second quarter of the given year. The second quarter is chosen as the reference period when studying wage distributions, as in this period wages are not affected by bonuses paid and transfers of non-entitlement components of wages to other quarters due to a change in legislation (in the case of a change in taxation, these components are always paid in the period when it is tax-advantageous).

2.2 Changes in the Czech and Slovak wage distribution in years 2016–2019

This chapter contains a summary of the development of wage distributions in the period under review, including the possible impact of the increase in the minimum wage on wage growth (see Table 1), which was significant in both countries in this period.

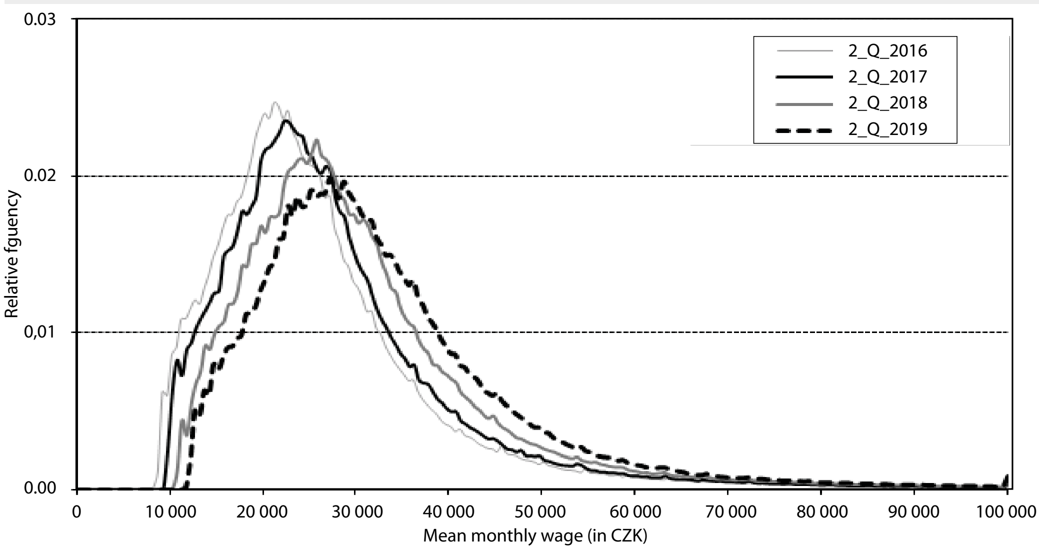
Table 1 Minimum wage in CZ and SK in years 2016–2019 (in CZK)

Country	Indicator	Year			
		2016	2017	2018	2019
CZ	Minimal wage (CZK)	9 900	11 000	12 200	13 350
SK		10 125	10 875	12 000	13 000

Source: Ministry of Labour and Social Affairs (CZ): <*www.mpsv.cz*>, Ministry of Labour, Social Affairs and Family (SK): <*www.employment.sk*>

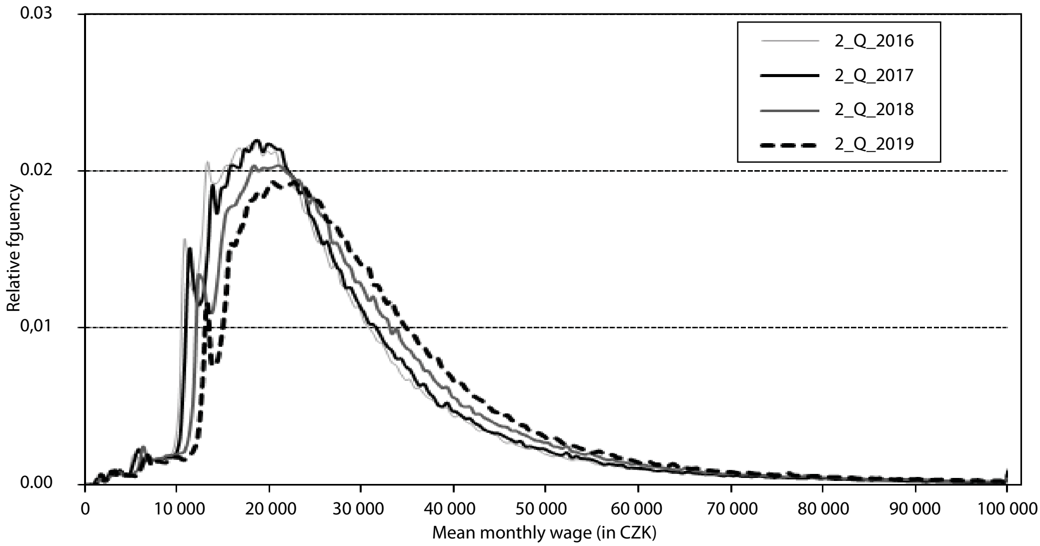
The wage distributions in both countries are strongly skewed from the right and indicate high wage equality (see the graphs in Figures 1 and 2).

Figure 1 Histogram of mean monthly wages in the second quarter wages in Czechia in years 2016–2019



Source: Own construction using interval data by company TREXIMA

Figure 2 Histogram of mean monthly wages in the second quarter wages in Slovakia in years 2016–2019



Source: Own construction using interval data by company TREXIMA

Wage levels and inequality are higher in Czechia than in Slovakia (see Table 2).

Table 2 Wage distribution in CZ and SK in years 2016–2019

Country	Mean monthly gross wage (CZK)	Period			
		2Q 2016	2Q 2017	2Q 2018	2Q 2019
CZ	1. decile	13 694	14 732	16 675	18 364
	1. quartile	18 141	19 302	21 465	23 448
	median	23 506	24 886	27 490	30 088
	3. quartile	30 710	32 361	35 625	39 080
	9. decile	41 779	44 082	47 888	52 099
SK	1. decile	12 778	13 256	14 398	15 830
	1. quartile	16 506	17 105	18 482	20 064
	median	22 493	23 079	24 901	26 828
	3. quartile	31 663	32 386	34 674	36 925
	9. decile	46 296	47 006	49 517	51 995

Source: Own construction from interval data by company TREXIMA

Both countries recorded high wage growth (most notably in 2018 and 2019), driven by minimum wage increases (see Table 1 and shift to the right in Figures 1 and 2). The increase was in lower wage bands and therefore there was a flattening and shift of wage distributions to the right (see Figures 1 and 2). The share of employees with an average monthly wage exceeding CZK 100 000 grew rapidly (over 3% in 2019 in both countries, see Figures 1 and 2).

2.3 Choice of wage distribution model

As anonymized microdata for scientific purposes from ISPV are not available (including that personal data covered by the GDPR), simulated data from the estimated parametric wage distribution from interval data must be used to study the proposed estimates.

Modelling of wage distributions in Czechia and Slovakia based on interval data from national ISPV statistical surveys has been the subject of a number of papers in which a suitable statistical methodology has been sought to find adequate models and changes associated with the transformation of both economies have been studied.

Marek (2010), Vrabec and Marek (2016) dealt with the choice of a suitable parametric model to capture the development of wage distributions. A three-parameter log-logistic distribution (see Formula (10)) has been shown as a suitable model, which will be used for simulations in this chapter due to its versatility for empirical data and computational simplicity. Bílková and Malá (2012), Bílková (2013) modelled wage distributions for individual industries using so-called L-estimates. Modelling of multivariate distributions (Malá, 2015) and finite mixtures of distributions (Malá, 2013; Marek and Vrabec, 2016) has proven to be a suitable methodology for modelling and analysing the development of wage distributions. Estimating the parameters of wage distributions by these models would be too complicated to generate the studied variables, so we chose a computationally simpler approach using a one-dimensional parametric model, namely a three-parametric log-logistic distribution with density:

$$f\left(y,\tau,\sigma,\delta\right)=\frac{\tau}{\sigma}\left(\frac{y-\delta}{\sigma}\right)^{\tau-1}\left(1+\left(\frac{y-\delta}{\sigma}\right)^{\tau}\right)^{-2},\ \ y\geq\delta>0,\tau>0,\sigma>0, \tag{10}$$

where $\tau > 0$ is a shape parameter, $\sigma > 0$ is a scale parameter and $\delta > 0$ je is a shift parameter.

The parameters of the three-parameter log-logistic distribution (10) were estimated by the moment method (see Table 3 for parameter estimates). All estimates were based on interval data in the range of CZK 0 to 100 000. Employees with an average monthly wage exceeding one hundred thousand. CZK, we have included in the last interval (CZK 99 501–100 000).

Table 3 Estimates of parameters of wage distribution in CZ and SK in 2016–2019

Country	Year	Parameter		
		$\hat{\tau}$	$\hat{\sigma}$	$\hat{\delta}$
CZ	2016	3.5	19 812	3 736
	2017	3.6	21 198	3 737
	2018	3.8	23 831	3 742
	2019	4.0	26 397	3 738
SK	2016	3.5	24 473	389
	2017	3.6	23 507	–12
	2018	3.7	25 378	–122
	2019	4.0	28 171	–980

Note: Model wage distribution is three-parametric log-logistic distribution (10), parameters estimated by moment methods form interval data by company TREXIMA.
Source: Own construction

3 SIMULATION STUDY

In the first subsection, we choose statistics for performance evaluation of estimates. In the second subsection, we propose the strategies for setting intervals (m, M) using prior information on minimal wage. In the third section numerical simulation results are presented and discussed. Because the aim of the paper is to find a general rule of thumb, the study is restricted period of rapid wage growth in years 2016–2019. The economic shocks due to Covid–19 pandemics and corresponding measures on labour market would complicate finding this rule of thumb.

3.1 Statistics for simulation evaluation

Assume, that is given simulation M replications are carried out and statistics S_i, $i = 1, 2, \dots, M$ (population mean) are in i^{th} replica estimated by $\hat{S}_i$, $i = 1, 2, \dots, M$ (sample mean). The statistics below is used to evaluate estimates $\hat{S}_i$, $i = 1, 2, \dots, M$.

To evaluate bias, mean percentage (MPE) is defined as:

$$MPE = \frac{1}{M}\sum_{i=1}^{M}\frac{\hat{S}_i - S_i}{S_i},\ S_i \neq 0,\ i = 1, 2, \dots, M. \tag{11}$$

To evaluate variance of estimates, median absolute percentage error (MdAPE) is defined (Hyndman a Koehler, 2006) as:

$$MdAPE = med\left(\left|\frac{\hat{S}_1 - S_1}{S_1}\right|, \left|\frac{\hat{S}_2 - S_2}{S_2}\right|, \dots \left|\frac{\hat{S}_M - S_M}{S_M}\right|\right),\ S_i \neq 0,\ i = 1, 2, \dots, M. \tag{12}$$

3.2 Strategies for setting intervals (*m*, *M*)

The aim of the application is to estimate a sensitive variable – the average monthly salary in the second quarter in a time series, which is a real application in the practice of state statistics or statistical agencies. The strategy differs according to the rules for choosing the interval (m, M) for the generation of random numbers, the value of the parameter α for the method using the knowledge of the random number is set to $\alpha = 0.5$ and 0.35 using recommendation of simulation study by Vozár (2024). The following strategies are proposed:

- S1: fixed interval values (m, M) for the whole period 2016–2019. The bounds are chosen ad hoc according to how we perceive too low or too high a wage.
- S2: fixed upper limit of the interval (m, M) for the whole period 2016–2019, the lower limit is equal to the minimum wage valid in the second quarter of the given year.

Multiple intervals (m, M) are to be evaluated for each strategy:

Table 4 Intervals (*m*, *M*) for random numbers

Strategy		Country	Year			
			2016	2017	2018	2019
S1	low	CZ, SK	(10 000, 50 000)			
	medium		(10 000, 60 000)			
	high		(12 000, 70 000)			
S2	low	CZ	(9 900, 50 000)	(11 000, 50 000)	(12 200, 50 000)	(13 350, 50 000)
	medium		(9 900, 60 000)	(11 000, 60 000)	(12 200, 60 000)	(13 350, 60 000)
	high		(9 900, 70 000)	(11 000, 70 000)	(12 200, 70 000)	(13 350, 70 000)
	low	SK	(10 125, 50 000)	(10 875, 50 000)	(12 000, 50 000)	(13 000, 50 000)
	medium		(10 125, 60 000)	(10 875, 60 000)	(12 000, 60 000)	(13000, 60 000)
	high		(10 125, 70 000)	(10 875, 70 000)	(12 000, 70 000)	(13 000, 70 000)

Source: Own construction

No that there is a bias-variance trade-off of population mean estimates (Antoch et al., 2022). The smaller range of the interval (m, M), the lower variance and mostly higher bias and vice versa.

3.3 Simulation study and results

We focus on a single combination of sample range and population size, namely sample range n = 1 000 and population size N = 1 000 000. All simulations were carried out by statistical freeware R (R Project, 2024). Wage data were generated by R package *flexsurv* (Jackson, 2016) following three-parametric log-logistic distribution (Formula 10) with parameters estimated by moment method (Table 3). The size of the sample is motivated, by the typical sample size for a national survey conducted by public opinion research agencies. We chose fixed population sizes for two reasons. The first reason is easier comparability of results in individual years, because this ensures the same size of the population and the same sampling ratio, namely one per mile. The second reason is the acceleration of simulations, and we have shown by numerical studies that for a sample ratio of one per mile, the results would not differ much from each other. We will limit ourselves to assessing the combination of methods and strategies for the choice of interval (m, M) (Table 4) from the point of view of impartiality and variability of estimates of average monthly wages. Effect of non-response is studied to provide benchmarks with direct questioning in this setting:

- Missing completely at random (MCAR), where 90 %, 80 % or 60 % respondents of sample answers.
- Missing not at random (MNAR), where 10 % of respondents with the highest respondents refuses to answer (we assume high values of wages as more sensitive)

We treat non-response by weighting, mean wage is estimated as sample mean of responses.

For wage data, the impact of systematic non-response is severe. The relative bias represents approximately one eighth of the average wage and an almost eleven-fold increase in variability compared to the 100% response in direct questioning (see Table 5).

Table 5 Effect non-response to bias and accuracy (N = 1 000 000, n = 1 000)

Country	Year	100% response (MdAPE)	Non-response model				
			MCAR (MdAPE)			MNAR ($n_r = 0.9n$)	
			$n_r = 0.9n$	$n_r = 0.8n$	$n_r = 0.6n$	MdAPE	MPE
CZ	2016	1.14	1.20	1.27	1.46	12.60	–12.59
	2017	1.10	1.16	1.23	1.41	12.15	–12.15
	2018	1.03	1.08	1.15	1.33	11.40	–11.40
	2019	1.03	1.08	1.15	1.33	11.40	–11.10
SK	2016	1.29	1.36	1.44	1.67	14.34	–14.34
	2017	1.26	1.32	1.41	1.61	13.95	–13.95
	2018	1.23	1.29	1.37	1.56	13.55	–13.54
	2019	1.14	1.19	1.27	1.46	12,61	–12,60

Source: Own construction

The S1 strategy with fixed intervals *(m, M)* is not satisfactory due to the high wage growth in the period under review, because at the end of the period the estimates $\overline{t}_{Y,(m,M)}^{HT,R}$ would be very biased (see Table 6). The S2 strategy, with an annual update of the lower limit of the interval with the applicable minimum wage, eliminates this bias only slightly (see Table 8). Moreover, the bias (MPE) changes over time with rising wage levels, leading to a strongly biased estimate of annual average wage growth.

For wage data, heuristics motivating estimated with use random number (Antoch et al., 2022) works very well. This leads to a significant reduction in the underestimation of the average wage (see Tables 6 and 8), only 1.7–2.5%, regardless of the value of the parameter α. To balance the bias and variance of estimates, it is necessary to choose the highest possible upper limit, i.e. CZK 70 000. If we choose an upper limit of only CZK 50 000 (the ninth decile), then the lower limit must be much higher than the values we consider compensating for the bias caused by wages exceeding this upper limit.

Considering the variability of estimates, a combination of the "high" variant with the highest upper bound and an estimate using the knowledge of random numbers is appropriate (see Table 7 and 9). It is also advisable to set the parameter $\alpha = 0.5$, i.e. the contribution to the variance caused by randomized response is constant, it does not depend on the value of the sensitive variable y_i. Then the bias of the estimates (MPE) is the same over time, which allows for unbiased estimates of year-on-year growth rate. Due to the sharp increase in wages in the period under review, we are inclined to the S2 strategy with an update of the lower bound of the interval *(m, M)*.

Table 6 Mean percentage error (MPE) – strategy S1

Estimate	Strategy for *(m, M)*	Country	Year			
			2016	2017	2018	2019
$\overline{t}_{Y}^{HT,R}$	all	CZ	–0.03	–0.01	–0.01	–0.01
	all	SK	–0.01	0.01	0.00	0.02
$\overline{t}_{Y,(m,M)}^{HT,R}$	low	CZ	–3.36	–8.20	–16.0	–22.8
	medium		–2.04	–6.92	–14.8	–21.7
	high		–1.32	–6.23	–14.2	–21.2
	low	SK	–10.0	–4.21	–10.3	–15.5
	medium		–8.82	–2.88	–9.10	–14.3
	high		–8.14	–2.16	–8.44	–13.7
$\overline{t}_{Y,0.35,(m,M)}^{HT,R}$	low	CZ	–3.38	–3.61	–4.01	–4.91
	medium		–2.06	–2.17	–2.32	–2.82
	high		–1.34	–1.37	–1.45	–1.75
	low	SK	–5.23	–4.07	–4.65	–4.71
	medium		–3.22	–2.38	–2.78	–2.72
	high		–2.08	–1.50	–1.74	–1.66
$\overline{t}_{Y,0.50,(m,M)}^{HT,R}$	low	CZ	–3.38	–3.60	–4.01	–4.90
	medium		–2.06	–2.16	–2.32	–2.81
	high		–1.34	–1.37	–1.45	–1.74
	low	SK	–5.24	–4.07	–4.65	–4.71
	medium		–3.22	–2.38	–2.79	–2.72
	high		–2.08	–1.50	–1.75	–1.65

Source: Own construction

Table 7 Median absolute percentage error (MdAPE) – strategy S1

Estimate	Strategy for (m, M)	Country	Year 2016	2017	2018	2019
$\overline{t}_{Y}^{HT,R}$	all	CZ	1.14	1.10	1.03	1.03
	all	SK	1.29	1.26	1.23	1.14
$\overline{t}_{Y,(m,M)}^{HT,R}$	low	CZ	3.37	8.17	16.03	22.83
	medium		2.40	6.90	14.84	21.73
	high		2.34	6.28	14.30	21.25
	low	SK	10.05	4.22	10.30	15.52
	medium		8.82	3.00	9.08	14.36
	high		8.15	2.68	8.49	13.74
$\overline{t}_{Y,0.35,(m,M)}^{HT,R}$	low	CZ	3.36	3.58	4.02	4.91
	medium		2.24	2.24	2.37	2.82
	high		1.98	1.95	1.89	1.99
	low	SK	5.25	4.08	4.68	4.69
	medium		3.27	2.47	2.82	2.72
	high		2.37	2.02	2.14	2.02
$\overline{t}_{Y,0.50,(m,M)}^{HT,R}$	low	CZ	3.38	3.60	4.01	4.92
	medium		2.24	2.24	2.36	2.82
	high		1.97	1.89	1.84	1.93
	low	SK	5.26	4.09	4.68	5.99
	medium		3.28	2.47	2.82	2.70
	high		2.35	2.00	2.11	1.97

Source: Own construction

Table 8 Mean percentage error (MPE) – strategy S2

Estimate	Strategy for (m, M)	Country	Year 2016	2017	2018	2019
$\overline{t}_{Y}^{HT,R}$	all	CZ	–0.03	0.00	–0.01	0.00
	all	SK	–0.01	0.01	0.002	0.015
$\overline{t}_{Y,(m,M)}^{HT,R}$	low	CZ	–3.37	–8.11	–15.7	–22.4
	medium		–2.05	–6.83	–14.6	–21.3
	high		–1.32	–6.16	–14.0	–20.8
	low	SK	–10.0	–4.13	–10.1	–15.1
	medium		–8.81	–2.81	–8.89	–13.9
	high		–8.14	–2.08	–8.22	–13.3
$\overline{t}_{Y,0.35,(m,M)}^{HT,R}$	low	CZ	–3.38	–3.55	–3.94	–4.81
	medium		–2.06	–2.10	–2.24	–2.72
	high		–1.34	–1.31	–1.38	–1.65
	low	SK	–5.22	–3.90	–4.33	–4.35
	medium		–3.21	–2.23	–2.47	–2.36
	high		–2.06	–1.35	–1.42	–1.29
$\overline{t}_{Y,0.50,(m,M)}^{HT,R}$	low	CZ	–3.39	–3.55	–3.94	–4.80
	medium		–2.07	–2.09	–2.24	–2.72
	high		–1.35	–1.30	–1.38	–1.64
	low	SK	–5.22	–3.90	–4.33	–4.35
	medium		–3.21	–2.23	–2.48	–2.35
	high		–2.06	–1.35	–1.42	–1.29

Source: Own construction

Table 9 Median absolute percentage error (MdAPE) – strategy S2

Estimate	Strategy for (m, M)	Country	Year			
			2016	2017	2018	2019
$\overline{t}_{Y}^{HT, R}$	all	CZ	1.14	1.10	1.03	1.03
	all	SK	1.29	1.26	1.23	1.14
$\overline{t}_{Y, (m, M)}^{HT, R}$	low	CZ	3.37	8.17	16.03	22.83
	medium		2.4	6.90	14.84	21.73
	high		2.34	6.28	14.30	21.25
	low	SK	10.05	4.22	10.30	15.52
	medium		8.82	3.00	9.08	14.36
	high		8.15	2.68	8.49	13.74
$\overline{t}_{Y, 0.35, (m, M)}^{HT, R}$	low	CZ	3.36	3.58	4.02	4.91
	medium		2.24	2.24	2.37	2.82
	high		1.98	1.95	1.89	1.99
	low	SK	5.25	4.08	4.68	4.69
	medium		3.27	2.47	2.82	2.72
	high		2.37	2.02	2.14	2.02
$\overline{t}_{Y, 0.50, (m, M)}^{HT, R}$	low	CZ	3.38	3.60	4.01	4.92
	medium		2.24	2.24	2.36	2.82
	high		1.97	1.89	1.84	1.93
	low	SK	5.26	4.09	4.68	5.99
	medium		3.28	2.47	2.82	2.70
	high		2.35	2.00	2.11	1.97

Source: Own construction

CONCLUSION

The aim of the paper was to find rules of thumbs for parameters or RRT for surveys regularly conducted in time to study both mean and year-on-year growth rate. Study focused on the case of sensitive variable with broad range of values, therefore the proposed rules were evaluated on Czech and Slovak wage data coming from period of rapid growth in years 2016–2019. Since the aim of the paper is to find a general rule of thumb, the study excludes the year of Covid-19 pandemics. The economic shocks due to Covid-19 pandemics and corresponding measures on labour market would complicate finding this rule of thumb.

We apply methods using of dichotomous responses by Antoch et. al. (2022), because they are designed for sensitive variable with broad range of values and they overcome many limitations of standard methods using scramble variables (Vozár and Marek, 2022). We apply both original method and method using random numbers.

Two strategies combining three levels of the range (low, medium and high width) of interval (m, M) provided six possible rules of thumb. The S1 strategy is using the same interval in all years. The S2 strategy is based on changing lower bounds m using information on current minimum wage. The upper bound M was constant because there was no prior information for that. Updating by inflation rate had no practical effect because of low inflation in years 2016–2019. We choose the intervals by subjective expert idea what was low and high wage in this time. For methods using random numbers values of parameter α were set to recommended values 0.50 or 0.35 by Vozár (2024). As benchmark direct questioning under missing completely random and missing not at random responses model were also evaluated.

The conclusion of the simulation study is as follows. The heuristics behind the method using knowledge of random numbers works well. The reduction of variance and bias of population mean estimates is high, which supports conclusions from Vozár (2022). Values of parameter α = 0.50 is a safe choice for any distribution of sensitive variable.

If the evolution of population mean in time is rapid, it is necessary to update the intervals (*m*, *M*) to avoid bias both in estimates of population mean and year-on-year growth rate. Interval bounds must be updated to avoid biases. The intervals (*m*, *M*) should be broad enough to cover as many units of population as possible. High variance and bias, if low non-response under missing not at random gives a convincing argument to use RRT instead of direct questioning. It is worth mentioning that the conclusion above was done by using both method on Czech and Slovak wage data in the period of growth. Therefore, further evaluation of more data sets and evolution patterns in time is recommended to refine the rule of thumb.

ACKNOWLEDGMENT
This work of was supported by the Institutional Support to Long-Term Conceptual Development of Research Organization, the Faculty of Informatics and Statistics of the University of Economics, Prague. The authors are grateful to two unknown reviewers for comments that considerably improved content 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>.

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>.

BÍLKOVÁ, D. (2013). Modelování mzdových rozdělení posledních let v České republice s využitím L-momentů a predikce mzdových rozdělení podle odvětví. *E+M. Ekonomie a Management*, XVI(4): 42–54.

BÍLKOVÁ, D., MALÁ, I. (2012). Application of the L-Moment Method when Modelling the Income Distribution in the Czech Republic [online]. *Austrian Journal of Statistics*, 41(2): 125–132. <http://www.stat.tugraz.at/AJS/ausg122/122Bilkova1.pdf>.

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

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>.

HYNDMAN, R. J., KOEHLE, A. B. (2006). Another look at measures of forecasts accuracy. *International Journal of Forecasting,* 22(4): 679–688.

JAKCSON, C. (2016). Flexsurv: a platform for parametric modelling in R. *Journal of Statistical Software*, 70: 1–33.

MALÁ, I. (2013). Použití konečných směsí logaritmicko-normálních rozdělení pro modelování příjmů českých domácností. *Politická ekonomie*, 61(3): 356–372.

MALÁ, I. (2015). Vícerozměrný pravděpodobnostní model rozdělení příjmů českých domácností. *Politická ekonomie*, 63(7): 895–908.

MAREK, L. (2010). Analýza vývoje mezd v ČR v letech 1995–2008. *Politická ekonomie*, 58(2):186–206.

MAREK, L., VRABEC, M. (2016). Using mixture density functions for modelling of wage distributions. *Central European Journal of Operations Research*, 24(4): 389–405.

R CORE TEAM (2024). *R: a language and environment for statistical computing.* Austria, Vienna: R Foundation for Statistical Computing.

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>.

VOZÁR, O., MAREK, L. (2023). Multicriteria Evaluation of Randomized Response Techniques for Population Mean [online]. *Statistika: Statistics and Economy Journal*, 103(4): 492–503. <https://doi.org/10.54694/stat.2023.32>.

VOZÁR, O. (2024). *Randomized Response Techniques for Population Mean*. Dissertation, Prague University of Economics and Business.

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>.

Shaping Inflation Expectations in the Czech Economy: a Case of Financial Analysts and Corporate Managers

Martin Mandel[1] | *Prague University of Economics and Business, Prague, Czechia*
Jan Vejmělek[2] | *Prague University of Economics and Business, Prague, Czechia*

Received 11.9.2024, **Accepted** (reviewed) 7.10.2024, **Published** 14.3.2025

Abstract

Inflation expectations play an important role in the transmission mechanism of inflation targeting in the context of the length and costs of the disinflationary process. The objective of our paper is to employ econometric analysis to verify whether financial analysts' and corporate managers' inflation expectations in Czechia (from Q3 1999 to Q2 2024) show basic features of rational expectations and what impact the past YoY CPI inflation rate, the CNB's inflation forecast and the CNB's inflation target have on their expectations. We find that the formation of financial analysts' and corporate managers' yearly inflation expectations with time horizons of one year and three years differs considerably. For corporate managers' inflation expectations, adaptive reasoning plays a more important role. Financial analysts take more account of the CNB's one-year inflation forecasts in forming their yearly expectations, while the inflation target, as an explanatory variable, is statistically significant only for their three-year inflation expectations. Neither group of respondents meets the required criteria for rational expectations in terms of the tests formulated by Pesaran (1987), and Fama (1965 and 1970). In particular, their yearly inflation expectations exhibit systematic errors. Surprisingly, the time series of financial analysts' inflation expectations contain a seasonal component.[3]

Keywords	DOI	JEL code
Inflation expectations, monetary policy, inflation targeting, Czech National Bank	*https://doi.org/10.54694/stat.2024.50*	*C12, B22, E58*

[1] Faculty of Finance and Accounting, Prague University of Economics and Business, W. Churchill Sq. 4, 130 67 Prague 3, Czechia. E-mail: martin.mandel@vse.cz.
[2] Faculty of Finance and Accounting, Prague University of Economics and Business, W. Churchill Sq. 4, 130 67 Prague 3, Czechia.
[3] This article was written with the support of institutional research VŠE FFÚ IP100040, Faculty of Finance and Accounting, Prague University of Economics and Business.

INTRODUCTION

Inflation expectations play an important role in the transmission mechanism of inflation targeting, in particular with regard to the length and cost of the disinflationary process. The Czech National Bank (CNB) has been pursuing a monetary policy based on explicit inflation target setting since 1998. From the perspective of the central bank, i.e. in terms of its model framework and the efficacy of its monetary policy, it is important to ascertain the manner in which inflation expectations are shaped in the various groups of market entities.

Simplifying somewhat, inflation expectations can be classified as backward-looking and forward-looking. Backward-looking expectations are typically associated with the hypotheses of adaptive and extrapolative inflation expectations, as proposed by Nerlove (1958 and 1983), and Metzler (1941). Rational inflation expectations (most notably Muth, 1961; and Sargent, 1986) and model-consistent inflation expectations (i.e. expectations that align with a specific inflation targeting model structure) encompass a combination of information about the past (the backward-looking component) and about the future (the forward-looking component).

Rational expectations assume not only the ready availability of the required information but also a perfect capacity to process it. A hard-to-sustain assumption of rational expectations is the notion that fully rational agents form their expectations based on an estimate of future developments, having knowledge of the model structure of the economy.[4] The majority of market entities operate within a relatively narrow economic space with which they have prior experience, acquiring and processing the necessary information within this context. In other areas of economic life, the rationality of market entities is quite bounded. They derive the future from a short-term backward-looking perspective, the psychological make-up of each individual is of great importance, and decision-making is heuristic.[5]

The construction of inflation targeting models usually works with the assumption that one part of the entities is backward-looking and the other part forward-looking consistently with the structure of the formulated inflation targeting model. Each central bank also asks itself to what extent its monetary policy is credible. In terms of inflation expectations, how much influence past inflation, the inflation target and the inflation forecast have on market entities' inflation expectations.

The objective of our paper is to employ statistical and econometric analysis to verify whether or not financial analysts' and corporate managers' inflation expectations show basic features of rational expectations, to what extent their inflation expectations are persistent, and what impact the past YoY CPI inflation rate (i.e. its last known value), the CNB's inflation forecast and the CNB's inflation target have on their expectations. The formation of inflation expectations is likely to differ between the various groups of respondents monitored by the CNB, i.e. managers of non-financial corporations and firms on the one hand and financial analysts on the other. The hypothesis can be formulated that in the above order of entities, there will be a shift from backward-looking to forward-looking expectations. The analysis will be performed on the following time series: a) financial analysts' expectations over one year and three years and b) expectations of non-financial corporations' and firms' managers over one year and three years.

The structure of our paper is as follows. First, an outline of referenced literature and the existing knowledge of inflation expectation analysis; then we describe (in Chapter 2) the basic properties of the time series under review and provide initial insights into the nature of financial analysts' and corporate

4 "The way expectations are formed depends specifically on the structure of the relevant system describing the economy" (Muth, 1961).

5 The notion of bounded rationality was developed by A. H. Simon (Simon, 1984) and further elaborated on by representatives of behavioural economics.

managers' inflation expectations. In Chapter 3, we conduct standard (indirect) tests of the rational expectations hypothesis as formulated by, in particular, Pesaran (1987), and Fama (1965 and 1970). In Chapter 4, we then use an econometric model to test how financial analysts' and corporate managers' inflation expectations are formed based on the CNB's database. We then formulate some findings and conclusions for the monetary policy.

1 REVIEW OF LITERATURE

The rational expectations hypothesis (Muth, 1961) posits that expectations are fully rational if all available information has been used in the optimal way. There should not be an alternative prediction model that has less variance in the actual forecast errors and has been built using the same set of available information. Actual forecast errors are random with a mean of zero and are due to random events (shocks) or unsystematic individual errors in individuals' forecasts. Pesaran (1987) formulated econometrically testable conditions for the validity of the rational expectations hypothesis:

1/ test of the unbiased expectation hypothesis,
2/ test of the properties of prediction errors, which should be orthogonal to the agents' information available at the time the prediction is made,
3/ test of the serial uncorrelated prediction error condition,
4/ efficiency test of the prediction, i.e. the prediction error must not be related to the predicted variables from previous periods.

The last mentioned test is most often conducted as a test of weak-form market efficiency (Fama, 1965 and 1970). In this paper, we apply this test, typically used to analyse expectations (predictions) in the capital and foreign exchange markets, to inflation expectations.

Adaptive expectations are formulated, for example, by Nerlove (1958 and 1983). "Purposeful economic agents have incentives to eliminate errors up to a point justified by the costs of obtaining the information necessary to do so... The most readily available and least costly information about the future value of a variable is its past value." (Nerlove, 1983).

In the first wave of interest in the formation of inflation expectations in the Czech literature, Tomšík and Mandel (2004) discuss how inflation expectations are formed on the basis of the adaptive component of expectations, the CNB's inflation forecast and the CNB's inflation target in the context of the credibility of the CNB's monetary policy. Filáček (2005) provides a detailed explanation of the various models of inflation expectations, including the distinction between adaptive expectations, adaptive learning and naive expectations. He then tests the significance of each form of expectations under variant shocks.[6] Vávra (2005) summarises international experience with business cycle-related tests of inflation expectations among households and economic experts. He discusses the statistical methodology of evaluating the results of business cycle-related surveys in Czechia conducted by the Czech Statistical Office (before 2005), which were, however, limited to only determining whether prices would increase, remain unchanged, or decrease.

Some findings can be noticed in international publications, which are important in the context of our subsequent analysis. Carrol (2003) analyzed inflation expectations based on newspaper predictions by professional financial analysts and a dataset collected by the University of Michigan's Survey Research Center on inflation expectations of US households. The author concludes that the model's assumptions about the existence of rational expectations are not entirely relevant, especially in the case of households. People only occasionally pay attention to new information. This inattention creates rigidity and adaptive expectations. The expectations of financial analysts are closer to rationality, which in retrospect can

[6] Options and discussion of the various forms of expectations in modelling, e.g. Brayton et al. (1997).

partially influence the expectations of households as well. Blanchflower and MacCoille (2009), in their extensive empirical analysis using the British population as an example, conclude that inflation expectations are significantly adaptive. They also find that individuals with higher education tend to have lower inflation expectations. Miyajima and Yetman (2018) analyze the characteristics of inflation expectations of households, businesses, financial analysts and trade unions in South Africa over the period 2000 to 2018. They conclude that inflation expectations are more anchored to a band inflation target of 3% to 6% in the case of longer-term predictions. The long-term inflation expectations of financial analysts lie within the range of the central bank's inflation target. Inflation expectations of businesses and unions remain above the upper bound of the official inflation target. Coibion et al. (2018) make an interesting finding that New Zealand firms devote few resources to collecting and processing inflation information, and inflation is not perceived as important for business decisions. However, this conclusion may be influenced by the fact that the Central Bank of New Zealand is successful in meeting its inflation target. Gerlain et al. (2019) conclude in their analysis of the inflation expectations of financial analysts from the Survey of Professional Forecasters database that a hybrid expectations model incorporating adaptive expectations better explains the inflation expectations of the surveyed group of respondents than a model with strictly rational expectations.

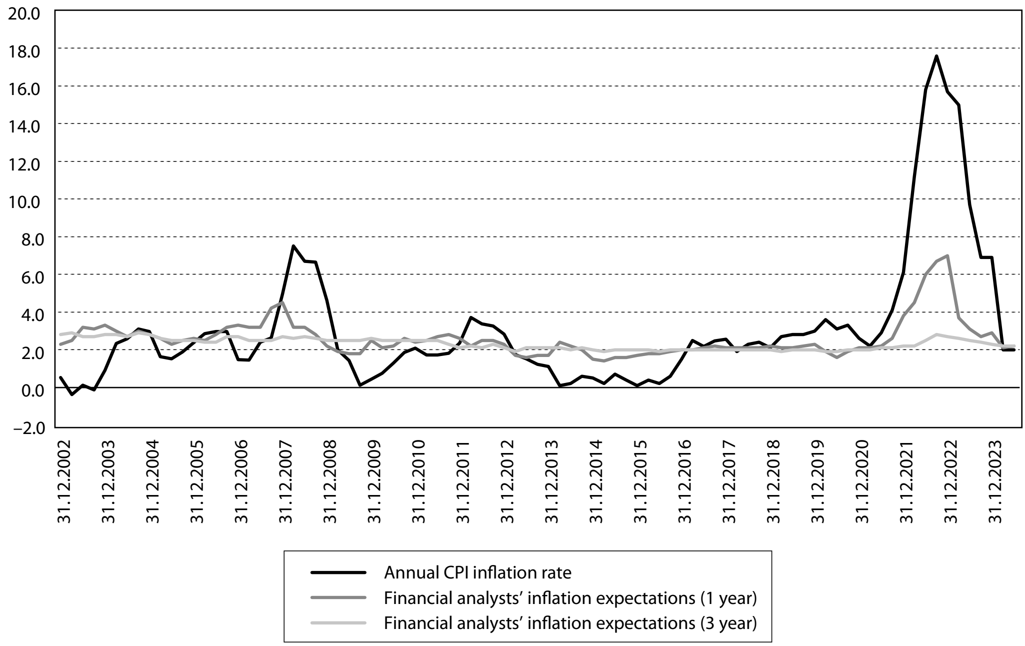
The most recent extensive research on this issue from the perspective of Czech circumstances was carried out by Brázdik et al. (2024). The authors analyse, among other things, the characteristics of inflation expectations in Czechia, highlighting the inflation year 2022. A comparison of financial analysts', non-financial enterprises' and households' inflation expectations shows that the distribution of households' one-year inflation expectations is significantly different (with a higher mean) from financial analysts' inflation expectations. In their analysis, they distinguish between the formation of inflation expectations (the strength of adaptive behaviour) in a low-inflation environment and a high-inflation environment. The authors further conclude that the adaptive expectations mechanism is more significant for one-year inflation expectations and less significant for three-year inflation expectations.

2 DESCRIPTIVE STATISTICS

Our empirical analysis is based on time series with a quarterly frequency (Figures 1 and 2). The basic set covers the period from Q3 1999 to Q2 2024 and comes from the CNB's ARAD database. The time series that enters our analysis is, with one exception, in the form of the YoY CPI inflation rate (π_t) at a quarterly frequency. The response variable in our econometric models is the expected YoY CPI inflation rate ($\pi_t^{e(t+n)}$) with time horizons of one year and three years for periods t to t+4 (time horizon one year) and for periods t+8 to t+12 (time horizon three years). The CNB surveys expected inflation periodically for two groups of entities (Figures 1 and 2): financial analysts (monthly frequency, 16 Czech analysts and 3 international analysts) and managers of firms and corporations (quarterly frequency, 118 enterprises until 2010 and a sample of 250 enterprises since 2011).[7] The CNB's inflation target (π^{T}) is also defined in the form of the YoY CPI inflation rate. The lagging YoY CPI inflation rate is the explanatory variable in the case of adaptive expectations modelling. Another explanatory variable in our econometric models is the CNB's inflation forecast with a time horizon of one year ($\pi_t^{P(t+4)}$). This time series does not take the form of a year-on-year inflation rate as in the previous cases but takes the form of a comparison of the average quarterly price level in the current quarter against the average quarterly price level in the same quarter of the previous year. This is a less commonly used method of calculating the inflation rate, which is not posted on the CZSO's website but is available in the CNB's ARAD database.

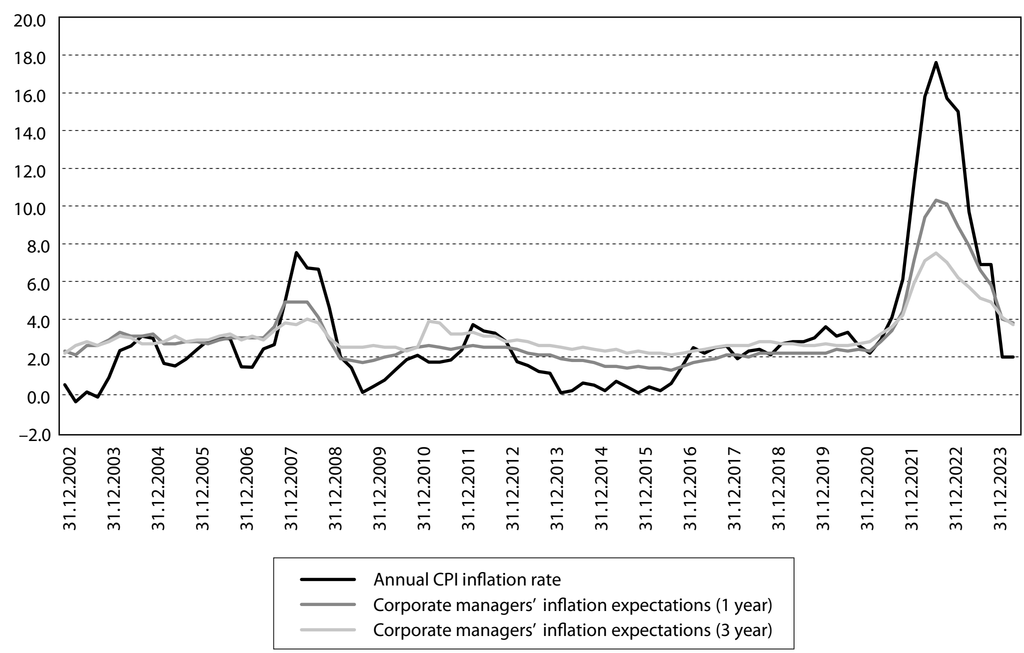
[7] The time series of households' inflation expectations in the CNB's ARAD database ends in 2007, and so we do not include it in our analysis. In general, however, these time series are characterised by large variance of the values obtained from respondents.

Figure 1 Financial analysts' inflation expectations (%)



Source: CZSO, CNB (ARAD), 2024

Figure 2 Corporate managers' inflation expectations (%)



Source: CZSO, CNB (ARAD), 2024

2.1 Tests of the seasonality and stationarity of time series

We used the Census X-12-ARIMA method (F-test, 1% significance level) to test for and then eliminate seasonality. Employing this method, we detected a seasonal component in the time series of the CNB's one-year inflation forecast and in the time series of financial analysts' inflation expectations with a horizon of one year and three years. The seasonal component is positive in Q3 and Q4 and negative in Q1 and Q2. At the same time, the seasonal component of financial analysts' yearly inflation expectation for one year is statistically significantly correlated with the CNB's inflation forecast (0.8638). The seasonal component was not found in the YoY CPI inflation rate published by the Czech Statistical Office (CZSO). This is the first observation indicating the different nature of the time series, i.e., on the one hand, the three time series representing inflation forecasts and inflation expectations and, on the other hand, the actual statistically found YoY CPI inflation rates. This observation is not consistent with the concept of rational expectations.

Unit root tests (Table 1) were performed using the Augmented Dickey-Fuller Test (ADF) and the ADF trend break test. Based on the model with a constant[8] we reject the hypothesis of the existence of a unit root for the time series of financial analysts' and corporate managers' one-year inflation expectations and financial analysts' three-year inflation expectations. Based on the model with a constant and a one-time break, we reject the hypothesis of the existence of a unit root in the case of the time series of year-on-year inflation rates and corporate managers' three-year inflation expectations. In the case of the CNB's inflation target, which gradually declined towards 2% over the period under review, we do not reject the hypothesis of the existence of a unit root at the 5% significance level.

Table 1 Unit root tests: Augmented Dickey-Fuller test (ADF)

Basic time series (Q3/1999–Q2/2024, quarterly frequency)	*t*-statistics (critical value 5%)	Prob.	Lag length, const., break date
YoY CPI inflation rate	–5.0802 (–4.4436)	< 0.01	4, const., Q2 2021
Financial analysts' inflation expectations* (1-year horizon)	–3.0603 (–2.8912)	0.0330	1, const.
Financial analysts' inflation expectations* (3-year horizon)	–3.6361 (–2.8909)	0.0066	0, const.
Corporate managers' inflation expectations (1-year horizon)	–4.0197 (–2.8912)	0.0020	1, const.
Corporate managers' inflation expectations (3-year horizon)	–5.3149 (–4.4436)	< 0.01	1, const., Q3 2021
CNB inflation forecast horizon* (1-year horizon)	–3.6729 (–2.8951)	0.0062	0, const.
CNB inflation target	–2.1629 (–2.8959)	0.2213	0, const.

Note: *seasonally adjusted time series.
Source: Author's calculations

[8] The algorithm for testing models with a trend and with a constant was chosen by Enders (2014).

2.2 Correlation analysis

The results of the correlation analysis in the observation period from Q3/1999 to Q2/2024 (quarterly frequency) suggest some characteristics of the formation of inflation expectations amongst financial analysts and corporate managers, which we will test in the selected models.

In the case of financial analysts, the correlation between the time series of the YoY CPI inflation rate (the last known value when forming expectations) and the expected year-on-year inflation rate with a time horizon of one year and three years is 0.7921 and 0.2588, respectively. However, the correlation of the current YoY CPI inflation rate with its values lagging one year and three years is only 0.3704 and 0.1250, respectively. Financial analysts' inflation expectations with time horizons of one year and three years are moderately strongly correlated at 0.5546. The correlation of the actually observed values of the YoY CPI inflation rate with time horizons of one year and three years is statistically insignificant at a 5% significance level (it is only 0.1223).

In the case of corporate managers, the correlation between the time series of the YoY CPI inflation rate (the last known value when forming expectations) and the expected year-on-year inflation rate with a time horizon of one year and three years is extremely high, i.e. 0.9501 and 0.9341 respectively. The correlation between their expectations of the year-on-year inflation rate with time horizons of one year and three years is also extremely high and amounts to 0.9725.

3 ECONOMETRIC ANALYSIS OF THE RATIONAL EXPECTATIONS HYPOTHESIS

The EViews 13 econometric software was used in the testing. The estimates presented in Tables 2 and 3 were made using the ARMA Generalized Least Squares (Gauss-Newton) method and the estimates of the final model presented in Table 4 use the least squares method.

We verify the problem of autocorrelation of the unsystematic component of individual models using the Durbin-Watson test (test for the existence of autocorrelation of first-order residuals), the results of which we publish. Higher-order autocorrelation was operationally tested using the Breusch-Godfrey autocorrelation test and the Ljung-Box test. The presence of heteroskedasticity was verified by means of the ARCH test.

3.1 Test of the unbiased expectations hypothesis

Under the unbiased expectations hypothesis, the inflation expectations of the tested groups are an unbiased predictor of actual future inflation,

$$\pi_{t+n} = \gamma_0 + \gamma_1 \pi_t^{e(t+n)} + \tau_t \, , \tag{1}$$

where τ is white noise and n is the number of quarters. Null hypotheses have parameters:

$\gamma_0 = 0$ and $\gamma_1 = 1$.

In the case of non-stationary time series, this hypothesis is usually tested using a relationship for the expected change in inflation:

$$\pi_{t+n} - \pi_t = \gamma_0 + \gamma_1 \left(\pi_t^{e(t+n)} - \pi_t \right) + \tau_t, \tag{2}$$

where τ is white noise and n is the number of quarters. Null hypotheses have parameters:

$\gamma_0 = 0$ and $\gamma_1 = 1$.

Table 2 Test of the unbiased expectations hypothesis (response variable: change in the YoY inflation rate $\pi_{t+n} - \pi_t$)

	Constant	Expected change in inflation	D-W stat.	Autocorrelation of higher order residues	*R* sq.
Financial analysts (1-year forecast, Q3/1999–Q1/2023)	0.3465 (0.2778)	1.2583* (8.7996)	1.8448	AR(1) 0.8772* (14.6295) MA(1) 0.414542* (3.6816)	0.8993
Financial analysts (3-year forecast, Q4/2002–Q1/2021)	–0.0616 (–0.0385)	1.0495 (5.7728)	1.8464	AR(1) 0.8774* (15.3030) MA(1) 0.5990* (6.3047)	0.8928
Corporate managers (1-year forecast, Q3/1999–Q1/2023)	–0.3182 (–0.2317)	1.8402* (9.3319)	1.7187	AR(1) 0.8931* (15.7894) MA(1) 0.4015* (3.4888)	0.9045
Corporate managers (3-year forecast Q4/2002–Q1/2021)	0.2829 (0.1793)	0.9328 (4.2831)	1.7461	AR(1) 0.8574* (12.9602) MA(1) 0.5762* (5.2628)	0.8806

Note: We show the values of the estimated parameters in the first row and the t-statistics in parentheses in the second row. *, **, *** denote parameters that are statistically significant at the 1%, 5% and 10% significance levels (constant from zero, expected change in inflation from one).
Source: Author's calculations

The unbiased expectation hypothesis was not confirmed (Table 2), although the regression coefficients before the constant are statistically insignificant, consistent with the unbiased expectation hypothesis. In the case of yearly inflation expectations of both financial analysts and corporate managers, the estimated coefficients for the expected change in inflation are statistically significantly different from one. In all the cases tested, the existence of AR(1) processes was shown at the 1% significance level, which is contrary to the unbiased expectation hypothesis. A statistically significant MA(1) process is consistent with the findings of Pesaran (1987), and Suk-Joong (1997) and is related to the higher frequency of observations (quarterly) compared with the length of inflation expectation (one year).

3.2 Testing the properties of prediction errors

The test of the properties of prediction errors can be based on the test for no serial dependence of prediction errors and the requirements of the hypothesis of weak-form market efficiency.

The test for no serial dependence verifies whether previous errors in expectations are used in the new predictions. In backward observation and testing, prediction errors should not be constant and should not depend on their previous values:

$$\pi_{t+n} - \pi_t^{e(t+n)} = \beta_0 + \sum_{m=1}^{k} \beta_{t-m}(\pi_{t+n-m} - \pi_{t-m}^{e(t+n-m)}) + \omega_{t+n}, \tag{3}$$

where ω_{t+n} is white noise, m is the number of quarterly lags. Null hypotheses have parameters: $\beta_0 = 0$ and $\beta_{\text{t-m}} = 0$.

The autocorrelation test of weak-form market efficiency (Fama, 1965) is based on the decomposition of actually observed changes into an expected part and an unexpected part. In the case of inflation expectations, this is the relationship:

$$\pi_{t+n} - \pi_t = (\pi_t^{e(t+n)} - \pi_t) + \left(\pi_{t+n} - \pi_t^{e(t+n)}\right). \tag{4}$$

Independence from past changes is required only in the case of the unexpected component (Fama, 1965), i.e. in the case of prediction errors. In contrast, an expected change can be correlated with a past change if the previous observed changes are persistent. The tested econometric equation has the following form. We test whether prediction errors can be explained by previous observed changes in inflation:

$$\pi_{t+n} - \pi_t^{e(t+n)} = \alpha_0 + \sum_{l=0}^{k} \alpha_{t-l}(\pi_{t-l} - \pi_{t-l-1}) + \vartheta_{t+n}, \tag{5}$$

where ϑ_{t+n} is white noise, l is the number of quarterly lags. Null hypotheses have parameters: $\alpha_0 = 0$ and $\alpha_{t-l} = 0$.

Based on both approaches for testing prediction errors, we formulate the following econometric equation:

$$\pi_{t+n} - \pi_t^{e(t+n)} = \alpha_0 + \sum_{l=0}^{k} \alpha_{t-l}(\pi_{t-l} - \pi_{t-l-1}) + \sum_{m=1}^{k} \beta_{t-m}(\pi_{t+n-m} - \pi_{t-m}^{e(t+n-m)}) + \varepsilon_{t+n}, \tag{6}$$

where ε_{t+n} is white noise. Null hypotheses have parameters: $\alpha_0 = 0$, $\alpha_{t-l} = 0$ and $\beta_{t-m} = 0$.

Table 3 lists the results of the tests for weak-form efficiency and the absence of serial dependence of errors in expectations. Financial analysts' and corporate managers' one-year and three-year inflation expectations have statistically insignificant coefficients for the constant at the 10% significance level. However, the other findings are not consistent with the required properties of the rational expectations model. Our tests show, at the 1% significance level, that prediction errors are related to prior prediction errors (t–1 and t–2 lags) for both groups of analysed respondents. In the case of one-year inflation expectations, our next finding is also inconsistent with the rational expectations hypothesis. A change in inflation lagging one period can (at the 1% level of statistical significance) explain some of the prediction errors in the case of financial analysts' as well as corporate managers' expectations. Let us add that the one-period lag of the change in inflation is the observed change in inflation at the time new inflation expectations are being formed.[9]

Table 3 Test for weak-form efficiency and absence of serial dependence (response variable: prediction error $\pi_{t+n} - \pi_t^{e(t+n)}$)

	Constant	Past change in inflation	Lagging prediction error (*t*–1 and *t*–2)	D-W stat.	*R* sq.
Financial analysts (1-year forecast, Q4/1999–Q1/2023)	–0.0512 (–0.5208)	0.5839* (4.5032)	1.3903* (16.990) –0.3828* (–4.2994)	2.0861	0.8931
Financial analysts (3-year forecast, Q4/1999–Q1/2021)	–0.0501 (–0.4317)	0.1451 (0.9038)	1.4276* (14.060) –0.5401* (–5.2881)	2.0215	0.9008
Corporate managers (1-year forecast, Q4/1999–Q1/2023)	–0.4687 (–0.4737)	0.8300* (5.7853)	1.4345* (19.223) –0.3761* (–4.2544)	1.9477	0.8940
Corporate managers (3-year forecast, Q4/1999–Q1/2021)	–0.0182 (–0.1294)	0.1015 (0.5621)	1.4532* (12.992) –0.5747* (–5.0944)	2.0111	0.8904

Note: We show the values of the estimated parameters in the first row and the t-statistics in parentheses in the second row. *, **, *** denote parameters that are statistically significant at the 1%, 5% and 10% significance levels.
Source: Author's calculations

4 SHAPING OF FINANCIAL ANALYSTS' AND CORPORATE MANAGERS' INFLATION EXPECTATIONS

Let us ask: what is the impact of lagging inflation expectations, past inflation (π_{t-m}), and the CNB's year-on-year inflation forecast ($\pi_t^{P(t+4)}$) and the CNB's inflation target (π^{T}) on the formation of financial analysts' and corporate managers' inflation expectations ($\pi_t^{e(t+n)}$)? We test the econometric model:[10]

[9] The last known value of the inflation rate as the only explanatory variable in the inflation expectations model is called the naive expectation. This is a special case of adaptive expectations.
[10] For comparison, e.g., Mandel (2009), Łyziak and Paloviita (2017), and Brázdik et al. (2024).

$$\pi_t^{e(t+n)} = c_0 + \sum_{l=1}^{k} \alpha_{t-l}\pi_{t-l}^{e(t+n-l)} + \sum_{m=1}^{k} \beta_{t-m}\pi_{t-m} + \vartheta_P \pi_t^{P(t+4)} + \gamma_T \pi_t^{T(t+n)} + \sum_{v=1}^{k} \varphi_{i,t-v} x_{i,t-v} + \epsilon_t \,, \tag{7}$$

where x_i represents the selected control variables[11] and ϵ_t is white noise.

Comparing the factors that inform financial analysts' and corporate managers' inflation expectations (Table 4), we can see considerable differences.

Table 4 Test of determinants of inflation expectations

	Constant	Lagging response variable	Past inflation (t–1, t–2)	CNB's inflation forecast (4Q)	CNB's inflation target	D-W stat.	*R* sq.
Financial analysts (1-year forecast, Q3/2002–2/2024)	0.2080*** (1.6408)	0.4364* (5.7447)	0.1125* (4.6533) –0.0621** (–2.3252)	0.4281* (7.3135)	Not categorised	1.9097	0.9059
Financial analysts (1-year forecast, Q1/2009–Q2/2024)		0.4997* (6.4346)	0.1305* (4.7922) –0.0800** (–2.6351)	0.4752* (6.9609)		2.0560	0.9236
Financial analysts (3-year forecast, Q3/2002–Q4/2023)	0.1882* (2.7014)	0.8586* (26.4767)		0.0532* (4.3953)	Not categorised	2.0964	0.9214
Financial analysts (3-year forecast, Q1/2007–Q4/2023)		0.8603* (20.8789)		0.0506* (4.0009)	0.0939** (2.1730)	2.0474	0.9002
Corporate managers (1-year forecast, Q3/2002–Q2/2024)		0.7487* (13.8052)	0.3156* (12.7839) –0.1972* (–4.9740)	0.1653* (4.4015)	Not categorised	1.8849	0.9781
Corporate managers (1-year forecast, Q1/2009–Q2/2024)		0.7272* (10.5030)	0.3702* (15.9450) –0.2083* (–4.5776)		0.1496* (3.2430)	2.0019	0.9860
Corporate managers (3-year forecast, Q3/2002–Q4/2023)	0.5503* (2.7967)	0.6869* (7.7809)	0.1862* (6.8203) –0.1021* (–3.1024)	0.0733*** (1.6307)	Not categorised	1.6659	0.9418
Corporate managers (3-year forecast Q1/2007–Q4/2023)		0.7200* (7.4554)	0.2152* (8.1085) –0.1300* (–3.8774)		0.3060* (2.8892)	1.7992	0.9472

Note: We We show the values of the estimated parameters in the first row and the t-statistics in parentheses in the second row. *, **, *** denote parameters that are statistically significant at the 1%, 5% and 10% significance levels. Explanatory variables with statistically insignificant parameters were not included in the estimated model.
Source: Author's calculations

[11] The control variables M2 monetary aggregate, loans to government, loans to residents, the CZK/EUR exchange rate, and the PRIBOR 3M interest rate at various lags were statistically insignificant.

In the case of the one-year horizon, corporate managers' inflation expectations have a significantly longer inertia (persistence) compared with financial analysts' expectations, while past inflation (i.e. the adaptive component of expectations) has a significantly heavier impact. However, the inertia of inflation expectations is partly technical and stems from the statistical properties of the time series of the YoY CPI inflation rate expectations with a quarterly frequency, which also contains nine 'old' values of expected month-on-month inflation in each new quarter.[12]

In contrast, financial analysts take more account of the CNB's one-year inflation forecasts when forming their year-on-year expectations. Larger analytical teams in financial institutions work with modelling approaches similar to those of the CNB in their forecasting, while smaller analytical teams at least regularly evaluate the CNB's inflation forecasts.

Financial analysts no longer take past inflation into account when forming their three-year inflation expectations. In the case of corporate managers, the respective coefficients are 'surprisingly' statistically significant, although the coefficients are lower than for one-year expectations. The coefficients for the lagging response variables, characterising the inertia of inflation expectations, are again high and statistically significant for both analysed groups.

Given that the CNB's inflation target for the period 1999–2023 is declining and thus non-stationary, its impact was tested only on a shortened price series when the CNB adopted the 2% target (i.e. from 2009 and from 2007). The impact of the inflation target on the formation of inflation expectations was not demonstrated only in the case of financial analysts' one-year forecasts. For both groups, the hypothesis that the inflation target has a stronger impact on the formation of three-year expectations than on one-year expectations was confirmed.

4.1 Lessons learned in the monetary policy context

Inflation expectations are not formed solely on the basis of the impact of macroeconomic fundamentals. Monetary policy, the central bank's inflation forecast, and the inflation target are factors that influence both inflation and the formation of inflation expectations. The high credibility of the central bank's inflation target, reaction function and monetary policy can 'anchor' inflation expectations; on the one hand, this is desirable in terms of reducing the monetary policy costs. From another perspective, however, the 'anchoring' of inflation expectations may be a manifestation of market entities' passivity in shaping their inflation expectations. The 'de-anchoring' may then take a larger leap when monetary policy and the inflation target lose credibility.

The adaptive nature of inflation expectations is normally regarded as a phenomenon that forces the central bank to adjust its interest rate policy more restrictively in an effort to 'break' the inflationary development towards the inflation target.[13] In our view, however, this conclusion is not generally valid. It is justified in cases where inflation is persistent in nature, is primarily driven by the central bank's too expansionary monetary policy, and where adaptive expectations have a long lag. In cases of external cost shocks and short lags of adaptive expectations (especially in the case of naive expectations), this simple 'textbook' rule is questionable. The precautionary motive of savings, which responds to the decline in the current real income and expected real income, contributes significantly to the restriction. Adaptive inflation expectations quickly dissipate as the main external cost impetus gradually recedes.[14]

[12] More detailed technical interpretation of this problem, e.g., Hassler and Demetrescu (2004), and Arlt (2023).

[13] This issue is analysed in terms of different types of shocks by Filáček (2005). In the case of the existence of adaptive inflation expectations, Orphanides and Williams (2003) stress the importance of combining an explicit inflation target with the central bank's active communication.

[14] The main criticism of this 'dovish' approach is usually waged from the perspective of the possibility of 'de-anchoring' inflation expectations, indexation to inflation (including wages), and the smoothing behaviour of households borrowing for consumption in times of falling income (given a higher permanent income).

In the context of the current discussion, Arlt (2023) also points out some important statistical problems associated with the use of the YoY CPI inflation rate. In the CNB's current monetary policy, this form of inflation measurement occurs in the setting of the inflation target and in the monitoring of financial analysts' and corporate managers' inflation expectations. The fact that a one-month shift in this indicator always incorporates the 11 old month-on-month values becomes a real monetary policy problem in the disinflationary process from high inflation towards the inflation target. In a disinflationary process, this inflation indicator creates a perception of long-lasting high inflation in some market entities. This causes political pressure on the central bank in the form of calls for a more restrictive monetary policy than is actually needed. The transition to monthly seasonally adjusted and then annualised inflation is, in theory, an appealing step in the right direction but it runs into the problem of the instability of seasonal coefficients. The central bank will find it difficult to explain to the public that in any given month, seasonally adjusted monthly annualised inflation surges significantly above the inflation target by 'many' percentage points. At the same time, in another month, the opposite deflationary problem will occur. In this case, the CNB has correctly chosen a communication strategy and argumentation using the inflation momentum (Adam and Michl, 2023), i.e. the rolling average of three consecutive annualised seasonally adjusted monthly price changes supplemented by decomposition into the various components of the consumer basket.

The CNB's forecasting model does not work with the YoY CPI inflation rate but with an average year-on-year inflation rate calculated by comparing the average quarterly price level in the current quarter with the average quarterly price level in the same quarter of the previous year. There is therefore an inconsistency between the definition of the CNB's inflation target (year-on-year inflation rate) and the construction of the CNB's forecasting model (average year-on-year inflation rate on a quarterly basis). This inconsistency is largely determined by the quarterly frequency of the national accounting data that enter the CNB's forecasting apparatus. However, it should be noted that the time series of the average year-on-year inflation rate on a quarterly basis is technically bound to lag behind the time series of the YoY CPI inflation rate.

CONCLUSION

Inflation expectations play an important role in the transmission mechanism of inflation targeting in the context of the length and costs of the disinflationary process. The purpose of our paper has been to verify, via econometric analysis, whether the inflation expectations of financial analysts and corporate managers in the Czech Republic (from Q3/1999 to Q2/2024) have the basic features of rational expectations and what influence the past YoY CPI inflation rate, the CNB's inflation forecast, and the CNB's inflation target have on their expectations.

We would like to point out that the CNB's inflation forecast does not take the form of a year-on-year inflation rate, which is the form of the CNB's inflation target and financial analysts' and corporate managers' inflation expectations under review, but takes the form of a comparison of the average quarterly price level in the current quarter against the average quarterly price level in the same quarter of the previous year.

We find that the formation of financial analysts' and corporate managers' yearly inflation expectations with time horizons of one year and three years is quite different. For corporate managers' inflation expectations, adaptive reasoning plays a more important role. Financial analysts take more account of the CNB's one-year inflation forecasts when forming their yearly expectations and the inflation target as an explanatory variable is statistically significant only for their three-year inflation expectations. The hypothesis that the inflation target has a stronger influence on the formation of three-year expectations than on one-year expectations has therefore been confirmed for both groups.

Neither group of the analysed respondents meets the required criteria for rational expectations in terms of the tests formulated by Pesaran (1987), and Fama (1965 and 1970). In particular, their yearly inflation expectations exhibit systematic errors. The correlations between their one-year and three-year inflation expectations are high and are completely inconsistent with the actually observed properties of time series of the YoY CPI inflation rate. Surprisingly, financial analysts' expectations (for both one year and three years) have a seasonal component, which is again inconsistent with the actually observed properties of time series of the YoY CPI inflation rate. The explanation seems to be the financial analysts' 'copying' of the CNB's yearly inflation forecast, which also contains this seasonal component.

References

ADAM T., MICHL, A. (2023). Inflation is declining – an interim assessment of the disinflationary process in the Czech Republic. *CNB Blog*, 15 May 2023.

ARLT, J. (2023). The problem of the year-on-year inflation rate indicator [online]. *International Journal of Finance & Economics*, 28(3): 2772–2788. <https://doi.org/10.1002/ijfe.2563>.

BLANCHFLOWER, D. G., MACCOILLE, C. (2009). The Formation of Inflation Expectations: an Empirical Analysis for the UK [online]. *NBER Working Paper Series*, No 15388, September 2009. <https://doi.org/10.3386/w15388>.

BRAYTON, F., MAUSKOPF, E., REIFSCHNEIDER, D., TINSLEY, P., WILLIAMS, J. (1997). The Role of Expectations in the FRB/US Macroeconomic Model [online]. *Federal Reserve Bulletin*, 83(April): 227–245. <https://doi.org/10.17016/bulletin.1997.83-4>.

BRÁZDIK, F., ŽÁČEK, J., KESELIOVÁ, T., MUSIL, K., ŠNOBL, R., ŠOLC, J., TVRZ, S. (2024). Understanding Inflation Expectations: Data, Drivers and Policy Implications. CNB, *Working Paper*, 3/2024.

CARROLL, C. D. (2003). Macroeconomic expectations of households and professional forecasters [online]. *Quarterly Journal of Economics*, 118(1): 269–298. <https://doi.org/10.1162/00335530360535207>.

COIBION, O., GORODNICHENKO, Y., KUMAR, S. (2018). How do firms form their expectations? New survey evidence [online]. *American Economic Review*, 108(9): 2671–2713. <https://www.jstor.org/stable/26528539>.

ENDERS, W. (2014). *Applied Econometric Time Series.* 4th Ed., Wiley, 496 p. ISBN 978-1-118-80856-6

FAMA, E. F. (1965). The Behavior of Stock-Market prices [online]. *The Journal of Business*, (38)1: 34–105. <https://doi.org/10.1086/294743>.

FAMA, E. F. (1970). Efficient Capital Markets: a Review of Theory and Empirical Work [online]. *The Journal of Finance*, (25)2: 383–417. <https://doi.org/10.1111/j.1540-6261.1970.tb00518.x>.

FILÁČEK, J. (2005). Inflation Expectations in Monetary Policy [online]. *Finance a úvěr (Finance and Credit)*, 55(7–8): 380–393. <http://ideas.repec.org/a/fau/fauart/v55y2005i7-8p380-394.html>.

GELAIN, P., ISKREV, N., LANSING, K. J., MENDICINO, C. (2019). Inflation Dynamics and Adaptive Expectations in an Estimated DSGE Model [online]. *Journal of Macroeconomics*, 59(C): 258–277. <https://doi.org/10.1016/j.jmacro.2018.12.002>.

HASSLER, U., DEMETRESCU, M. (2004). Spurious Persistence and Unit Roots Due to Seasonal Differencing: the Case of Inflation Rates [online]. *Journal of Economics and Statistics*, 225(4): 413–426. <https://doi.org/10.1515/jbnst-2005-0404>.

KIM, S.-J. (1997). Testing the Rationality of Exchange Rate and Interest Rate Expectations. An Empirical Study of Australian Survey-based Expectations [online]. *Applied Economics*, 29(8): 1011–1022. <https://doi.org/10.1080/000368497326408>.

ŁYZIAK, T., PALOVIITA, M. (2017). Anchoring of inflation expectations in the euro area: Recent evidence based on survey data [online]. *European Journal of Political Economy*, 46(C): 52–73. <https://doi.org/10.1016/j.ejpoleco.2016.11.001>.

MANDEL, M. (2009). Milestones of Czech Monetary Policy. In: *20 let finančních a bankovních reforem v České republice* (*20 Years of Financial and Banking Reforms in the Czech Republic*). Prague: Oeconomica, 11 November 2009, 13–19.

METZLER L. A. (1941). The Nature and Stability of Inventory Cycles [online]. *Review of Economics and Statistics*, 23(3): 113–129. <https://doi.org/10.2307/1927555>.

MIYAJIMA, K., YETMAN, J. (2018). Inflation Expectations Anchoring across Different Types of Agents: the Case of South Africa. *IMF Working Paper 2018/177*, Washington, DC: International Monetary Fund. ISBN 9781484372043

MUTH, J. F. (1961). Rational Expectations and the Theory of Price Movements [online]. *Econometrica*, 29(3): 315–335. <https://doi.org/10.2307/1909635>.

NERLOVE, M. (1958). Adaptive Expectations and Cobweb Phenomena [online]. *Quarterly Journal of Economics*, 72(2): 227–40. <https://doi.org/10.2307/1880597>.

NERLOVE, M. (1983). Expectations, Plans, and Realizations in Theory and Practice [online]. *Econometrica*, 51(5): 1251–79. <https://doi.org/10.2307/1912274>.

ORPHANIDES, A., WILLIAMS, J. C. (2003). Imperfect Knowledge, Inflation Expectations, and Monetary Policy [online]. *NBER Working Paper Series*, No 9884. <https://doi.org/10.3386/w9884>.

PESARAN, M. H. (1987). *The Limits to Rational Expectations*. New York: Basil Blackwell, 325 p. ISBN 0-631-14036-0

SARGENT, T. (1986). *Rational Expectations and Inflation.* New York: Harper and Row, 212 p. ISBN 978-0060457419
SIMON A. H. (1984). *Models of Bounded Rationality (Economic Analysis and Public Policy).* MIT Press, 504 p. ISBN 9780262690867
TOMŠÍK, V., MANDEL, M. (2004). Formation of Inflation Expectations in the Czech Economy. *Bankovnictví (Banking)*, 12(8): 15–17.
VÁVRA, M. (2005). Inflation Expectations. *Research Paper*, 1/2005, Prague: Ministry of Finance of the Czech Republic.

Making Sense of Exchange Rate Pressures in Macroeconomic Statistics

Anna Drahozalová[1] | *Charles University, Prague, Czechia*
Václav Rybáček[2] | *University of Economics and Management, Prague, Czechia*

Received 2.7.2024 (revision received 30.8.2024), **Accepted** (reviewed) 30.9.2024, **Published** 14.3.2025

Abstract

The paper explores new statistical approaches for studying links between cross-border financial flows and the exchange rate. The paper first discusses limitations of the balance of payments statistics in analysing foreign exchange market imbalances by explaining the methodological principles that drive a wedge between balance of payments transactions and cross-border flows. Then it continues with a discussion on the analytical potential of the so-called monetary presentation of the balance payments. Visual analysis tentatively suggests that in the case of the CZK/EUR currency pair, net external flows (an aggregate identified by the monetary presentation) can explain reasonably well past exchange rate developments, if adjusted for central bank intervention.[3]

Keywords	***DOI***	***JEL code***
Balance of payments, monetary presentation, monetary aggregates, foreign exchange	*https://doi.org/10.54694/stat.2024.33*	*F31, F32, E00*

INTRODUCTION

In 2022, elevated inflation engulfed the Czech economy, invoked by a multitude of factors among which the aftermath of the COVID-19 crisis, the outbreak of the war in Ukraine and loose monetary policy in the preceding period feature high on the list. Aiming to curb growing prices, the Czech National Bank raised interest rates to 7%, which provided an incentive for profit-seeking investors to move their capital into assets denominated in the Czech koruna. In result, the large 2022 current account deficit (4.9% of GDP) was accompanied by the Czech koruna appreciating on the foreign exchange market (hereinafter: "FX market") . Such evolution of the current account and the exchange rate seems to be rather counter-intuitive.

For decades, the current account balance has been seen as a proxy of the demand and supply of foreign exchange (Dornbusch and Fischer, 1980). As proposed by Gourinchas and Rey (2007), the intertemporal approach to the current account suggests that high current account deficits today will

[1] Charles University, Institute of Economic Studies, Opletalova 26, 110 00 Prague 1, Czechia. E-mail: anna.drahozalova@cnb.cz.
[2] University of Economics and Management, Nárožní 2600, 158 00 Prague 5, Czechia. E-mail: vaclav.rybacek@vsem.cz.
[3] The views expressed in this paper are those of the authors and do not necessarily reflect those of the Czech National Bank.

need to be compensated for in the future. That could happen via trade or valuation channels, in both of which depreciation of the currency plays a vital role. In fact, at times these theories were elaborated upon, capital flows recorded on the financial account were relatively unimportant and thus current account balance could have served as a proxy for FX market imbalances relatively well. However, this is not currently the case. The binding capital controls that were present during the 20th century[4] were lifted in the late 1990s[5]; capital is therefore broadly free to move across jurisdictions around the world.

Indeed, according to the BIS survey (BIS, 2022), global FX trading rose to $7.5 trillion per day in April 2022, with 14% annual growth. To put this figure into perspective, the world trade reached $25.3 trillion in the whole year 2022 (WTO, 2023). Not least, foreign direct investment (FDI) flows increased globally from 0.55% of the world's GDP in the 1970s and 1980s to 3.11% in the 2000s and early 2010s[6] (World Bank).[7] The financial integration was further supported by the emergence and growth of international organizations such as the International Bank for Reconstruction and Development (IBRD), whose task is to promote the flow of long-term loans to less developed countries. By 2010, the average daily turnover on the foreign exchange market was 36 times the balance of trade according to King et al. (2012). Comparing these volumes suggests that FX changes are dominated by capital flows rather than by traded commodities or income flows. In other words, with global FX trading clearly surpassing the world trade volume, most FX movements must be tied up with financial transaction (Miller, 2002). As Miller (2002) further points out, economic fundaments such as economic growth dominate in the long run, while short term drift in the spot rate depends rather on the interest rate differential.

Rather weaker link between the current account balance and the evolution of the exchange rate creates a need to find another statistical tool how to analyse demand and supply for foreign exchange. In fact, strengthening interconnectedness of global financial markets introduced increased complexity to the financial account. Consequently, when analysing the BoP to determine the excess demand for foreign currency, the financial account cannot be overlooked, as doing so would result in biased conclusions.

In 2008, the ECB paper presenting so-called monetary presentation of the balance of payments has been published, suggesting an alternative presentation meant to complement the analysis of the monetary aggregates development. Reflecting on the expanding volume of capital flows highly surpassing the flow recorded in the current account, the paper suggests a statistical system monitoring domestic monetary development in relation to external transactions. The aim is to identify transactions contributing to the net external assets of domestic banks and hence having an impact on the level of money holdings in the domestic economy (Be Duc et al., 2008). Although meant for monetary aggregates analysis, the monetary presentation could potentially be useful in analysing supply and demand on the FX market, for it better reflects FX relevant transactions compared to the current account alone (Drahozalová, 2023).

In the following chapters, we will thus firstly explain the classification of transactions in the balance of payments model, then we will describe the basic structure of the monetary presentation of the balance of payments and, in the last chapter, we will analyse the development of the key elements of the monetary presentation in Czechia.

---

4 During the 20th century capital controls were implemented for various reasons, starting with the Great Depression (1930s) to Bretton Woods system (1944–1971) and Cold War era (1945–1990s).

5 The effectiveness and duration of capital controls varied across countries and time periods under consideration, however from the late 1990s the trend has generally been toward liberalization and the relaxation of capital controls, driven by globalization and the increased integration of financial markets.

6 The period considered here ends with the year 2017, while the peak in FDI flows was in 2007 (5.33% GDP). From 2018 the worlds' FDI is on a decline as increased economic uncertainty (US-China trade war, covid pandemic or Russian invasion of Ukraine) dampens investor confidence.

7 The respective data were retrieved from the World Bank database in June 2023: *<https://data.worldbank.org/indicator/BX.KLT.DINV.WD.GD.ZS>*.

1 BALANCE OF PAYMENTS AND THE FX MARKET

Macroeconomic model named “Balance of Payments” (thereafter “BoP”) represents a statistical statement covering transactions and positions between residents and non-residents over a certain period. Practical compilation of the BoP follows the accounting rules laid down in the Balance of Payments Manual whose sixth edition is currently in effect (thereinafter “BPM6”). Despite being named as “Balance of payments”, the model however does not capture only transactions involving the payment of money but any transactions that reflect underlying resource flows.

The fundamental principle of the BoP compilation is double-entry accounting. Each transaction is recorded twice in the BoP: once as a credit entry and once as a debit entry. This dual principle ensures that the total value of credits is always equal to the total value of debits, resulting in a net balance of zero for all BoP entries. Furthermore, records entering the BoP must follow precise rules. Typically, credit entries record non-financial transactions such as exports of goods, income receivable from cross-border work, one-time transfers from abroad, as well as financial transactions involving a decrease in financial assets or an increase in financial liabilities. Conversely, debit entries represent the opposite.[8]

For example, when a domestic resident imports goods from abroad, the debit entry recorded on the current account represents the obligation to pay the rest of the world, while the credit entry recorded on the financial account mirrors it as the payment to foreign exporter is executed. Furthermore, the obligation to pay essentially represents the demand for foreign currency, while the entitlement to a payment from abroad represents the supply of foreign currency.[9] This simple concept can be expanded to include other current account items, such as trade in services, claims arising from cross-border work, or returns on capital. Assuming that all claims and obligations must be settled immediately, the current account deficit indicates a current excess in demand for foreign currency over its supply. In fact, at times this theory was elaborated upon, capital flows recorded on the financial account were relatively unimportant and thus current account balance could serve as a proxy for FX market imbalances relatively well.

1.1 Accrual principle of the BoP weakens the link between current account and the exchange rate

Nevertheless, in the previous example, we assumed that the transactions in the BoP are recorded on a payment basis, where revenues are recognized when cash is received and expenses are recognized when cash is paid. However, this simplified approach does not fully account for the common business practices, when payments are made in advance or deferred, as well as for the development of new financial products designed to facilitate international business activities, including trade credit arrangements. To sum up, the payment approach fails to recognize outstanding obligations or non-cash transactions.

To provide a fuller picture of the international flows of economic values, the BoP methodology employs an accrual approach recording transactions at the time they are incurred, regardless of the timing of cash flows (BPM6, par. 3.35). For instance, if goods are purchased on trade credit payable in three months, the corresponding entry reflects the import of goods (at the time of change of ownership) as a debit on the current account, while the financial account recognizes the trade credit due in three months.[10] Although the accrual principle allows us to consider outstanding obligations, barter trade, or transactions without a quid pro quo, it further complicates the situation. Under this principle, the current account includes not only exchange rate-relevant items that signal the current demand or supply of currency but also

[8] Debit entries record e.g. import of goods, income payable, gifts to abroad, increase in financial assets or decrease in financial liabilities.

[9] Here we need to differentiate between current and future demand for (supply of) currency. If the payment is deferred, the transaction implies a future demand for or supply of foreign currency.

[10] Note that this type of entry signals future (rather than current) demand for foreign currency.

outstanding obligations that contribute to future exchange rate movements. Additionally, it encompasses barter trade and transactions without a quid pro quo, which have no impact on the FX market, neither in the present nor in the future.[11]

1.2 Furthermore, the current account itself does not fully capture all cross-border financial flows

The differentiating between credit and debit transactions helps to explain the overall balance of accounting entries in the BoP. Analytically, in order to get better understanding of the imbalances in the transactions vis-à-vis the rest of the world, another breakdown of transactions is to be considered. The macroeconomic models such as national accounts or balance of payments make distinction between non-financial and financial transactions.[12] The former are often referred to as "above the line", while the latter "below the line". The transactions "below the line" can be further broken down into those associated with or financing the transaction "above the line", e.g. the receipt of money for selling goods abroad, and those having the nature of pure financial transactions. By drawing the line between both types of transactions, we arrive at the well know BoP identity:

$$CA + KA = FA, \tag{1}$$

where CA denotes the current account balance, KA refers to the capital account balance and FA to the financial account balance.[13] Leaving aside the item "errors and omissions", which conceptually reflect imperfections in the compilation process and data sources, we can formulate that FA encompasses both pure financial transactions and those that are a mere reflection of the transactions "above the line". In case of pure financial transactions,[14] the investing party expects the financial means originally provided to be returned at some point of time in the future, along with a certain compensation being regularly paid during the lifetime of financial instrument (more commonly in the form of interests or dividends). Leaving aside the income element, these transactions themselves exert no direct impact on the current account balance but increasing/decreasing demand/supply on the exchange rate on the FX market. Therefore, while being recorded out of the scope of the current and the capital accounts, these transactions may exert a strong pressure on the FX market.

The other group of financial transactions is associated with the transactions recorded on the current and the capital account where the paying party does not expect these financial means to be returned. In the logic outlined above, these transactions may be referred to as "not-purely financial transactions". This is the case of payments for the goods exported or imported, or for the payments of primary or secondary income. This group of transactions is simply mirroring the non-financial transactions

[11] It is worth noting that the aforementioned difficulties can be effectively mitigated if a country operates within a fixed exchange rate regime. Under such a framework, measuring the excess demand for foreign currency becomes more direct, precisely corresponding to the volume of foreign exchange that the central bank must introduce into the market to maintain exchange rate stability. In essence, when the demand for foreign currency exceeds its supply, the central bank intervenes by purchasing the excess liquidity from the market while simultaneously providing foreign currency to meet the market's needs. Conversely, a surplus supply of foreign currency prompts the opposite response. While examining external imbalances is an important task in the presence of the fixed exchange rate, we often wonder about the implications in the absence of central bank interventions. This is where the monetary presentation of the BoP, which is discussed further in the text, can provide valuable insights.

[12] Being consistent with the SNA and other macroeconomic statistics, this presentation is commonly referred to as „standard presentation" (BPM6, par. 14.15).

[13] The balances on both side of identity 1 are commonly referred to as net lending (+)/net borrowing (−) either from non-financial transactions (capital and current account) or from financial transactions (financial account). However, further in the text, the sum of current and capital account balance (left-hand side) will be referred to as „external balance".

[14] Such as purchases of securities, provisions of loans or depositing money in the bank.

and showing how the transactions "above the line" were financed whether in cash or in other ways. The fact that transactions do not need to take the form of cash transfer only brings us to another element weakening the linkage between the BoP fundamentals and the FX rate movements which is the capturing of non-cash transactions in the system. Methodologically, the involvement of non-cash transactions is necessitated by the application of so-called accrual principle, as prescribed in the BPM7 (par. 3.34).[15] Following this requirement of the methodology, all economic events are to be recorded in the period in which they occurred, irrespective of whether a corresponding flow of cash was observable or not. If no cash flow followed the event, the time lag between the occurrence of this economic event and corresponding cash payments is bridged by recording of a non-cash transaction such as loan, trade credit, other receivable, etc.

From the FX market perspective, the application of accrual principle weakens the explanatory power of the BoP balancing items, such as current account balance, towards the FX movements. The reason is obvious, if no cash is transferred between both parties concerned, no pressure on the exchange rate on the FX market is exerted as no currency conversion was carried out. Transactions with no cash movement constitute only future demand or supply of the currency.[16] The occurrence of purely accrual transactions therefore suggests that e.g. current accounts surplus does not automatically implies a pressure on the domestic currency to appreciate and, if the situation is reversed, depreciation of the domestic currency does not necessarily follow current account deficit. The link between the external balance and FX changes is further weakened by imputed values such as reinvested earnings or the EU flows as, following the agreement between economic policy authorities, the financial means from the EU do not enter the FX market but are converted by the central bank instead.

To sum up, there are many methodological aspects of the model that step in between the current account balancing items and the FX rate movements. That is the application of the accrual principle, which does not allow for a straightforward interpretation of the current account balance, paired with the fact that financial transactions, that are not to be found on the current account, play an equal role in FX determination. These methodological conceptions help in explaining seemingly puzzling evolution of the FX and the current account in the Czech Republic over the year 2022 where the Czech economy experienced a massive current account deficit accompanied by appreciating foreign exchange rate. In the following section, we use the monetary presentation of the balance of payments to better understand these concepts within a common methodological framework.

2 THE MONETARY PRESENTATION OF THE BALANCE OF PAYMENTS

The monetary presentation of the BoP was designed to analyse the monetary aggregate M3 and money creation in the economy through the intersection of two different statistics: consolidated balance sheet items of monetary financial institutions (MFIs)[17] and balance of payments statistics. By examining the intersection of a bank's net external assets, an external counterpart to the M3 aggregate (Aguilar et al., 2020), with the balance of payments statistics, we can examine changes in money supply in detail.

To detect the impact of the BoP on the monetary aggregate M3, it is necessary to decompose it for individual economic sectors which is another way of analysing the BoP (BPM6, par. 14.18). However, unlike the sector accounts in national accounts, where the balancing item on the capital account (B.9n) and on the financial account (B.9f) are to be balanced for each individual economic sector, this is essentially

[15] Except for the BoP, the application of accruals is applied also for other major macroeconomic models as national accounts or government finance statistics.

[16] Or no demand or supply of the currency at all, as is the case for barter trade and transactions without a quid pro quo.

[17] According to Regulation (EU) 2021/379 of the ECB, the term „monetary financial institutions" covers central banks, deposit taking corporations and money market funds (see Article 2 of the Regulation).

not the case in the BoP. Transactions in the current or in the capital accounts, and corresponding financial transactions in the financial accounts, are in the BoP routinely performed by different economic sectors, and recorded accordingly.[18]

For this purpose, we can decompose the right-hand side of Formula (1) in the following way:

$$CA + KA = FA_{MFI} + FA_{nonMFI} \,. \qquad (2)$$

Sectoral split by the type of the resident enables compilers and users to monitor the economic position of individual sectors towards the rest of the world. The BoP statistics here follows the classification as laid down in the manuals of national accounts SNA2008 and ESA2010. For the monetary presentation of the BoP, the sector dimension is essential as it distinguishes between monetary and non-monetary sectors. The former, which factually embraces sectors entitled to issue money (monetary and financial institutions entitled to issue money, commonly referred to as "MFI"), covers the sub-sectors S121, S122 and S123[19] as these are defined in the SNA/ESA/BPM. The latter, i.e. "non-MFI" therefore encompasses all other economic sectors which may be in the positions of money-holding or money-neutral sectors.

The sectoral breakdown allows us to separate the financial account for MFIs, which summarizes all transactions on the financial account made by MFIs and non-MFIs. Formula (2) can be further adjusted:

$$CA + KA - FA_{nonMFI} = FA_{MFI} \,. \qquad (3)$$

The right-hand side (FA_{MFI}) provides a conceptual link between the two statistical sets, i.e. the BoP statistics and the aggregate M3. Concretely, the aggregate FA_{MFI} can be equated to the transaction component of the changes in the net external assets[20] (hereinafter NEA) which is meant to quantify the inflow and outflow of money in/from the economy from an external perspective.

NEA represent a consolidated MFI's net position, i.e. the assets minus the liabilities vis-à-vis the rest of the world. External assets include loans and advances to customers, investments in securities, and cash and balances with other banks. External liabilities encompass deposits from customers, borrowings from other financial institutions, and other forms of funding obtained by the bank from external sources. In all cases, the counterparty is always a non-resident. By subtracting external liabilities from external assets, we arrive at NEA, a variable showing the net result of the MFIs interaction with the rest of the world.

What is important for our purposes is that transactions captured in NEA represent an inflow and outflow of money from an external sector, therefore affecting domestic money supply. Formally, we can write:

$$M3 = Credit + NEA + OA - CGD - LTFL \,, \qquad (4)$$

where the aggregate Credit covers loans granted to residents[21] as well as holdings of securities issued by residents. The item OA denotes "other net assets",[22] CGD refers to "deposits of central government"

[18] Suppose a household wishing to make an online purchase of a book from abroad. Once the product is delivered and paid for out of the buyer´s deposit held in the domestic bank, the corresponding entry is booked in the current account (import) and the households accounts (consumption of imported good). In the financial accounts, the counterparty of a decrease in the household´s financial asset is, however, the domestic bank, i.e. a resident. The final settlement of this purchase would then be recorded in the BoP as a transaction between domestic bank and its counterparty abroad, generally a correspondent bank where resident banks hold their accounts.

[19] See par. 2.67, ESA2010.

[20] One of the counterparts to monetary aggregate M3.

[21] Private sector and general government.

[22] Such as fixed assets, reverse repos, etc., in the stylized MFI consolidated balance sheet.

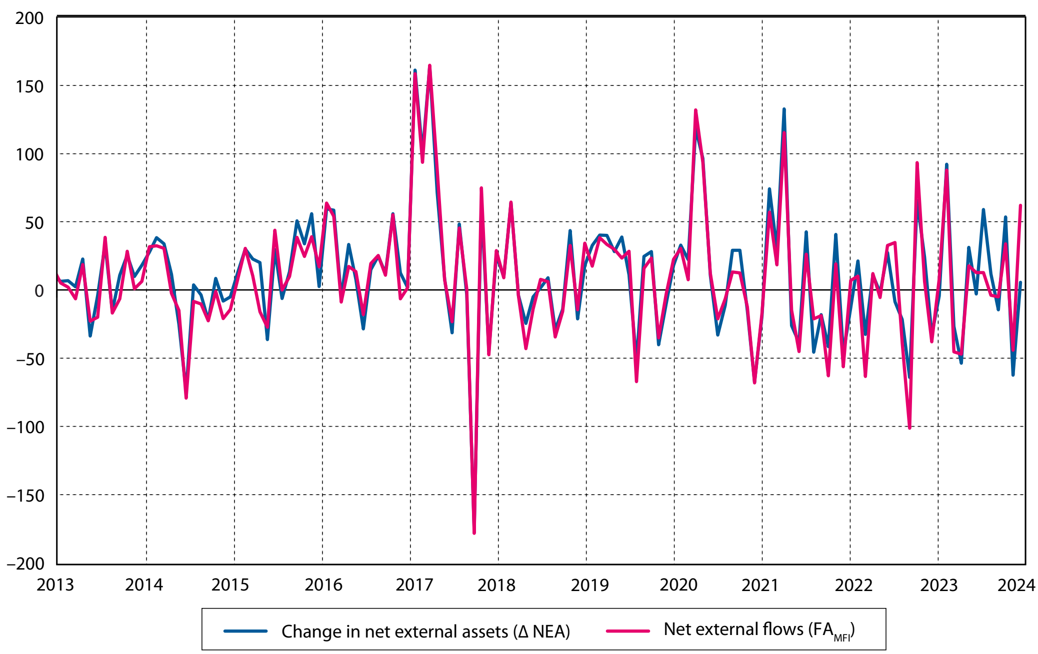
and LTFL to long-term financial liabilities such as long-term debt securities, capital and reserves or long-term deposits of government. Importantly, equation 4 holds both for positions and for net transactions (Aguilar et al., 2020). By deriving respective transactions from corresponding changes in the stocks of NEA, a conceptual link between the BoP and NEA is found and quantified.

The mechanism enabling this breakdown is based on the intersection of the two aforementioned statistics. While MFIs document all transactions – both domestic and external – in which they are involved, the BoP encompasses all transactions with the rest of the world for both MFI and non-MFI sectors. Therefore, by definition, the common intersection refers to the external transactions conducted by the MFI sector. We can therefore link the change in the stock item "NEA" in Formula (4) to the flows item "FA_{MFI}" in Formula (3). Then, we can further substitute in the following way to arrive at a formula showing the changes in M3:

$$\Delta M3 = \Delta Credit + CA + KA - FA_{nonMFI} + \Delta OA - \Delta CGD - \Delta LTFL . \quad (5)$$ [23]

Figure 1 shows change in the net external assets compared to net external flows as identified by the monetary presentation. We can see that the two time series are almost identical, except for minor inconsistencies that stem from different statistical approaches used to compile these statistics.

Figure 1 Change in net external assets and net external flows, Czechia, CZK bn, 2013–24

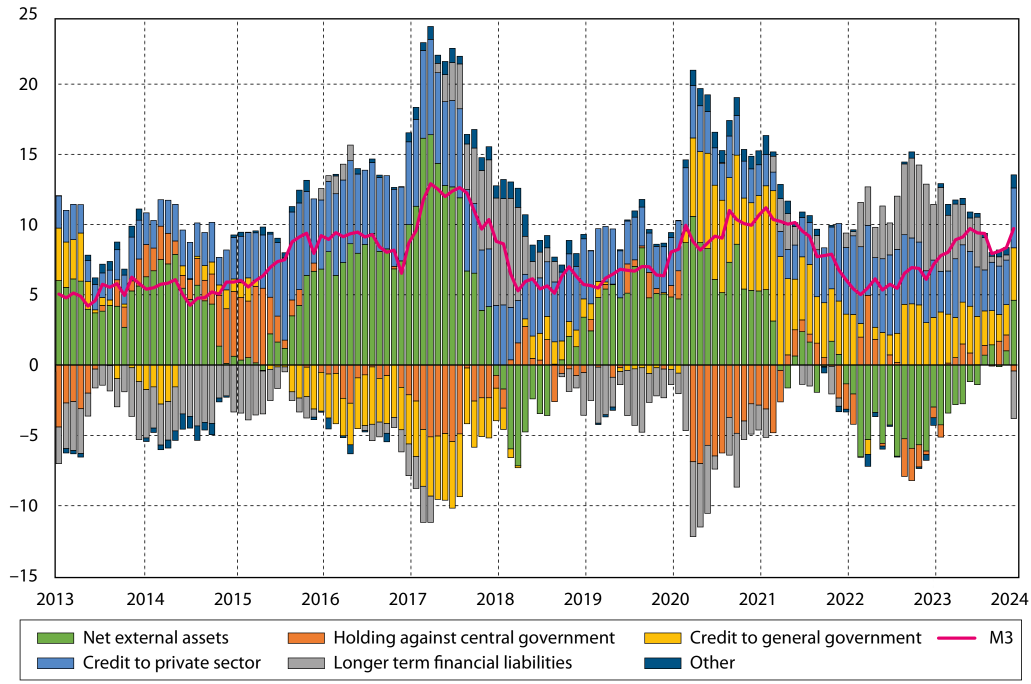


Source: CNB, own calculations

[23] Whether a transaction will impact M3 also depends on how it affects the MFI consolidated balance sheet as a whole. For instance, when a resident sells foreign share to a domestic bank, this leads to an increase in NEA. If the proceeds are then invested into long-term government bonds, where the government holds the money raised on its deposits in a domestic bank, than an increase in NEA will be offset by a corresponding increase in LFTL. As both changes cancel each other, M3 will remain unaffected.

As NEA is one of the counterparts to the M3 aggregate, separating FA_{MFI} from the rest of the BoP can helps us explain the changes in the money supply from the perspective of external developments. The following chart illustrates the contributions of NEA to the dynamics of the M3 monetary aggregate. Figure 2 depicts the annual percentage changes in M3 and the contribution of its counterparts to these changes, including the NEA aggregate. Pronounced contributions are observable within the period leading up to the second quarter of 2017, during which the central bank exited its intervention regime. The end of the currency cap regime was preceded by a substantial influx of capital as foreign investors heavily engaged in investing into domestic assets, anticipating future currency appreciation. Simultaneously, the current account registered a surplus, thereby providing an additional impulse to the expansion of the money supply. Nonetheless, the inherent characteristics of these transactions are not readily apparent when relying only on the MFI's statistics. Consequently, for a more detailed analysis, it becomes pivotal to explore the relationship between the evolution of the NEA and the FA_{MFI}.

Figure 2 Changes in monetary aggregate M3 (in %) and contribution of its counterparts (in p.p.), Czechia, 2013–24



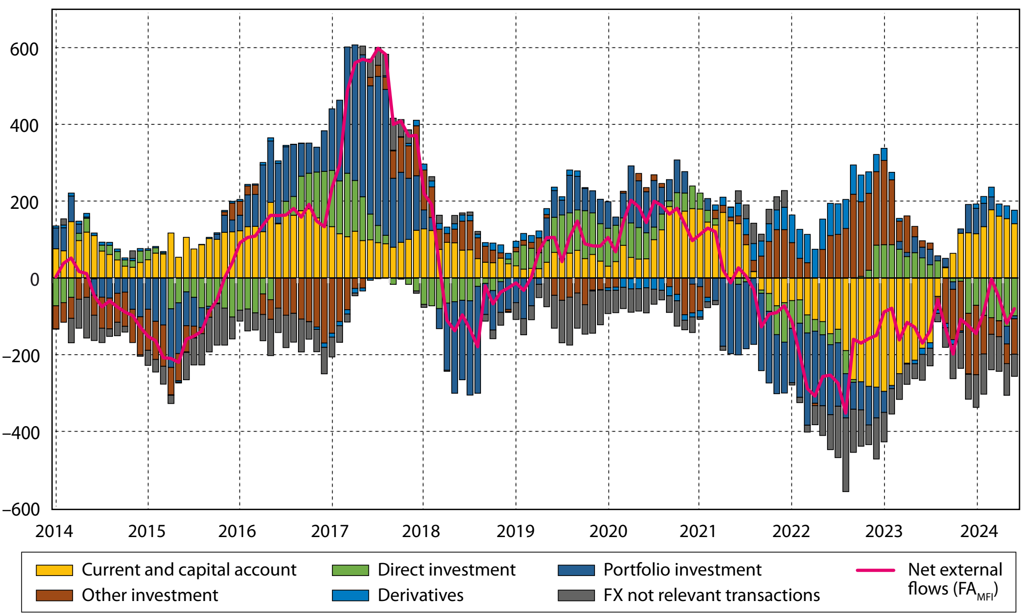
Source: CNB, own calculations

2.1 The evolution of monetary presentation of the BoP

Using the link between the change in NEA and FA_{MFI}, we can now dissect the growth rate of M3 during the currency cap regime in detail. Figure 3 clearly illustrates that within the first half of the intervention regime, capital flew out of the economy as the outlook of depreciated currency dissuaded foreign investors from investing in domestic assets. Conversely, the weakened currency bolstered the current account, which recorded stable surpluses during that period. The trend reversed at the beginning of 2016, as domestic assets became more appealing with the approaching end of the currency cap regime and anticipated currency appreciation. The influx of capital into the domestic economy during the final year of the intervention regime can largely be attributed to portfolio investments,[24] complemented by positive external balances and the inflow of foreign direct investments. This trend, despite the ongoing external balance surpluses, shifted in the latter half of 2017 (i.e. after the intervention regime was discontinued), primarily due to a slowdown in capital inflow.

After years of stable capital inflow, the situation substantially changed in 2021 when money inflow came to a halt.[25] The initial outflow of portfolio investments was swiftly followed by accumulated trade balance deficits, driven by the surge in energy prices. Additionally, exports suffered from supply chain bottlenecks and decreased external demand. The outflow of capital was somewhat mitigated by other investments, where the positive contributions to M3 development can be largely attributed to foreign loans taken by the government, such as the SURE loan. Moreover, the interest rate differential and the appreciated domestic currency made it profitable for firms to secure their financial needs in foreign currency. Simultaneously, the positive contributions of foreign direct investments largely stemmed from inflows of debt capital, as foreign direct investors were trying to exploit the high interest rate differential within their affiliated enterprises, and from reinvested earnings, which are concurrently offset by their counterparts recorded on the current account.[26]

Figure 3 Monetary presentation of the BoP (12m moving sum), CZK bn, Czechia, 2013–24



Note: i) Net inflow of capital into the Czech Republic is denoted with positive values. ii) Correction of FX not relevant transactions includes income from EU purchased by CNB, returns on reserve assets (both otherwise present on the current and capital account), reinvested earnings (moved from direct investment for better clarity), and the item errors and omissions.
Source: CNB, own calculations

[24] The biggest influx of capital was, in fact, due to non-residents opening deposit accounts in the Czech banks. However, the monetary presentation does not fully reflect these operations, which is a weakness that we will address in the following chapters.

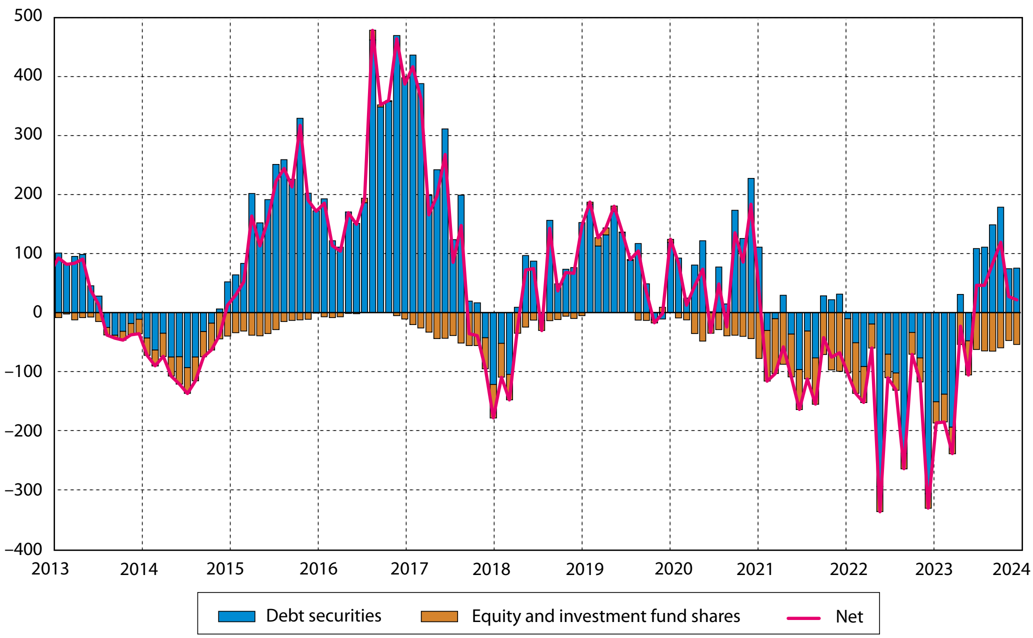
[25] Similarly to the currency cap regime, there was a sharp increase in non-residents' holdings in Czech banks due to expectations of earlier tightening of monetary policy conditions in Czechia compared to the Eurozone and the USA.

[26] Reinvested earnings is an imputed item which does not give rise to any cash movement so it is not FX relevant and correctly consolidated within the LHS of the monetary presentation (see Formula 3).

Analysis above clearly indicates that the flow of money between economies is not dominantly driven by transactions recorded in the non-financial part of the balance of payments (the current and the capital accounts). Especially portfolio investment generally represents an important driving force behind the movements of money, without the Czech economy being an exception. Let´s thus take a closer look at the composition of portfolio investment in terms of instruments. Portfolio investments are generally defined as cross-border transactions (and positions) involving tradable securities (debt and equity) which do not fall within the group of direct investments (BPM6, par. 6.54). Portfolio investments are commonly referred to as "hot capital" seeking a quick profit from interest rate differentials or exchange rate changes. For that reason, they are of particular interest of policy makers as any change in the market sentiment or policy setting may trigger a capital flight, i.e. a situation of an abrupt outflow of money with a pressure on the domestic exchange rate to depreciate.

The following Figure 4 depicts the composition of portfolio investments flows in the Czech Republic. As it appears, the dynamics are predominantly driven by transactions with debt securities while investments in shares remain relatively limited. The chart suggests that the portfolio investment flows stabilized following the period of the FX interventions conducted by the central bank until mid-2017.

Figure 4 Monetary presentation of the BoP (12m moving sum), CZK bn, Czechia, 2013–24



Note: Net inflow of capital into Czechia is denoted with positive values.
Source: CNB, own calculations

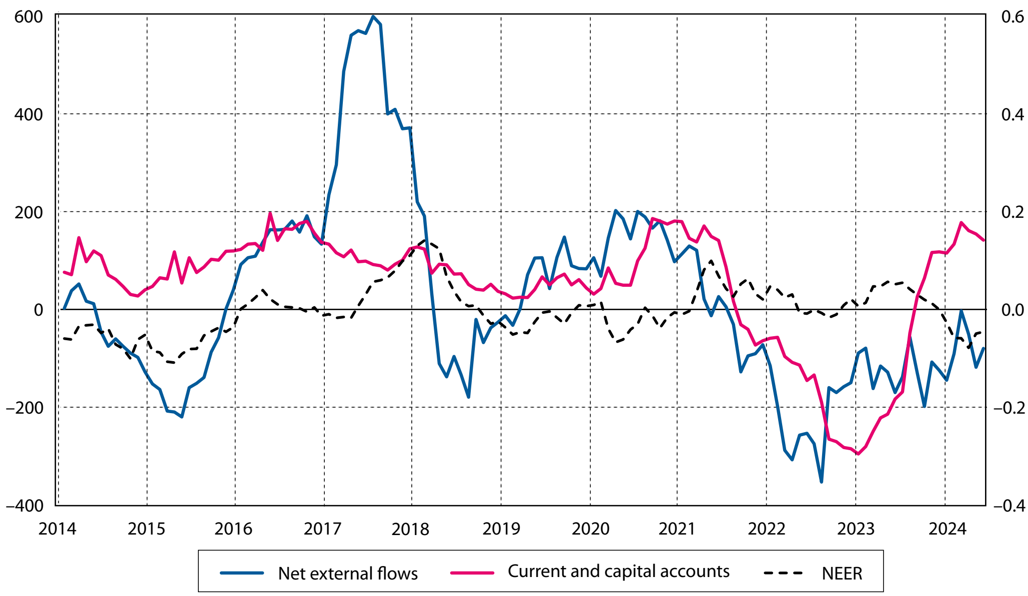
However, speculative capital also flowed into the domestic economy in the form of short-term bank loans or financial derivatives, typically secured by the MFI sector. Since these transactions are directly conducted with the banking sector, affecting both external assets and liabilities of MFIs, they do not contribute to the growth of M3 and are appropriately consolidated on the right-hand side of the monetary presentation (see Formula 3). Nevertheless, these transactions also play a role in determining the exchange rate movements.

2.2 The monetary presentation and the exchange rate

To investigate the driving forces behind the FX changes, we can first compare the net external flows as identified by the monetary presentation of the BoP with fluctuations in the exchange rate. In our analysis, we opt for the nominal effective exchange rate (NEER)[27] to better reflect the currency pressures stemming from the BoP. To reflect the currency pressures accrued over a one-year period we illustrate year-on-year changes, where positive values indicate appreciation and negative values denote depreciation.

The identity between the change in the net external assets and net external flows ensures that the monetary presentation serves as a valuable and reliable tool for explaining M3 changes from an external perspective.[28] However, if we aim to utilize net external flows as an estimate of the excess demand for currency over its supply, certain limitations come into play. Using the currency cap regime during 2013–2017 as an illustrative example, we observe in Figure 5 a substantial influx of capital (demand for koruna) accompanied by minimal exchange rate fluctuations. This mild appreciation is primarily attributable to the exchange rate with the U.S. dollar,[29] as the exchange rate with the euro was capped during that period. Nevertheless, the largest inflow of capital originated from euro area countries, driven by speculation on expected exchange rate appreciation. Furthermore, most transactions on the current account were also conducted in euros, as the euro area is the Czech Republic's largest trading partner.

Figure 5 NEER (y-o-y changes, rhs), net external flows (12m moving sum, CZK bn) and current and capital account (12m moving sum, CZK bn), Czechia, 2013–24



Source: CNB, own calculations

[27] Weights used for NEER calculation by the Czech Statistical Office are based on the share of a trading partner's total turnover to the overall trade turnover of Czechia. However, the objective of the monetary presentation is to simultaneously consider both international trade and capital market, where the relative shares of traded currencies may differ. Additionally, trading partners do not always represent the currency in which the trade is conducted (for instance, purchases of Russian oil and petroleum are predominantly transacted in U.S. dollars). Therefore, we have recalculated the NEER using weights as suggested by the turnover on the FX market, which is published quarterly by the CNB.

[28] The minor discrepancies between both series are merely a result of different statistical approaches.

[29] The weight of U.S. dollar in the constructed NEER index represent about 40% in that period.

The reason for such significant capital influx and simultaneously minimal FX changes lies in the intervention regime itself. For capital flows not to influence the exchange rate, the central bank must have provided additional liquidity to the market by offering to purchase excess euros from financial institutions. However, this is not captured by the monetary presentation, as it represents transactions between two MFI sectors – commercial banks and the central bank – and is therefore consolidated on the right-hand side of Formula (3).

However, official interventions that are explicitly defined by the central bank, such as a currency cap or other forms of commitments where the central bank pledges to intervene in the market as needed, are not the only means by which a central bank can influence the FX market. In fact, any action – or inaction – by the central bank on the FX market can be considered a form of intervention. We identify three distinct ways in which CNB engages in FX market operations.

The CNB operates under a managed float regime, where it may buy or sell foreign currency to mitigate short-term fluctuations, prevent excessive depreciation or appreciation, or achieve specific economic objectives, such as controlling inflation or boosting exports. For example, between 2013 and 2017, the CNB intervened to weaken the exchange rate at the zero lower bound, and in 2022, it intervened to manage high inflation. This type of intervention, where the central bank utilizes its reserve assets as a monetary policy tool to achieve specific short- to mid-term objectives, is considered standard practice under a managed floating exchange rate regime. From a broader perspective, interventions can also be understood as any action where the central bank systematically prevents the domestic currency from moving in a particular direction, such as purchasing the income from European Union. Finally, both inaction and action can be viewed as a form of intervention in the FX market. Over the past years, the CNB has accumulated substantial reserve assets, generating significant revenue. Whether these flows should be converted into domestic currency – following standard market behaviour – or be retained as reserves depends on whether we consider the central bank as a unique market participant with the discretion to either sell or hold its income. Over the last decade, the CNB has primarily accumulated these revenues, with the exception of last year, when it began selling its income on the FX market.

The way we conceptualize interventions is crucial, as it determines how they are treated in the monetary presentation. As previously discussed, the type of intervention deemed standard under a managed floating exchange rate regime involves a transaction between two (MFIs) sectors, which is consequently netted on the rhs of Formula (3). The selling of revenues gained from reserve assets is recorded similarly in the monetary presentation, with one additional entry. Since the sale of these funds originates from actual revenues accrued on reserve assets, these flows are also simultaneously recorded on the current account under primary income. We default to treating these flows as being accumulated into reserves. However, if the central bank opts to sell these revenues, we classify it as an intervention in the FX market.[30] Consequently, revenues on reserve assets must be subtracted from the current account. This is because when these revenues are accumulated, they do not influence the exchange rate. Conversely, when they are sold in the FX market by the central bank, we categorize them as standard interventions and move them to the response side – an EMP index that will be discussed further in this paper.

Finally, income from the European Union, recorded under both the current and capital account, represents a standard economic transaction that would exert an appreciating effect on the exchange rate. Therefore, when the CNB converts this income into domestic currency, it can also be considered a form of intervention. Following the logic outlined earlier, it would be consistent to retain this income on the current and capital account and treat the CNB's purchases as an intervention.

[30] This is consistent with the central bank's own perception of reserve income sales.

However, for the sake of simplicity, we have chosen to subtract these incomes from the lhs of Formula (3) for two reasons. First, unlike income from reserve assets, the CNB itself does not perceive this as an intervention. Second, the purchases of EU income by the CNB have been carried out based on a prior agreement with the government since the Czech Republic joined the EU. This agreement remains unchanged, in contrast to the treatment of income from reserve assets, where the CNB has full discretion. The policy regarding the latter has been altered several times over the past 20 years and more closely resembles standard interventions.

Following the adjustments outlined above, we derive a variable that reflects currency pressures originating from the economy, which the central bank can choose to address. The central bank has the option to intervene in the full amount (keeping the exchange rate fixed), partially intervene (allowing the exchange rate to adjust for the remaining imbalance), or not intervene at all (letting the exchange rate fully accommodate the imbalance). To capture this dynamic, we construct an Exchange Market Pressure (EMP) index.[31] This index measures the excess supply or demand in the FX market, which must be balanced either through adjustments in the exchange rate or through the central bank's interventions. The general form of the EMP index can be expressed as follows:

$$\text{EMP} = \frac{\sigma R}{\sigma R + \sigma e} \cdot \frac{\triangle et}{et - 1} - \frac{\sigma e}{\sigma R + \sigma e} \cdot \frac{\triangle Rt}{Rt - 1}\,, \tag{6}$$

where e_t is the exchange rate, R_t are central bank's foreign currency reserves, σ_e and σ_R are standard deviations of respective series and *t* denotes the time. The interpretation is as follows: Net external flows variable captures pressures arising from imbalances in the FX market. These pressures can be fully or partially mitigated by interventions from the central bank. If the central bank opts not to intervene, these pressures will be entirely absorbed by adjustments in the exchange rate.

We can see from Formula (6) that outflow of capital (EMP index is positive) is reflected through exchange rate depreciation (i.e. an increase in e_t) or through the purchases of the excess liquidity from the market by the central bank (i.e. decrease in reserve assets R_t).[32] Moreover, to ensure that both components contribute equally to the overall index, both the exchange rate and FX interventions are scaled using precision weights.[33]

Figure 6 depicts the EMP index (plotted with the opposite sign for convenience, i.e., positive values signal appreciation pressures), alongside both net external flows and the external balance (current and capital account). The data in the chart reveals that net external flows could potentially serve as a better indicator for currency pressures compared to the external balance alone.

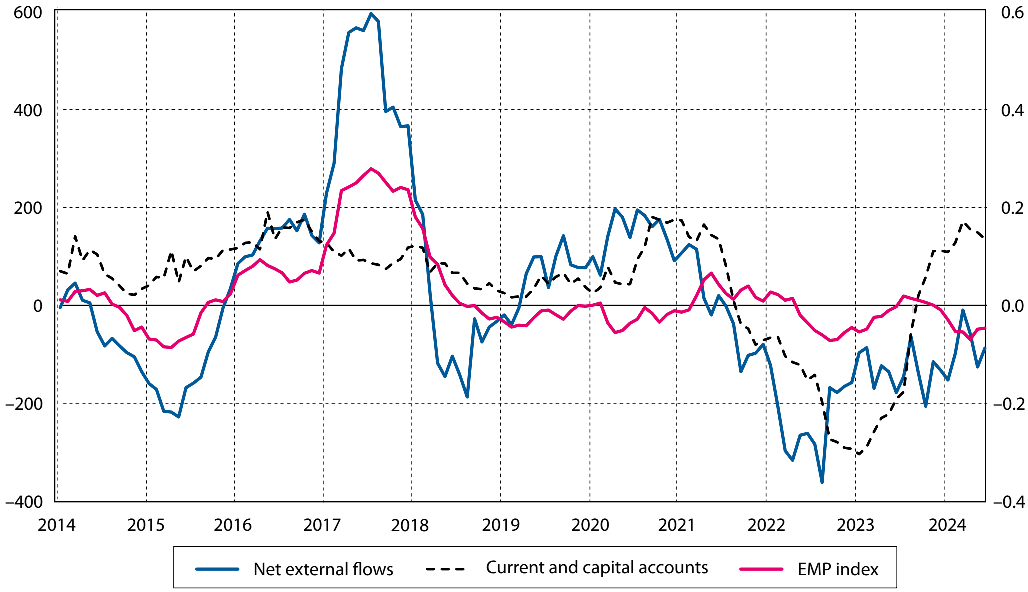
Focusing on the latest development, at first glance, it might be striking that from the beginning of 2021 up to mid-2023, NEER almost constantly appreciated amid large current account deficits, and as our analysis shows, somewhat milder capital inflows. The reason is that to keep the exchange rate from depreciating (and thus to fight high inflation), the central bank intervened in the foreign exchange market and purchased additional liquidity from the market by selling euros. Indeed, EMP index shows depreciation pressures in this period as illustrated in Figure 6.

[31] The EMP index was developed to determine and to analyze a currency crisis in more detail as the usual data on realized capital flows are generally imprecisely measured, incomplete and often available only in quarterly frequency (Goldberg and Krogstrup, 2018).

[32] It is problematic to rely on the difference over time in the market value of reserves because their value does not change only as a result of interventions. Instead it is influenced by revaluation changes generated by price or exchange rate movements, the accumulated return on reserves, foreign currency transactions with the government (incl. purchasing the income from the EU) and other factors.

[33] Precision weights weight the individual variables by the relative shares of their standard deviations to level the different volatilities with which they enter the overall index.

Figure 6 EMP index (y-o-y changes, rhs), net external flows (12m moving sum, CZK bn) and current and capital account (12m moving sum, CZK bn), Czechia, 2013–24



Source: CNB, own calculations

Although the monetary presentation seems to be more instrumental in explaining the FX movements compared to current account itself, there are still several limitations making the link between the net external flows and the FX changes weaker.

First problem relates to the fact that the BoP is based on the sector approach, not on the currency approach. Thus transactions and positions between residents and non-residents can be equally well denominated in foreign currencies instead of in the domestic currency (CZK). However, in most of the cases the domestic currency is demanded in the market regardless of the currency in which the invoices are issued. In instances where exports are denominated in the domestic currency, foreign importers demand domestic currency to meet their obligations for debt repayment. Conversely, when export invoices are denominated in foreign currency, domestic exporters demand the domestic currency in order to cover their domestic expenses (e.g. labour or material costs).[34]

A second issue arises when banks do not serve as intermediaries in payments, such as when non-residents take out loans from domestic banks or open deposit accounts within these banks. Consider a scenario where a non-resident opens a bank account denominated in Czech koruna at a domestic bank. The domestic bank receives their deposits in euros, which leads to an increase in the bank's external assets. However, at the same time, the bank's external liabilities also increase since the owners of these financial assets are non-residents. As a result, NEA remain unchanged, despite an actual inflow of financial capital and the subsequent currency appreciation.

[34] Consider a firm that conducts all its external operations in euros. If the firm purchases materials from abroad for 1 billion CZK and sells its products abroad for 2 billion CZK, then the domestic value added (and net trade balance) equals 1 billion CZK. One could argue that a positive change in NEA therefore indicates appreciation pressures, even though no currency conversion took place. However, the domestic value added consists of domestic inputs (labor and capital) that are paid in CZK. Similarly, residents working abroad may receive their compensation in foreign currency, but since most of their expenses are related to the domestic economy, they will likely convert it into Czech koruna.

Following this event, two possible scenarios may occur. In the first scenario, non-residents decide to use these funds immediately, for example, to purchase government bonds. In this case the external liabilities of the banks decrease (while domestic deposits increase). This results in a rise in NEA, with both events occurring within the same period, leading to currency appreciation accompanied by an increase in NEA.

In the second scenario, if these two events occur with a time lag and are recorded in different periods, they may each provide misleading signals regarding exchange rate pressures. Initially, the NEA remains unchanged despite the currency conversion, and later, the NEA increases without any real currency pressures.

Both of these factors weaken the relationship between net external flows and the resulting exchange rate pressures. This issue tends to be particularly pronounced during turbulent times when specific economic events may trigger sudden and massive outflows (or inflows) of short-term capital that may not be captured by the net external flows variable. One example is the currency cap regime during which NEA registered inflows of hundreds of billions of Czech koruna (Figure 6), while the amount of capital the central bank absorbed into its reserves during that time was more than double that amount. With this in mind, it can be concluded that part of the money inflow was in the form of short-term deposits and hence not captured by the monetary presentation. As a result, the net external flows variable does not serve as a precise indicator of currency pressures, although it may be more instrumental in explaining exchange rate changes than the current account alone.

CONCLUSION

The recent developments in the Czech economy have provided evidence indicating a relatively weaker link between the exchange rate and the external balance, defined as the balance of the current and capital accounts. The weaker explanatory power of the external balance concerning FX changes may be partially addressed by the monetary presentation of the Balance of Payments (BoP), an analytical tool designed to explain contribution of net external assets to the growth rate of the monetary aggregate M3 from the BoP perspective. The monetary presentation serves as a more convenient tool for identifying FX-relevant transactions recorded in the BoP. However, even this improved analytical tool has its limitations. It does not encompass the full scope of FX-relevant transactions, such as transactions within the monetary financial institutions (MFI) sector. To address these limitations, the Exchange Market Pressure (EMP) index emerges as a potentially better indicator of exchange rate movements for measuring currency pressures arising from FX demand and supply imbalances.

ACKNOWLEDGMENT

The authors would like to thank to Martin Kábrt, a former employee of the Czech National Bank, for his valuable and insightful comments on a draft of the paper, which significantly assisted in refining the argument presented in the paper as well as the factual elements.

References

AGUILAR, C. P., SOARES, R. O., ADALID, R. (2020). *Revisiting the monetary presentation of the euro area balance of payments*. ECB Occasional Paper Nr. 238.

BANK FOR INTERNATIONAL SETTLEMENT. (2023). *Triennial Central Bank Survey of foreign exchange and Over-the-Counter (OTC) derivatives markets in 2022* [online]. Basel. <https://www.bis.org/statistics/rpfx22.htm>.

BE DUC, L., MAYERLEN, F., SOLA, P. (2008). *The monetary presentation of the euro area balance of payments*. ECB Occasional Paper Nr. 96.

DRAHOZALOVA, A. (2023). Stačí sledovat peníze? Vnější peněžní toky a měnový kurz [online]. In: *Zpráva o vývoji platební bilance 2022*, Prague: Czech National Bank. <https://www.cnb.cz/export/sites/cnb/cs/menova-politika/.galleries/zpravy-o-vyvoji-platebni-bilance/zprava_o_vyvoji_platebni_bilance_2022.pdf>.

DORNBUSCH, R., FISCHER, S. (1980). Exchange Rates and the Current Account. *American Economic Review*, 70(5): 960–971.
EUROPEAN COMMISSION. (2013). *European System of Accounts (ESA 2010)*. Luxembourg: Publication Office of the European Union. ISBN 978-92-79-31242-7
GOLDBERG, L. S., KROGSTRUP, S. (2018). *International capital flow pressures*. National Bureau of Economic Research.
GOURINCHAS, P., REY, H. (2007). International Financial Adjustments. *Journal of Political Economy*, 115(4): 665–703.
INTERNATIONAL MONETARY FUND. (2013). *Balance of Payments and International Investment Position Manual*. Washington.
KING, M. R., OSLER, C., RIME, D. (2012). Foreign Exchange Market Structure, Players, and Evolution [online]. In: JAMES, J., MARSH, I. W., SARNO, L. (eds.) *Handbook of exchange rates*, John Wiley & Sons, Inc., 1–44. <https://doi.org/10.1002/9781118445785.ch1>.
MANKIW, G. (2020). *Macroeconomics*. 5th Ed. Cengage Learning EMEA. ISBN 147376856X
MILLER, N. (2002). *Balance of payments and exchange rate theories*. Edward Elgar. ISBN 1840649542
UNITED NATIONS. (2008). *System of National Accounts 2008*. European Commission, International Monetary Fund, Organization for Economic Co-Operation and Development, United Nations, World Bank. ISBN 978-92-1-161522-7
WORLD TRADE ORGANIZATION. (2023). *Global Trade Outlook and Statistics* [online]. Geneva. <https://www.wto.org/english/res_e/booksp_e/trade_outlook23_e.pdf>.

The Sustainability Pension Sub-Index in the Context of the Indicators Weights

Jolana Gubalová[1] | *Matej Bel University, Banská Bystrica, Slovakia*
Petra Medveďová | *Matej Bel University, Banská Bystrica, Slovakia*
Jana Špirková | *Matej Bel University, Banská Bystrica, Slovakia*

Received 26.7.2024 (revision 26.9.2024), **Accepted** (reviewed) 4.10.2024, **Published** 14.3.2025

Abstract

Pension systems are sensitive to economic and demographic challenges. The sustainability sub-index is essential for ensuring the stability and reliability of pension systems for both current and future pensioners. It assists policymakers and stakeholders in identifying areas where reforms may be necessary to enhance the sustainability of pension policies. This contribution focuses on determining the weights for individual indicators of the sustainability sub-index using several methods of weight determination in multi-criteria decision-making, specifically subjective methods. The weights are determined using four methods: Saaty's exact and approximate method, Thurstone's method of pair comparison, and Best-Worst method. It also provides sustainability sub-index values for selected European countries included in Mercer's score determination, as well as for Slovakia, which is not yet included in Mercer's evaluations. Thurstone's method of pair comparison appears to be the method most consistent with Mercer's methodology in determining the weights for indicators of the sustainability sub-index.

Keywords	***DOI***	***JEL code***
Sustainability, indicators, weights, Saaty's method, Best-Worst method, Thurstone's method of pair comparison	*https://doi.org/10.54694/stat.2024.40*	*H55, J26, C18*

INTRODUCTION

The Global Pension Index is an annual report that assesses the retirement income systems of different countries. It was started by the Mercer consulting firm in 2009 and, over time, gained participation from various stakeholders, including governments, financial, and academic institutions, Mercer (2009). In 2023, the Mercer CFA Institute Global Pension Index 2023 compares retirement income systems in 47 countries around the world Mercer (2023). The ranking includes countries with advanced economies and less developed countries. It is difficult to estimate based on which key the countries are selected; there are missing some economically advanced countries such as Luxembourg or politically influential countries,

[1] Faculty of Economics, Matej Bel University in Banská Bystrica, Tajovského 10, 975 90 Banská Bystrica, Slovakia. Corresponding author: e-mail: jolana.gubalova@umb.sk, phone: (+421)904821609.

as well as Ukraine and Russia, are missing. The index ranks countries according to several factors to provide information on the effectiveness of their pension systems. It considers variables such as retirement income levels, coverage, and the regulatory framework. The aim is to highlight strengths and weaknesses in pension systems around the world. All data used come from the Pensions at a Glance 2021, OECD (2021), life expectancy 2024 and 2053, age dependency 2053, and total fertility rate 2021-2020 data were from the United Nations World Population Prospects 2022, United Nations (2022). Countries with higher scores are considered to have stronger pension systems. The index helps policy-makers and stakeholders identify areas for improvement in pension policies. It serves as a reference for comparing pension systems across countries. The Global Pension Index is widely recognized and used by governments, organizations, and researchers. Its findings contribute to discussions about retirement planning and policy reforms around the world.

The Global Pension Index evaluates three main factors: the adequacy, sustainability, and integrity of pension systems. Sub-index adequacy assesses the retirement income levels and the extent to which pension systems provide sufficient financial support to retirees. However, the sub-index of integrity evaluates the regulatory framework and governance of pension systems, focussing on transparency, accountability, and protection against fraud or mismanagement. Each index value represents a score between 0 and 100. Each question is scored for each system with a minimum score of 0 and a maximum score of 10, Mercer (2023).

The main reason we offer this paper is that the individual indicators in all three sub-indices are assigned weights based on their importance. Although the authors state that the weights of individual indicators were determined subjectively. We believe that a full description of the determination of weights is certainly beyond the scope of their publications. Therefore, from the position of examining the importance of individual indicators, we have decided to offer the determination of weights specifically for the sustainability sub-index. We decided to start with this sub-index because it has nine indicators. Experts in determining scales assume that 9 indicators represent a relatively high number when a person can recognize the importance of the mentioned indicators. Of course, in the future, we will also have to deal with a larger number of them.

The sustainability sub-index focusses on the future and uses various indicators that will affect the likelihood that the current system will be able to provide benefits in the future. The sub-index includes the economic importance of the private pension system, the level of its financing, the length of expected retirement now and in the future, the rate of participation of the elderly population in the labour market, the current level of public pension expenditure and government debt, and the level of real economic growth. The most important indicators that influence this sub-index are the coverage of private pension plans, demographic factors, and the level of pension assets as a share of GDP.

The paper is organized as follows. Section 1 – *Preliminaries* recalls 9 indicators that make up the sub-index of sustainability, Mercer (2023). They are used in the form of questions as presented by Mercer. In addition, they are supplemented with additional information so that their mutual importance and comparability are clearly evident to the reader. Section 2 – *Methods and Methodologies* introduces our own approach to determining the weights of individual indicators using the Saaty's exact and approximate method, the Thurstone's method of pair comparison, and finally the Best-Worst method. Section 3 – *Results* gives sustainability index scores for selected European countries that are included in Mercer's research and also for Slovakia, which is not yet included in the mentioned publication. Section 4 – *Discussion* selects the method for determining the weights of the individual indicators that most closely aligns with Mercer's methodology. Last section – *Conclusion* talks about our next plans and the possibility of a solution procedure in examining and determining the scores of individual indicators using new methods that use fuzzy numbers.

1 PRELIMINARIES

The sustainability sub-index assesses the long-term viability of pension systems. It takes into account factors such as demographic trends, financial stability, and the ability to meet future pension obligations. Sustainability evaluates whether pension systems can adapt to changes in the demographic composition of the population, such as ageing populations and declining birth rates. It also examines the balance between contributions and expenditures to ensure the sustainable operation of pension systems.

1.1 Indicators of the sustainability pension sub-index

In Mercer's original text Mercer (2023), individual indicators are written in the form of questions. We quote them as follows.

Question S1: What proportion of the working age population are members of retirement savings plans? Question S1 assesses the proportion of the working population that is a member of a retirement savings plan. This measures the degree of participation of the working population in retirement savings schemes, which may indicate the level of preparedness of the population for future pension needs.

Question S2: What is the level of pension assets, expressed as a percentage of GDP, held in private pension arrangements, public pension reserve funds, protected book reserves, and pension insurance contracts? Question S2 assesses the level of pension assets, expressed as a percentage of GDP, that are held in private pension schemes, public reserve funds, protected book reserves, and pension insurance contracts. In this way, the total value of pension assets is measured compared to the economic performance of the country. A higher proportion may indicate a stronger and more stable pension system with sufficient funds for future pensions.

Question S3:
a. What is the life expectancy at the current state pension age?
b. What is the projected life expectancy at the expected state pension age in 2053 (that is, in 30 years' time)? This calculation allows for an improvement in mortality.
c. What is the projected old-age dependency ratio in 2053?
d. What is the estimated total fertility rate (TFR) for 2021–2025?

To a. The first question aims to determine how long people live on average when they start receiving state pension benefits. This helps to understand the duration for which pensions might need to be paid out.

To b. By projecting life expectancy in 2053, the question seeks to estimate how long people will live in the future, allowing policymakers to anticipate the financial needs and sustainability of pension systems.

To c. The projected old-age dependency ratio in 2053 is intended to assess the balance between the working-age population and retirees, helping to gauge the strain on pension systems and the economy.

To d. The question about TFR for 2021–2025 aims to understand the current trend in childbirth rates, which influences the future size of the workforce and, consequently, the financial stability of pension systems. The previous four sub-questions were grouped into one question to comprehensively assess various aspects of the pension system's sustainability and demographics. By covering life expectancy, dependency ratio, and fertility rate in a single question, policy-makers can understand the interconnected factors influencing the future of pension systems. This consolidated approach provides a holistic view of the challenges and opportunities facing pension schemes, allowing for more informed decision-making.

Question S4: What is the level of mandatory contributions that are set aside for future retirement benefits (that is, funded), expressed as a percentage of the annual wage for a full-time median income earner?

This may include mandatory employer and/or employee contributions paid into funded public benefits (that is, social security) and/or retirement benefits from the private sector. Question S4 assesses the level of mandatory contributions for future pensions (i.e., funded), expressed as a percentage of the annual salary for a full-time worker with a middle income. This may include mandatory employer and/or employee contributions paid into funded public pension benefits (i.e., social security) and/or private sector pension benefits.

Question S5: What is the labour force participation rate for people 55 to 64 years of age? What is the labour force participation rate for people 65 years or older?

Question S5 determines the labour force participation rate for a person aged 55 to 64 years and also for people 65 years and older. This assesses the proportion of people in specific age groups that are still active on the labour market. Higher labour force participation rates may indicate higher work participation of older individuals, which may have an impact on a country's economy and social stability. This assessment helps gauge the extent to which older people contribute to the economy. Higher participation rates among older age groups may suggest increased economic productivity and social stability, highlighting the potential impact on a country's labour market dynamics and overall well-being. The older generation (65+) can contribute to improving the economy by continuing to work, volunteering, mentoring, investing wisely, and participating in community activities. Here, two scenarios can occur. If the older generation continues to work, this can potentially increase the rate of participation in the labour force, leading to higher economic productivity and GDP growth. It can also ease pressure on pension schemes by delaying the payment of pension benefits. However, it can also limit employment opportunities for younger generations, which can lead to intergenerational competition in the labour market. However, if the older generation cares for their grandchildren, this can allow parents to work or take advantage of educational opportunities, potentially increasing labour force participation and economic performance. It can also strengthen family bonds and reduce parental childcare costs. However, it may limit the direct economic contribution of the older generation and potentially increase the financial burden on the pension and healthcare systems if they retire earlier. Ultimately, the optimal scenario depends on balancing the economic benefits of increased participation in the labour force with the social benefits of intergenerational support and family bonds.

Question S6: What is the level of adjusted government debt (being the gross public debt reduced by the size of any sovereign wealth funds that are not set aside for future pension liabilities), expressed as a percentage of GDP? What is the level of public expenditure on pensions expressed as a percentage of GDP, averaged over the latest available figure and the projected figure for 2050?

Question S7: In respect of private pension arrangements, are older employees able to access part of their retirement savings or pension and continue working (for example, part time)? If so, can employees continue to contribute and accrue benefits at an appropriate rate?

Question S8: What is the real economic growth rate averaged over seven years (namely, the last four years and projected for the next three years)?

Question S9: Is it a requirement for the pension plan's trustees/fiduciaries to consider environmental, social and governance (ESG) issues in developing their investment policies or strategies? If not a requirement, is it encouraged by the relevant pension regulator?

2 METHODS AND METHODOLOGIES

Although the company Mercer mentions the weights of individual indicators in its annual reports, it does not mention where or by what methods these weights were determined.

We have decided to look at the importance of individual indicators from the perspective of several methods for determining weights in multi-criteria decision-making, which belong to the group of so-called subjective methods. In our research, we used the Saaty's exact and approximate method, or the Thurstone's method of pair comparison, and the so-called Best-Worst method. Saaty emphasizes that the human mind is capable of meaningfully comparing approximately 7 ± 2 indicators, Saaty (1977). That is why we also chose the sustainability sub-index, which includes 9 indicators.

Table 1 Share of positive answers to job search questions and item-response probabilities

Scale	Numerical evaluation alternative *i* from *j*	Reciprocal alternative *i* from *j*
Extremely preferred	9	1/9
Between very strong and extremely	8	1/8
Very strongly preferred	7	1/7
Between strong and very strong	6	1/6
Very strongly preferred	5	1/5
Between moderate and strong	4	1/4
Moderately preferred	3	1/3
Between equal and moderate	2	1/2
Equal importance	1	1

Source: Authors' work based on Saaty (2005)

The first step is to create a square table that has as many rows and columns as we have indicators. We write the values from 1 to 9 in the table so that we express the relative importance of the row indicator compared to the column one. To make the data consistent, it is obvious that we express the relative "unimportance" as the inverse value. To verify the validity of Table 1, it is necessary to calculate the so-called consistency index first using the formula:

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

where: λ_{max} is the largest positive eigenvalue of the matrix, n is the number of the indicators.

We can calculate the Consistency ratio CR as follows:

$$CR = \frac{CI}{RI}, \tag{2}$$

where: RI is the random index, which can be found e.g. in Mu et al. (2017).

For the table to be valid, with respect to Mu et al. (2017) the value *CR* must not exceed 0.10. The individual weights in Table 2 can be calculated using the solver function in MS Office Excel, where the Saaty optimization criterion (3) is minimized:

$$\min \sum_{i=1}^{n}\sum_{j=1}^{n}\left(s_{i,j}-\frac{v_i}{v_j}\right)^2, \tag{3}$$

with conditions:

$$v_1, v_2, \ldots, v_n > 0 \wedge \sum_{i=1}^{n} v_i = 1, \tag{4}$$

where $s_{i,j}$ are individual matrix elements.

In this part, we introduce specific data on the basis of which we determined the weights of individual indicators. Since none of the methods for determining the weights is a priori superior and none can be preferred, it will finally use the weights of the indicators as the average value of the weights determined by the selected methods.

2.1 Exact Saaty's method

To determine the relative importance of each indicator with maximum precision, we began by ranking them based on our subjective yet professional and thorough research. We consider indicator S1 to be the most important, as it reflects the level of participation of the working population in private pension schemes. In our view, private savings will play a critical role in providing greater financial security during retirement. This is closely linked to indicator S2, which represents a form of national wealth. In third place, we ranked indicator S8, which reflects economic growth over a given period and forecasts future growth. Next in importance is indicator S6, which pertains to government debt and public expenditures, as well as their projected future trends. Following this, we placed indicator S4 in fifth position, as it highlights the amount of mandatory contributions set aside for retirement. While life expectancy S3 is undoubtedly significant in determining pension payments, we anticipate that its rate of increase will not be as extreme moving forward. Hence, we placed S3 in sixth place. In our opinion, it is closely related to indicator S5, which measures the labour force participation of older individuals. Finally, we ranked indicator S9 last, as we believe that while ESG issues may currently be a popular topic of discussion, they do not yet have a tangible impact on the sustainability of pension systems.

The importance of each indicator is ranked as shown in Table 2. To determine the weights with the greatest accuracy, we utilized the full range of values from 1 to 9. Since the indicators are listed from most to least important, we can easily apply Saaty's relative importance scale, as presented in Table 1.

In our case, Saaty's optimization function (3) acquires the value 37.3645. The maximum positive eigenvalue of the matrix is λmax = 9.4015042, Brunner (2008), and the consistency index according to (1) is *CI* = 0.050188, and the consistency ratio (2) with the random index 1.45 is on the level of *CR* = 0.0346. This means that we can consider our data to be consistent and the respective weights to be relevant.

Table 2 Exact Saaty's method

i, j	S_1	S_2	S_8	S_6	S_4	S_3	S_5	S_7	S_9	v_i
S_1	1	2	3	4	5	6	7	8	9	0.25
S_2	$\frac{1}{2}$	1	2	3	4	5	6	7	8	0.22
S_8	$\frac{1}{3}$	$\frac{1}{2}$	1	2	3	4	5	6	7	0.18
S_6	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{2}$	1	2	3	4	5	6	0.13
S_4	$\frac{1}{5}$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{2}$	1	2	3	4	5	0.08
S_3	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{2}$	1	2	3	4	0.05
S_5	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{2}$	1	2	3	0.04
S_7	$\frac{1}{8}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{2}$	1	2	0.03
S_9	$\frac{1}{9}$	$\frac{1}{8}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{2}$	1	0.02
v_i	0.25	0.22	0.18	0.13	0.08	0.05	0.04	0.03	0.02	1.00

Source: Authors' work

2.2 Approximate Saaty's method

If we do not have the solver function available, we can also use the proximate Saaty's method. An approximate determination of the weights was derived from the logarithmic least-squares method, and the resulting values are not very different from the weights obtained from more precise procedures Boďa et al. (2021). In the approximate Saaty's procedure, the average multiple importance is determined for each criterion using the geometric mean.

Table 3 Approximate Saaty's method

i, j	S_1	S_2	S_8	S_6	S_4	S_3	S_5	S_7	S_9	p_i	v_i
S_1	1	2	3	4	5	6	7	8	9	4.1472	0.31
S_2	$\frac{1}{2}$	1	2	3	4	5	6	7	8	3.0008	0.22
S_8	$\frac{1}{3}$	$\frac{1}{2}$	1	2	3	4	5	6	7	2.1131	0.16
S_6	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{2}$	1	2	3	4	5	6	1.4592	0.11
S_4	$\frac{1}{5}$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{2}$	1	2	3	4	5	1.0000	0.07
S_3	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{2}$	1	2	3	4	0.6853	0.05
S_5	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{2}$	1	2	3	0.4732	0.04
S_7	$\frac{1}{8}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{2}$	1	2	0.3324	0.02
S_9	$\frac{1}{9}$	$\frac{1}{8}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{2}$	1	0.2411	0.02
Sum										13.4596	1.00

Source: Authors' work

Values of p_i are determined by formula:

$$p_i = \sqrt[n]{\prod_{j=1}^{n} s_{i,j}} \,. \tag{5}$$

From the pi values, the normalized weights are usually determined by the formula:

$$v_i = \frac{p_i}{\sum_{i=1}^{n} p_i} \,. \tag{6}$$

2.3 Thurstone's method of pair comparison

The Thurstone's method of pair comparison or pairwise comparison method was first introduced by Thurstone (1927). In the method, it is assumed that it is possible to evaluate the mutual significance of indicators in pairs. The decision-maker always compares pairs of indicators and decides which is more significant. He then writes his choice of couple preferences into the table using the following three options:

1. assigns the value $s_{i,j} = 1$ if the *i*-th is more significant, i.e., the line indicator.
2. assigns the value $s_{i,j} = 0$ if the *j*-th is more significant, that is, the column indicator.
3. assigns the value $s_{i,j} = 0.5$ if the indicator of rows and columns is equally significant Boďa et al. (2021).

Table 4 Thurstone's method of pair comparison

i, j	S_1	S_2	S_8	S_6	S_4	S_3	S_5	S_7	S_9	w_i	w_i^*	v_i
S_1		1	1	1	1	1	1	1	1	8	9	0.20
S_2	0		1	1	1	1	1	1	1	7	8	0.18
S_8	0	0		1	1	1	1	1	1	6	7	0.16
S_6	0	0	0		1	1	1	1	1	5	6	0.13
S_4	0	0	0	0		1	1	1	1	4	5	0.11
S_3	0	0	0	0	0		1	1	1	3	4	0.09
S_5	0	0	0	0	0	0		1	1	2	3	0.07
S_7	0	0	0	0	0	0	0		1	1	2	0.04
S_9	0	0	0	0	0	0	0	0		0	1	0.02
Sum											45.00	1.00

Source: Authors' work

After obtaining the comparison matrix, the number of pairs preferences assigned is added by individual lines, and thus the number of preferences $w_i = \sum_i s_{i,j}$ is obtained for the indicator *i-th*. Subsequently, we will determine the weight of the *i*-th indicator using the formula:

$$v_i = \frac{w_i}{\sum_{i=1}^{n} w_i}. \tag{7}$$

If some value of w_i is equal to 0, as in our case, we can add, e.g. to each value of w_i value 1. Then we can calculate the nonzero normalized weights of individual indicators replacing w_i by w_i^*

2.4 Best-Worst method

The Best-Worst method is, similarly to the previous methods, based on the gradual comparison of pairs of indicators. It was developed by Rezaei (2016a) and is believed to be capable of providing reliable scales in less time. In addition, it turns out that within a few years this method has be- come, thanks to its properties, very popular in multi-criteria decision-making, Brunelli et al. (2019), Mi et al. (2019).

The method is implemented in several steps:

1. Determine a set of decision criteria. In our case we have 9 indicators of the sustainability sub-index.
2. Determine the best and worst criteria. If more than one criterion is considered to be the best or the worst, one can be arbitrarily chosen, Rezaei (2015). Among 9 indicators, the most significant – Best and the least significant – Worst are identified. The best is $S_{1,}$ and the worst is S_9.
3. The preference for the best indicator will be gradually expressed in comparison with other indicators using the cardinal scale, which was also used in the previous methods. Again, a value of 1 represents agreement in importance. In this way, we get the vector $A_B = (s_{B1}, ...s_{Bn})$ of the most significant indicator to the others. In our case, the vector $A_B = (1, 2, 3, 4, 5, 6, 7, 8, 9)$, see Table 5.
4. The preference of the worst indicator will gradually be expressed in comparison with other indicators. This is how we get the vector $A_W = (s_{1W}, ...s_{nW})$ of the least significant indicator for the others. This vector is $A_W = (9, 8, 7, 6, 5, 4, 3, 2, 1)$, and you can see it written in Table 5.

Table 5 Best-Worst method

Names of criteria	**S_1**	**S_2**	**S_8**	**S_6**	**S_4**	**S_3**	**S_5**	**S_7**	**S_9**
Select the best	S_1								
Select the worst	S_9								
Best to others A_B	1	2	3	4	5	6	7	8	9
Others to the worst	A_W								
S_1	9								
S_2	8								
S_8	7								
S_6	6								
S_4	5								
S_3	4								
S_5	3								
S_7	2								
S_9	1								
Optimal weights v_i^*	0.31	0.19	0.06	0.08	0.05	0.10	0.05	0.13	0.03

Source: Authors' work

5. Find the optimal weights $\left(v_1^*, v_2^*, \ldots, v_n^*\right)$.
The optimal weights are determined using an optimization mini-max model:

$$\min \max \left\{ \left| \frac{v_B}{v_i} - s_{Bi} \right|, \left| \frac{v_i}{v_W} - s_{iW} \right| \right\}, \tag{8}$$

with conditions $\sum_{i=1}^{n} v_i = 1$ and $v_1, v_2, \ldots, v_n > 0$..

This model is converted to the following model:

$$\min \xi , \tag{9}$$

under conditions:

$$\left| \frac{v_B}{v_i} - s_{Bi} \right| \le \xi , \tag{10}$$

$$\left| \frac{v_i}{v_W} - s_{iW} \right| \le \xi , \tag{11}$$

and $\sum_{i=1}^{n} v_i = 1$ and $v_1, v_2, \ldots, v_n > 0$. Solving this model, the optimal weights $\left(v_1^*, v_2^*, \ldots, v_n^*\right)$ are obtained.

6. The last step is to calculate the level of consistency using a robust index called consistency ratio *CR* which is given by:

$$CR = \frac{\xi^*}{CI}, \tag{12}$$

where: *CI* is consistency index which is for $n = 9$ indicators on the level of 5.23, Rezaei (2015).

Using solver Rezaei (2016b) we obtained optimal value $\xi^* = 0.06839945$, hence $CR = 0.0131$. The consistency ratio is a number from the interval [0, 1], and the smaller it is, the more reliable the results.

Let us compare the consistency ratio determined using the Saaty's method and the Best-Worst method. Based on Saaty's method, $CR = 0.0346$, and using the Best-Worst method, this value is $CR = 0.0131$. Also, based on our calculations, the statement Rezaei (2015) that the Best-Worst method leads to a more consistent comparison that gives more reliable results is confirmed.

Based on our personal experience, we characterize individual methods as indicated in Table 6.

Table 6 Features, advantages and disadvantages of selected methods

	Exact Saaty's method	Approximate Saaty's method	Thurstone's method of pair comparison	Best-Worst method
Complexity	✓	✓	✓	✓
Time consumption	–	✓	✓	–
Subjectivity	✓	✓	✓	✓
Interpretation problems	–	–	–	–
Coherence	–	–	✓	✓

Source: Authors' work

3 RESULTS

We would like to emphasize that the weights of individual indicators were determined based on our professional expertise in the area of pension system sustainability. Additionally, since Slovakia was not included in the list of countries for which Mercer calculates the total pension index, we independently determined Slovakia's individual score. We utilized the available sources and methodologies from Mercer (2023) in our calculations. The remaining scores were sourced directly from Mercer (2023). Our objective was to derive the weights of sustainability indicators using selected methodologies and to compare them with the Mercer approach. We can confirm that our assessment of the importance of sustainability indicators closely aligns with Mercer's findings.

Table 7 presents the European countries for which Mercer annually publishes a sustainability sub-index, now supplemented by Slovakia's sub-index, which has not yet been included among the monitored countries. The countries are ranked in descending order according to their sustainability sub-index values. For better clarity, Figure 1 provides a graphical representation of the individual sub-index values.

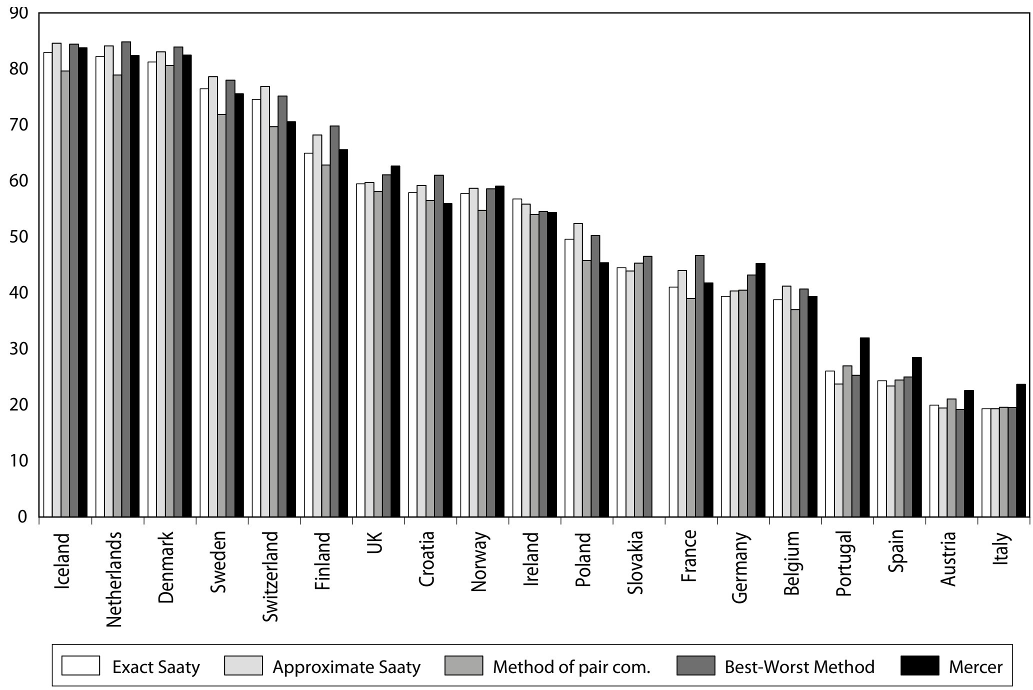
According to our assessment, Slovakia ranks 11th out of the 19 monitored countries. Notably, some developed Western European countries, such as France, Germany, and Belgium, ranked behind Slovakia. Among the V4 countries, Poland's placement in 10th position was also a surprising result.

Table 7 The sustainability pension sub-index of the selected methods, maximal score 100

Country	w.r.to Exact Saaty's method	w.r.to Approximate Saaty's method	w.r.to Thurstone`s method of pair comparison	w.r.to Best-Worst method	w.r.to Mercer (2023) weights	Average
Iceland	82.97	84.63	79.67	84.46	83.80	82.93
Netherlands	82.25	84.13	78.95	84.85	82.40	82.55
Denmark	81.26	83.1	80.62	83.94	82.50	82.23
Sweden	76.49	78.66	71.86	77.99	75.60	76.25
Switzerland	74.59	76.90	69.72	75.16	70.60	74.09
Finland	64.98	68.23	62.87	69.82	65.60	66.48
UK	59.51	59.74	58.10	61.14	62.70	59.62
Croatia	57.96	59.20	56.53	61.04	56.00	58.68
Norway	57.78	58.69	54.78	58.62	59.10	57.47
Ireland	56.82	55.86	54.03	54.58	54.40	55.32
Poland	49.62	52.46	45.79	50.30	45.40	49.54
Slovakia	44.52	43.93	45.32	46.55	–	45.08
France	41.07	44.03	39.04	46.73	41.80	42.72
Germany	39.40	40.36	40.51	43.22	45.30	40.87
Belgium	38.83	41.21	37.03	40.73	39.40	39.45
Portugal	26.07	23.73	26.99	25.32	32.00	25.53
Spain	24.32	23.42	24.46	25.00	28.50	24.30
Austria	19.96	19.48	21.07	19.21	22.60	19.93
Italy	19.33	19.34	19.61	19.54	23.70	19.46

Source: Authors' work using basic score data from Mercer (2023)

Figure 1 The sustainability pension sub-index in selected European countries



Source: Authors' work

4 DISCUSSION

Using the four selected methods, we obtained sustainability sub-index values that were very similar to those determined by Mercer, as shown in Table 7. Our next step is to select the method for determining the weights of the individual indicators that most closely aligns with Mercer's methodology. To do this, we will apply two criteria: the relative or absolute differences and the ranking of the sustainability sub-index values determined by the selected methods compared to those determined by Mercer.

For the first criterion, we calculated the absolute differences between the sustainability sub-index values using the data in Table 7. As illustrated in Table 8, based on the sum of both absolute and relative differences, the Exact Saaty's method produced the smallest sum of differences. According to this criterion of relative differences, the Exact Saaty's method aligns most closely with Mercer's methodology, followed by Thurstone's pairwise comparison method, the Best-Worst method, and the Approximate Saaty's method.

Applying the second criterion, we first established the rankings of the countries across all methods, as presented in Table 9. Table 10 then shows the differences in rankings compared to those determined by Mercer. This table also highlights the number of countries whose rankings differ in each method, revealing that Thurstone's pairwise comparison method has the fewest discrepancies. Based on this criterion of ranking differences, Thurstone's method of pairwise comparison aligns most closely with Mercer's methodology, followed by the Best-Worst method, the Exact Saaty's method, and the Approximate Saaty's method.

Using Thurstone's method, only a third of the 18 monitored countries experienced changes in their rankings – 3 countries improved their positions (Denmark, Croatia, Austria), while 3 others (Iceland, Norway, Italy) saw their rankings decline.

Table 8 Difference between the sustainability pension sub-index (method – Mercer)

Country	Exact Saaty's method – Mercer	Approximate Saaty's method – Mercer	Thurstone`s method – Mercer	Best-Worst method – Mercer	Average – Mercer
Iceland	0.83	0.83	4.13	0.66	0.87
Denmark	1.24	0.60	1.88	1.44	0.27
Netherlands	0.15	1.73	3.45	2.45	0.15
Sweden	0.89	3.06	3.74	2.39	0.65
Switzerland	3.99	6.30	0.88	4.56	3.49
Finland	0.62	2.63	2.73	4.22	0.88
UK	3.19	2.96	4.60	1.56	3.08
Norway	1.32	0.41	4.32	0.48	1.63
Croatia	1.96	3.20	0.53	5.04	2.68
Ireland	2.42	1.46	0.37	0.18	0.92
Poland	4.22	7.06	0.39	4.90	4.14
Germany	5.90	4.94	4.79	2.08	4.43
France	0.73	2.23	2.76	4.93	0.92
Belgium	0.57	1.81	2.37	1.33	0.05
Portugal	5.93	8.27	5.01	6.68	6.47
Spain	4.18	5.08	4.04	3.50	4.20
Italy	4.37	4.36	4.09	4.16	4.24
Austria	2.64	3.12	1.53	3.39	2.67
Sum of absolute differences	45.15	60.05	51.61	53.95	41.74
Sum of relative differences	0.74	0.98	0.84	0.88	0.68

Source: Authors' work using basic score data from Mercer (2023)

Table 9 Ranking of countries according to the sustainability pension sub-index values

Country	Exact Saaty's method	Approximate Saaty's method	Thurstone`s method	Best-Worst method	Mercer	Average
Iceland	1	1	2	2	1	1
Denmark	3	3	1	3	2	3
Netherlands	2	2	3	1	3	2
Sweden	4	4	4	4	4	4
Switzerland	5	5	5	5	5	5
Finland	6	6	6	6	6	6
UK	7	7	7	7	7	7
Norway	9	9	9	9	8	9
Croatia	8	8	8	8	9	8
Ireland	10	10	10	10	10	10
Poland	11	11	11	11	11	11
Germany	13	14	12	13	12	13
France	12	12	13	12	13	12
Belgium	14	13	14	14	14	14
Portugal	15	15	15	15	15	15
Spain	16	16	16	16	16	16
Italy	18	18	18	17	17	18
Austria	17	17	17	18	18	17

Source: Authors' work

Under the criterion of ranking changes, approximately half of the countries saw an improvement in their rankings. For all methods, each country's ranking shifted by only one position relative to Mercer's ranking, with two exceptions: the Netherlands improved by two places in the Best-Worst method, and Germany's ranking deteriorated by two places in the Approximate Saaty's method.

Table 10 Difference between places in the ranking (Method – Mercer)

Country	Exact Saaty's method – Mercer	Approximate Saaty's method – Mercer	Thurstone`s method – Mercer	Best-Worst method – Mercer	Average – Mercer
Iceland	0	0	1	1	0
Denmark	1	1	1	1	1
Netherlands	1	1	0	2	1
Sweden	0	0	0	0	0
Switzerland	0	0	0	0	0
Finland	0	0	0	0	0
UK	0	0	0	0	0
Norway	1	1	1	1	1
Croatia	1	1	1	1	1
Ireland	0	0	0	0	0
Poland	0	0	0	0	0
Germany	1	2	0	1	1
France	1	1	0	1	1
Belgium	0	1	0	0	0
Portugal	0	0	0	0	0
Spain	0	0	0	0	0
Italy	1	1	1	0	1
Austria	1	1	1	0	1
Number of countries with different positions	8	9	6	7	8

Source: Authors' work

Table 11 summarizes the evaluation of the most suitable method when both criteria are applied simultaneously. In the table, the following designations are used for the individual methods: a rating of 1 indicates the most similar method, while a rating of 4 signifies the least similar.

The method that aligns most closely with Mercer's methodology is Thurstone's pairwise comparison method, followed by the Exact Saaty's method, the Best-Worst method, and lastly, the Approximate Saaty's method. According to both Thurstone's and the Exact Saaty's methods, Slovakia ranks 12th among the selected European countries, while the Best-Worst and Approximate Saaty's methods place Slovakia in 13th position.

Table 11 Summary of method evaluation

Criterion	Exact Saaty's method	Approximate Saaty's method	Thurstone`s method	Best-Worst method
Number of countries with different positions in the compared rankings	3	4	1	2
Sum of relative differences	1	4	2	3
Sum	4	8	3	5
Ranking	2	4	1	3

Source: Authors' work

CONCLUSION

In our contribution, we provided both psychological and economic perspectives on the sustainability of pension systems. Our work was motivated by Mercer, which does not include our home country, Slovakia, among the evaluated countries. Consequently, we decided to establish the sustainability sub-index value for Slovakia as well. Additionally, we presented our calculations for the weights of individual sustainability indicators using four different methods.

It is important to note that we should not view these individual weights as precisely determined figures; rather, they invite discussion and potential adjustments. We can confirm that our weight determinations are relatively consistent with those established by Mercer.

To select the method that aligns most closely with Mercer's methodology, we applied two criteria: the comparison of relative or absolute differences and the ranking of sustainability sub-index values derived from our chosen methods against those determined by Mercer. When using both criteria simultaneously, the method most similar to Mercer's approach is Thurstone's pairwise comparison, followed by the Exact Saaty's method, the Best-Worst method, and lastly, the Approximate Saaty's method.

According to both Thurstone's method and the Exact Saaty's method, Slovakia ranks 12th among the 19 selected European countries. In contrast, the Best-Worst and Approximate Saaty's methods place Slovakia in 13th position.

Looking ahead, we plan to establish weights not only for the sustainability sub-index but also for the remaining two sub-indices – Adequacy and Integrity. Our next challenge will be to determine these weights using new procedures, specifically the Fuzzy Best-Worst Multi-Criteria Decision-Making Method as described by Guo et al. (2021), and Dong (2021), along with the insights from Rezaei (2020). Our overarching ambition is to collaborate with Mercer.

ACKNOWLEDGMENT

The work was supported by the Slovak Scientific Grant Agency VEGA No. 1/0124/24, Slovakia in the Context of the Pension Index and Population Aging Index – A Future Perspective.

We would like to express our gratitude to the anonymous reviewers for their very valuable comments and advice that significantly improved our ideas.

References

BOĎA, M., ÚRADNÍČEK, V. (2021). *Stochastic analysis of uncertainty in significance weights in multi-aspect evaluation.* Belianum, Matej Bel University in Banská Bystrica.

BRUNELLI, M., REZAEI, J. (2018). A multiplicative best-worst method for multi- criteria decision making [online]. *Operations Research Letters*, 47(1): 12–15. <https://doi.org/10.1016/j.orl.2018.11.008>.

BRUNNER, A. (2017). *Calculate eigenvalues and eigenvectors* [online]. <https://www.arndt- bruenner.de/mathe/scripts/engl eigenwert2.htm>.

DONG, J., WAN, S., CHEN, S. M. (2021). Fuzzy best-worst method based on triangular fuzzy numbers for multi-criteria decision-making [online]. *Information Sciences*, 547: 1080–1104. <https://doi.org/10.1016/j.ins.2020.09.014>.

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>.

GUO, S., ZHAO, H. (2017). Fuzzy best-worst multi-criteria decision-making method and its applications [online]. *Knowledge-Based Systems*, 121: 23–31. <https://doi.org/10.1016/j.knosys.2017.01.010>.

MERCER. (2009). *Melbourne Mercer Global Pension Index* [online]. Melbourne Centre for Financial Studies. [cit. 20.6.2024]. <https://parlinfo.aph.gov.au/parlInfo/search/display/display.w3p;query= Id:%22library/lcatalog/00460314%22;src1=sm1>.

MERCER. (2023). *Mercer CFA Institute Global Pension Index* [online]. CFA Institute, Monash University, Monash Business School. [cit. 20.6.2024]. <https://www.mercer.com/globalpensionindex>.

MI, X., TANG, M., LIAO, H., SHEN, W., BENJAMIN, L. (2019). The state-of-the-art survey on integrations and applications of the best-worst method in decision making: Why, what, what for and what's next? [online]. *Omega*, 87: 205–225. <https://doi.org/10.1016/j.omega.2019.01.009>.

MU, E., PEREYRA-ROJAS, M. (2017). *Practical Decision Making: an Introduction to the Analytic Hierarchy Process (AHP) Using Super Decisions V2*. Springer International Publishing, Cham.

OECD. (2021). *Pensions at a Glance 2021: OECD and G20 Indicators* [online]. Paris: OECD Publishing. <https://doi.org/10.1787/ca401ebd-en>.

REZAEI, J. (2015). Best-worst multi-criteria decision-making [online]. *Omega*, 53: 49–57. <https://doi.org/10.1016/j.omega.2014.11.009>.

REZAEI, J. (2016). Best-worst multi-criteria decision-making method: Some properties and a linear model [online]. *Omega*, 64: 126–130. <https://doi.org/10.1016/j.omega.2015.12.001>.

REZAEI, J. (2020). A Concentration Ratio for Nonlinear Best Worst Method [online]. *International Journal of Information Technology & Decision Making*, 19(3): 891–907. <https://doi.org/10.1142/S0219622020500170>.

REZAEI, J. (2024). *BWM-Solver-4.xlsx* [online]. [cit. 3.6.2024]. <https://bestworstmethod.com/software/>.

SAATY, T. L. (1977). A scaling method for priorities in hierarchical structures [online]. *Journal of Mathematical Psychology*, 15(3): 234–281. <https://doi.org/10.1016/0022-2496(77)90033-5>.

SAATY, T. L. (2009). *Theory and applications of the analytic network process: decision making with benefits, opportunities, costs, and risks*. RWS publications.

THURSTONE, L. L. (1927). A law of Comparative Judgment. *Psychology Review*, 34: 273–286.

UNITED NATIONS. (2022). *World Population Prospects 2022* [online]. Department of Economic and Social Affairs Population Division. [cit. 3.7.2024] <https://population.un.org/wpp/>.

WANG, J. J., JING, Y. Y., ZHANG, CH. F. (2009). Weighting methodologies in multicriteria evaluations of combined heat and power systems [online]. *International Journal of Energy Research*, 33: 1023–1039. <https://doi.org/10.1002/er.1527>.

Structural Modeling of Health Services' Quality Level as a Determinant of User Satisfaction of the Tertiary Health Care

Aldin Brajić[1] | *International University of Travnik, Travnik, Bosnia and Herzegovina*
Saliha Brajić[2] | *NGO EconVita, Zenica, Bosnia and Herzegovina*

Received 14.5.2024 (revision 6.9.2024), **Accepted** (reviewed) 24.10.2024, **Published** 14.3.2025

Abstract

Tertiary level of health care should provide highly specialized health services, that include the most complex methods and procedures of diagnosis, treatment and rehabilitation. The aim of the research is to examine the impact of the quality of health services on user satisfaction of medical services at the tertiary level of health care. For the purposes of this research, clinical centers of the tertiary health care level operating in the territory of Bosnia and Herzegovina were selected, which also represents the basic set of research. The survey was conducted on a sample of 1 022 users of health services provided by clinical centers in the territory of Bosnia and Herzegovina, where the cities represented the strata in the research: Sarajevo, Banja Luka, Tuzla, Mostar and Foča. The results indicate a strong influence of independent constructs on dependent constructs, that is, the quality of health services has a strong influence on the level of user satisfaction with (non)medical services.

Keywords	*DOI*	*JEL code*
Quality of health services, user satisfaction, tertiary level of health care, structural equation modeling	*https://doi.org/10.54694/stat.2024.26*	*C38, C83, I11, M31*

INTRODUCTION

Service quality and customer satisfaction are often used as indicators of competitiveness. However, their mutual relationship is relatively unclear. Namely, in some studies these two concepts were used as synonyms (Zeithaml, Berry and Parasuraman, 1993), while in other studies a distinction was made

[1] International University of Travnik, Aleja konzula, Meljanac bb, 72270, Travnik, Bosnia and Herzegovina. E-mail: brajicaldin@gmail.com, phone: (+387)66991393. ORCID: <https://orcid.org/0000-0002-3838-7373>.
[2] NGO EconVita, Zenica, Bosnia and Herzegovina. E-mail: saliha.cabro@yahoo.com.

between these concepts (Cronin and Taylor, 1992). Although today the dominant concepts of satisfaction (disconfirmation paradigm) and service quality (gap models) start from two different paradigms, both use expectations and perceptions as key determinants in their explanation, which is the reason why two comparative groups of satisfaction researchers have developed in the literature users and service quality researchers.

Regardless of the obvious overlaps of the previously mentioned concepts, some important differences were also established between them. Very often, the subject of controversy is the question of what precedes what, user satisfaction with service quality or quality precedes satisfaction? According to a growing number of authors, these two concepts are largely related. Of course, they differ from each other in the duration of the experience with the service, in the levels of expectations, the degree of affectivity and depending on stability of duration of the relationship between the user and service provider (Snoj, Savić and Rajtmajer, 1999). According to the integral approach, advocated by Klaus (Snoj, 1995), service quality could be understood as the value of the service for the user. It is thought to become more congruent with longer-term attitudes (Stafford, Stafford and Wells, 1998). On the other hand, satisfaction is considered more of an emotional reaction to the experience with a product or service, quite similar to the individual emotional state of mood deterioration that forms the basis of the level of regret. Cronin and Taylor (Oliver, 1993) conducted an empirical test of the reciprocity of satisfaction and quality in several service industries. Their research showed the fact that service quality can be seen as one of the determinants of user satisfaction, and satisfaction itself affects the user's future purchasing decisions. Oliver points out in his studies that satisfaction is the result of the user's overall reaction, and that it can have a potential impact on future perceptions of quality (Oliver, 1993). On the other hand, some authors in their research point out that satisfaction precedes the perceived quality of services (Bolton and Drew, 1991). However, regardless of the divided attitudes in the earlier period, today the prevailing opinion is that quality precedes satisfaction (Mikulić, 2007). Accordingly, this paper has the task of examining whether the quality of health services has an impact on the achieved level of satisfaction of users of tertiary level health care, both medical and non-medical services.

1 LITERATURE REVIEW

Among the first research efforts on the quality of services at the tertiary level of health care, the research of Alaloola and Albedaiwi (2008) who surveyed a total of 1983 users on the case study of the King Abdulaziz Medical Center in Riyadh stands out. A significant degree of satisfaction was expressed regarding the comfort of the hospital rooms (88.5%), the temperature of the rooms (78.1%), the emergency call of the staff (87.9%), the cleanliness of the rooms (79.6%) and respect from the staff (87.4%). On the other hand, dissatisfaction was expressed regarding the clarification of procedures (57.2%) and presentation by the doctor (59.1%).

A comparative study of satisfaction with the quality of health services at the tertiary level of health care by 383 users and 162 nurses, using the SERVQUAL survey questionnaire, was carried out by Nashrath, Akkadechanunt and Chontawan (2011) as a case study of a clinical hospital in the Maldives. The uniqueness of this research is reflected both in the consistency of the application of the quality dimensions according to SERVQUAL, and in the examination of the gap analysis between the two categories. The results of the research are also extremely interesting, since they show a certain overlap in the perceptions of users and nurses regarding the poorly rated quality dimensions, especially regarding the "identification" dimension. The authors believe that feedback from nurses can be an excellent input for creating a quality management system.

Lee and Kim (2017) attempted a comparative analysis of health service quality measurements in a clinical hospital in Seoul, on two sample groups: current users and their family members, and former users and the general public. It is a study on two different samples of different sizes – the first sample of 365 and the second sample of 232 participants. The survey is based on the HEALTHQUAL questionnaire with dimensions: empathy, safety, tangibility, efficiency and improvement of health services. Testing was performed through t-test, ANOVA and confirmatory factor analysis (SEM). The results of the research showed that there is no statistically significant difference between the presented samples, and it is recommended to act simultaneously on all five dimensions of the quality of health services.

In their research on the evaluation of the quality of health services at the tertiary level of health care in India, Natarajappa and a group of authors (2020) distinguished 13 dimensions of the quality of health services, namely: 1) reception, 2) social responsibility, 3) staff behavior, 4) service quality and service availability, 5) confidence, 6) continuity, 7) communication, 8) environment, 9) treatment costs, 10) customer loyalty, 11) hospital discharge, 12) medical services, and 13) overall services. The presented dimensions are mainly derived by the authors from the SERVQUAL questionnaire, creating their own model of quality management at the tertiary level of health care. On a sample of 30 users (patients), the authors come to the conclusion that the highest level of correlation of the dimension "user loyalty" was achieved with the dimensions "overall service" and "discharge from the hospital", although a positive correlation is also noticeable in other dimensions.

In a sample of 410 tertiary level health care users in Bangladesh, Dilshad et al. (2020) confirms the existence of a growing concern on the part of users regarding the level of quality of health services of public clinical centers. The overall level of satisfaction with all services, technical equipment and interpersonal relations was at an extremely low level.

We can see a somewhat different approach to examining the satisfaction of users of health services in a study of a tertiary clinical hospital in Karachi (Pakistan). The researchers, on their own created a questionnaire, examined the satisfaction of a total of 173 users, excluding users of maternity, psychiatry and chemotherapy departments. What is extremely interesting in this research is the fact that satisfaction was measured by the frequency of problems that occurred in certain departments. The less often the problems happened, the more satisfied the users were and vice versa (Imam et al. 2007).

Garg et al. (2014) conducted a user satisfaction survey of a tertiary specialist hospital in India. The research was carried out by surveying users who were hospitalized during the two months of monitoring. For the survey questionnaire, they used the previously created Canadian questionnaire for measuring satisfaction with medical services (The Northwest Territories Hospital satisfaction questionnaire). More than 88% of users rated the services as excellent and good. The areas in which dissatisfaction was noted were the cleanliness, especially in the toilets, and the quality of the food served to the users. Also, the research results point to the need to develop the soft-skills of the medical staff, in order to better understand the users.

In South Korea, a more extensive study was conducted in 29 outpatient clinics of the University Clinical Center in Seoul with a sample of 1 194 users. The authors made a significant effort in creating measurement scales based on the theoretical framework of Donabedian and the National User Satisfaction Index questionnaire. The research makes a clear distinction between health service quality and satisfaction. The shortcoming of the research is that the quality of the health service was developed in 5 constructs (doctor's examination, services of nurses/technicians, technical services, amenities and physical surroundings), while satisfaction was measured exclusively by one question: "Are you satisfied with the overall health service of this institution (Ham, et al. 2015)?"

Kulkarni (2018) developed his own scale for measuring the satisfaction of users of the tertiary level of health care, inspired by the HCAHPS survey questionnaire. The scale was tested on a sample of 100 randomly selected users of a tertiary clinical hospital in Maharashtra (India). The results of the descriptive analysis showed that the users were satisfied with the availability of services, the professionalism of care, the waiting time for an examination, the behavior of consultants, medical and other staff. Overall satisfaction level of 73% was excellent or good and 22% average. Dissatisfaction was expressed mostly with regard to toilets and drinking water.

Another study on the satisfaction of users of tertiary care in Pakistan was carried out by Maroof et al. (2019). Like their predecessors, this team created its own satisfaction measurement scale. The questionnaire consisted of 38 questions, and the statements used to measure satisfaction were formulated with a negative sign. Using a Likert scale from 1 – completely disagree, to 5 – completely agree. A total of 110 users gave their ratings. The results showed that there is significant user dissatisfaction and that their needs are not being met at an adequate level.

A very interesting survey was conducted on a sample of as many as 136 hospitals of the tertiary level of health care in China, which showed that the users were mostly satisfied with the level of services provided. Through statistical analysis, 12 variables were singled out: 1) Medical skill of the doctor, 2) Inquiry about the medical history and current situation of the user, 3) Convenience of using the elevator in the hospital, 4) Feeling of respect from the medical staff, 5) Timely instructions from the staff, 6) Explanation of treatment and medication, 7) Waiting time before consultation, 8) Waiting time for medical examination, 9) Privacy protection, 10) Waiting time for bill payment, 11) Bathroom cleanliness, and 12) Drinking water supply in waiting areas. The identified variables were then allocated into four categories: 1) Waiting time, 2) Service and treatment, 3) Costs and 4) Environment. The research showed that Chinese users were most satisfied with the category "Service and treatment", and somewhat less with "Waiting time", "Costs" and "Environment", for which a greater degree of investment by the management of clinical centers is recommended (Hu, 2019.). By the way, this is one of the rare ones among hospital researches with the author's own measurement scales.

By using Gaps Model of Service Quality and the SERVQUAL instrument Ozretić et al. (2020) performed an analysis of deviations in the perception and expectations of users of university clinics regarding the quality of health services. Although the data was collected for each of the 18 departments, it was established that there was a significant deviation of the variables at the level of the entire clinical center. The largest gaps were indentified "responsiveness" and "tangibility".

Ojeniweh et al. (2021) investigate the satisfaction of 141 users of tertiary level health care in Nigeria. The highest degree of satisfaction was achieved with communication, technical equipment and interpersonal aspects of the health service, contrary to the previously mentioned research. On the other hand, dissatisfaction was expressed in terms of costs and waiting time for service provision.

2 RESEARCH METHODOLOGY

Within the elaboration of certain theoretical and methodological origins of the observed problem, and certain applied considerations, the work used: hypothetical-deductive method, method of induction and deduction, method of analysis and synthesis and statistical methods (descriptive and inferential statistics) with a systematic approach to research.

2.1 Sample and data collection

Data collection was carried out through primary research among University-Clinical Centers in Bosnia and Herzegovina. In the process of empirical research, a survey questionnaire was used as a method for data collection. Content validity is ensured by the use of validated measuring instruments, and by consulting a group of experts when formulating, translating and adjusting the measurement scales. The questionnaire used to measure the various variables in the predicted models consists of a series of questions. The questions were chosen based on a systematic review of the literature, the subject of which was the quality of health services and user satisfaction.

The survey questionnaire also contains a group of questions related to the sociodemographic characteristics of the respondents. The questions were mostly taken over and adapted from earlier validated empirical research.

Content and nomological validity was conducted during and after operationalization of measurement models. All indicators are thoroughly checked for wording, specificity and sentence length to ensure relevance to the context in which the research is conducted. A panel of experts from the academic community in Bosnia and Herzegovina checked the content validity and relevance of the measuring instrument. They received the questionnaires via e-mail, and were able to fill out the questionnaire, as well as send written comments to the indicators of the measured constructs. Minor changes, i.e. rewording, were made to several indicators based on comments received from panel experts.

The invitation to participate in the research was distributed by registered mail to the addresses of all clinical centers, as well as by e-mail. The survey was conducted using a combination of field research and online via Google Forms. Field data collection was carried out by expert and trained persons who had the necessary information in case of ambiguities of respondents. The survey questionnaires that were submitted online contained additional information for each question, and contact information was provided in case of additional questions and ambiguities from respondents. Also, in order to avoid missing data in Google survey questionnaires, the questions were arranged as mandatory, that is, they could not be skipped. Data collection was carried out in the period from July 2021 to February 2022. For the purposes of the research, clinical centers of the tertiary level of health care operating in the territory of Bosnia and Herzegovina were selected, which also represents the basic set of research. Data collection was carried out on the basis of a stratified sample, since it belongs to the category of random samples and allows to evaluate the degree of reliability of drawing conclusions about the investigated parameters. The survey was conducted on a sample of 1 022 users of health services of clinical centers in Bosnia and Herzegovina with an extremely high return rate (70%), with the cities representing the strata in the survey: Sarajevo, Banja Luka, Tuzla, Mostar and Foča. In Bosnia and Herzegovina there are the following tertiary health care institutions: University Clinical Center Sarajevo; University Clinical Center of Republika Srpska Banja Luka; University Clinical Center Tuzla; University Clinical Hospital Mostar and University Hospital Foča. Individual filling in of questionnaires took 7 minutes. All statements of the constructs were measured with a Likert scale ranging from 1 – "completely disagree" to 5 – "completely agree". The second group of questions related to the sociodemographic characteristics of the respondents.

2.2 Structural model and hypotheses

The basic structural model consists of the independent variable "quality of health services" and the dependent variable "user satisfaction", which are connected to the following hypotheses:

H1: There is a statistically significant influence of the quality of health services on the level of user satisfaction with medical services in Bosnia and Herzegovina.

- H1a: Tangibility has a statistically significant effect on the level of user satisfaction with medical services.
- H1b: Reliability has a statistically significant effect on the level of user satisfaction with medical services.
- H1c: Response has a statistically significant effect on the level of user satisfaction with medical services.
- H1d: Safety has a statistically significant effect on the level of user satisfaction with medical services.
- H1e: Empathy has a statistically significant effect on the level of user satisfaction with medical services.

H2: There is a statistically significant influence of the quality of health services on the level of user satisfaction with non-medical services in Bosnia and Herzegovina.

- H2a: Tangibility has a statistically significant effect on the level of user satisfaction with non-medical services.

H2b: Reliability has a statistically significant effect on the level of user satisfaction with non-medical services.
H2c: Response has a statistically significant effect on the level of user satisfaction with non-medical services.
H2d: Safety has a statistically significant effect on the level of user satisfaction with non-medical services.
H2e: Empathy has a statistically significant effect on the level of user satisfaction with non-medical services.

Quality of health services (QHS) is a second-order construct consisting of a total of 23 statements, and it consists of five first-order constructs, of which five statements measure the construct tangibility, six statements measure the construct reliability, three statements measure the construct response, four statements measure the construct safety and five statements measure the construct empathy.

Table 1 Instruments for measuring the quality of health services

Dim.	Subdimension	Code	Indicators (assertions)
QHS	Tangibility	Tangibility_1	The health institution has modern equipment.
		Tangibility_2	The exterior and interior of the healthcare facility is visually acceptable.
		Tangibility_3	The employees of the health care facility look neat.
		Tangibility_4	The accessories and devices of the healthcare facility are clean.
		Tangibility_5	The health institution has equipment and facilities in accordance with the services it provides.
	Reliability	Reliability_1	In the health facility, examinations, treatments and treatment services are quick and precise.
		Reliability_2	User review schedule is on time.
		Reliability_3	The service procedure is performed correctly on the first attempt.
		Reliability_4	Ease of contacting hospital staff.
		Reliability_5	The health institution insists on providing a health service without errors.
		Reliability_6	The employees of the health care facility have the knowledge to respond to the user's inquiry.
	Response	Response_1	Employees of the health care facility warn when the user needs help.
		Response_2	User complaints are resolved successfully and promptly.
		Response_3	Employees of the health care facility provide clear and understandable information.
	Safety	Safety _1	Sufficient attention is paid to the user.
		Safety_2	Employees of the health facility are available when needed by the user.
		Safety_3	The employees of the healthcare facility are capable of analyzing the user's illness.
		Safety_4	The medical staff accurately and precisely treats the user's ailments.
	Empathy	Empathy_1	Employees show extreme patience in dealing with users.
		Empathy_2	The employees are friendly and hospitable.
		Empathy_3	Users can easily file complaints.
		Empathy_4	Moral support is provided to users.
		Empathy_5	Services are provided to all users regardless of social status.

Source: Own construction

User satisfaction (US) is a second-order construct consisting of a total of 30 statements, and it consists of two first-order constructs: user satisfaction with medical services, which is measured by 15 statements, and user satisfaction with non-medical services, which is measured by 15 statements.

Table 2 Instruments for measuring user satisfaction

Dim.	Subdimension	Code	Indicators (assertions)
US	User satisfaction with medical services	USMS1	I am satisfied with the reception upon arrival at the health facility.
		USMS2	I am satisfied with the presentation of doctors and nurses/technicians in the health institution.
		USMS3	I am satisfied with the clarity of the information provided about upcoming procedures and interventions by doctors and nurses/technicians.
		USMS4	I am satisfied with the length of the conversations that the doctors and nurses/technicians spent with me.
		USMS5	I am satisfied with the professional approach of the doctors and nurses/technicians in the health facility.
		USMS6	I am satisfied with the doctor's ability to diagnose the health problem.
		USMS7	I am satisfied with the expediency (quickly provided services and no waiting).
		USMS8	I am satisfied with the explanation for the delay in the ordered examination.
		USMS9	I am satisfied with the success of the treatment.
		USMS10	Satisfied with the treatment process.
		USMS11	The doctors and nurses/techs did everything possible to ease my pain.
		USMS12	Before I receive the medicine, the doctors and nurses/technicians explain what it is for and the possible side effects.
		USMS13	The doctors and nurses/technicians have provided written information about the symptoms I have or recommendations that I must follow after I leave the healthcare facility.
		USMS14	After leaving the health care facility, I understand my health condition and the procedures I am responsible for implementing for the benefit of my health.
		USMS15	Every experience I had with a healthcare facility has met my expectations in terms of medical services.
	User satisfaction with non-medical services	USNMS1	I am satisfied with the resolution of the complaint.
		USNMS2	I am satisfied with the hospital environment.
		USNMS3	I am satisfied with the accommodation services.
		USNMS4	I am satisfied with the food services.
		USNMS5	I am satisfied with the prices of health services.
		USNMS6	I am satisfied with the low level of corruption in the institution where I stayed.
		USNMS7	I am satisfied with the clarity of information provided about upcoming procedures by non-medical staff.
		USNMS8	I am satisfied with the application of information technologies (e-cards, e-orders, e-prescriptions, etc.) in the institution where I stayed.
		USNMS9	I am familiar with the rights arising from compulsory health insurance, and they refer to the right to use health care and the right to certain financial benefits and assistance.
		USNMS10	I am familiar with the rights arising from extended health insurance.
		USNMS11	I am familiar with the rights arising from voluntary health insurance for myself and my family.
		USNMS12	I am satisfied with the method of financing health institutions and health services.
		USNMS13	I am satisfied with the payment of medicines, surcharges, co-payments, additional payments, etc.
		USNMS14	I am satisfied with the friendliness of the staff.
		USNMS15	Every experience had with a healthcare facility has met my expectations, from the aspect of non-medical services.

Source: Own construction

For structural equation modeling analysis we used IBM SPSS AMOS 21.0. The steps for structural equation modeling (SEM) analysis are: (1) Descriptive statistical analysis, (2) Data testing and verification, (3) Assessment of model fit, (4) Model reliability testing, (5) Model validity testing, and (6) Assessment of structural relationships / hypothesis testing. All the mentioned methodological steps of testing are carried out on the sample itself, since the shape of the structural model that is ultimately tested depends on the character of the sample. Therefore, each individual step will be explained in chapter 3 Results.

Finally, the greatest contribution of this research is the simultaneous examination of user satisfaction with health services for all five functioning health institutions of the tertiary level of health care: University Clinical Center Sarajevo, University Clinical Center of Republika Srpska Banja Luka, University Clinical Center Tuzla, University Clinical Hospital Mostar and University Hospital Foča. In this way, a higher level of research objectivity was ensured and the making of generalizing conclusions at the level of Bosnia and Herzegovina, making this research endeavor unique in these areas.

3 RESULTS

3.1 Descriptive statistical analysis

Based on the descriptive analysis, we concluded that among the respondents, women dominated, which make up 58.6% of the sample, while men make up the remaining 41.4%. Looking at the age structure, the largest number of respondents are between the ages of 21 and 29, 253 of them, or 24.8%, followed by the age group between 30 and 39, relatively 23.5%, that is, 240 in absolute terms. Furthermore, we have 191 respondents between the ages of 40 and 49, i.e. 18.7%, while the least number of respondents are over 60, 97 of them, i.e. 9.5%. The majority of respondents have completed secondary vocational education (45%) and belong to the category of workers (37.2%).

The arithmetic mean of all indicators for the independent construct *"quality of healthcare services"* ranges from 2.80 to 3.78 (on a scale of 1 to 5, 1 – completely disagree, 5 – completely agree). The mode and median are 3, and the standard deviations range between 1.138 and 1.398. The arithmetic mean of all indicators of the dependent construct *"user satisfaction"* ranges from 2.64 to 3.51. The mode and median are 3, and the standard deviations range between 1.209 and 1.407.

3.2 Data testing and verification

Multivariate analysis was performed using the *Mahalanobis Distance test*, the calculation of which enables the identification of outliers through an approximate test of statistical significance. After converting the *Mahalanobis Distance test* to its probabilities, we eliminated all observations that are less than or equal to 0.001 ($p \leq 0.001$). The total number of outliers among users of health services was 136, leaving 886 respondents in the sample.

Based on the results of the *Kolmogorov-Smirnov* and *Shapiro-Wilk tests,* $p < 0.001$ we concluded that the null hypothesis is rejected, that is, that the assumption of normal distribution is not satisfied. However, consulting the literature resulted in the conclusion that data deviations from the assumption of normality do not represent a problem, if the analysis uses the Maximum Likelihood (ML) estimation method. Namely, Nwabueze et al. (2009) confirmed in their study that the ML method is robust in five different distributions. A similar conclusion was reached by Fuller and Hemmerle (1966), confirming the robustness of the ML technique in six distributions, including the one in which the assumption of normality of the data was violated. In order to test the data for the presence of homoscedasticity, the Breusch-Pagan test was applied, where the null hypothesis assumes the presence of homoscedasticity. Based on the obtained results, we concluded that the statistical significance of the χ^2 test is < 0.001. Therefore, the results indicate a violation of the assumption of homoscedasticity.

The analysis of the Scatter plot concluded that the data are linearly distributed, thereby fulfilling the assumption of data linearity. Potential multicollinearity was examined using the tolerance index

(TOL) and the variance inflation factor (VIF). For all variables, TOL values are greater than 0.1, and all VIF values are less than 10. Previously, it led to the conclusion that no variable causes the problem of multicollinearity in the research.

3.3 Assessment of model fit

In order to examine the suitability of the model, a confirmatory factor analysis of the measured models of the quality of health services (QHS) and user satisfaction (US) was carried out. However, the obtained results did not show satisfactory suitability of the model. Accordingly, the model was respecified. After analyzing the modification index, it was determined that some statements have too high a correlation, that is, they measure almost the same concept. By reviewing the standardized factor loadings, it was observed that all the values of the manifest variables are above the recommended value of 0.5, and therefore we kept all the variables of the observed models. Based on the above, the CFA analysis was repeated, and based on the obtained results, we concluded that the models achieved a good level of suitability. Namely, the GOF indicators are above/below the recommended limit values.

Table 3 Evaluation of the suitability of the measurement model (GOF), the quality of health services (QHS) and the satisfaction of users of health services (US)

Measures	Threshold value	QHS construct	US construct
P-value	> 0.05	0.001	0.001
Minimum Discrepancy Function by degrees of freedom divided (CMIN/df)	< 5	4.522	4.815
Root mean square error of approximation (RMSEA)	< 0.08	0.063	0.066
Standardized root mean squared residual (SRMR)	< 0.09	0.0248	0.0345
Comparative Fit Index (CFI)	> 0.90	0.967	0.953
Normed Fit Index (NFI)	> 0.90	0.957	0.942
Relative Fit Index (RFI)	> 0.90	0.95	0.934
Incremental Fit Index (IFI)	> 0.90	0.967	0.953
Tucker Lewis – non-normed fit index (TLI – NNFI)	> 0.90	0.961	0.947
Goodness of Fit Index (GFI)	> 0.90	0.902	0.869
Adjusted Goodness of Fit Index (AGFI)	> 0.80	0.874	0.843
Parsimonious Normed Fit Index (PNFI)	> 0.50	0.814	0.838
Parsimonious Comparative Fit Index (PCFI)	> 0.50	0.821	0.848

Source: Own construction

Based on the data from the Table 3, we can conclude that the absolute, parsimonious and incremental indicators are acceptable and therefore suitable for the QHS construct. Chi square coefficient (CMIN) is significant and is 972 274 with 215 degrees of freedom (df), while CMIN/df = 4 522. Considering the sample of 886 observations and 23 manifest variables, it is expected that the chi square test will be significant. Absolute suitability indicators include the value of the square root of the standard error of assessment (RMSEA), which should be less than 0.08, which is 0.063, as well as the standardized root mean square (SRMR), whose value should be less than 0.09. and is 0.0248, which additionally confirms the suitability of the measurement model. Parsimony indicators such as AGFI = 0.874 and PNFI = 0.814 are above the recommended values and also indicate a good fit of the model. The values of the incremental indicators – the standardized fit index NFI = 0.957 and the comparative fit index CFI = 0.967 are above the limit of 0.9, and indicate a satisfactory fit of the model.

We can also conclude that absolute, parsimonious and incremental indicators are acceptable and therefore suitable for the US construct. Chi square coefficient (CMIN) is significant and is 1 863 353 with 387 degrees of freedom (df), while CMIN/df = 4 815. Considering the sample of 886 observations and 30 manifest variables, it is expected that the chi square test will be significant. Absolute suitability indicators include the value of the square root of the standard error of assessment (RMSEA), which should be less than 0.08, which is 0.066, as well as the standardized root mean square (SRMR), whose value should be less than 0.09. and is 0.0345, which additionally confirms the suitability of the measurement model. Parsimony indicators such as AGFI = 0.843 and PNFI = 0.838 are above the recommended values and also indicate a good fit of the model. The values of the incremental indicators – the standardized fit index NFI = 0.942 and the comparative fit index CFI = 0.953 are above the limit of 0.9 and indicate a satisfactory fit of the model.

While evaluating the suitability of the model, it is always necessary to display the chi-square value with the associated p-value. However, the sensitivity of the chi-square test to the sample size is high, given that its value increases along with the sample size. Higher chi-square values are associated with lower p-values, which indicate a statistically significant result, that is, poor model fit. This may encourage the use of smaller samples, which may be counterproductive, as they may only ostensibly mask poorer fit indicators and provide less precise parameter estimates. Also, when the variables do not satisfy multivariate normality and/or when the sample is small, then the test size is unlikely to follow a chi-square distribution. With such models, depending on the method and degree of deviation from normality, wrong conclusions are more common, so the chi-square can be overestimated and thus show the model worse than it is in reality, or else the chi-square can be underestimated and show the model better than it is. In addition, the chi-square increases with the increase in the number of manifest (observed) variables (indicators), i.e. it also depends on the complexity of the model. Therefore, more complex models with more parameters will show a smaller chi-square value, due to the reduction in the number of degrees of freedom.

Therefore, the acceptable value of this indicator can be the result of adding free parameters to the model, and not exclusively the result of a correctly specified model. Due to the limitations of the chi-square test with its associated p-value, it is recommended to use additional indicators of model suitability during the analysis, instead of relying on only one indicator. It is useful to combine the so-called goodness-of-fit and badness-of-fit measures, where higher values in the former indicate better suitability, and worse in the latter. According to Sharif and Sharif (2018) for the model to be eligible it is necessary to show:

- the results of the chi-square test, even if the p-value is significant;
- at least three incremental suitability indicators, the values of which are greater than 0.90 (CFI, NFI, TLI, etc.);
- RMSEA although the values are not good, and the acceptable value of RMSEA is < 0.08;
- SRMR whose value should be less than 0.08.

3.4 Model reliability testing

To test the reliability of measuring models, the quality of health services and the satisfaction of users of health services, the following were used: Cronbach alpha coefficient, Standardized Cronbach alpha coefficient, compozite reliability (CR) and maximal reliability [MaxR(H)]. Based on the obtained results, we established that all indicators in the models consistently represent the corresponding latent construct or factor, that is, that the measurement models are reliable.

Table 4 CFA results for the measurement model of health service quality

Dim.	Subdimension	Code	Factor loads	Cronbach alpha	Standardized Cronbach alpha	CR	MaxR(H)
QHS	Tangibility	Tangibility_1	0.791	0.911	0.912	0.901	0.907
		Tangibility_2	0.776				
		Tangibility_3	0.761				
		Tangibility_4	0.803				
		Tangibility_5	0.879				
	Reliability	Reliability_1	0.829	0.940	0.940	0.936	0.938
		Reliability_2	0.815				
		Reliability_3	0.858				
		Reliability_4	0.834				
		Reliability_5	0.886				
		Reliability_6	0.836				
	Response	Response_1	0.864	0.900	0.900	0.901	0.901
		Response_2	0.854				
		Response_3	0.882				
	Safety	Safety _1	0.898	0.934	0.934	0.930	0.933
		Safety_2	0.901				
		Safety_3	0.839				
		Safety_4	0.867				
	Empathy	Empathy_1	0.897	0.939	0.939	0.940	0.945
		Empathy_2	0.911				
		Empathy_3	0.815				
		Empathy_4	0.896				
		Empathy_5	0.829				

Source: Own construction

Cronbach's alpha coefficient was used to show the reliability of the "quality of health services (QHS)" scale. Cronbach's alpha coefficient for the first-order latent variables tangibility, reliability, response, safety and empathy ranges from 0.900 to 0.940, which shows excellent reliability and internal consistency for this sample. In addition, the value of the "Standardized Cronbach coefficient alpha", which also ranges from 0.900 to 0.940, further confirms the reliability of the scale. Therefore, it can be concluded that all indicators in the model consistently represent the appropriate latent construct or factor, i.e. that the measurement model "quality of healthcare services" is reliable.

Based on the data presented in Table 5, we can draw the same conclusions regarding the reliability of the US model. Since Cronbach's alpha coefficient and Standardized Cronbach coefficient alpha for User satisfaction with medical services are 0.976 and for User satisfaction with non-medical services are 0.964, it can be concluded that all indicators in the model consistently represent the appropriate latent construct or factor, i.e. that the measurement model "user satisfaction" is reliable.

Table 5 CFA results for the measurement model of user satisfaction

Dim.	Subdimension	Code	Factor loads	Cronbach alpha	Standardized Cronbach alpha	CR	MaxR(H)
US	**User satisfaction with medical services**	USMS1	0.835	0.976	0.976	0.976	0.977
		USMS2	0.863				
		USMS3	0.879				
		USMS4	0.864				
		USMS5	0.900				
		USMS6	0.870				
		USMS7	0.847				
		USMS8	0.799				
		USMS9	0.865				
		USMS10	0.889				
		USMS11	0.878				
		USMS12	0.809				
		USMS13	0.796				
		USMS14	0.846				
		USMS15	0.886				
	User satisfaction with non-medical services	USNMS1	0.866	0.964	0.964	0.962	0.967
		USNMS2	0.846				
		USNMS3	0.829				
		USNMS4	0.808				
		USNMS5	0.798				
		USNMS6	0.819				
		USNMS7	0.874				
		USNMS8	0.780				
		USNMS9	0.684				
		USNMS10	0.672				
		USNMS11	0.658				
		USNMS12	0.752				
		USNMS13	0.729				
		USNMS14	0.846				
		USNMS15	0.881				

Source: Own construction

3.5 Model validity testing

To determine the convergent validity of the factors in the models, indicators of standardized factor loading and average variance extracted (AVE) were calculated, and finally we compared the values of CR and AVE. Based on Tables 4 and 5, we can state that all factor loadings are statistically significant, at a significance level of 1%. The standardized factor loadings of all indicators are greater than 0.5, which indicates the fact that they well reflect the latent variable they measure. In Table 6 we compared the values of CR and AVE.

Table 6 Results of convergent validity of measurement models of quality of health services and user satisfaction

Dimension	Subdimension	CR	AVE	CR > AVE
QHS	Tangibility	0.901	0.645	Fulfilled
	Reliability	0.936	0.711	Fulfilled
	Response	0.901	0.751	Fulfilled
	Safety	0.930	0.768	Fulfilled
	Empathy	0.940	0.758	Fulfilled
US	User satisfaction with medical services	0.976	0.732	Fulfilled
	User satisfaction with non-medical services	0.962	0.628	Fulfilled

Source: Own construction

Based on the data from the Table 6, we can see that all AVE values are above the recommended threshold value of 0.5 (from 0.628 to 0.768), which indicates that each construct in the model explains at least 50% of the variance in its indicators. The highest explanation of the variance is among the indicators of the latent variable of security (0.768, i.e. 76.8%). AVE is a more conservative assessment of the validity of the measurement model, and it is considered that based on the CR indicator, a conclusion can also be made about the convergent validity of the model. After all, this indicator also indicates, as we previously stated, the reliability of the construct, where the lower limit of good reliability is the value 0.7, which is satisfied in this model (from 0.901 to 0.976). Since all values of CR > AVE, it can be concluded that the measurement model meets the conditions of convergent validity. According to all analyzed indicators of validity and reliability, it was concluded that the selected indicators explain well the factor (latent variable) they represent. Discriminant validity was tested using Fornell-Larcker criteria, HTMT and HTMT2.

Table 7 Results of discriminant validity of measurement models of quality of health services and satisfaction of users of health services using Fornell-Larcker criteria

Fornell-Larcker

Latent variables	Tangibility	Reliability	Response	Security	Empathy
Tangibility	0.803				
Reliability	0.883***	0.843			
Response	0.852***	0.983***	0.867		
Security	0.848***	0.971***	1.000***	0.877	
Empathy	0.819***	0.911***	0.960***	0.956***	0.871

Fornell-Larcker

Latent variables	User satisfaction with medical services	User satisfaction with non-medical services
User satisfaction with medical services	0.856	
User satisfaction with non-medical services	0.890***	0.793

Source: Own construction

Table 7 shows the values of the square root of the AVE index, while below are the correlations between the constructs, on the basis of which discriminant validity testing can be performed. Testing is done by comparing the square root of the AVE index value with the correlation of the given construct with all other constructs. Based on the presented results, we can conclude that the discriminant validity of the measuring model of the QHS and US is impaired, given that many values on the diagonal are smaller than the correlation coefficients in the relevant rows and columns.

As the discriminant validity according to the Fornell-Larcker criterion was violated, it was retested using the HTMT and HTMT2. Namely, in the simulation study conducted by Henseler, Ringle and Sarstedt (2015) it was shown that the traditional criterion for determining discriminant validity was not effective in finding the problem of discriminant validity when it really exists. The Fornell-Larcker criterion successfully identified 139 problems in only 15% of cases. For this reason, a new measure of discriminant validity was proposed, called the ***heterotrait-monotrait ratio*** (HTMT). After calculating the discriminant validity using HTMT and HTMT2, we can say that it is still impaired, because the ratio of individual constructs hovered around the value of 1 (from 0.80 to 0.99). Of all the latent constructs in the QHS, Response achieved the highest values (0.97 and 0.99). In this case, the literature suggests that the construct should be removed and its indicators should also be removed or joined to other constructs, and that the testing should be repeated. Respecting the content of the indicator, we attached Response_1 and Response_3 to the latent construct Safety, and we attached Response_2 to the construct Empathy.

Table 8 Results of discriminant validity of measurement models of quality of health services and satisfaction of users of health services using HTMT and HTMT2

HTMT

Latent variables	Tangibility	Reliability	Security	Empathy
Tangibility				
Reliability	0.84			
Security	0.84	0.92		
Empathy	0.80	0.90	0.91	

HTMT2

Latent variables	Tangibility	Reliability	Security	Empathy
Tangibility				
Reliability	0.84			
Security	0.84	0.92		
Empathy	0.80	0.90	0.91	

HTMT

Latent variables	User satisfaction with medical services	User satisfaction with non-medical services
User satisfaction with medical services		
User satisfaction with non-medical services	0,87	

HTMT 2

Latent variables	User satisfaction with medical services	User satisfaction with non-medical services
User satisfaction with medical services		
User satisfaction with non-medical services	0,87	

Source: Own construction

The next step was to find the manifest variables of one latent construct that are highly correlated with the manifest variables of other latent constructs. Based on a detailed analysis, the existence of a high correlation (correlations above 0.7) with the manifest variables of other latent constructs was identified for the following manifest variables: Safety_2, Empathy_2, Security_1, Reliability_5, Response_2 which was added to latent construct Empathy and Response_3 which was added to latent construct Safety, and therefore they were eliminated. Within each latent construct, the condition of having at least three manifest variables is met.

After the transformation, we repeated the testing, and the final results are shown in the Table 8. The results showed that values of HTMT and HTMT2 are around the threshold value of 0.90, indicating the absence of discriminant validity problems in the models, with the QHS measurement model retaining four latent constructs of the first order: tangibility, reliability, security and empathy. Namely, smaller values of HTMT and HTMT2 show that the correlations between indicators measuring different constructs are smaller compared to the correlations between indicators measuring the same construct. Precisely from this comes the conclusion that the discriminant validity in the models has been confirmed, that is, it has been established that the constructs are mutually different and that their associated indicators measure them well. In the case of US, the aforementioned problem did not manifest itself, and therefore no transformation was performed.

3.6 Assessment of structural relationships/hypothesis testing

Based on the conducted analyses, we concluded that all measurement models met the assumed criteria of suitability, reliability and validity, and as such were the subject of analysis and hypothesis testing. The basic work model is a recursive structural model, in which all paths go from predictor to dependent variables, which means that no two-way relationships are defined. The basic structural model now consists of four exogenous (independent) latent constructs: tangibility, reliability, security and empathy (the result we obtained based on discriminant validity). Furthermore, the structural model is made up of two endogenous (dependent) latent constructs: the level of user satisfaction with medical services and the level of user satisfaction with non-medical services. The effects of exogenous on endogenous latent constructs are defined by hypotheses. Looking at the basic structural model of the subject research, it can be concluded that the model consisted of 2 latent constructs of the second order with 6 latent constructs of the first order and a total of 47 manifest variables. The total number of parameters is 1 128 (based on the formula: $p \cdot (p + 1)/2$), the number of parameters to be calculated is 131, while the number of degrees of freedom is 997 (1 128 – 131). Assessment of the suitability of the basic model was carried out using the GOF indicator. GOF indicators are above/below the recommended limit values.

Based on the tabular presentation, we can conclude that the absolute, parsimonious and incremental indicators are acceptable and thus we have confirmed the suitability of the basic structural model. Chi square coefficient (CMIN), as an absolute indicator, is significant and amounts to 3 860 119 with 997 degrees of freedom (df), while CMIN/df = 3 872. Considering the sample of 886 observations and 47 manifest variables, it is expected that the chi square test will be significant. The values of RMSEA, which should be less than 0.08 and equal to 0.057, and SRMR, whose value should be less than 0.09 and equal to 0.0324, are classified as absolute indicators of suitability, which additionally confirms the suitability of the measurement model. Parsimony indicators such as AGFI = 0.797, PNFI = 0.851 and PCFI = 0.868 are above the recommended values and also indicate a good fit of the model. The values of incremental indicators – NFI = 0.923, CFI = 0.941, TLI – NNFI = 0.936, RFI = 0.916 and IFI = 0.941 are above the limit of 0.9, and indicate a satisfactory fit of the model. Based on the presented results of the suitability of the model, it is possible to conclude that the proposed model of the influence of the quality of health services on the level of user satisfaction with (non)medical services meets the methodological requirements.

Table 9 GOF fit index of the structural model

Measures	Threshold value	Structural model
P-value	> 0.05	0.001
Minimum Discrepancy function by degrees of freedom divided (CMIN/df)	< 5	3.872
Root mean square error of approximation (RMSEA)	< 0.08	0.057
Standardized root mean squared residual (SRMR)	< 0.09	0.0324
Comparative Fit Index (CFI)	> 0.90	0.941
Normed Fit Index (NFI)	> 0.90	0.923
Relative Fit Index (RFI)	> 0.90	0.916
Incremental Fit Index (IFI)	> 0.90	0.941
Tucker Lewis – Non-Normed Fit Index (TLI – NNFI)	> 0.90	0.936
Goodness of Fit Index (GFI)	> 0.90	0.82
Adjusted Goodness of Fit Index (AGFI)	> 0.80	0.797
Parsimonious Normed Fit Index (PNFI)	> 0.50	0.851
Parsimonious Comparative Fit Index (PCFI)	> 0.50	0.868

Source: Own construction

Given that the overall suitability of the model is acceptable, it is possible to approach the analysis of the structural part of the model, with the aim of assessing whether the proposed theoretical relations, that is, hypotheses are supported in the specific research context of Bosnia and Herzegovina. In the analysis, it is necessary to check the signs of the parameters, and the statistical significance of the parameters measured by the t-value. The results of the path analysis, for the H1 and H2 hypothesis, are presented in the following Table 10.

Table 10 Presentation of the results of testing the basic structural model

Hypotheses	Non-standard. evaluation parameters	Standard. evaluation parameters	t	P	R^2	Cohen's f
H1a: tangibility → USMS	0.199	0.184	3.730	0.001	0.843	2.32
H1b: reliability → USMS	–0.034	–0.030	–0.226	0.821		
H1d: safety → USMS	0.619	0.597	3.743	0.001		
H1e: empathy → USMS	0.189	0.197	2.711	0.007		
H2a: tangibility → USNMS	0.030	0.028	0.443	0.658	0.886	2.78
H2b: reliability → USNMS	0.892	0.795	4.098	0.001		
H2d: safety → USNMS	–0.951	–0.921	–3.651	0.001		
H2e: empathy → USNMS	0.457	0.477	5.111	0.001		

Source: Own construction

When it comes to the results of testing the H1 and H2 hypotheses, it can be confirmed that six of the eight sub-hypotheses are supported by this study. The results show that tangibility ($t = 3.730$, $p < 0.01$), security ($t = 3.743$, $p < 0.01$) and empathy ($t = 2.711$, $p < 0.01$) have a statistically significant positive impact on the level of customer satisfaction medical services, while reliability does not have a statistically significant positive influence on the level of user satisfaction with medical services ($t = -0.226$, $p > 0.1$). According to the obtained results, it is obvious that the users of the tertiary level of health care do not sufficiently perceive the provided reliability as a dimension of the quality of the health service, however, although such user perception was identified, it does not affect the overall level of satisfaction with medical services. Therefore, other isolated dimensions of the quality of health services at the tertiary level of health care have a much greater impact on user satisfaction. Of course, this should not mean for management that reliability improvement activities should not be affirmed.

The results further indicate that reliability ($t = 4.098$, $p < 0.01$) and empathy ($t = 5.111$, $p < 0.01$) have a statistically significant positive influence, while security ($t = -3.651$, $p < 0.01$) has statistically significant, but negative impact on the level of user satisfaction with non-medical services. The results of the latent construct tangibility at the tertiary level of health care show that it has no statistically significant effect on the level of satisfaction with non-medical services ($t = 0.443$, $p > 0.1$). The reason for this lies in the fact that tangibility is directly related to the implementation of quality medical services, and is very little represented, that is, noticeable in the case of non-medical services. In the case of the latent construct safety, a statistically significant negative impact on the level of satisfaction with non-medical services is noticeable, which certainly has its logical basis. Namely, in order to provide a higher level of security such as better quality medicines, especially in the case of more complex and demanding hospital conditions, the users have to pay a higher price and undergo more complex internal procedures in clinical centers, which directly contributes to a decrease in their satisfaction and vice versa.

The table also shows indicators of the coefficient of determination (R^2) and effect size and for the calculation of which Cohen's f indicator was used. As a "rough rule" it can be said that R^2 values of 0.25–0.50 indicate weak, 0.50–0.75 moderate, and above 0.75 high explanatory power of the model (Hair et al., 2019). However, for social sciences, lower limit values of the coefficient of determination (0.02, 0.13 and 0.26) were proposed. Therefore, when interpreting R^2, one should always take into account the context of the research, i.e. the scientific discipline in which the model is observed (Sarstedt, Ringle and Hair, 2017). Looking at the explanatory power of the estimated basic structural model in Table 10, it can primarily be determined that the coefficients of determination (R^2) are high. For both endogenous constructs, the indicators are higher from the stated limit, which confirms the high explanatory power of the estimated model. The determination coefficients show that a high proportion of the variance in the endogenous constructs (84.3% and 88.6%) is explained by the combination of the influence of exogenous latent variables on the endogenous latent variable bias towards complex models, an *adjusted coefficient of determination* R^2_{adj} can be calculated. Adjusted coefficients of determination (R^2_{adj}) of the estimated basic structural model, which can be used to compare different models, have similar values as the coefficient of determination (84.2% i 88.5%). Additionally, the change in R^2 value when a particular exogenous construct is omitted from the model can be used to estimate the magnitude of its effect on the endogenous construct. This measure is called the effect size (f), and it assesses how strongly an exogenous construct contributes to explaining a certain endogenous construct in terms of R^2 (Avkiran, 2018), which is calculated according to the following formula (Cohen, 1988):

$$\textbf{Cohen's } \boldsymbol{f} = \sqrt{\frac{\mathbf{R}^2}{\left(1-\mathbf{R}^2\right)}} \,. \tag{1}$$

Guidelines for interpreting effect sizes are as follows: 0.01 – weak effect, 0.20 – moderate effect, and 0.40 strong effect. Based on the tabular presentation, we can state a strong influence of independent constructs on the dependent construct, that is, the quality of health services has a strong influence on the level of user satisfaction with medical services (f = 2.32) and on the level of user satisfaction with non-medical services (f = 2.78). Based on the basic structural model and empirical research, the following table presents the hypotheses.

Table 11 Conclusions of hypothesis testing – basic structural model

Hypotheses	Content of the hypothesis	Conclusion of the analysis
H1a	Tangibility has a statistically significant effect on the level of user satisfaction with medical services.	**Confirmed**
H1b	Reliability has a statistically significant effect on the level of user satisfaction with medical services.	**Not confirmed**
H1d	Safety has a statistically significant effect on the level of user satisfaction with medical services.	**Confirmed**
H1e	Empathy has a statistically significant effect on the level of user satisfaction with medical services.	**Confirmed**
H1	***There is a statistically significant influence of the quality of health services on the level of user satisfaction with medical services in Bosnia and Herzegovina.***	**Confirmed**
H2a	Tangibility has a statistically significant effect on the level of user satisfaction with non-medical services.	**Not confirmed**
H2b	Reliability has a statistically significant effect on the level of user satisfaction with non-medical services.	**Confirmed**
H2d	Safety has a statistically significant effect on the level of user satisfaction with non-medical services.	**Confirmed**
H2e	Empathy has a statistically significant effect on the level of user satisfaction with non-medical services.	**Confirmed**
H2	***There is a statistically significant influence of the quality of health services on the level of user satisfaction with non-medical services in Bosnia and Herzegovina.***	**Confirmed**

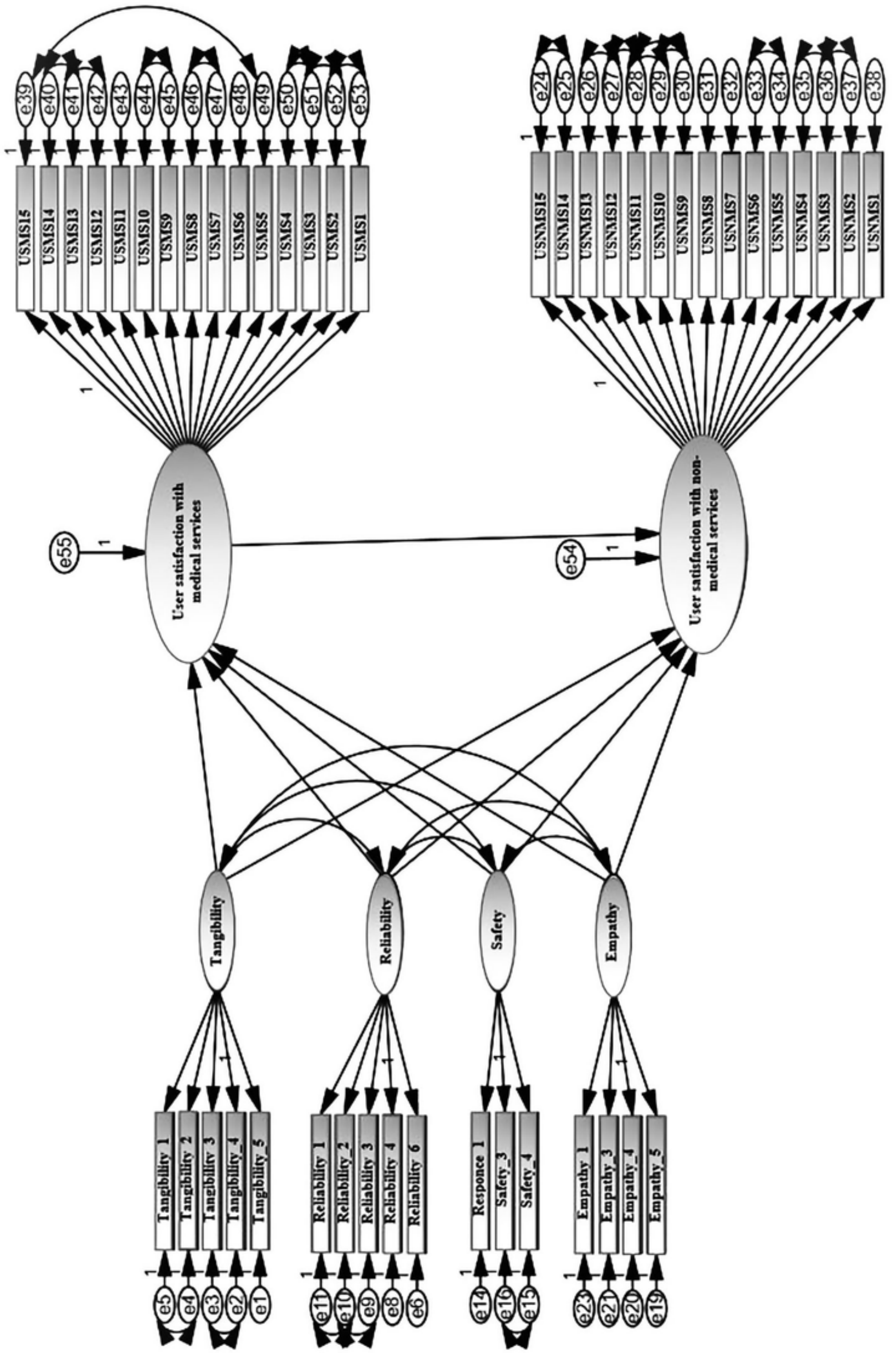
Source: Own construction

Considering the complex connections that can be analyzed by SEM, it is common to visually present the so-called model path diagram. The symbolism in the diagram is as follows (Ho et al., 2012):

- Manifest variables are shown as squares or rectangles;
- Latent variables are shown as circles or ellipses;
- The assumed influence of one variable on another (direct effect) is shown by a straight arrow with one tip;
- Covariances (in the non-standardized solution) and correlations (in the standardized solution) between independent (exogenous) variables are shown by a curved line with arrows.

In addition to the above, error components of manifest and latent variables are also displayed. SEM model parameters include: direct influences on endogenous variables (either from exogenous or other endogenous variables), factor loadings that connect indicators with the corresponding latent variable, and variances and covariances of exogenous variables (Kline, 2011). The following Figure shows the structural model obtained after testing the reliability, validity and hypotheses.

Figure 1 Conclusions of hypothesis testing – basic structural model



Source: Own construction

4 DISCUSSION

A direct comparison of the obtained results with previous research is not possible due to the use of modified measuring scales, as due to the fact that very little research is based on comparative SEM analysis.. However, we can indirectly make a comparison with research based on SERVQUAL or HEALTHQUAL analyses of tertiary levels of health care. The confirmation of the hypotheses about the influence of the quality dimension on user satisfaction (medical and non-medical services) is primarily in accordance with previous research on this topic (Nashrath et al., 2011; Lee and Kim, 2017; Natarajappa et al., 2020). However, there are certain deviations. Thus, Nashrath et al. (2011) point out that Reliability is the greatest source of user satisfaction of tertiary level health care, while the results of the presented research indicate that Reliability is not the source of user satisfaction with medical services, but it is in the case of non-medical services. Also, in their research, Response achieved the lowest values, and as we saw in the results of our research, Response did not correspond to the construct, and was excluded from the hypothesis testing process.

Although it is based on the application of the HEALTHQUAL questionnaire, the research from Lee and Kim (2017) can be partially compared with this one since there are common quality dimensions such as Empathy, Safety and Tangibles. Using confirmatory factor analysis, these authors proved that it was necessary to act simultaneously on all dimensions of the quality of the tertiary level of health care in order to achieve user satisfaction, while in the research in Bosnia and Herzegovina it was established that Tangibility has no effect on satisfaction with non-medical services and Reliability on medical services.

In their work, Natarajappa et al., (2020) focused more on the impact of medical and non-medical dimensions of quality on user loyalty. Although our research does not deal directly with the loyalty of users of the tertiary level of health care, taking into account the theoretical assumption that satisfaction is a prerequisite for loyalty, we can state that the conclusions of these two studies coincide.

While interpreting the results, it is necessary to take into account certain limitations, as well as recommendations for future research arising from the aforementioned limitations. The first limitation is reflected in the limitations of literature bases, which can lead to the exclusion of certain publications that would have a significant impact on the formation of a theoretical concept and its empirical testing. Another limitation relates to the time frame for conducting the empirical research. Considering the spatial limitation of research at the level of Bosnia and Herzegovina, it should certainly be aimed at confirming the created structural model and research at the level of other countries. Such a comparative analysis would contribute to greater objectivity and originality of the presented research. Also, it is recommended to carry out identical research at other levels of health care in order to examine the (non)consistency of differences in the importance of individual dimensions of the quality of health services to user satisfaction. Finally, it is suggested to check the measuring scales on the same or a similar sample in order to test the suitability, reliability, validity and objectivity of future research on the impact of the quality of health services on the satisfaction of users.

CONCLUSION

The focus of business in the twenty-first century is to achieve quality in order to satisfy the increasing demands of users. Quality becomes a basic assumption for the survival and development of any organization. Quality management in healthcare is gaining more and more importance, although it has not been fully explored. Health care institutions, especially public ones, by their very nature are not profit-oriented. However, the essence of their existence is the provision of the highest quality health services to the population and the well-being of society as a whole. In order to achieve the primary goal of their existence, it is necessary to introduce systems, internal processes, organization of people and other resources in a way that will contribute to user satisfaction and the fulfillment of their mission. Achieving the satisfaction of users of health services is not an easy task, considering that it is a question of the results of treatment and the safety of the health of each individual user.

The problem of user satisfaction of health services becomes more complex with the increasing level of health care provided to the user. Therefore, the main motive was aimed at providing assistance to the management of health institutions that operate at the tertiary level of health care in identifying key factors, that is, quality dimensions that contribute the most to the satisfaction of their users. Since satisfaction is a broad and multidimensional concept, a distinction was made between satisfaction with medical and non-medical services. In this way, the direct effects of certain dimensions of quality on the main (medical) services and on secondary (non-medical) services are shown, which together form the overall perception of clinical centers, that is, tertiary hospitals. In the previous publication, the authors primarily focus on the medical aspects of the service, ignoring the fact that admission, discharge and other non-medical aspects have an impact on the overall level of satisfaction of users of health services. This problem was overcome through scientific work and an original and unique construct of the quality of health services and user satisfaction at the tertiary level of health care was created by applying SEM analysis.

The construct confirmed the persistence of the cause-and-effect relationship between the quality of health services and the level of satisfaction with both medical and non-medical services. The conclusion of the construct is the persistence of the inextricable link between the mentioned variables, and that management must simultaneously look at all aspects of the services provided in tertiary level health care institutions with additional engagement in improving the quality dimensions that contribute most to the satisfaction of health care users.

Considering that the statistically significant influence of the quality of health services on the satisfaction of users of health services has been proven, the inevitable conclusion is the necessity of improving working conditions as an initial assumption of user satisfaction. The improvement of working environment conditions can be realized through a greater degree of integration of employees in business decision-making, valorization of their opinions, and enabling open and closer interaction between employees, as well as between management and employees. Employees of healthcare institutions are an inexhaustible source of information and recommendations for improving the quality of services and the satisfaction of users of healthcare services, and are an excellent indicator of successful quality management in healthcare.

References

ALALOOLA, N. A., ALBEDAIWI, W. A. (2008). Patient satisfaction in a Riyadh Tertiary Care Centre. *International Journal of Health Care Quality Assurance*, 21(7): 630–637.

AVKIRAN, N. K. (2018). Rise of the partial least squares structural equation modeling: an application in banking. In: AVKIRAN, N. K., RINGLE, C. M. *Partial least squares structural equation modeling-Recent Advances in Banking and Finance,* Cham: Springer, 1–29.

BOLTON, R. N., DREW, J. H. (1991). A Longitudinal analysis of the impact of service changes on customer attitudes. *Journal of Marketing,* 55: 365–384.

COHEN, J. (1988). *Statistical power analysis for the behavioral sciences*. 2nd Ed. New Jersey: Lawrence Erlbaum.

CRONIN, J. J., TAYLOR, S. A. (1992). Measuring service quality: a reexamination and extension. *Journal of Marketing*, 6: 55–68.

DILSHAD, S., AKHTAR, A., HUDA, S. S., SAMAD, N. (2020). Assessment of Healthcare Service Quality: Tertiary Care Hospitals of Dhaka City. *Global Issues and Innovative Solutions in Healthcare, Culture, and the Environmen*t, 271–291.

FULLER, E. L., HEMMERLE, W. J. (1966). Robustness of the maximum-likelihood estimation procedure in factor analysis. *Psychometrika*, 31(2): 255–266.

GARG, N., GUPTA, S. K., MAHESH, R. (2014). Patient Satisfaction Survey at a Tertiary Care Speciality Hospital. *Int. J. Res. Foundation Hosp. Healthc Adm*, 2(2): 79–83.

HAIR, J. F., RISHER, J. J., SARSTEDT, M., RINGLE, C. M. (2019). When to use and how to report the results of PLS-SEM [online]. *European Business Review*, 31(1): 2–24, <https://doi.org/10.1108/EBR-11-2018-0203>.

HAM, H. S., PECK, E. H., MOON, H. S., YEOM, H. A. (2015). Predictors of Patient Satisfaction withTertiary Hospitals in Korea [online]. *Hindawi Publishing Corporation Nursing Research and Practice*, 1–6. <https://doi.org/10.1155/2015/749754>.

HO, M.-H. R., STARK, S., CHERNYSHENKO, O. (2012). *Graphical representation of structural equation models using path diagrams.* Psychology Faculty Publications.

HU, L., DING, H., HU, G., WANG, Z., LIU, S., LIU, Y. (2019). How perceived quality of care affects outpatient satisfaction in China: a cross-sectional study of 136 tertiary hospitals. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*, 56: 1–8.

IMAM, S. Z., SYED, K. S., ALI, S. A., ALI, S. U., FATIMA, K., GILL, M., HASSAN, M. O., HASHMI, S. H., SIDDIQI, M.T., KHAN, H. M., JAMEEL, O. F. (2007). Patients' satisfaction and opinions of their experiences during admission in a tertiary care hospital in Pakistan – a cross sectional study [online]. *BMC Health Services Research*, Vol. 7. <https://doi.org/10.1186/1472-6963-7-161>.

HENSELER, J., RINGLE, C., SARSTEDT, M. (2015). A New Criterion for Assessing Discriminant Validity in Variance-based Structural Equation Modeling [online]. *Journal of the Academy of Marketing Science*, 43: 115–135. <https://doi.org/10.1007/s11747-014-0403-8>.

KLINE, R.B. (2011). *Principles and Practice of Structural Equation Modeling*. 3rd Ed. New York: The Guilford Press.

KULKARNI, S. K., (2018). A study of patient satisfaction level in Out Patient Department (OPD) in a Tertiary Care Hospital in Maharashtra. *Journal of Dental and Medical Sciences,* 17(3).

LEE, D., KIM, K. K. (2017). Assessing healthcare service quality: a comparative study of patient treatment types. *Int. J. Qual. Innov.,* 3(1): 1–15.

MAROOF, B., TASNEEM, S. S., JABEEN, R. J., TASKEEN, R. (2019). The Assessment of Patient's Satisfaction and Opinions of their Experience during Admission in the Tertiary Care Hospital [online]. *International Journal of Academic Research Business and Social Sciences,* 9(5): 555–577. <http://dx.doi.org/10.46886/IJARP/v9-i5/5905>.

MIKULIĆ, J. (2007). Quality Management of Airportservices – an Analysis of the Multifactorstructure of Customer Satisfaction. *Market*, Vol. XIX(1), Zagreb: Faculty of Economics.

NASHRATH, M., AKKADECHANUNT, T., CHONTAWAN, R. (2011). Perceived nursing service quality in a tertiary care hospital Maldives [online]. *Nursing & Health Sciences*, 13(4): 495–501. <https://doi.org/10.1111/j.1442-2018.2011.00648.x>.

NATARAJAPPA, SH., RAVI, M. D., CHANDRASHEKAR, B. R., NAGESH, P. (2020). Evaluation of Service Quality in a Tertiary Care Hospital Using a Novel Patient Satisfaction Survey Tool in Mysore City, India – a Pilot Study. *Indian Journal of Public Health Research & Development,* 11(7): 858–865.

NWABUEZE, C. J., ONYEAGU, S. I., IKPEGBU, O. (2009). Robustness of the maximum likelihood estimation procedure in factor analysis. *African Journal of Mathematics and Computer Science Research,* 2(5): 81–87.

OJENIWEH, O., POPOOLA, O., ADESINA, A. D., MARIERE, U., ROTIFA, S. (2021). Patient satisfaction with physical therapy care, in a tertiary healthcare facility: an evaluation of quality of physical therapy services provided in Federal Medical Centre [online]. *Yenagoa, Bayelsa State, Yen Med*, 3(2): 116–123. <https://doi.org/10.13140/RG.2.2.31742.10567>.

OLIVER, R. L. (1993). A conceptual model of service quality and service satisfaction: compatible goals, different concepts. *Advances in Service Marketing and Management*, 2: 65–85.

OZRETIĆ, D. Đ., ŠKARE, V., ČERFALVI, V., BENCEKOVIĆ, Ž., KOMARAC, T. (2020*).* Assessment of the Quality of Public Hospital Healthcare Services by using SERVQUAL [online]. *Acta Clin. Croatia*, 59(2): 285–293. <https://doi.org/10.20471/acc.2020.59.02.12>.

SARSTEDT, M., RINGLE, C. M., HAIR, J. F. (2017). *Partial least squares structural equation modeling*. In: HOMBURG, C., KLARMANN, M.,VOMBERG, A. E. *Handbook of market research*, Cham: Springer.

SHARIF, P., SHARIF, N. (2018). *Structural Equation Modeling with AMOS.* Teheran: Artin Teb.

SNOJ, B. (1995). The profiles of importance of service quality components in health spas [online]. *Der Markt,* 34(2): 95–104. <https://doi.org/10.1007/BF03032046>.

SNOJ, B., SAVIĆ, V., RAJTMAJER, B. (1999). Measuring the perceived quality of ISS Servisystem from Maribor. *Akademija MM*, 5: 33–42.

STAFFORD, M., STAFFORD, T. F., WELLS, B. P. (1998). Determinants of customer satisfaction in the auto casualty claims process. *The Journal of Services Marketing*, 12(6).

ZEITHAML, V. A., BERRY, L. L., PARASURAMAN, A. (1993). The nature and determinants of customer expectations of service [online]. *Journal of the Academy of Marketing Science*, 21(1): 1–12. <http://dx.doi.org/10.1177/0092070393211001>.

Digital Country Rankings for the Visegrád Group Countries with DEA and TOPSIS

Zoltán Bánhidi[1] | *Budapest University of Technology and Economics, Budapest, Hungary*
Imre Dobos[2] | *Budapest University of Technology and Economics, Budapest, Hungary*
Noémi Kalló[3] | *Budapest University of Technology and Economics, Budapest, Hungary*
Ariella Janka Tarjáni[4] | *Budapest University of Technology and Economics, Budapest, Hungary*

Received 14.6.2024, **Accepted** (reviewed) 1.8.2024, **Published** 14.3.2025

Abstract

Our paper is based on the five principal dimensions of the International Digital Economy and Society Index (I-DESI), but instead of using the original scoring model based on arbitrary pre-determined weights, we apply more objective ranking methods that use the statistical properties of the data series to determine where the Visegrád Group (V4) countries (Czechia, Hungary, Poland and Slovakia) stand in terms of digital development among the countries of the European Union and other developed countries in the data set. The ranking is performed using the DEA-CWA (Data Envelopment Analysis/Common Weights Analysis) method (with six models) and the TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) method. Although the resulting weight vectors differ significantly from the arbitrary weights set by the European Commission, the country rankings remain similar, displaying relatively little sensitivity to the weighting method chosen.

Keywords	***DOI***	***JEL code***
Data Envelopment Analysis, information and communication technology, International Digital Economy and Society Index, TOPSIS, Visegrád Group	*https://doi.org/10.54694/stat.2024.32*	*C44, L38*

[1] Department of Economics, Budapest University of Technology and Economics, 1111 Budapest, Műegyetem rkp. 3, Hungary. Corresponding author: e-mail: banhidi.zoltan@gtk.bme.hu, phone: (+36)14631185. ORCID: <https://orcid.org/0000-0003-0262-5197>.

[2] Department of Economics, Budapest University of Technology and Economics, 1111 Budapest, Műegyetem rkp. 3, Hungary. E-mail: dobos.imre@gtk.bme.hu, phone: (+36)14631971. ORCID: <https://orcid.org/0000-0001-6248-2920>.

[3] Department of Management and Business Economics, Budapest University of Technology and Economics, 1111 Budapest, Műegyetem rkp. 3, Hungary. E-mail: kallo.noemi@gtk.bme.hu, phone: (+36)14631057. ORCID: <https://orcid.org/0000-0003-3193-081X>.

[4] Department of Management and Business Economics, Budapest University of Technology and Economics, 1111 Budapest, Műegyetem rkp. 3, Hungary. E-mail: tarjani.janka@gtk.bme.hu, phone: (+36)14632782. ORCID: <https://orcid.org/0000-0002-3912-4022>.

INTRODUCTION

The International Digital Economy and Society Index (I-DESI) is an indicator system published every two years (from 2016), which is prepared by independent experts (research companies) at the request of the European Commission. Its objective is to provide a comprehensive assessment of the position of the European Union on the "road towards a digital society and economy" compared to non-EU economies.

The defining principle in the methodology is that the indicator system reflects the results of the European Commission's original DESI (Digital Economy and Society Index) indicator, i.e. I-DESI also measures the progress of digital transformation (as DESI does for EU member countries) and expands it to non-EU countries using substitute indicators, which are available for them. DESI and I-DESI are indicator systems combining individual indicators with predefined weights and using similar but not identical scoring models to evaluate and rank individual countries based on their digital performance. Consequently, minor inconsistencies can be observed in the country rankings and scores within the DESI and I-DESI databases, even when non-EU countries are excluded from the analysis. Nevertheless, the authors of I-DESI implemented strategies to ensure a high degree of alignment between the two. To this end, they conducted correlation and covariance tests of indicators, sub-dimensions, and dimensions with the objective of "[minimising] covariance in any new or replacement indicators" (European Commission, 2021: 40).

The 2020 edition of I-DESI measures the achievements in the field of digitisation in 45 countries (of which 27 are EU members, 18 are non-EU countries) in five main policy areas, as shown in Table 1 (European Commission, 2021). In the following sections, we will use the abbreviations for the dimensions, as shown in Table 1.

Table 1 Dimensions of I-DESI 2020 and their interpretations

Dimension	Abbrev.	Interpretation of the given dimension
Connectivity (internet access)	CNC	Availability and quality of broadband network infrastructure.
Human capital	HUC	Level of digital skills necessary to access opportunities offered by the digital society.
Use of internet	UOI	Use of internet services by citizens (consumers).
Integration of digital technology	IDT	Digitisation of businesses and online sales channels.
Digital public services	DPS	Digitisation of public services, focusing on e-government services.

Source: Based on European Commission (2021)

In this study, we used the data of these main dimensions of the year 2018 as a basis for the evaluation of the countries and the calculation of the DEA and TOPSIS efficiencies. It is worth noting that, despite the authors' desire to include more recent data, the most up-to-date edition of I-DESI available at the time of analysis was the 2020 edition. While a call for tenders for the 2022 edition of I-DESI is publicly available (European Commission, 2022), and the publication itself is believed to exist in an eBook form (WorldCat, 2022), it has been inaccessible for an unknown reason.

In the I-DESI 2020 study, the methodology for aggregating indicators into sub-dimensions, sub-dimensions into dimensions, and dimensions into the overall index follows a simple bottom-up approach (scoring model based on subjective weights). Weighted arithmetic averages are applied following the structure of the index to calculate the top-level I-DESI score using Formula (1):

$$IDESI = CNC \cdot 0.25 + HUC \cdot 0.25 + UOI \cdot 0.15 + IDT \cdot 0.2 + DPS \cdot 0.15 \,. \tag{1}$$

It should be noted, however, that since 2021, the Use of Internet (UOI) dimension has been removed from the (EU-only) DESI, and the structure has been simplified, with the remaining four dimensions having equal weights of 25%. Given that the I-DESI was designed to be an international extension of the original index, it is reasonable to expect that these changes would be reflected in the I-DESI 2022 study and all future editions.

The paper is structured as follows. In the next section, we provide a brief overview of the literature related to the DESI and I-DESI indicator systems. In the subsequent sections, we present the methodology of the models used, the Data Envelopment Analysis (DEA), the models based on it, and the TOPSIS method, as well as the results, highlighting the position of the four Visegrád Group (V4) countries (Czechia, Hungary, Poland, Slovakia). Finally, we conclude our study.

1 LITERATURE SURVEY

There are several indicator systems that aim to characterise the degree of digital transformation and digital development at the level of country groups and countries. In this paper, we focused on presenting the main findings in the literature review related to DESI and I-DESI.

As these indices measure the digital economy performance of a specific set of countries, their analyses can have different goals. According to our literature review, these goals are (I) analysing the indices, their dimensions, and methodologies, (II) evaluating the performance of one or some included countries or (III) some other important issues.

Many studies have analysed the indices' different dimensions and their relationships with statistical methods (e.g. correlation and factor analysis). Multivariate statistical methods were used by Bánhidi et al. (2020), and the authors found that the dimensions of DESI are closely related. Tokmergenova et al. (2021) analysed the multicollinearity among the dimensions of I-DESI and found some redundancy among them. These findings are important not only for further analysis of the digital economy's performance but also show that these characteristics can only be developed with a comprehensive strategy.

Other researchers considered these indices as competitors or measures complementary to others. Kotarba (2017) discusses the most important indicator systems of digital development, including DESI, showing the main similarities and differences, and making suggestions for their development. Esses and Szalmáné Csete (2022) examined the digital transformation and sustainability transition of the European capital cities using the Cities In Motion Index (CIMI), with the Sustainable Development Goals (SDG) and the Human Development Index (HDI). They also developed a modified CIMIS index extended with sustainability to rank the cities and assess their strengths and weaknesses.

Focusing on one or some countries, many (I-)DESI-related analyses were done (performing e.g. cluster analyses, data envelopment analyses). Bánhidi et al. (2020) ranked the EU-member countries based on DESI data using many methods. Tarjáni et al. (2023) compared the 27 EU Member countries and 18 non-EU participants based on I-DESI data with discriminant and variance analysis. Bánhidi and Dobos (2021), and Bánhidi et al. (2021) examined Russia's position and performance with a DEA/CI model. Moroz (2017) used the DESI and the Networked Readiness Index (NRI) to evaluate Poland's situation and the dynamics of its development. Laitsou et al. (2020) analysed Greece's digital performance and possible development strategies using the Gompertz model.

Different important parts of the economy and phenomena in our society can be analysed with DESI and I-DESI measures, which lie somewhat outside our analysis's scope but should be mentioned as they highlight the following analysis possibilities. Roukanas (2021) analysed the innovation capabilities of countries based (among others) on DESI, I-DESI data. Skare et al. (2023) meanwhile examined the impact of digital technology on SMEs using DESI data. Finally, Basol and Yalçin (2021) analysed the effects of DESI indicators, especially on the labour market indicators.

2 METHODS

The first DEA model was initiated by Charnes et al. (1978) and was followed by countless other model variants. When using the International Digital Economy and Society Index (I-DESI), the ideal value of the given criterion for each of the five dimensions is the highest possible, so these criteria can be considered outputs in the DEA model. Since the denominator, i.e. the weighted criterion of this DEA model is equal to one, there are no explicit input criteria. In the literature, this latter model is called a DEA model without explicit inputs (DEA/WEI) and/or a DEA-type composite indicators (DEA/CI) method (Bánhidi and Dobos, 2023; Cherchye, 2008; Dobos and Vörösmarty, 2014).

The DEA/WEI model was first developed by Fernandez-Castro and Smith (1994), then applied by Despotis (2005) and Liu and Peng (2008) to practical problems. Due to the model's shape, the earlier mentioned composite indicators were used.

Table 2 summarises the advantages and disadvantages of DEA models. The table is based on the results of Iqbal and Lerme (1997).

Table 2 Advantages and disadvantages of DEA models

<table>
<tr><th>Advantages of DEA</th><th>Disadvantages of DEA</th></tr>
<tr><td>Versatility: DEA can handle multiple inputs and outputs simultaneously, making it suitable for complex scenarios where various factors influence efficiency.</td><td>Sensitivity to Input and Output Selection: The results in DEA can be sensitive to the choice of inputs and outputs. Proper analysis of their relative importance is crucial before conducting DEA.</td></tr>
<tr><td>Returns to Scale Consideration: DEA accounts for returns to scale when calculating efficiency. It allows us to assess efficiency changes based on the size and output levels of an organisation.</td><td>Lack of Statistical Inference: DEA does not provide statistical significance tests, limiting our ability to make formal statistical inferences.</td></tr>
<tr><td>Non-parametric Approach: DEA does not require specific functional forms or assumptions about the production process, making it flexible and applicable across different industries.</td><td>Assumption of Constant Returns to Scale: While DEA considers returns to scale, it assumes constant returns to scale for all decision-making units (DMUs), which may not always hold in practice.</td></tr>
<tr><td>Competitor Analysis: By evaluating comparative advantage and disadvantage, DEA provides insights into organisations' performance relative to its peers.</td><td rowspan="2">Efficiency Frontier Ambiguity: The efficiency frontier (the boundary of efficient DMUs) may be ambiguous due to the non-parametric nature of DEA.</td></tr>
<tr><td>Strategic Alliances: DEA helps assess the feasibility and desirability of forming strategic alliances.</td></tr>
</table>

Source: Based on Iqbal and Lerme (1997)

The basic DEA model for ranking cannot be fully applied here because the DEA efficiency of several decision-making units (DMUs, i.e., the countries) can reach an optimum, i.e., a single value; this means that these DMUs cannot be ranked clearly. At the same time, we also have the problem that individual DMUs achieve different efficiencies that can be calculated for every country. This also means that the weights of DMUs cannot be clearly determined in classical DEA models. Therefore, another method must be found that evaluates all possible DMUs with the same weight.

One of the first such applications is the Common Weights Method (CWA) model, first used by Podinovski and Athanassopoulos (1998). The problem is named the MaxiMin DEA model. In this DEA type model, we first look for the country that gives the smallest efficiency for a predefined weight vector, and after that find the weight vector that optimises this minimum. The name of this procedure refers to this hierarchical optimisation procedure, where we first perform minimisation and only then maximise the DEA efficiency.

To solve the problem, it is necessary to find a DEA procedure that evaluates all countries with the same weight. This second procedure is called the method of common weights. The already mentioned MaxiMin model can also be classified in this group. Liu et al. (2011) suggested an easy form of the procedure. The model in this form is a linear programming problem searching for common weights with boundary conditions, including efficiency constraints.

The third type of DEA/CWA model determines common weights by compromise programming. This method was initiated by Kao and Hung (2005). The goal functions here can be e.g. Manhattan, Euclidean, or Chebyshev functions, which are distance functions. However, to apply these methods, a nadir and/or ideal point must be determined in the set of the efficiency of the countries. A possible nadir efficiency can be equal to zero vector. The ideal is the previously defined DEA efficiencies or the maximum efficiency achievable for each DMU, i.e. one. In this study, only the two ideal efficiencies are used. The ideal point requires finding the weights for which the distance between the ideal point and the set of weights is minimal. In the case of non-ideal points, we find the minimal distance between the nadir point and the set of all possible weights. The $\boldsymbol{E}^*$ vector represents the efficiencies that show the optimal efficiency of each country in all possible DEA models.

In the DEA method, a different weighting system is assigned to each DMU. Therefore, the search for common weights often appears in the literature. Roll and Golani (1993) solve the problem of finding a common weight by restricting them. In Kao and Hung (2005), a method of compromise programming (CPM) was proposed to find the common weights with solutions of non-linear programming problems.

The inequality system (2)–(3) determines the set of possible weights. The first of the two systems of inequalities (2) specifies the upper limit of the common weights, while the second (3) specifies the non-negativity condition. The number of DMUs is p, and the vector $\boldsymbol{y}_j$ is the evaluation of the j-th DMU, i.e. country. The vectors $\boldsymbol{y}_j$ can be gathered in the matrix $\mathbf{Y}$. The vector $\mathbf{u}$ contains the DEA weights. The DEA/WEI weights are equal to the $\mathbf{u} \cdot \mathbf{Y}$ vector.

$$\mathbf{u} \cdot \mathbf{y}_{\mathrm{j}} \leq 1; \mathrm{j} = 1, 2, ..., \mathrm{p}. \tag{2}$$

$$\mathbf{u} \geq 0. \tag{3}$$

The goal functions of the possible Date Envelopment Analysis models are shown in Table 3.

Table 3 Goal functions of the used DEA models

DEA Models	Goal function ($F_i(\mathbf{u})$)	Literature
MaxiMin model (1.)	$F_1(\mathbf{u}) = \min\limits_{1\, jp} \mathbf{u} \cdot \boldsymbol{y}_j \rightarrow \max$	Podinovski and Athanassopoulos (1998)
CWA model (2.)	$F_2(\mathbf{u}) = \mathbf{u} \cdot \mathbf{Y} \cdot \mathbf{1} \rightarrow \max$	Liu and Peng (2008)
CPM with **1** (1; 1; 1;…) as ideal point (3., 4.)	$F_3(\mathbf{u}) = d_2(\mathbf{u} \cdot \mathbf{Y}; \mathbf{1}) \rightarrow \min$ $F_4(\mathbf{u}) = d_{+\infty}(\mathbf{u} \cdot \mathbf{Y}; \mathbf{1}) \rightarrow \min$	Kao and Hung (2005)
CPM with $\boldsymbol{E}^*$ ideal points (5., 6.)	$F_5(\mathbf{u}) = d_2(\mathbf{u} \cdot \mathbf{Y}; \boldsymbol{E}^*) \rightarrow \min$ $F_6(\mathbf{u}) = d_{+\infty}(\mathbf{u} \cdot \mathbf{Y}; \boldsymbol{E}^*) \rightarrow \min$	Kao and Hung (2005)

Source: Own elaboration based on Kao and Hung (2005), Liu and Peng (2008), and Podinovski and Athanassopoulos (1998)

Similar results for both the summation **1** and the DEA efficiency ***E****** vectors are given for the result of the CWA of the Manhattan distances (Toloo, 2014). The distance functions of CPM are (4)–(5):

$$\text{Euclidean distance } (k = 2)\text{: } d_2\left(\mathbf{u}\cdot\mathbf{Y};\boldsymbol{E}\right) = \sqrt{\sum_{j=1}^{p}(\mathbf{u}\cdot\mathbf{y}_j - E_j)^2}. \tag{4}$$

$$\text{Chebyshev distance } (k = +\infty)\text{: } d_{+\infty}\left(\mathbf{u}\cdot\mathbf{Y};\ \boldsymbol{E}\right) = \max_{1\le j\le p}\left|\mathbf{u}\cdot\mathbf{y}_j - E_j\right|. \tag{5}$$

As defined before, vector ***E*** is an ideal efficiency vector that can be equal to the efficiency vectors ***E**** or **1**.

TOPSIS is a decision-theory method based on a simple geometric approach that ranks alternatives according to their similarity to an ideal solution. It has gained popularity due to its relative simplicity, stability, and moderate computational complexity. It has been widely used in various domains (e.g. Omurbek et al., 2021; Shaktawat and Vadhera, 2021), and its effectiveness has been demonstrated in numerous studies. The main limitation of the method is its limited applicability to dynamic problems with critical time dimensions and close interdependence between ranking criteria. However, these limitations are not relevant to our study as we worked with a cross-sectional database.

3 RESULTS AND DISCUSSION

3.1 Ranking with DEA/WEI and CWA methods

Before presenting the results of the calculations, we show that the CWA model in Table 3 and the Manhattan distance CPM analysis do not differ from each other, i.e. they result in the same weighting system. Therefore, of the two models, it is sufficient to solve the CWA model. Since the weighted sum of the predetermined **1** and ***E**** vectors does not depend explicitly on the weights, we optimise the expression **–u·Y·1**, which means that the negative of a linear function must be optimised. It also means we have got the CWA model back. Hence, minimising the Manhattan distance models also leads to the optimisation problem of the CWA model.

The mathematical model of the six common weighting problems can be characterised as (6)–(8):

$$\mathbf{u}\cdot\mathbf{y}_j \le 1;\ j = 1, 2, \ldots, p. \tag{6}$$

$$\mathbf{u} \ge 0. \tag{7}$$

$$\boldsymbol{F_i}\left(\mathbf{u}\right) \to \min / \max,\ \ i = 1, 2, \ldots, 6. \tag{8}$$

Problems (i) 1 and 2 are maximised, while problems 3, 4, 5 and 6 are minimised. The mathematical model of the functions is shown in Table 3. All calculations were performed using Microsoft Excel 2016, except the calculations of the correlation coefficients, which were done in IBM SPSS 28. For the sake of brevity, only the results (in terms of weights and efficiencies) are presented in the paper; the detailed calculations are omitted. The efficiencies of the six DEA models (6)–(8) [i = 1, 2, …, 6] can be found in Table 4. The table shows the efficiency solutions obtained using the six common weighting methods, with the Visegrád Group (V4) countries highlighted in grey.

Table 4 Efficiencies for common weights methods [data of the 2020 I-DESI]

Country	I-DESI	MaxiMin	DEA CWA	Euclidean E*	Euclidean 1	Chebyshev E*	Chebyshev 1	TOPSIS	DEA efficiency
Australia	0.599	0.894	0.899	0.898	0.899	0.866	0.894	0.645	0.934
Austria	0.519	0.810	0.826	0.824	0.826	0.773	0.810	0.503	0.831
Belgium	0.489	0.832	0.836	0.836	0.836	0.808	0.832	0.467	0.853
Brazil	0.365	0.634	0.629	0.628	0.629	0.544	0.634	0.246	0.663
Bulgaria	0.400	0.801	0.805	0.804	0.805	0.698	0.801	0.287	0.805
Canada	0.553	0.824	0.805	0.807	0.805	0.829	0.824	0.582	0.844
Chile	0.353	0.699	0.706	0.705	0.706	0.633	0.699	0.194	0.707
China	0.463	0.766	0.771	0.769	0.771	0.681	0.766	0.406	0.777
Croatia	0.348	0.739	0.752	0.751	0.752	0.655	0.739	0.204	0.760
Cyprus	0.471	0.855	0.849	0.847	0.849	0.746	0.855	0.425	0.855
Czechia	0.476	0.812	0.822	0.821	0.822	0.767	0.812	0.425	0.822
Denmark	0.695	1.000	1.000	1.000	1.000	1.000	1.000	0.837	1.000
Estonia	0.572	0.869	0.861	0.861	0.861	0.844	0.869	0.602	0.911
Finland	0.683	0.953	0.966	0.966	0.966	0.995	0.953	0.809	1.000
France	0.575	0.928	0.913	0.913	0.913	0.882	0.928	0.583	1.000
Germany	0.579	0.844	0.863	0.863	0.863	0.866	0.844	0.623	0.886
Greece	0.403	0.799	0.790	0.788	0.790	0.679	0.799	0.307	0.799
Hungary	0.411	0.726	0.734	0.733	0.734	0.685	0.726	0.309	0.739
Iceland	0.619	0.939	0.977	0.975	0.977	0.948	0.939	0.672	1.000
Ireland	0.597	0.835	0.849	0.848	0.849	0.848	0.835	0.650	0.869
Israel	0.584	0.744	0.759	0.759	0.759	0.818	0.744	0.633	0.946
Italy	0.382	0.792	0.777	0.777	0.777	0.686	0.792	0.265	0.792
Japan	0.577	1.000	1.000	1.000	1.000	0.961	1.000	0.607	1.000
Latvia	0.412	0.750	0.752	0.752	0.752	0.704	0.750	0.327	0.770
Lithuania	0.437	0.827	0.849	0.846	0.849	0.721	0.827	0.366	0.849
Luxembourg	0.620	0.887	0.911	0.910	0.911	0.893	0.887	0.707	0.922
Malta	0.478	0.934	0.933	0.932	0.933	0.831	0.934	0.428	0.934
Mexico	0.370	0.624	0.614	0.613	0.614	0.561	0.624	0.245	0.679
Netherlands	0.682	0.881	0.886	0.887	0.886	0.952	0.881	0.813	1.000
New Zealand	0.542	0.846	0.844	0.844	0.844	0.821	0.846	0.550	0.846
Norway	0.638	0.919	0.908	0.909	0.908	0.930	0.919	0.737	0.985
Poland	0.364	0.730	0.720	0.718	0.720	0.617	0.730	0.235	0.730
Portugal	0.409	0.774	0.760	0.761	0.760	0.727	0.774	0.313	0.774
Romania	0.417	0.738	0.749	0.747	0.749	0.644	0.738	0.324	0.749
Russia	0.427	0.640	0.631	0.630	0.631	0.604	0.640	0.351	0.728
Serbia	0.378	0.673	0.685	0.683	0.685	0.590	0.673	0.240	0.686
Slovakia	0.389	0.718	0.718	0.717	0.718	0.650	0.718	0.269	0.727
Slovenia	0.469	0.793	0.801	0.800	0.801	0.744	0.793	0.404	0.801
South Korea	0.544	0.952	0.917	0.918	0.917	0.872	0.952	0.545	1.000
Spain	0.466	0.825	0.808	0.808	0.808	0.736	0.825	0.412	0.845
Sweden	0.650	0.922	0.953	0.952	0.953	0.946	0.922	0.752	0.977
Switzerland	0.656	0.914	0.947	0.947	0.947	0.973	0.914	0.736	1.000
Turkey	0.336	0.585	0.572	0.572	0.572	0.541	0.585	0.187	0.585
United Kingdom	0.593	0.905	0.902	0.903	0.902	0.913	0.905	0.656	0.927
United States	0.710	0.960	0.975	0.975	0.975	0.987	0.960	0.870	1.000

Source: Own elaboration based on the database of European Commission (2021)

If we take into consideration the relationship between the efficiencies calculated by CWA procedure and the efficiency of DEA, there are two types of connections.

Table 5 presents the correlation coefficients. Since they are continuous variables, i.e. the statistical scale is minimum interval scale; the Pearson correlation can be used. They show high correlations between the I-DESI and the weights of the DEA efficiencies with values above 0.8. We investigated if it is worth to determine efficiencies without determining the DEA efficiencies. Then if we consider the DEA efficiency as an ideal (virtual) DMU, these efficiencies must first be calculated, which implies solving p linear programming (LP) problems. Calculating the CWA efficiencies is the solution of an auxiliary LP, i.e. $p + 1$ LP models have to be solved. On the other hand, if we only determine the maximally achievable efficiency, i.e. the individual efficiency, the mileage from the possible efficiency, then only one (non)linear programming problem must be solved.

Table 5 Pearson correlation between the I-DESI and the six DEA/CWA efficiencies

Pearson	Manhattan	Euclidean	Chebyshev
E^*	0.823*	0.845	0.952
1	0.842	0.842	0.823

Note: * the solution of the MaxiMin model is given and not the Manhattan distance.
Source: Own elaboration

Results in Table 5 show that the Chebyshev distance is minimising the distance from the ideal efficiencies vector $\boldsymbol{E^*}$. A correlation of 0.952 is achieved, which is considered a very strong relationship. The efficiency with the biggest correlation from vector **1** gives a value of 0.842, which is also considered relatively high. The two computational efficiencies do not show significant difference from the I-DESI efficiencies. This may lead to the results that it is unnecessary to calculate all p DEA efficiencies, which may save time and costs not solving p LP problems.

Kendall's *tau*-b shows the linear relationship between the rankings (Table 6). This correlation indicates a strong stochastic (linear) relationship when greater than 0.7 and it is close to this value for the I-DESI data. The Chebyshev distance leads to the highest correlation among the six common weightings, with a value of 0.826. This also shows the greatest linear relationship with DEA efficiency.

Table 6 Kendall's *tau*-b correlation between I-DESI and the DEA/CWA models

Kendall	Manhattan	Euclidean	Chebyshev
E^*	0.669*	0.693	0.826
1	0.689	0.687	0.669

Note: * the solution of the MaxiMin model is given and not the Manhattan distance.
Source: Own elaboration

Finally, we compare the weight vectors of the six DEA models. As shown in Table 7, the MaxiMin and the Chebyshev model with the unit vector distance produced identical results, as did the Euclidean distance model and the DEA/CWA model. This means that the six models gave only four different solutions for the weight vectors. Table 7 also contains the weight vectors defined by the European Commission. While the Commission's weight vectors are balanced, in all our models the weight vector of the Connectivity dimension is the largest, with at least 60% (but in most models it receives a weight of around 90%). At the same time, the UOI dimension determines efficiency in only one model (and with a very small contribution).

Table 7 I-DESI and DEA common weights of each dimension and their normalised ratio

Weights	CNC	HUC	UOI	IDT	DPS
I-DESI	*0.25*	*0.25*	*0.15*	*0.20*	*0.15*
MaxiMin	1.247	0.000	0.000	0.000	0.108
	0.920	*0.000*	*0.000*	*0.000*	*0.080*
Chebyshev 1	1.247	0.000	0.000	0.000	0.108
	0.920	*0.000*	*0.000*	*0.000*	*0.080*
Euclidean 1	1.246	0.156	0.000	0.000	0.000
	0.889	*0.111*	*0.000*	*0.000*	*0.000*
DEA CWA	1.246	0.156	0.000	0.000	0.000
	0.889	*0.111*	*0.000*	*0.000*	*0.000*
Euclidean E*	1.243	0.145	0.000	0.006	0.005
	0.888	*0.104*	*0.000*	*0.004*	*0.004*
Chebyshev E*	0.944	0.000	0.029	0.279	0.127
	0.685	*0.000*	*0.021*	*0.202*	*0.092*

Source: Own elaboration

The weights of the model with the smallest Chebyshev distance measured by DEA efficiency are the closest to the common weights given by the European Commission, although they are much less balanced than those. Connectivity is by far the most weighted in this model as well. Nevertheless, the deviation of the weight vectors from the weights determined by the Commission does not usually have a significant effect on the ranking position of individual countries. Czechia is ranked 24th in four out of the six DEA models (in line with the Commission's ranking), and 25th in the remaining two models. Hungary is ranked 37th or 38th in five of the six models above, and 33rd in the sixth (and the Commission's ranking). Poland is ranked 38th in three models, 37th in two models, but only 40th in the sixth model and 42nd in the original ranking. Finally, Slovakia is ranked 39th in five out of the six DEA models and 37th in the remaining one and the Commission's scoring model. The larger discrepancy between the rankings of Hungary and Poland in the DEA models and in the scoring model may be attributable to the considerable, perhaps excessive, emphasis placed on the Connectivity dimension in the DEA models and their relative strengths in this domain of digital development.

3.2 Ranking with the TOPSIS method

Before the TOPSIS method is briefly presented, Table 8 summarises the advantages and disadvantages of the procedure. The description was summarised based on Shih (2022). The TOPSIS method determines the order of countries in three steps described below (readers can find a more detailed description of the method in Bánhidi and Dobos, 2021).

Table 8 Advantages and disadvantages of TOPSIS models

Advantages of TOPSIS	Disadvantages of TOPSIS
Comprehensive Comparison: TOPSIS allows decision-makers to evaluate and compare multiple alternatives when faced with various criteria. It provides a holistic view by considering all relevant factors simultaneously.	*Assumption of Monotonicity*: TOPSIS assumes that the criteria are monotonically increasing or decreasing. In practice, this may not always hold true, especially when dealing with complex and dynamic systems.
Compensatory Aggregation: Unlike non-compensatory methods that use hard cut-offs, TOPSIS allows trade-offs between criteria. A poor result in one criterion can be offset by a good result in another. This realistic modelling accounts for the complexities of real-world decision-making.	*Normalisation Challenges*: Normalisation is necessary because criteria often have incongruous dimensions. However, selecting an appropriate normalisation method can be challenging. Linear normalisation and vector normalisation are commonly used approaches.
Geometric Distance Calculation: TOPSIS calculates the geometric distance between each alternative and the ideal solution (both positive and negative). This approach ensures that the chosen alternative is close to the positive ideal solution and far from the negative ideal solution.	*Subjectivity in Weight Assignment*: Assigning weights to criteria involves subjectivity. Different decision-makers may assign different weights, leading to variations in the final rankings.
Weighted Criteria: The method accommodates weighted criteria. Decision-makers can assign weights to each criterion based on their relative importance. These weights influence the overall ranking of alternatives.	*Sensitive to Outliers*: TOPSIS is sensitive to outliers. Extreme values in the data can significantly impact the results, affecting the overall ranking.

Source: Based on Shih (2022)

Step 1: Data normalisation

Let us assume that the data of criterion i is contained in the vector $\boldsymbol{x}_i$ according to each country (based on European Commission, 2021). Then data transformation is then performed as follows:

$$y_{ji} = \frac{x_{ji} - x_j^{min}}{x_j^{max} - x_j^{min}}, \left(j = 1,2,\ldots,n;\ i = 1,2,\ldots,m \right), \tag{9}$$

where the minimum and maximum values of criterion i are x_j^{min} and x_j^{max}, n is the number of DMUs, and m denotes the number of criteria. With this determination, the calculated values of the individual criteria were converted to the interval [0,1] for every country. The values of the new vectors are y_{ji}.

Step 2: Determining the objective weights

We determine the weight of the variables using the entropy-based method (Zou et al., 2006). The conversion formula is as follows:

$$H_i = -\frac{1}{\ln(n)} \cdot \sum_{j=1}^{n} \frac{y_{ji}}{\sum_{j=1}^{n} y_{ji}} \cdot \ln\left(\frac{y_{ji}}{\sum_{j=1}^{n} y_{ji}} \right), \left(i = 1,2,\ldots,m \right). \tag{10}$$

The new weights are:

$$w_i = \frac{1 - H_i}{n - \sum_{i=1}^{m} H_i}, \left(i = 1,2,\ldots,m \right). \tag{11}$$

The weighted normalised values are z_{ji} which are equal to: $z_{ji} = w_i \cdot y_{ji}$. Then the ideal and lowest points are determined using the z_{ji} values.

Step 3: Calculating the TOPSIS efficiencies
Thirdly, based on the newly calculated information, the efficiency index is determined using the highest, the ideal (I_i) and the lowest, the nadir (N_i) points, which are determined as follows:

$$I_i = \max_{j=1,2,\ldots,n} z_{ji},\ N_i = \min_{j=1,2,\ldots,n} z_{ji},\ (i = 1,2,\ldots,m). \tag{12}$$

The distance of the j-th country from the highest ideal and the lowest nadir point is calculated as follows:

$$d_j^I = \sqrt{\sum_{i=1}^{n}\left(z_{ji} - I_i\right)^2},\ d_j^N = \sqrt{\sum_{i=1}^{n}\left(z_{ji} - N_i\right)^2},\ (j=1,2,\ldots,n). \tag{13}$$

The last calculation is to determine the E_j TOPSIS efficiencies, which represent the proportion of the distance from two specified points:

$$E_j = \frac{d_j^N}{d_j^I + d_j^N},\ (j = 1,2,\ldots,n). \tag{14}$$

We present the results of the calculations performed on the dataset using Excel 2016. For the sake of brevity, we have omitted the detailed calculations. The calculated objective weights are listed in Table 9, while the TOPSIS efficiencies themselves are listed in Table 4.

Table 9 Calculated TOPSIS weights of digital dimensions

	CNC	HUC	UOI	IDT	DPS
Weights	0.172	0.196	0.200	0.249	0.184

Source: Own elaboration

Among the weights of the dimensions, IDT's is the largest (with a weight of almost 25%), followed by UOI and HUC (with a weight of around 20%). This indicates that the most advanced countries in terms of education are ranked highest. The DPS and CNC dimensions are given slightly less weight than these. Compared with the weight vector determined by the European Commission, the most striking difference is that the Commission gave more than 20% of the weight to the dimension of basic infrastructure and competencies (CNC, HUC), while TOPSIS gave the highest weight to enterprise applications (IDT). However, it may be worth recalling that in the post-2021 editions of the DESI, the structure of the overall index has been changed, and all four (remaining) principal dimensions, including IDT, have the same weight (25%, which in this case is almost equal to the entropy-based weight). Among the V4 countries, Czechia has the highest TOPSIS ranking (24th), followed by Hungary (34th) Slovakia (37th) and Poland (42nd). These are the same as the original positions, except for Hungary, which ranks 33rd according to the I-DESI scoring model.

First, the ranking obtained with I-DESI weights is compared with the TOPSIS ranking with Kendall's *tau*-b correlation. The value of the correlation is high, with a value of 0.966. This correlation indicator is also significant. Interestingly, despite the difference in weights, the rankings are very similar.

Table 10 shows the correlation between the efficiencies calculated by CWA and the TOPSIS efficiency of using Kendall's *tau*-b linear relationship. The highest correlation (0.812) is obtained again in the case of the maximal solution of the Chebyshev distance by minimising the distance from the ***E**** vector, which shows a strong stochastic relationship.

Table 10 Kendall's *tau*-b correlation between the TOPSIS and DEA/CWA efficiencies

Kendall	Manhattan	Euclidean	Chebyshev
*E**	0.669*	0.691	0.812
1	0.686	0.685	0.669

Note: * the solution of the MaxiMin model is given and not the Manhattan distance.
Source: Own elaboration

In Table 11, the 45 countries of the database (I-DESI report) were divided into 9 groups of five countries based on the ranking of digital development evaluated based on TOPSIS efficiencies, of which the most developed (with a TOPSIS efficiency higher than 0.75) are the United States, Denmark, the Netherlands, Finland and Sweden. The least developed group includes Poland, Serbia, Croatia, Chile and Turkey. Slovakia is in the 8th group with Bulgaria, Italy, Brazil and Mexico. Hungary was placed in the 7th group, which also includes Latvia, Romania, Portugal and Greece meaning that the country and even more the companies in the country have started on the path to digital maturity. Among the V4 countries, Czechia, together with Austria, Belgium, Malta and Cyprus, is in the highest group (5th), which roughly corresponds to an average level of development among the countries in the dataset, in contrast to the other V4 countries, which are in the bottom three groups. However, it is important to emphasise that the I-DESI only includes countries that are either in the EU itself or are considered to be its competitors, and in a global context even these countries would be considered relatively advanced in terms of digital maturity.

Table 11 TOPSIS efficiencies of I-DESI countries by efficiency category

Tier	Best eff.	Worst eff.	Countries in each tier (grouped by TOPSIS efficiencies)
1	0.870	0.752	United States, Denmark, Netherlands, Finland, Sweden
2	0.737	0.656	Norway, Switzerland, Luxembourg, Iceland, United Kingdom
3	0.650	0.607	Ireland, Australia, Israel, Germany, Japan
4	0.602	0.545	Estonia, France, Canada, New Zealand, South Korea
5	0.503	0.425	Austria, Belgium, Malta, Cyprus, Czechia
6	0.412	0.351	Spain, China, Slovenia, Lithuania, Russia
7	0.327	0.307	Latvia, Romania, Portugal, Hungary, Greece
8	0.287	0.245	Bulgaria, Slovakia, Italy, Brazil, Mexico
9	0.240	0.187	Serbia, Poland, Croatia, Chile, Turkey

Source: Own elaboration

CONCLUSION

The results of our DEA models do not differ significantly from the results of the original I-DESI (scoring) model (Hungary's ranking, for example, is never worse than 38th). However, the weights are very different from those defined by the European Commission. Based on the Pearson correlations, the two different calculation efficiencies used in the models do not show significantly different solutions. This may indicate that it is unnecessary to determine all p DEA efficiencies, which may lead to time and cost savings. The highest correlation coefficient was obtained for Chebyshev distance by minimising the distance from the efficiency vector $\boldsymbol{E}^*$, both in the case of the Pearson correlation and the Kendall *tau*-b linear relationship.

Our weights calculated using the TOPSIS method are relatively balanced, like the original weight vector determined by the European Commission (in contrast to the DEA models), but they place more emphasis on the use of digital technologies for business and private purposes than the Commission which gives the greatest weight to the dimension of basic infrastructure and competences, and smaller weights to the dimensions of various applications (this has changed in the post-2021 editions of DESI). The resulting ranking closely resembles the Commission's original ranking, as determined by the I-DESI scoring model (all V4 countries except for Hungary have the same ranking). However, in the DEA rankings, Hungary is typically ranked lower than in the original model, whereas Poland is ranked higher. We posit that this is the joint consequence of the digital development profiles of these two countries being slightly disparate from their peers (despite a similar overall level) and the DEA models placing considerable emphasis on the Connectivity dimension at the expense of the other dimensions.

Although the rankings obtained were relatively similar to those generated by the Commission's original scoring model, we consider the results derived from methods with objective weights to be valuable for assessing their findings' reliability and providing a useful reference point. Compared to the other Visegrád Group (V4) countries, Czechia is relatively advanced, as it could be characterised as having an average level of development. In contrast, the other V4 countries could be considered as laggards. In a global context, however, even these countries would be considered relatively advanced in terms of digital maturity.

Based on our findings, the TOPSIS methodology appears to be more suitable than DEA-based methods for determining digital country rankings due to its more even distribution of weights. However, methods based on the DEA approach require less methodological consideration and are easier to calculate than those based on TOPSIS. Regarding future research avenues, it may be beneficial to explore hybrid models combining expert judgment with data-driven approaches for determining weights, with the aim of ensuring objectivity and relevance.

References

BÁNHIDI, Z., DOBOS, I., NEMESLAKI, A. (2020). What the overall Digital Economy and Society Index reveals: a statistical analysis of the DESI EU28 dimensions [online]. *Regional Statistics*, 10(2): 42–62. <https://doi.org/10.15196/RS100209>.

BÁNHIDI, Z., DOBOS, I. (2021). *A DEA módszertan alkalmazása rangsorolásra az EU-28 és Oroszország példáján* (Applying DEA methodology for ranking – the example of the EU-28 and Russia) [online]. *Szigma*, 52(4): 383–400.

BÁNHIDI, Z., DOBOS, I. (2021). A digitális fejlődés rangsorolása a DEA-típusú összetett indikátorok és a TOPSIS módszerével (Statistical analysis of the European Union's digital economy and society index) [online]. *Statisztikai Szemle*, 99(3): 253–265. <https://doi.org/10.20311/stat2020.2.hu0149>.

BÁNHIDI, Z., DOBOS, I. (2023). A Data Envelopment Analysis model for ranking digital development in the countries of the European Union without explicit inputs and common weights analysis [online]. *Decision Analytics Journal*, 6: 100167. <https://doi.org/10.1016/j.dajour.2023.100167>.

BÁNHIDI, Z., DOBOS, I., TOKMERGENOVA, M. (2021). Russia's Place Vis-à-Vis the EU28 Countries in Digital Development: a Ranking Using DEA-Type Composite Indicators and the TOPSIS Method [online]. In: *Digitalization, Digital Transformation and Sustainability in the Global Economy*, 135–146, Springer, Cham. <https://doi.org/10.1007/978-3-030-77340-3_11>.

BASOL, O., YALÇIN, E. (2021). How Does the Digital Economy and Society Index (DESI) Affect Labor Market Indicators in EU Countries? [online]. *Human Systems Management*, 40: 503–512. <https://doi.org/10.3233/HSM-200904>.

CHARNES, A., COOPER, W. W., RHODES, E. (1978). Measuring the efficiency of decision making units [online]. *European Journal of Operational Research*, 2(6): 429–444. <https://doi.org/10.1016/0377-2217(78)90138-8>.

CHERCHYE, L., MOESEN, W., ROGGE, N., VAN PUYENBROECK, T., SAISANA, M., SALTELLI, A., LISKA, R., TARANTOLA, S. (2008). Creating composite indicators with DEA and robustness analysis: the case of the Technology Achievement Index [online]. *Journal of the Operational Research Society*, 59(2): 239–251. <https://doi.org/10.1057/palgrave.jors.2602445>.

DESPOTIS, D. K. (2005). A reassessment of the human development index via data envelopment analysis [online]. *Journal of the Operational Research Society*, 56(8): 969–980. <https://doi.org/10.1057/palgrave.jors.2601927>.

DOBOS, I., VÖRÖSMARTY, G. (2014). Green supplier selection and evaluation using DEA-type composite indicators [online]. *International Journal of Production Economics*, 157: 273–278. <https://doi.org/10.1016/j.ijpe.2014.09.026>.

ESSES, D., SZALMÁNÉ CSETE, M. (2022). A digitális átalakulás és a fenntarthatósági átmenet összefüggéseinek értékelése az Európai Unió fővárosaiban (Assessing the links between the digital transformation and the sustainability transition in the capitals of the European Union) [online]. *Területi Statisztika*, 62(6): 683–697. <https://doi.org/10.15196/TS620603>.

EUROPEAN COMMISSION. (2021). *International Digital Economy and Society Index 2020* [online]. SMART 2019/0087, prepared for the European Commission DG Communications Networks, Content & Technology, Tech4i2. <https://ec.europa.eu/newsroom/dae/redirection/document/72352>.

EUROPEAN COMMISSION. (2022). *International Digital Economy and Society Index (I-DESI) 2022: Call for Tenders* [online]. [cit. 6.8.2024]. <https://digital-strategy.ec.europa.eu/en/funding/international-digital-economy-and-society-index-i-desi-2022>.

FERNANDEZ-CASTRO, A., SMITH, P. (1994). Towards a general non-parametric model of corporate performance [online]. *Omega*, 22(3): 237–249. <https://doi.org/10.1016/0305-0483(94)90037-X>.

IQBAL ALI, A., LERME, C. S. (1997). Comparative advantage and disadvantage in DEA [online]. *Annals of Operations Research*, 73: 215–232. <https://doi.org/10.1023/A:1018929228294>.

KAO, C., HUNG, H. (2005). Data envelopment analysis with common weights: the compromise solution approach [online]. *Journal of the Operational Research Society*, 56(10): 1196–1203. <https://doi.org/10.1057/palgrave.jors.2601924>.

KOTARBA, M. (2017). Measuring digitalization – key metrics [online]. *Foundations of Management*, 9(1): 123–138. <https://doi.org/10.1515/fman-2017-0010>.

LAITSOU, E., KARGAS, A., VAROUTAS, D. (2020). Digital Competitiveness in the European Union Era: the Greek Case [online]. *Economies*, 8(4): 85. <https://doi.org/10.3390/economies8040085>.

LIU, F. H. F., PENG, H. H. (2008). Ranking of units on the DEA frontier with common weights [online]. *Computers & Operations Research*, 35(5): 1624–1637. <https://doi.org/10.1016/j.cor.2006.09.006>.

LIU, W. B., ZHANG, D. Q., MENG, W., LI, X. X., XU, F. (2011). A study of DEA models without explicit inputs [online]. *Omega*, 39(5): 472–480. <https://doi.org/10.1016/j.omega.2010.10.005>.

MOROZ, M. (2017). The level of development of the digital economy in Poland and selected European countries: a comparative analysis [online]. *Foundations of Management*, 9(1): 175–190. <https://doi.org/10.1515/fman-2017-0014>.

OMURBEK, N., AKCAKAYA, O., AKCAKAYA. E. D. U. (2021). Integrating cluster analysis with MCDM methods for the evaluation of local agricultural production [online]. *Croatian Operational Research Review*, 2(2021): 105–117. <https://10.17535/crorr.2021.0009>.

PODINOVSKI, V. V., ATHANASSOPOULOS, A. D. (1998). Assessing the relative efficiency of decision making units using DEA models with weight restrictions [online]. *Journal of the Operational Research Society*, 49(5): 500–508. <https://doi.org/10.1057/palgrave.jors.2600543>

ROLL, Y., GOLANY, B. (1993). Alternate methods of treating factor weights in DEA [online]. *Omega*, 21: 99–109. <https://doi.org/10.1016/0305-0483(93)90042-J>.

ROUKANAS, S. (2021). Measuring Innovation of Countries [online]. *KnE Social Sciences*, 5(9): 157–189. <https://doi.org/10.18502/kss.v5i9.9892>.

SHAKTAWAT, A., VADHERA, S. (2021). Ranking of hydropower projects based on sustainability criteria in India using multicriteria decision making methods [online]. *Croatian Operational Research Review*, 12(2021): 75–90. <https://10.17535/crorr.2021.000>.

SHIH, H-S. (2022). TOPSIS Basics [online]. In: *TOPSIS and its Extensions: a Distance-Based MCDM Approach, Studies in Systems, Decision and Control*, Vol 447, Springer, Cham. <https://doi.org/10.1007/978-3-031-09577-1_2>.

SKARE, M., DE OBESSO, M. D. L. M., RIBEIRO-NAVARRETE, S. (2023). Digital transformation and European small and medium enterprises (SMEs): a comparative study using digital economy and society index data [online]. *International Journal of Information Management*, 68. <https://doi.org/10.1016/j.ijinfomgt.2022.102594>.

TARJÁNI, A. J., KALLÓ, N., DOBOS, I. (2023). Evaluation of Digital Development Based on the International Digital Economy and Society Index 2020 Data [online]. *Statistika*, 103(3): 355–373. <https://doi.org/10.54694/stat.2023.21>.

TOKMERGENOVA, M., BÁNHIDI, Z., DOBOS, I. (2021). Analysis of I-DESI dimensions of the digital economy development of the Russian Federation and EU-28 using multivariate statistics [online]. *Вестник Санкт-Петербургского университета. Экономика*, 37(2): 189–204. <https://doi.org/10.21638/spbu05.2021.201>.
TOLOO, M. (2014). *Data Envelopment Analysis with Selected Models and Applications*. SAEI, Vol. 30, Ostrava: VŠB-TU.
WORLDCAT. (2022). *International Digital Economy and Society Index 2022: Final Report* [online]. Catalogue entry [cit. 6.8.2024]. <https://search.worldcat.org/title/1347440917>.
ZOU, Z. H., YUN, Y., SUN, J. N. (2006). Entropy method for determination of weight of evaluating indicators in fuzzy synthetic evaluation for water quality assessment [online]. *Journal of Environmental Sciences*, 18(5): 1020–1023. <https://doi.org/10.1016/S1001-0742(06)60032-6>.

A Quantile Regression Modelling Approach to Study Gender Wage Gap in India

Samapriya Trivedi[1] | *University of Lucknow, Lucknow, India*
Shambhavi Mishra[2] | *University of Lucknow, Lucknow, India*

Received 5.4.2024 (revision received 30.7.2024), **Accepted** (reviewed) 10.8.2024, **Published** 14.3.2025

Abstract

The Indian labour market exhibits significant gender wage disparities, particularly among regular/salaried employees and casual workers. To study these disparities comprehensively, we present a dual-methodological approach by combining Quantile Regression (QR) and Melly-Machado-Mata (MMM) decomposition. Using secondary data from India's Periodic Labour Force Survey (PLFS) 2020–21, the study highlights the intricate interplay of various demographic, personal, and occupational characteristics on wage distributions. The findings highlight the persistence of the gender wage gap across different quantile levels for both employment types. The decomposition results reveal that discrimination significantly contributes to the wage gap, particularly at lower income levels, indicating a "sticky floor" effect for regular/salaried employees. Conversely, casual workers face a consistent wage gap across all quantiles, with discrimination remaining a crucial factor. This research highlights the robustness and precision of QR modelling and decomposition, providing a comprehensive framework for scientifically assessing the gender-based wage gap and exploring policy interventions to address these inequalities.

Keywords	***DOI***	***JEL code***
Quantile regression, decomposition, Oaxaca-Blinder, Melly-Machado-Mata, PLFS	*https://doi.org/10.54694/stat.2024.18*	*C10, C40, C50, J31*

INTRODUCTION

The persistent wage gap remains a prominent concern within labour economics, reflecting its complex nature as a multifaceted bias embedded within market mechanisms. This discrepancy in earnings, often measured by comparing wages across groups differentiated by gender, education, and other factors,

[1] Department of Statistics, University of Lucknow, Lucknow, Uttar Pradesh, 226007, India. E-mail: s.p.trivedi796@gmail.com.
[2] Department of Statistics, University of Lucknow, Lucknow, Uttar Pradesh, 226007, India. Corresponding author: e-mail: shambhavimishra.lko@gmail.com. ORCID: <https://orcid.org/0000-0002-7057-067X>.

carries a profound weight in academic inquiry and professional practice. While human capital differences and market discrimination contribute to the gap, its persistence warrants further investigation even after controlling for such factors. The gender wage gap exemplifies this complexity, highlighting the need to address its multifaceted nature. Factors such as gender, caste, religion, region, and education intertwine to depress females' earning potential, even within comparable roles and skills. India's labour market also exhibits a stark "duality" between regular or salaried and casual workers. Regular, also called salaried employment, offers stability, social security, and compliance with labour norms, while casual employment entails instability, low wages, and limited entitlements. Despite recent growth, casual workers' wages remain significantly lower (India Wage Report, 2018). Also, despite laws demanding equal pay, discriminatory practices stubbornly stand in the way of fairness for females. Further interventions are crucial to dismantle discriminatory practices and achieve true economic parity for all.

Previous analyses of wage distribution, primarily relying on models based on means, have yielded limited insights into the intricate dynamics underlying these distributions. QR, as highlighted by Roger Koenker and Gilbert Bassett (Koenker and Bassett, 1978), offers a more comprehensive approach. It reveals how the influence of several factors on wages changes across the spectrum, from the lowest earners to the highest. Unlike ordinary least squares (OLS) regression, QR estimates the entire conditional wage distribution, not just the mean. This approach enables robust analysis of skewed wage data and outliers, providing valuable insights into distributional inequality and heterogeneity. While regression models offer useful insights into the relationship between wages and several factors, they often fail to explain the observed wage differences between groups comprehensively. Decomposition techniques like the Oaxaca-Blinder (OB) method (Blinder, 1973; Oaxaca, 1973) and Machado-Mata (MM) decomposition (Machado and Mata, 2005) address this limitation by statistically dissecting these differences. These techniques divide the wage disparity into two parts: the explained component, which addresses variances in personal attributes between groups, and the unexplained component, which signifies the gap that may be attributed to discrimination or undisclosed factors. This breakdown helps us identify the wage gap and measure the impact of several factors, providing insights for addressing inequalities and discriminatory practices.

This study employs QR, enabling a comprehensive analysis of how various worker and labour market characteristics influence the entire distribution of employee wages in India. This approach transcends the limitations of traditional mean-based analysis, which only captures the average impact. Additionally, decomposition techniques are utilised to dissect the relative contributions of observable and unobservable factors to wage disparities across different quantile levels, potentially revealing a "sticky floor" and the glass ceiling effects (Das, 2018). This nuanced approach offers valuable insights into the intricate dynamics of the disparities in workers' wages for men and women in an Indian context, potentially holding broader applicability for wage analysis across other countries. Furthermore, our methodological approach facilitates optimal model selection for future studies, crucial for informing policy interventions to narrow India's persistent wage gap.

1 LITERATURE SURVEY

Since the mid-1990s, a pronounced gender wage gap in India has sparked a considerable academic interest in gender-based wage discrimination. Economists and statisticians have extensively researched labour market discrimination within the context of developing countries, particularly India.

A substantial body of research has examined India's wage structure, exploring a range of factors influencing wages and the challenges of discrimination faced by minority groups and females. Kingdon and Unni (2001) analysed 1987–88 National Sample Survey Office (NSSO) data, finding a significant disparity in education returns. Building on this, Madheswaran and Attewell (2007) emphasised occupational discrimination over wage discrimination for disadvantaged groups like scheduled tribes and castes. Agrawal (2013) used 2005 India Human Development Survey (IHDS) data to show that while endowment differences partly explained the wage gap between social groups, labour market discrimination was the primary driver. According to Chakraborty and Mukherjee (2014), a significant gap in wages between different genders in India's industries and professions using NSSO's Employment-Unemployment Survey (EUS) 2009–10 data, indicating wage discrimination against women in rural and urban areas. Sengupta and Das (2014) demonstrated that economically disadvantaged caste women and religious minorities faced greater discrimination. Duraisamy and Duraisamy (2016) analysed the NSSO data from 1983 to 2012, finding a decreasing raw wage gap and signs of gender convergence in productive characteristics. Pala and Nongspung (2022) identified slow wage growth for regular workers but noted faster long-term growth for casual workers, suggesting convergence. These studies underscore the continued existence of the gender wage gap, highlighting its variable impact across different social groups and the intricate interplay between discriminatory practices and individual characteristics. Madan and Mor (2022) analysed the persistence of the gender earnings gap in India using the PLFS 2017−18 dataset, employing Generalized Linear Models (GLM) and Analysis of Covariance to estimate marginal mean earnings, revealing a significant gender earnings gap across occupational groups and work statuses, with males earning 1.744 times more than females, even after controlling for education.

Previous studies on wages in India have relied on mean-based methodologies, overlooking the labour market's nuanced dynamics and varied composition. These approaches have disregarded extreme wage structures and failed to encompass the entirety of the wage distribution for different employment types. This research gap emphasises the necessity for an all-inclusive assessment that includes the entire workforce spectrum and addresses gender disparities within these segments. Studies have advocated the use of QR modelling to obtain a deeper insight into wage differentials, particularly those concerning gender (Fitzenberger et al., 2021; Waldmann, 2018). Khanna (2012) identified a persistent "sticky floor" effect in India's wage distribution, where gender pay gaps are more pronounced at lower income levels. Deshpande and Sharma (2015) corroborated this finding, revealing significant income disparities within India's wage structure. Their research underscored the "sticky floor" phenomenon, hindering labour market access for low-income earners while also unveiling a glass ceiling impeding women's advancement toward higher-paying positions. Azam (2012) utilised the MM procedure to scrutinise the evolution of urban wage structures in India between 1983 and 2004, drawing on NSSO data. Azam and Prakash (2015) extended this approach to investigate public-private wage differentials within India's labour market in 2004–05, again relying on NSSO data. Mitra (2016) employed Augmented Mincerian equations to analyse how education and other factors affect the salaries of different worker groups (regular/casual, male/female) in India, highlighting the interplay between education and earnings across various segments. Sengupta and Puri (2021) examined the gender pay gap using NSSO data from the same period, employing OLS and linear QR methods. While their OLS decomposition provided insights at the mean level, it was limited in distributional analysis. To overcome these constraints, our study suggests using advanced decomposition techniques, such as MM or MMM, to illuminate how this gap affects wage distribution entirely. This approach would deepen our understanding of this complex phenomenon.

Drawing upon the extensive unit-by-unit data representing the entire nation from the PLFS conducted by the Ministry of Statistics and Programme Implementation (MoSPI), our study examines gender-based wage disparities between the regular/salaried and casual employees in India. Launched in 2017, the PLFS is an annual survey implemented by the Government of India's National Statistical Office (NSO). To ensure the findings are generalisable to the national demographics, the PLFS incorporates survey weights within its analysis. This rigorous design has established the PLFS as a valuable source for investigating employment trends, income patterns, and wage disparities within the Indian labour market (Pala and Nongspung, 2022).

2 METHODS

2.1 Quantile Regression Model

Quantile regression, recognised for its robustness and flexibility compared to traditional OLS regression, is frequently employed in wage analysis to explore the explanatory variables that heterogeneously affect conditional wage distribution. It measures how explanatory variables affect a specific part of the dependent variable's distribution without assuming a particular shape for that distribution (Waldmann, 2018). QR offers several advantages, including reduced sensitivity to outliers and misspecified error distributions commonly encountered in wage data (Huang et al., 2017; Patidar et al., 2023). Additionally, it can handle situations where the error variance varies with the explanatory variables, a scenario where OLS estimates may lack reliability (Porter, 2014). Although other techniques, such as Generalized Least Squares (GLS) and sandwich estimators, offer robust alternatives for heteroscedasticity, they have limitations. GLS is a statistical method used to estimate parameters in linear regression models, particularly in the presence of heteroscedasticity or autocorrelation among residuals. By transforming the original data, GLS addresses these issues, allowing for the efficient application of OLS on the transformed data. The sandwich estimator is another robust technique for estimating the variance of parameter estimates in regression models. It is commonly used alongside GLS to provide robust standard errors that remain valid even when assumptions like homoscedasticity are violated. However, GLS has limitations, as it assumes a specific form of heteroscedasticity or correlation among residuals and necessitates data transformation to satisfy OLS assumptions. Similarly, sandwich estimators are primarily applied to adjust standard errors in mean regression models. In contrast, quantile regression presents a robust alternative, with a more comprehensive data distribution analysis, by providing insights beyond the mean, which is the focus of both GLS and sandwich estimator methods.

The equation below estimates the QR model's coefficients:

$$y_i = X_i \beta_\theta + \mu_{\theta i}, \tag{1}$$

with:

$$Q_{\theta i}\left(y_i | X_i\right) = X_i \beta\left(\theta\right), \tag{2}$$

here: y_i is ln (daily pay) and X_i is the covariates related to workers, β is the coefficient vector, θ represents the specified quantile of the wage distribution ($0 < \theta < 1$) and $\mu_{\theta i}$ is the random error term which accounts for erratic components in y_i.

The θ^{th} QR estimator, $\hat{\beta}(\theta)$ minimises over $\beta(\theta)$, for the value of β. In this case, the objective function is:

$$\theta\left(\beta\left(\theta\right)\right) = \sum_{i \in \{i: y_i \geq X_i \beta\}}^{N} \theta \left| y_i - X_i \beta_\theta \right| + \sum_{i \in \{i: y_i < X_i \beta\}}^{N} \left(1-\theta\right) \left| y_i - X_i \beta_\theta \right|. \tag{3}$$

The estimation method for the QR relies on a linear programming approach. In STATA, the "QREG" command is used for this analysis, where the minimisation problem is formulated as a linear programming problem. This approach is consistent with the methodology suggested by Armstrong et al. (1979) and comprehensively described by Koenker (2005). It uses the simplex method, which iteratively improves the objective function at each step until the optimal solution is achieved. Thus, the coefficients from the QR model reveal how various factors influence wages at different points in the wage distribution, demonstrating how these factors' effects vary across different wage levels (Mitra, 2016).

Koenker and Machado (1999) developed a quantile-specific goodness-of-fit measure for QR models. This metric addresses the limitations of traditional, global measures by assessing model fit at individual quantiles, enabling a more localised evaluation of model performance. The pseudo-R², which ranges from 0 to 1, is calculated using the Residual Absolute Sum of Weighted Differences (RASW) and the Total Absolute Sum of Squared Differences (TASW) according to the provided formula for the specific quantile (θ):

$$pseudo\, R_{\theta}^{2} = 1 - \frac{RASW_{\theta}}{TASW_{\theta}}\ , \tag{4}$$

where:

$$RASW_{\theta} = \sum_{y_i \geq X_i \hat{\beta}(\theta)} \theta \left| y_i - X_i \hat{\beta}_{\theta} \right| + \sum_{y_i < X_i \hat{\beta}(\theta)} (1-\theta) \left| y_i - X_i \hat{\beta}_{\theta} \right|,$$

$$TASW_{\theta} = \sum_{y_i \geq \theta} \theta \left| y_i - \hat{\theta} \right| + \sum_{y_i < \theta} (1-\theta) \left| y_i - \hat{\theta} \right|.$$

In the above equations, $X_i \hat{\beta}_{\theta}$ represents the predicted dependent variable for the i^{th} recording at quantile °θ and $\hat{\theta}$ denotes the estimated quantile value. $RASW_{\theta}$ differs from the absolute quantile function presented in Formula (3) as it is utilized to assess the QR model is good fit. In contrast, the absolute quantile function is integral to the optimisation process for estimating the model parameters.

It is essential to acknowledge that various methodologies exist for analysing wage distribution, each with limitations and unique features. Unconditional Quantile Regression (UQR) assesses covariates' influence on the dependent variable's unconditional quantiles, which might offer a limited understanding of the conditional distribution of wages (Adireksombat et al., 2010). The Heckman Correction Model addresses sample selection bias but is sensitive to the choice of instruments and the specification of the selection equation. GLMs primarily focus on the mean of the outcome variable, potentially overlooking significant distributional aspects of the wage gap and often relying on assumptions about the distribution of error terms, which may not always be valid (Madan and Mor, 2022). Kernel regression, being non-parametric, can model complex relationships but does not offer the same level of detail about the conditional distribution of wages as QR. Therefore, the application of QR to study the gender wage distribution among the two categories of employees is well-justified.

2.2 Oaxaca-Blinder Decomposition

The OB Decomposition method extends regression analysis by decomposing the average difference in outcome between two groups into attributable components related to group differences in independent variable endowments and the effects of these variables. This method employs OLS regressions for each group separately, predicting the dependent variable using the same explanatory variables. The decomposition then isolates the contributions of endowment and coefficient effects. Endowment (explained) effects capture the influence of disparities in the groups' average levels of explanatory variables. Conversely, coefficient (unexplained) effects isolate how the explanatory variables impact the outcome variable. However, each group's impact is measured separately based on their respective coefficients estimated in the regressions.

The approach entails calculating wage equations independently for individuals belonging to the male (m) and female (f) groups:

$$y_{gi} = \beta_{g0} + \sum_{i=1}^{p} X_{gki}\beta_{gk} + \mu_{gi} \,, \tag{5}$$

where g represents the two groups (male and female), k represents the independent variable, while all other variables maintain the same meanings as defined in Formula (1). Considering that the residuals from OLS regression have a mean of zero, the equation below calculates the discrepancy in average wages across both genders by comparing the predicted wages for each group:

$$\bar{y}_m - \bar{y}_f = \left(\hat{\beta}_{m0} + \sum_{k=1}^{p} \bar{X}_{mk}\hat{\beta}_{mk} \right) - \left(\hat{\beta}_{f0} + \sum_{k=1}^{p} \bar{X}_{fk}\hat{\beta}_{fk} \right). \tag{6}$$

The assumption that the non–discriminatory wage framework applied to males was used to construct a counterfactual (CF) average wage for females using the coefficients estimated for males:

$$CF_f = \hat{\beta}_{m0} + \sum_{k=1}^{p} \bar{X}_{fk}\hat{\beta}_{fk} \,. \tag{7}$$

We can estimate the effect of discrimination on wages by creating a hypothetical scenario where male employees earn the same as female employees for observable characteristics. Now, adding and subtracting Formula (7) from Formula (6), we get:

$$\bar{y}_m - \bar{y}_f = \left(\hat{\beta}_{m0} - \hat{\beta}_{f0} \right) + \sum_{k=1}^{p} \bar{X}_{fk} \left(\hat{\beta}_{mk} - \hat{\beta}_{fk} \right) + \sum_{k=1}^{p} \left(\bar{X}_{mk} - \bar{X}_{fk} \right) \hat{\beta}_{mk} \,. \tag{8}$$

The first two terms decompose the coefficient effects, reflecting wage discrimination through discrepancies in returns for each gender's attributes (separate regression coefficients). These disparities create a wage gap despite controlling for average covariate levels. The final term isolates the unexplained portion due to the unequal distribution of individual characteristics across genders. It captures the average log wage difference attributable to gender differences in average covariate levels (Deshpande et al., 2017).

2.3 Melly-Machado-Mata Decomposition

This study used Melly's refined MM methodology to achieve a quantile-specific decomposition of the gender-based employment wage gap. This approach breaks down the observed disparity into components specific to each wage distribution quantile. Doing so separates the influence of individual worker characteristics from the impact of wage structures associated with those characteristics. This facilitates an in-depth wage disparity analysis across the whole wage distribution for male and female employees, which moves beyond the limitations of mean-focused decomposition techniques, revealing the dynamic nature of the disparity at diverse wage spectrum quantiles (Azam, 2012).

From Formula (2), for each group, the conditional quantile function can be stated as:

$$Q_\theta \left(y_{gi} | X_{gi} \right) = X_{gi}\beta_g \left(\theta \right); \; \theta \epsilon \left(0.1 \right). \tag{9}$$

In accordance with MM methodology, the following steps outline the decomposition process:

i. Generate a random sample from a uniform distribution $U[0.1]$. This step leverages the probability integral transformation theorem, which asserts that if U is a uniform random variable on $[0.1]$, then $F^{-1}(U)$ follows the distribution F. By applying the inverse cumulative distribution function (CDF) F^{-1} of the wage distribution to these uniform random variables, the transformation $F^{-1}(\theta_i)$ yields the conditional quantiles of wages. This process effectively simulates a sample from the estimated conditional salary distribution, given a set of covariates.
ii. Estimate n distinct QR coefficient vectors for males and females separately.
iii. Draw independent random samples with replacement from the covariate distributions of males and females.
iv. Construct the counterfactuals by multiplying various combinations of estimated quantile coefficients and respective covariate distributions across genders, i.e., $y_j^{cf} = \tilde{X}_j^m \beta_{\mu j}^f$.

The final decomposition model is as follows:

$$\hat{Q}_m(\theta)-\hat{Q}_f(\theta)=\left(\hat{Q}_m(\theta)-\hat{Q}_{cf}(\theta)\right)+\left(\hat{Q}_{cf}(\theta)-\hat{Q}_f(\theta)\right). \tag{10}$$

The observed disparity can be divided into two components: the characteristics (explained) component, shown as the initial term on the right, measures how much of the difference can be accounted for by characteristic variations. It is calculated as the difference between the quantile regression estimates for males and the counterfactual distribution, which represents what the female wage distribution would be if females had the same characteristics as males but were paid according to the male wage structure. The second component is the coefficients (unexplained), which describe the remainder of the difference that cannot be explained by the measured characteristics and represent discrimination or bias, indicating that even if females had the same characteristics as males, they would still face a wage gap due to differences in how these characteristics are valued in the labour market (Deshpande et al., 2017; Khanna, 2012). As specified in Formula (10), the MM decomposition model utilises coefficient estimates obtained from quantile regression for both male and female employees at selected quantiles of the wage distribution. In contrast, the OB model, outlined in Formula (8), employs a linear regression framework that concentrates on the mean of the wage distribution. While both decompositions are used to analyse the gender wage gap, the OB model emphasises mean differences, whereas the MM decomposition, with its use of quantile-specific coefficient estimates, provides a more comprehensive analysis across the entire distribution, capturing variations at different quantiles.

The MM decomposition provides valuable insights into wage inequality. However, its reliance on computationally expensive Monte Carlo simulations for counterfactual wage estimation is a significant limitation. Melly proposed an alternative quantile-based decomposition approach that addresses this limitation by directly integrating the conditional wage distribution across the relevant variable space, eliminating the need for computationally expensive simulations. This methodological improvement allows for a more efficient and statistically robust analysis of wage inequality. Melly's framework disaggregates wage inequality at specific wage distribution quantiles into distinct characteristic and coefficient components. The characteristic component captures disparities that can be credited to differences in worker features, while the coefficients component isolates disparities in wage returns due to other factors. Crucially, the quantile-based approach demonstrates convergence to the MM results under the assumption of infinitely many simulations, confirming its accuracy and computational efficiency.

For alternative approaches, the following suggestions could be considered for different studies within the same domain. One method proposed by Ñopo (2008) involves decomposing the wage gap by comparing individuals with similar characteristics. This method overtly explains the differences in the supports of the characteristics' distribution. However, it may encounter issues related to dimensionality and the availability of suitable instruments (Ñopo et al., 2012). Additionally, copula analysis within the decomposition framework can enhance the accuracy and flexibility of gender wage gap analysis by effectively modelling the dependence structure between variables and adjusting for sample selection bias. This approach is particularly useful for addressing sample selection issues and providing a more flexible and accurate decomposition of wage gaps across different quantiles. This is crucial for accurately estimating the wage distribution, especially when dealing with non-random selection into the labour force (Arellano and Bonhomme, 2017; Biewen and Erhardt, 2021).

2.4 Gini coefficient of inequality

The Gini coefficient, a statistical tool, is the most utilised measure of inequality. It effectively summarises the extent of a population's income disparity. An alternative method to calculate the Gini coefficient is the relative mean absolute difference, which is complementary to the traditional Lorenz curve-based method. However, the most used formula for the Gini coefficient, introduced by Stephen P. Jenkins in 1999, is more computationally efficient and captures the same concept as the relative mean absolute difference by utilising the ranks of incomes (Jenkins, 1999). This formula, employed in our study, is given by:

$$G = 1 + \frac{1}{N} - \frac{2}{aN^2}\sum_{i=1}^{n}\left(N - i + 1\right)w_i, \tag{11}$$

here: N represents the total number of employees, w is the i^{th} and j^{th} individuals' incomes, and a is the arithmetic mean of the income (Gazeley et al., 2018). The Gini coefficient, which measures relative rather than absolute inequality, spans from 0 to 1. A value of 0 suggests perfect equality, and a 1 signifies complete inequality.

3 RESULTS

3.1 Data description

This research utilised Unit-Level data from the PLFS Schedule 10.4 (first visit) for the period between July 2020 and June 2021, conducted by the NSO and obtained from MoSPI, to evaluate the gender-based wage gap in India's regular/salaried and casual-basis employees. Worker classifications followed the NSSO activity status definitions, i.e., regular/salaried employees receive fixed wages (not based on daily contracts), while casual workers work irregularly and are paid per day or contract. Self-employment was excluded due to challenges in separating profit and wage components.

We analysed factors influencing the distribution of daily wages of workers in both genders aged 15–59 in regular/salaried and casual employment in India. We used the natural logarithm of daily wage, derived from weekly salary and days worked reported in the NSO, as the dependent variable. To understand the effects on wage distribution across the spectrum, we systematically assessed predictor impacts at five quantiles (10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th}) for both regular/salaried employees and casual workers. This approach revealed key distributional characteristics, such as shape, spread, and central tendency, facilitating a comprehensive analysis of associated differentials and enhancing study clarity and consistency.

The QR models and decomposition of the wage gap for the regular/salaried employees and casual workers included various individual characteristics, such as age, residential location, social group, marital status, educational levels (general and technical), and occupational characteristics (types of occupation and industry). Age and its squared term (divided by 100) were included to capture non-linear effects on wages (Bai and Veall, 2023; Si and Li, 2023). The National Classification of Occupations-2004 from the Directorate General of Employment and Training was aggregated into three major categories: white-collar (NCO 1 to 4), blue-collar (NCO 5 to 9), and agricultural (NCO 6) occupations (Hnatkovska et al., 2012). Similarly, the National Industrial Classification (MoSPI, 2008) was further grouped into five major types with NIC codes: Production and Extraction (NIC 1 to 3), Infrastructure and Utilities (NIC 4 to 6), Goods and Service Distribution (NIC 7 to 9), Knowledge and Service-Based (NIC 10 to 15), and Public and Social Services (NIC 16 to 20). These classifications encompass most regular/salaried and casual employment occupations and industries in India. The categorical variables were used as dummy variables in the regression modelling approach, with one category as the reference (Alkharusi, 2012). All analyses used Stata v. 13 (StataCorp, 2013) on Windows x64, incorporating sample weights for population representativeness.

After data pre-processing to eliminate missing wage/salary entries, 63 704 respondents (23.92% female, 76.08% male) were retained for model fitting. The table in the Annex presents the detailed distribution of male and female workers across various characteristics. Daily wage disparity persisted between genders across both regular/salaried employees and casual worker categories. Male regular/salaried workers earned a significantly higher average daily wage (Rs. 682.84) than females (Rs. 540.62). A similar disparity was observed for casual workers, with males earning Rs. 357.68 on average and females earning Rs. 225.47. A Mann-Whitney test showed significant gender wage disparities among India's regular/salaried and casual workers (p-value<0.001). Regular employees had a Gini coefficient of 0.450, indicating moderate income inequality. In contrast, casual workers had a more equitable distribution with a Gini coefficient of 0.234. Female regular employees had a higher wage disparity (Gini = 0.509) than male regular employees (Gini = 0.425). This difference was not seen among casual workers, where males (Gini = 0.210) and females (Gini = 0.201) had lower Gini coefficients than regular employees, suggesting a more equitable wage distribution, especially for females.

3.2 Wage effects across quantiles

QR models were separately estimated for male and female regular/salaried employees, as presented in Tables 1 and 2, and for male and female casual workers, as presented in Tables 3 and 4. The models utilised the daily wage's natural logarithm as the dependent variable and included the same independent predictors. Applying a logarithmic transformation to daily wages ensures that the estimated coefficients represent the change (percentage) in the daily wages log for a one-unit alteration in a continuous independent variable and the percentage difference in the log of daily wages between the reference category and the category in question for categorical independent variables. They were exponentiated to revert the coefficients to the original scale (e^{β}). The exponentiated coefficients were then interpreted as the percentage increase, as $(e^{\beta}-1)\times100\%$, in the dependent variable attributable to the dummy variable. In cases where there was a decrease, it was interpreted as a percentage decrease, disregarding the sign. This method ensures that the interpretation of the coefficients remains consistent with the original scale of the dependent variable, thereby providing meaningful insights into the effects of the independent variables across different quantiles of the wage distribution.

The wage-age analysis revealed a concave relationship for all workers, with wages initially increasing before decelerating with age. This trend was more pronounced for females and in higher wage quartiles. The wage-age profile also varied across the wage distribution, particularly for male casual workers in the top half. Furthermore, the study found a significant urban wage premium for regular employees. Males experienced a 23.9% increase, while females had a 26.4% increase at the 10th percentile, which reversed at the 90th percentile to 13.5% (males) and 21.9% (females). Conversely, casual workers had a lower raw wage premium, ranging from 8.0% to 22.8% for males and 9.6% to 18.8% for females at higher percentiles. Regular employees, especially females, gained more across the wage spectrum than male casual workers. These findings highlight the gender-specific impact of urbanisation on wages across different wage levels and job types, supporting prior research on rural wage disparities, particularly for females (Dutta, 2006; Khanna, 2012).

The analysis revealed persistent wage disparities across caste groups. Scheduled Tribes (STs), Scheduled Castes (SCs), and Other Backward Classes (OBCs) earned less compared to the 'Others' category. Male STs experienced a 13.7% wage disadvantage among regular employees, while female STs displayed a surprising 19.1% wage advantage at the lowest income levels. SCs faced the most significant gaps, with male incomes lagging 5% to 10% and female incomes falling behind 6% to 11%. OBCs exhibited minor discrepancies. In the casual workforce, STs were the most disadvantaged, with male and female incomes falling behind the 'Others' category by 22.9% to 9.4% and 11% to 33.6% from the lower-to-upper wage spectrum, respectively. This suggests that STs faced greater disparities in casual employment, while SCs encountered more challenges in regular jobs. OBCs performed comparatively better in both job types. These findings underscore the persistence of wage gaps for socially disadvantaged groups in India, highlighting the complex interplay between worker type, gender, and income level, aligning with Madheswaran and Attewell's (2007) findings, who observed similar disparities among reserved categories.

Table 1 Quantile Regression Model for male regular/salaried employees

Independent variables	Percentiles				
	θ = 0.10	θ = 0.25	θ = 0.50	θ = 0.75	θ = 0.90
Age	0.049*** (0.01)	0.026*** (0.01)	0.015** (0.01)	0.014*** (0.00)	0.024*** (0.01)
$Age^2/100$	–0.053*** (0.01)	–0.018* (0.01)	0.002 (0.01)	0.008 (0.01)	–0.000 (0.01)
Residential area (rural)					
Urban	0.214*** (0.02)	0.191*** (0.02)	0.177*** (0.01)	0.152*** (0.01)	0.127*** (0.02)
Social group (others)					
STs	–0.147** (0.05)	–0.080* (0.03)	–0.032 (0.02)	0.011 (0.02)	0.038 (0.03)
SCs	–0.104*** (0.03)	–0.065** (0.02)	–0.060*** (0.02)	–0.055** (0.02)	–0.057** (0.02)
OBCs	–0.053* (0.02)	–0.055*** (0.02)	–0.052** (0.02)	–0.045** (0.02)	–0.035 (0.02)
Marital status (never married)					
Currently married	0.167*** (0.03)	0.153*** (0.02)	0.135*** (0.02)	0.132*** (0.02)	0.098*** (0.02)
Widowed	0.076 (0.04)	–0.030 (0.12)	0.006 (0.08)	0.033 (0.04)	–0.017 (0.07)
Divorced/separated	–0.117 (0.24)	–0.025 (0.15)	–0.117*** (0.03)	–0.063 (0.13)	0.013 (0.23)
General education (illiterate)					
No formal schooling	0.150* (0.06)	0.047 (0.05)	0.183*** (0.04)	–0.032 (0.04)	–0.253 (0.48)
Up to primary school	0.022 (0.05)	0.059 (0.04)	0.082** (0.03)	0.087*** (0.03)	0.141** (0.04)
Middle school	0.133** (0.05)	0.158*** (0.04)	0.241*** (0.02)	0.244*** (0.02)	0.286*** (0.03)
Secondary school	0.281*** (0.05)	0.280*** (0.04)	0.332*** (0.03)	0.364*** (0.03)	0.399*** (0.03)
Higher secondary	0.274*** (0.05)	0.286*** (0.04)	0.393*** (0.03)	0.454*** (0.03)	0.529*** (0.03)
Graduate	0.439*** (0.05)	0.477*** (0.04)	0.613*** (0.03)	0.693*** (0.03)	0.708*** (0.03)
Postgraduate & above	0.598*** (0.06)	0.766*** (0.06)	0.903*** (0.04)	0.940*** (0.04)	0.951*** (0.04)
Technical education (not received)					
Have technical edu.	0.254*** (0.04)	0.235*** (0.03)	0.236*** (0.03)	0.191*** (0.02)	0.249*** (0.03)
Occupation (agricultural)					
White-collar	0.225 (0.16)	0.383*** (0.04)	0.352*** (0.03)	0.322*** (0.09)	0.330*** (0.03)
Blue-collar	0.038 (0.16)	0.110** (0.04)	0.052* (0.02)	0.026 (0.08)	0.033 (0.02)
Industry (public & social services)					
Production & extraction	0.260*** (0.04)	0.176*** (0.03)	0.005 (0.03)	–0.020 (0.02)	–0.043 (0.02)
Infrastructure & utilities	0.378*** (0.04)	0.264*** (0.04)	0.130*** (0.04)	0.097** (0.03)	0.167*** (0.05)
Goods & services dist.	0.171*** (0.04)	0.105*** (0.03)	–0.056* (0.03)	–0.082*** (0.02)	–0.058* (0.03)
Knowledge & service based	0.323*** (0.04)	0.285*** (0.03)	0.190*** (0.03)	0.220*** (0.02)	0.254*** (0.02)
Intercept	3.720*** (0.21)	4.375*** (0.11)	4.942*** (0.10)	5.262*** (0.11)	5.285*** (0.10)
pseudo-R^2	0.133	0.181	0.268	0.348	0.343

Notes: * for p<.05, ** for p<.01, *** for p<.001. Standard errors are reported in parentheses, estimated using 300 iterations. Reference categories for independent variables are indicated in parentheses.
Source: Author's calculation based on PLFS (2020–21) dataset using Stata v.13

Table 2 Quantile Regression Model for female regular/salaried employees

Independent variables	Percentiles				
	θ = 0.10	**θ = 0.25**	**θ = 0.50**	**θ = 0.75**	**θ = 0.90**
Age	0.082*** (0.02)	0.068*** (0.01)	0.044*** (0.01)	0.011 (0.01)	0.015 (0.01)
$Age^2/100$	−0.096*** (0.02)	−0.076*** (0.02)	−0.040** (0.01)	0.005 (0.02)	0.009 (0.02)
Residential area (rural)					
Urban	0.234*** (0.04)	0.282*** (0.03)	0.304*** (0.03)	0.254*** (0.03)	0.198*** (0.03)
Social group (others)					
STs	0.175** (0.06)	0.037 (0.06)	0.028 (0.04)	−0.011 (0.06)	−0.010 (0.11)
SCs	−0.064 (0.06)	−0.118*** (0.03)	−0.085* (0.04)	−0.067* (0.03)	−0.122** (0.05)
OBCs	0.030 (0.04)	−0.076* (0.03)	−0.033 (0.03)	−0.093*** (0.03)	−0.154*** (0.04)
Marital status (never married)					
Currently married	0.000 (0.05)	0.120*** (0.04)	0.160*** (0.04)	0.226*** (0.04)	0.097* (0.04)
Widowed	0.150* (0.07)	0.228*** (0.06)	0.258*** (0.06)	0.316*** (0.05)	0.166** (0.06)
Divorced/separated	−0.030 (0.10)	0.055 (0.07)	−0.102 (0.07)	0.178 (0.19)	0.084 (0.14)
General education (illiterate)					
No formal schooling	0.229 (0.19)	−0.149 (0.11)	0.373*** (0.04)	0.015 (0.79)	0.169 (0.14)
Up to primary school	0.164 (0.09)	0.094* (0.04)	0.170*** (0.05)	0.160* (0.08)	0.187 (0.10)
Middle school	0.242** (0.09)	0.357*** (0.04)	0.332*** (0.04)	0.305*** (0.06)	0.300** (0.11)
Secondary school	0.488*** (0.08)	0.554*** (0.05)	0.552*** (0.05)	0.494*** (0.05)	0.498*** (0.12)
Higher secondary	0.644*** (0.10)	0.647*** (0.04)	0.701*** (0.06)	0.730*** (0.08)	0.793*** (0.12)
Graduate	0.853*** (0.09)	0.934*** (0.06)	1.055*** (0.06)	1.163*** (0.07)	1.159*** (0.11)
Postgraduate & above	1.126*** (0.10)	1.280*** (0.08)	1.508*** (0.07)	1.500*** (0.07)	1.377*** (0.11)
Technical education (not received)					
Have technical edu.	0.318*** (0.07)	0.328*** (0.05)	0.328*** (0.04)	0.267*** (0.04)	0.255*** (0.06)
Occupation (agricultural)					
White-collar	0.192 (0.51)	0.228 (0.45)	0.129 (0.13)	0.146 (0.08)	0.375* (0.15)
Blue-collar	−0.041 (0.51)	0.027 (0.45)	−0.148 (0.13)	−0.200** (0.08)	−0.046 (0.15)
Industry (public & social services)					
Production & extraction	0.648*** (0.05)	0.528*** (0.03)	0.465*** (0.03)	0.362*** (0.03)	0.274*** (0.03)
Infrastructure & utilities	0.492*** (0.04)	0.376** (0.13)	0.692*** (0.11)	0.603*** (0.08)	0.479*** (0.04)
Goods & services dist.	0.429*** (0.06)	0.437*** (0.04)	0.386*** (0.04)	0.299*** (0.04)	0.241*** (0.04)
Knowledge & service based	0.413*** (0.04)	0.486*** (0.04)	0.500*** (0.04)	0.406*** (0.03)	0.357*** (0.04)
Intercept	2.317*** (0.61)	2.768*** (0.52)	3.607*** (0.24)	4.647*** (0.25)	4.825*** (0.28)
pseudo-R^2	0.211	0.238	0.298	0.372	0.353

Notes: * for p<.05, ** for p<.01, *** for p<.001. Standard errors are reported in parentheses, estimated using 300 iterations. Reference categories for independent variables are indicated in parentheses.
Source: Author's calculation based on PLFS (2020–21) dataset using Stata v.13

Our study also identified a marriage premium for married regular employees of both genders compared to unmarried individuals. The premium was more pronounced for low-income males, decreasing from 18.2% at the 10th percentile to 10.3% at the 90th. Married females also experienced a wage advantage, ranging from 12.7% to 25.4%, with a higher premium observed in the upper-income bracket. Conversely, widowed and divorced/separated individuals did not experience significant effects, except for widowed females, who saw a larger premium, ranging from 16.2% to 37.2%. This suggests a potential role of job security in the form of wage advantage for widowed females. Interestingly, among casual workers, marital status had the opposite effect. Married males experienced a wage decrease, ranging from 4.1% to 15.2% at higher percentiles, while married females saw a positive effect at the same percentiles. These findings support Spence's (1974) theory that marriage, especially for employees in regular/salaried employment, can signal stability and lead to a wage premium, with variations based on gender and employment type.

QR models revealed a significant positive effect of education levels on wages. Individuals with higher education, including those with informal schooling, consistently earned more than those without formal education. Among regular/salaried workers, male higher-secondary school graduates enjoyed a wage premium of 31.5% to 69.7%, while females experienced a larger premium of 90.4% to 121.0%. Postgraduates had even greater advantages, with males experiencing a premium of 81.8% to 158.8% and females a premium of 208.3% to 296.3%. Education also positively affected casual workers, primarily influencing male wages, while only informal education significantly impacted female casual workers' wages. Among male casual workers, postgraduate degrees offered the highest advantage, ranging from 10% to 26.6%. These findings underscore the significance of education for regular/salaried employees, particularly females who benefit significantly. However, for casual workers, especially females, educational attainment had minimal wage impact, suggesting a need for further research and targeted interventions. This aligns with Chakraborty and Mukherjee's (2014) study, which found that education equalised wages for both genders, enabling females to earn more. In regular employment, technical education increased earnings for both genders, particularly lower-income females. However, this trend was less pronounced for casual workers, where only males significantly benefited from technical education.

The study revealed significant wage variations across occupations, employment types, and genders. For regular/salaried employees, white-collar occupations had a more substantial and consistent positive impact on males across percentiles, with 46.7% at the 25th percentile decreasing to 39.1% at the 90th percentile, compared to females, where the effect was significant only at the 90th percentile. Blue-collar occupations showed a negative impact of 18.1% for females at the 75th percentile, while the impact was more negligible and less consistent for males, with 11.6% at the 25th percentile.

Among casual workers, males experienced significant negative impacts from both white-collar (19.3% at the 10th percentile) and blue-collar occupations, particularly at the lower (19.3% at the 10th and upper (20.7% at the 90th) percentiles. For females, the impact of white-collar occupations was mixed, with a positive impact at the 10th percentile (24.0%) and 90th percentile (23.5%) but negative at the median (20.0%), while blue-collar occupations consistently showed negative impacts, with significant decreases at the 25th and 75th percentiles (16.9% and 14.2%, respectively). Gender differences in the two occupational categories' impact on earnings compared to agricultural occupations were evident in both employment categories. Regular/salaried males benefited more from white-collar occupations than females, while casual males faced more negative impacts from both occupation types compared to females.

Table 3 Quantile Regression Model for male casual workers

Independent variables	Percentiles				
	θ = 0.10	**θ = 0.25**	**θ = 0.50**	**θ = 0.75**	**θ = 0.90**
Age	0.010*** (0.00)	0.014*** (0.00)	0.027*** (0.00)	0.031*** (0.00)	0.043*** (0.00)
$Age^2/100$	−0.012*** (0.00)	−0.016*** (0.00)	−0.029*** (0.00)	−0.030*** (0.01)	−0.041*** (0.01)
Residential area (rural)					
Urban	0.077*** (0.02)	0.074*** (0.01)	0.124*** (0.01)	0.163*** (0.01)	0.205*** (0.01)
Social group (others)					
STs	−0.260*** (0.01)	−0.170*** (0.01)	−0.137*** (0.02)	−0.119*** (0.03)	−0.099*** (0.03)
SCs	−0.004 (0.01)	0.006 (0.01)	−0.001 (0.01)	−0.026 (0.02)	−0.032 (0.03)
OBCs	−0.024*** (0.01)	0.000 (0.01)	0.024 (0.01)	0.056** (0.02)	0.081** (0.03)
Marital status (never married)					
Currently married	−0.003 (0.01)	0.002 (0.01)	−0.042* (0.02)	−0.084*** (0.02)	−0.165*** (0.02)
Widowed	−0.008 (0.02)	−0.040 (0.08)	−0.113 (0.06)	−0.060 (0.08)	−0.186*** (0.03)
Divorced/separated	−0.312*** (0.03)	−0.113 (0.18)	−0.011 (0.04)	−0.087* (0.04)	−0.079* (0.04)
General education (illiterate)					
No formal schooling	0.005 (0.04)	−0.072 (0.08)	0.022 (0.09)	−0.060* (0.03)	0.012 (0.08)
Up to primary school	0.005 (0.01)	0.017* (0.01)	0.040** (0.01)	0.092*** (0.02)	0.120*** (0.02)
Middle school	0.018 (0.01)	0.035*** (0.01)	0.058*** (0.01)	0.133*** (0.02)	0.172*** (0.02)
Secondary school	0.030** (0.01)	0.075*** (0.01)	0.114*** (0.01)	0.187*** (0.02)	0.216*** (0.03)
Higher secondary	0.021** (0.01)	0.025 (0.02)	0.072*** (0.02)	0.160*** (0.02)	0.197*** (0.04)
Graduate	0.042*** (0.01)	0.087*** (0.01)	0.094* (0.04)	0.127** (0.04)	0.157** (0.05)
Postgraduate & above	0.100 (0.15)	0.025 (0.03)	0.098*** (0.02)	0.264*** (0.05)	0.266*** (0.04)
Technical education (not received)					
Have technical edu.	0.131 (0.07)	0.190*** (0.02)	0.292*** (0.07)	0.407*** (0.06)	0.300*** (0.05)
Occupation (agricultural)					
White-collar	−0.215*** (0.04)	−0.085 (0.05)	−0.101 (0.12)	0.063 (0.10)	0.021 (0.04)
Blue-collar	−0.215*** (0.01)	−0.116*** (0.03)	−0.104* (0.04)	−0.178* (0.09)	−0.232*** (0.02)
Industry (public & social services)					
Production & extraction	−0.118*** (0.01)	−0.178** (0.05)	−0.150** (0.06)	−0.101*** (0.02)	0.001 (0.07)
Infrastructure & utilities	0.134*** (0.01)	0.073 (0.05)	0.054 (0.05)	0.077*** (0.02)	0.181* (0.07)
Goods & services dist.	0.037* (0.02)	0.023 (0.05)	0.012 (0.06)	0.064 (0.03)	0.161* (0.07)
Knowledge & service based	−0.108 (0.22)	−0.013 (0.04)	−0.005 (0.09)	−0.045 (0.04)	−0.045 (0.08)
Intercept	5.417*** (0.05)	5.443*** (0.07)	5.356*** (0.09)	5.449*** (0.11)	5.364*** (0.10)
pseudo-R^2	0.101	0.125	0.110	0.085	0.128

Notes: * for p<.05, ** for p<.01, *** for p<.001. Standard errors are reported in parentheses, estimated using 300 iterations. Reference categories for independent variables are indicated in parentheses.
Source: Author's calculation based on PLFS (2020–21) dataset using Stata v.13

Table 4 Quantile Regression Model for female casual workers

Independent variables	Percentiles				
	θ = 0.10	θ = 0.25	θ = 0.50	θ = 0.75	θ = 0.90
Age	0.008 (0.01)	0.002 (0.00)	−0.000 (0.02)	0.017** (0.01)	0.010 (0.01)
$Age^2/100$	−0.013 (0.01)	−0.003 (0.01)	0.000 (0.02)	−0.026*** (0.01)	−0.013 (0.01)
Residential area (rural)					
Urban	−0.003 (0.02)	0.092** (0.03)	0.182 (0.15)	0.172*** (0.02)	0.149*** (0.01)
Social group (others)					
STs	−0.117* (0.06)	−0.015 (0.06)	−0.105 (0.17)	−0.214*** (0.04)	−0.409*** (0.06)
SCs	0.033 (0.04)	0.090 (0.06)	0.000 (0.08)	−0.148*** (0.02)	−0.275*** (0.05)
OBCs	−0.054 (0.05)	−0.006 (0.06)	−0.000 (0.07)	−0.153*** (0.02)	−0.266*** (0.05)
Marital status (never married)					
Currently married	−0.031 (0.04)	−0.003 (0.04)	0.000 (0.09)	0.057** (0.02)	0.092** (0.03)
Widowed	−0.043 (0.05)	−0.006 (0.05)	0.000 (0.10)	0.034 (0.03)	0.080* (0.04)
Divorced/separated	−0.000 (0.04)	−0.008 (0.09)	0.000 (0.16)	0.053 (0.11)	0.235 (0.25)
General education (illiterate)					
No formal schooling	0.279 (0.16)	0.094*** (0.03)	−0.118 (.)	−0.291** (0.09)	−0.176 (0.18)
Up to primary school	−0.049* (0.02)	−0.005 (0.02)	−0.000 (0.05)	−0.010 (0.02)	0.010 (0.01)
Middle school	−0.027 (0.03)	−0.004 (0.03)	−0.000 (0.05)	−0.055 (0.04)	−0.003 (0.01)
Secondary school	−0.091 (0.11)	−0.012 (0.04)	−0.000 (0.10)	0.000 (0.02)	0.021 (0.08)
Higher secondary	−0.036 (0.06)	−0.001 (0.07)	−0.000 (0.12)	−0.006 (0.14)	0.108 (0.07)
Graduate	0.021 (0.07)	0.088 (0.15)	−0.223 (0.15)	−0.107 (0.12)	−0.085 (0.12)
Postgraduate & above	0.277 (1.13)	0.173 (.)	−0.105 (0.78)	−0.178 (0.62)	−0.731 (1.75)
Technical education (not received)					
Have technical edu.	0.001 (0.52)	−0.006 (0.17)	0.329 (0.82)	0.172 (0.18)	0.809 (1.03)
Occupation (agricultural)					
White-collar	0.215 (0.87)	−0.094 (0.09)	−0.223 (0.18)	−0.123 (0.46)	0.211 (0.23)
Blue-collar	−0.077 (0.05)	−0.185*** (0.03)	−0.223 (0.18)	−0.153*** (0.02)	−0.004 (0.02)
Industry (public & social services)					
Production & extraction	0.338*** (0.07)	−0.001 (0.11)	−0.000 (0.05)	−0.162 (0.10)	−0.157*** (0.02)
Infrastructure & utilities	0.449*** (0.06)	0.185 (0.11)	0.223 (0.13)	0.057 (0.10)	0.108** (0.04)
Goods & services dist.	0.550*** (0.08)	0.181 (0.12)	0.318** (0.11)	0.102 (0.10)	0.098 (0.12)
Knowledge & service based	0.746 (0.41)	0.414* (0.17)	0.247 (0.58)	−0.025 (0.13)	−0.169* (0.09)
Intercept	4.628*** (0.16)	5.185*** (0.16)	5.521*** (0.29)	5.671*** (0.15)	5.861*** (0.15)
pseudo-R^2	0.021	0.049	0.033	0.072	0.111

Notes: * for p<.05, ** for p<.01, *** for p<.001. Standard errors are reported in parentheses, estimated using 300 iterations. Reference categories for independent variables are indicated in parentheses.
Source: Author's calculation based on PLFS (2020–21) dataset using Stata v.13

The public and social services sectors served as a reference when comparing industry types for employees. Among male regular/salaried employees, the impact of working in the Production and Extraction industries was only significant at lower percentiles, i.e., 29.7% at the 10th and 19.2% at the 25th percentiles. For females, the impact remained consistently positive throughout the wage distribution, starting at 91.2% at the 10th percentile, which decreased to 31.5% at the 90th percentile. For males in the Infrastructure & Utilities industries, the effect was positive across all percentiles, beginning at 45.9% at the 10th percentile, which decreased to 18.2% at the 90th percentile. Similarly, the effect was positive for females throughout the wage distribution, with an exceptionally positive impact of 99.8% at the median. In the Goods & Services Distribution industries, the impact for males was positively higher at lower percentiles (18.6% at the 10th percentile) but became negative at higher percentiles, reaching –5.6% at the 90th percentile. This suggested that males earned more than those in Public & Social Services at lower percentiles but less at higher percentiles. This effect remained consistently positive for females, starting at 53.6% at the 10th percentile but decreasing to 27.3% at the 90th percentile. For males in Knowledge & Service Based industries, the impact was positive across all percentiles, starting at 38.1% at the 10th percentile and reaching 28.9% at the 90th percentile. The effect was also positive for females, starting at 51.1% at the 10th percentile and decreasing to 42.9% at the 90th percentile. These results indicated that both male and female regular/salaried employees in these industries earned more than those in Public & Social Services, with variations across percentiles. Females tended to experience higher positive impacts compared to males, particularly in the Production & Extraction and Goods & Services Distribution industries. For male casual workers in the Production & Extraction industries, effects were negative across all percentiles, with the most significant decline of 16.3% at the 25th percentile and a slight positive effect of 0.1% at the 90th percentile. The effects were positive for females, with the highest increase of 40.2% at the 10th percentile and a decrease of 14.5% at the 90th percentile. In the Infrastructure & Utilities industries, male casual workers experienced mostly positive effects, with the highest increase of 19.8% at the 90th percentile and a significant positive effect of 14.3% at the 10th percentile. Female workers also saw positive effects at lower percentiles, with the highest impact of 56.7% at the 10th percentile, which, however, decreased to 10.3% at the 90th percentile. Among male casual workers in the Goods & Services Distribution industries, the effects were positive, with the highest increase of 17.5% at the 90th percentile, in contrast to females, where the highest advantage was 73.3% at the 10th percentile. In the Knowledge & Service-Based industries, male casual workers faced insignificant negative effects across all percentiles, and the results of female casual workers in those industries were mixed, i.e., major advantage at the 25th percentile but negative at the 90th percentile of the wage distribution. Overall, female casual workers experienced more positive effects from industry variables compared to male casual workers, particularly at lower percentiles. Thus, the differences in impact across percentiles indicated that industry effects varied significantly between genders and across different income levels.

The analysis of pseudo-R^2 values revealed an improved model fit at higher quantiles, indicating better predictions for higher response percentiles. Both regular employee groups (male and female) displayed higher pseudo-R^2 values than casual workers, suggesting a superior model fit for regular employees. Specifically, male regular employees had values ranging from 0.133 to 0.343, while females ranged from 0.211 to 0.353. In contrast, casual workers showed lower values, with males ranging from 0.101 to 0.128 and females from 0.021 to 0.111. However, the QR models for regular females and casual males had slightly higher average pseudoR^2 values, suggesting the potential for marginally better fits in these specific subgroups.

The fitted QR models effectively yielded statistically significant and accurate parameter estimates even under non-normal error terms and heteroscedasticity, aligning with Chen and Chalhoub-Deville (2014). Notably, a substantial portion of the significant estimates from both models exhibited exceptionally small standard errors, some even approaching or exceeding the precision of the median. This finding supports the suitability of QR models for analysing wage data, which often features non-normality and outliers, where traditional methods might be less reliable.

3.3 Decomposition of wage differentials

The results from Table 5, obtained through the OB and MMM decomposition, show a statistically significant male worker advantage over females. This is indicated by positive coefficients in the differences, coefficient, and characteristics components for both regular/salaried employees and casual workers.

Table 5 Wage differential decomposition across genders for regular/salaried employees and casual workers

Components	OB	MMM				
		θ = 0.10	θ = 0.25	θ = 0.50	θ = 0.75	θ = 0.90
Regular/salaried employees						
Difference	0.458*** (0.02)	0.709*** (0.01)	0.593*** (0.01)	0.473*** (0.01)	0.315*** (0.01)	0.161*** (0.01)
Characteristics	(0.104***) (0.02)	0.171*** (0.04)	0.152*** (0.02)	0.105*** (0.02)	0.040*** (0.03)	0.023* (0.03)
Coefficients	(0.354***) (0.02)	0.538*** (0.03)	0.441*** (0.02)	0.367*** (0.02)	0.271*** (0.03)	0.137*** (0.03)
Casual workers						
Difference	0.457*** (0.01)	0.349*** (0.01)	0.455*** (0.01)	0.493*** (0.01)	0.471*** (0.01)	0.498*** (0.01)
Characteristics	(0.093***) (0.01)	0.034*** (0.02)	0.081*** (0.02)	0.138*** (0.03)	0.142*** (0.01)	0.145*** (0.02)
Coefficients	(0.364***) (0.01)	0.315*** (0.02)*	0.374*** (0.02)	0.355*** (0.01)	0.329*** (0.01)	0.353*** (0.02)

Notes: * for p<.05, ** for p<.01, *** for p<.001. Standard errors using bootstrapping are reported in parentheses.
Source: Author's calculation based on PLFS 2020–21 dataset using Stata v.13

The OB decomposition analysis revealed a noteworthy wage differential between male and female workers in both employment types (regular: 0.458, casual: 0.457). The raw wage gap, expressed as a percentage, was 58.1% for regular/salaried employees and 57.9% for casual workers. Worker characteristics explained a portion of these gaps (11% in regular, 9.7% in casual employment). However, discrimination emerged as the primary factor in both groups, accounting for 42.5% and 43.9% of the unexplained raw wage gap among regular/salaried and casual employees, respectively. Notably, the level of discrimination appeared similar across both worker types despite minor differences in explained factors. However, given the potential limitations of the OB decomposition, such as its possible inability to capture nuanced discrimination due to non-linear relationships and interactions between variables, further analysis was conducted using the MMM decomposition.

The MMM decomposition, employing QR methodology, investigated the gender wage disparity across different percentiles. The examination of regular/salaried workers showed a significant variation in the wage disparity across the wage range. The disparity decreased from 0.709 at the 10^{th} to 0.160 at the 90^{th} percentile. This translated into a significant advantage for males, with a raw gap of 103.2% at the 10^{th} percentile, narrowing to 17.4% at the 90^{th} percentile. The observed wage distribution pattern suggested a "sticky floor" phenomenon, leading to more pronounced wage gaps for female workers, particularly at lower income levels, in contrast to male workers. These findings are consistent with prior research indicating heightened discriminatory practices against females in low-wage occupations

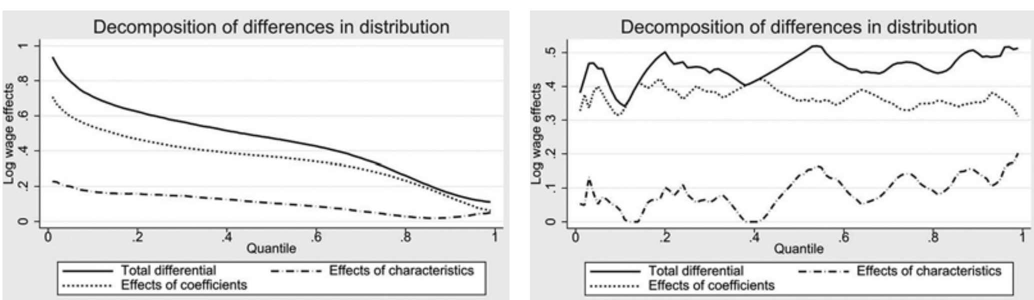
(Chakraborty and Mukherjee, 2014; Das, 2018). Discrimination was most pronounced at the 10th percentile, reaching 0.538, corresponding to a raw wage gap of 71.3%. Discrimination remained significant even at the median (0.473, 60.5% raw gap). Although the overall gap decreased toward the higher percentiles, discrimination persisted. High earners still encountered a 31.1% gap (75th percentile) and a 14.7% gap (90th percentile) due to discrimination in raw terms. The study also examined how characteristics contribute to the wage gap, revealing that their maximum contribution was observed at the 10th percentile, accounting for 18.6% of the overall gap. However, their influence diminished rapidly thereafter, reaching only 2.3% at the 90th percentile. This suggests that for high earners, discrimination alone could explain all the remaining wage disparity. Figure 1(a) replicated previous findings, illustrating a narrowing wage gap with increasing income. Lower-income quantiles exhibited larger disparities, influenced by both observed and unobserved characteristics, with the latter playing a more prominent role across all income strata.

The examination of the gender wage differentials among casual workers using the MMM decomposition showed a clear contrast with regular/salaried employees. The overall gap was lowest at the 10th percentile (0.349) but increased steadily towards the 90th percentile (0.498), with a slight dip at the 75th percentile (0.471). This translated to a raw wage gap of 41.8% at the 10th percentile, peaking at 64.5% at the 90th percentile before dropping back to 60.2% at the 75th percentile. Like regular workers, the gap for casual workers is statistically significant at all percentiles (10th to 90th). However, unlike the "sticky floor" effect or the "glass ceiling" barrier observed for regular workers, the casual worker gap exhibited a unique upward trend, indicating a potentially different form of discrimination females face in casual work. The unexplained wage differential component, which may indicate potential bias, was the dominant factor across all income percentiles. This unexplained component increased from 0.315 at the 10th percentile (37% of the raw gap) to 0.353 at the 90th percentile (42.3% of the raw gap). In contrast, the explained portion, attributed to differences in worker characteristics, was significantly smaller, ranging from 0.034 at the 10th percentile (3.5% of the raw gap) to 0.145 at the 90th percentile (15.6% of the raw gap). This suggests that discrimination plays a dominant role in the wage gap, particularly for higher-wage earners. Further research is necessary to comprehensively understand the specific forms of discrimination experienced by females in casual work, alongside the systemic and cultural mechanisms that perpetuate such inequities. Figure 1(b) supports this interpretation, revealing nuanced differences in the decomposition. While unexplained factors remain significant, their relative contribution remains higher than observed for regular workers. This finding suggests that workers' characteristics have a more significant impact on elucidating wage disparity in casual employment.

Figure 1 Plot of the MMM decomposition results for the male and female wage differential

a) Regular/salaried employees

b) Casual workers



Source: Author's calculation based on PLFS 2020–21 dataset using Stata v.13

The utilisation of the MMM decomposition, based on the QR framework and implemented across various percentiles, demonstrated statistically significant reductions in standard errors for both coefficient estimates and raw wage differences compared to those derived from the OLS-based OB decomposition. This increased accuracy significantly strengthened the robustness of the results, thereby enhancing the credibility and trustworthiness of the conclusions regarding the extent and factors influencing the identified gender pay disparity.

CONCLUSION

The QR modelling approach unveiled key factors influencing wage structures for male and female workers, distinguished by employment type (regular vs casual). Socioeconomic indicators, such as social group, marital status, and education, significantly impacted wages across income levels. Urban workers, irrespective of gender or employment type, earned more than their rural counterparts. Lower social groups faced wage disadvantages compared to higher ones, particularly SCs in regular employment and STs in casual work. Marriage positively impacted regular employees' wages, with widowed females even experiencing an unexpected advantage, although this effect was less pronounced for casual workers. Education strongly improves wages, especially for regular employees, with higher qualifications leading to more significant gains than casual workers. Gender interacted with education, benefiting regular female employees more from higher qualifications than males, a trend not observed for casual female workers. White-collar jobs offered the highest wages for regular males, while blue-collar jobs also positively influenced their earnings. Female regular employees showed no significant differences across occupational categories. Agriculture was the most lucrative sector for casual workers, while white-collar work offered no advantage, and blue-collar jobs were associated with lower wages than agriculture. Specific industries further complicated the picture, with some offering wage premiums and others leading to lower earnings for both employment types. These findings underscore the complex interrelationship of social, economic, and sectoral factors in shaping wage structures for diverse worker segments.

The MMM analysis revealed significant gender wage gaps among regular/salaried and casual workers across all percentiles. For regular/salaried employees, the largest gaps were observed at lower percentiles, ranging from 103.2% to 17.4%, indicating a "sticky floor" effect. This suggests that females in lower wage brackets face obstacles to career advancement due to discriminatory practices. Worker characteristics played a minor role, with discrimination being the main driver. Among casual workers, the wage gap was significant across all percentiles, ranging from 41.8% to 64.5%. However, no clear "sticky floor" or "glass ceiling" pattern indicated pervasive discrimination at all wage levels. The analysis attributes the wage gap among casual workers primarily to gender-based discrimination rather than differences in characteristics.

Our study explored the potential of QR methodology to unveil non-linear wage patterns across income distribution for both regular/salaried and casual workers, irrespective of gender. Compared to traditional linear regression, QR models in the study have performed in several key areas. First, QR demonstrates greater adaptability to workforce heterogeneity, accommodating non-linear relationships between wages and explanatory variables. Second, it demonstrates robustness to outliers inherent in wage data, yielding more precise and reliable estimates. These findings suggest QR's superior applicability for labour market analysis. QR can inform targeted policy development by enhancing our understanding of wage structures. Further insights into the gender wage gap can be gained through the MMM decomposition within the QR framework. In contrast to OB decomposition, our study found that the MMM resulted in smaller standard errors and enabled the examination of wage distribution across different percentiles. This allowed for a more detailed understanding of the wage gap, highlighting possible "glass-ceiling" or "sticky-floor" effects within the distribution. This solidifies its appeal for investigating wage discrimination in future research as well.

This study's findings emphasise the importance of accurately selecting appropriate methodologies to measure and analyse wage disparities in the labour market. Applying QR and MMM decomposition techniques demonstrates their potential as a robust framework for future research in this domain. These methods offer powerful tools to dissect the multifaceted nature of wage inequality, including the gender wage gap. QR and the MMM decomposition pave the way for developing more informed policy interventions to achieve greater labour market equity by facilitating a deeper understanding of these complex phenomena.

ACKNOWLEDGMENTS
Samapriya Trivedi is a recipient of the Indian Council of Social Science Research Doctoral Fellowship. His article is largely an outcome of his doctoral work sponsored by ICSSR. However, the responsibility for the facts stated, opinions expressed and the conclusions drawn are entirely that of the authors.

References

ADIREKSOMBAT, K., FANG, Z., SAKELLARIOU, C. (2010). The Evolution of gender wage Differentials and Discrimination in Thailand: 1991–2007 An application of Unconditional Quantile Regression [online]. *MPRA Paper*. <https://ideas.repec.org/p/pra/mprapa/27516.html>.

AGRAWAL, T. (2013). Gender and caste–based wage discrimination in India: Some recent evidence [online]. *Journal for Labour Market Research*, 47(4): 329–340. <https://doi.org/10.1007/s12651–013–0152-z>.

ALKHARUSI, H. (2012). Categorical variables in regression analysis: a comparison of dummy and effect coding [online]. *International Journal of Education*, 4(2): 202. <https://doi.org/10.5296/ije.v4i2.1962>.

ARELLANO, M., BONHOMME, S. (2017). Quantile selection models with an application to understanding changes in wage inequality [online]. *Econometrica*, 85(1): 1–28. <https://doi.org/10.3982/ecta14030>.

ARMSTRONG, R. D., FROME, E. L., KUNG, D. S. (1979). A revised simplex algorithm for the absolute deviation curve fitting problem [online]. *Communications in Statistics. Simulation and Computation*, 8(2): 175–190. <https://doi.org/10.1080/03610917908812113>.

AZAM, M. (2012). Changes in wage structure in urban India, 1983–2004: a quantile regression decomposition [online]. *World Development*, 40(6): 1135–1150. <https://doi.org/10.1016/j.worlddev.2012.02.002>.

AZAM, M., PRAKASH, N. (2015). A distributional analysis of public-private wage differential in India [online]. *LABOUR*, 29(4): 394–414. <https://doi.org/10.1111/labr.12068>.

BAI, Y., VEALL, M. R. (2023). Minimum wages and mental health: Evidence from Canada [online]. *SSM. Mental Health*, 3: 100187. <https://doi.org/10.1016/j.ssmmh.2023.100187>.

BIEWEN, M., ERHARDT, P. (2021). Arhomme: An implementation of the Arellano and Bonhomme (2017) estimator for quantile regression with selection correction [online]. *The Stata Journal*, 21(3): 602–625. <https://doi.org/10.1177/1536867x211045516>.

CHAKRABORTY, S., MUKHERJEE, S. (2014). Gender wage gap in the Indian labour market: Evidence from the NSS 66th round data. *The Indian Journal of Labour Economics*, 57(2): 259–280.

CHEN, F., CHALHOUB-DEVILLE, M. (2014). Principles of quantile regression and an application [online]. *Language Testing*, 31(1): 63–87. <https://doi.org/10.1177/0265532213493623>.

DAS, P. (2018). Wage gap and employment status in Indian labour market: Quantile based counterfactual analysis. *World Journal of Applied Economics*, 4(2): 117–134. <https://doi.org/10.22440/wjae.4.2.4>.

DESHPANDE, A., GOEL, D., KHANNA, S. (2017). Bad Karma or Discrimination? Male–Female Wage Gaps Among Salaried Workers in India [online]. *World Development*, 102: 331–344. <https://doi.org/10.1016/j.worlddev.2017.07.012>.

DESHPANDE, A., SHARMA, S. (2015). Disadvantage and discrimination in self-employment: Caste gaps in earnings in Indian small businesses [online]. *Small Business Economics*, 46(2): 325–346. <https://doi.org/10.1007/s11187–015–9687–4>.

DURAISAMY, M., DURAISAMY, P. (2016). Gender wage gap across the wage distribution in different segments of the Indian Labour Market, 1983–2012: Exploring the glass ceiling or sticky floor phenomenon [online]. *Applied Economics*, 48(43): 4098–4111. <https://doi.org/10.1080/00036846.2016.1150955>.

DUTTA, P. V. (2006). Returns to education: New evidence for India, 1983–1999 [online]. *Education Economics*, 14(4): 431–451. <https://doi.org/10.1080/09645290600854128>.

FITZENBERGER, B., KOENKER, R., MACHADO, J., MELLY, B. (2021). Economic applications of quantile regression 2.0 [online]. *Empirical Economics*, 62(1): 1–6. <https://doi.org/10.1007/s00181–021–02186–1>.

GAZELEY, I., HOLMES, R., NEWELL, A., REYNOLDS, K. P., RUFRANCOS, H. G. (2018). Inequality among European working households, 1890–1960. *Social Science Research Network*. <https://doi.org/10.2139/ssrn.3137485>.

HNATKOVSKA, V., LAHIRI, A., PAUL, S. (2012). Castes and labour mobility [online]. *American Economic Journal: Applied Economics*, 4(2): 274–307. <https://doi.org/10.1257/app.4.2.274>.
HUANG, Q., ZHANG, H., CHEN, J., HE, M. (2017). Quantile Regression Models and their Applications: a Review [online]. *Journal of Biometrics & Biostatistics*, 8(3). <https://doi.org/10.4172/2155–6180.1000354>.
INDIA WAGE REPORT (2018). *International Labour Organisation* [online]. [cit. 25.12.2023]. <https://www.ilo.org/wcmsp5/groups/public/−−−asia/−−−ro−bangkok/−−−sro−new_delhi/documents/publication/wcms_638305.pdf>.
JENKINS, S. P. (1999). INEQDECO: Stata module to calculate inequality indices with decomposition by subgroup [online]. *Statistical Software Components*. <https://ideas.repec.org/c/boc/bocode/s366002.html>.
KHANNA, S. (2012). Gender wage discrimination in India: Glass ceiling or sticky floor? [online]. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2115074>.
KINGDON, G. G., UNNI, J. (2001). Education and Women's Labour Market Outcomes in India [online]. *Education Economics*, 9(2): 173–195. <https://doi.org/10.1080/09645290110056994>.
KOENKER, R. (2005). *Quantile regression* [online]. New York: Cambridge University Press. <https://doi.org/10.1017/cbo9780511754098>.
KOENKER, R., BASSETT, G. (1978). Regression quantiles [online]. *Econometrica*, 46(1): 33–50. <https://doi.org/10.2307/1913643>.
KOENKER, R., MACHADO, J. A. (1999). Goodness of fit and related inference processes for quantile regression [online]. *Journal of the American Statistical Association*, 94(448): 1296–1310. <https://doi.org/10.1080/01621459.1999.10473882>.
MACHADO, J. A., MATA, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression [online]. *Journal of Applied Econometrics*, 20(4): 445–465. <https://doi.org/10.1002/jae.788>.
MADAN, S., MOR, S. (2022). Is Gender Earnings Gap a Reality? Signals from Indian Labour Market [online]. *Statistika: Statistics ans Economy Journal*, 102(2): 168–183. <https://doi.org/10.54694/stat.2021.19>.
MADHESWARAN, S., ATTEWELL, P. (2007). Caste Discrimination in the Indian Urban Labour Market: Evidence from the National Sample Survey [online]. *Economic and Political Weekly*, 42(41): 4146–4153. <http://www.jstor.org/stable/40276549>.
MELLY, B. (2006). *Estimation of Counterfactual Distributions Using Quantile Regression* [online]. University of St.Gallen Discussion Paper. <https://www.alexandria.unisg.ch/handle/20.500.14171/83172>.
MITRA, A. (2016). Education and earning linkages of regular and casual workers in India: a quantile regression approach [online]. *Journal of Social and Economic Development*, 18(1−2): 147–174. <https://doi.org/10.1007/s40847−016−0029−4>.
ÑOPO, H. (2008). Matching as a Tool to Decompose Wage Gaps [online]. *The Review of Economics and Statistics*, 90(2): 290–299. <https://doi.org/10.1162/rest.90.2.290>.
ÑOPO, H., DAZA, N., RAMOS, J. (2012). Gender earning gaps around the world: a study of 64 countries [online]. *International Journal of Manpower*, 33(5): 464–513. <https://doi.org/10.1108/01437721211253164>.
OAXACA, R. (1973). Male-female wage differentials in urban labour markets [online]. *International Economic Review*, 14(3): 693–709. <https://doi.org/10.2307/2525981>.
PALA, V., NONGSPUNG, A. (2022). Trends in wages of regular wage employees and casual workers in Indian states [online]. *The Indian Journal of Labour Economics*, 65(3): 667–687. <https://doi.org/10.1007/s41027-022-00386-6>.
PATIDAR, V. K., WADHVANI, R., SHUKLA, S., GUPTA, M., GYANCHANDANI, M. (2023). Quantile Regression Comprehensive in Machine Learning: a Review [online]. In: *2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS),* 1–6. <https://doi.org/10.1109/SCEECS57921.2023.10063026>.
PORTER, S. R. (2014). Quantile regression: Analyzing changes in distributions instead of means [online]. *Higher Education: Handbook of Theory and Research*, 335–381. <https://doi.org/10.1007/978−3−319−12835-1_8>.
SENGUPTA, A., DAS, P. (2014). Gender Wage Discrimination across Social and Religious Groups in India: Estimates with Unit Level Data [online]. *Economic and Political Weekly*, 49(21): 71–76. <http://www.jstor.org/stable/24479554>.
SENGUPTA, P., PURI, R. (2021). Gender pay gap in India: a reality and the way forward – an empirical approach using quantile regression technique [online]. *Studies in Microeconomics*, 10(1): 50–81. <https://doi.org/10.1177/2321022221995674>.
SI, X., LI, M. (2023). Impact of the internet use on informal workers' wages: Evidence from China [online]. *PloS One*, 18(5): e0285973. <https://doi.org/10.1371/journal.pone.0285973>.
SPENCE, M. (1973). Job Market Signaling. *The Quarterly Journal of Economics*, 87(3): 355–374. <https://doi.org/10.2307/1882010>.
STATACORP (2013). *Stata Statistical Software: Release 13.1*. College Station TX, StataCorp LLC.
Unit Level data of Periodic Labour Force Survey (PLFS) July 2020–June 2021 [online]. Ministry of Statistics and Program Implementation, Government of India. <https://mospi.gov.in/PLFS-July20-June21>.
WALDMANN, E. (2018). Quantile regression: A short story on how and why [online]. *Statistical Modeling*, 18(3–4): 203–218. <https://doi.org/10.1177/1471082x18759142>.

ANNEX

Table A1 Distribution of male and female workers (in percent) based on different characteristics

Variables/characteristics	Regular/salaried employees		Casual workers	
	Male	**Female**	**Male**	**Female**
	(N = 29 691)	**(N = 9 727)**	**(N = 18 773)**	**(N = 5 513)**
Residential location				
Rural	76.5	23.5	72.8	27.3
Urban	75.8	24.2	83.5	16.5
Social group				
STs	67.6	32.4	67.4	32.6
SCs	73.4	26.6	75.2	24.9
OBCs	77.2	22.8	73.9	26.1
Others	78.1	21.9	82.3	17.7
Marital status				
Never married	81.4	18.6	91.2	8.8
Married currently	78.2	21.8	75.4	24.6
Widowed	14.8	85.2	20.4	79.6
Divorced/separated	32.1	67.9	38.3	61.7
General education				
Illiterate	52.0	48.0	59.0	41.0
No formal schooling	77.6	22.4	76.0	24.0
Up to primary school	72.6	27.4	75.6	24.4
Middle school	82.7	17.3	84.9	15.1
Secondary school	81.5	18.5	86.6	13.4
Higher secondary	81.1	18.9	88.8	11.2
Graduate	77.8	22.2	91.3	8.7
Postgraduate or above	67.4	32.6	89.5	10.6
Technical education				
Not received	76.4	23.6	74.4	25.6
Have technical education	74.3	25.7	94.5	5.5
Occupations				
White-collar	68.9	31.2	75.8	24.2
Blue-collar	80.9	19.1	74.5	25.5
Agricultural	80.8	19.2	73.6	26.4
Industries				
Production and extraction	85.2	14.8	60.3	39.7
Infrastructure & utilities	94.1	5.9	89.0	11.0
Goods & service distribution	91.6	8.5	93.8	6.2
Knowledge & service-based	80.9	19.1	79.5	20.5
Public and social services	44.8	55.2	60.3	39.8

Source: Author's calculation based on PLFS 2020–21 dataset using Stata v.13

Household Finance and Consumption Survey (HFCS) in Czechia

Jiří Vopravil[1] | *Czech Statistical Office, Prague, Czechia*
Táňa Dvornáková[2] | *Czech Statistical Office, Prague, Czechia*
Barbora Linhartová Jiřičková[3] | *Czech Statistical Office, Prague, Czechia*

Received 18.7.2024, **Accepted** (reviewed) 6.9.2024, **Published** 14.3.2025

Abstract

Household Finance and Consumption Survey (HFCS) is a survey focused mainly on mapping financial and assets situation of households. In Czechia, the survey has been conducted annually since 2020 under the name Finanční situace domácností (FSD). The article focuses on the implementation of the survey in the Czech setting, mainly describing the cooperation between the Czech Statistical Office and the Czech National Bank. A particular attention is paid to the methodological aspect of FSD as well as to the survey's main outcomes, primarily the net wealth indicator. The integration of FSD into the EU-SILC (European Union – Statistics on Income and Living Conditions) survey enables annual data collection and publication of main results, as well as the reduction of respondent burden.

Keywords	***DOI***	***JEL code***
Household Finance and Consumption Survey (HFCS), net wealth, household survey, EU-SILC	*https://doi.org/10.54694/stat.2024.38*	*D14, G51, E21*

INTRODUCTION

The *Household Finance and Consumption Survey* (abbreviated as 'HFCS') is a survey of which the national Czech version is conducted annually by the Czech Statistical Office (CZSO) in cooperation with the Czech National Bank (CNB) under the name *Finanční situace domácností* (abbreviated as 'FSD' in Czech). The results of the survey map the situation of Czech households in terms of their financial and assets situation.

The survey is mainly focused on data on the households' ownership and estimated value of assets, household members' use of financial products, and their potential debts by their amount and type.

[1] Household Surveys Department, Czech Statistical Office, Na padesátém 3268/81, 100 82, Prague 10, Czechia. E-mail: jiri.vopravil@csu.gov.cz.
[2] Household Surveys Department, Czech Statistical Office, Na padesátém 3268/81, 100 82, Prague 10, Czechia. E-mail: tana.dvornakova@csu.gov.cz.
[3] Household Surveys Department, Czech Statistical Office, Na padesátém 3268/81, 100 82, Prague 10, Czechia. E-mail: barborajirickova@seznam.cz.

FSD results in unique data which cannot be obtained by any other means than direct household interviewing, as the CZSO does not yet have access to administrative data for use in household surveys. One of the key outcomes of the survey is the so-called *net wealth* indicator, which is calculated from data collected in the survey, specifically from the value of household's assets and debt.

The aim of this paper is primarily to describe the methodological aspects of the Czech module of the survey, focusing on the specifics of this implementation of the international HFC survey. Similarly to the Household Budget Survey (HBS; in the Czech setting Statistika rodinných účtů, abbreviated as 'SRÚ'), the FSD is conducted by being integrated into the Living Conditions Survey, which is the national version of the European Union – Statistics on Income and Living Conditions (hereafter 'SILC') survey, i.e. on a subset of households randomly selected for SILC. This model of survey offers a number of benefits which are further described in the article.

Households' net wealth can then be classified according to the number of persons living in the household, the size of the municipality, the legal form of housing use (in private ownership or rented) and income quintiles. Furthermore, due to the internationally comparable methodology of HFCS, FSD results provide a comparison with the situation in other euro-area countries surveyed.

1 IMPLEMENTATION OF HFCS IN CZECHIA

In 2005, the European Central Bank began discussing the introduction of a project to create a research network on household wealth and consumption in the Euro area by collecting data via household surveys. However, initial attempts to introduce a project resulting in micro data on household net worth had already begun a few years earlier. In 2006, a working group on the subject met for the first time as the Household Finance and Consumption Survey Task Force (Brandolini, 2023).

The first wave of the Household Finance and Consumption survey organized by the European Central Bank took place in 2013, the next wave was conducted in 2017. The reference period is three years, however, due to the COVID-19 pandemic, the intended 2020 wave was postponed and took place in 2021 (ECB, 2024). The next wave of the survey was carried out in 2023 and the following is scheduled to take place in 2026.

1.1 Cooperation between CZSO and CNB

In Czechia, the national module of HFCS, Finanční situace domácností (FSD), is conducted by the Czech Statistical Office in cooperation with the Czech National Bank. The cooperation between the two institutions on the preparation of the survey began in 2018. In the March of 2018, the working group focusing on the implementation of the survey was established, as well as the main tasks of the parties involved. The Czech Statistical Office primarily ensures the collection of data in the field by a network of internal interviewers, and related issues.

In particular, the CZSO has created a uniform introductory part of the questionnaire containing data on household composition, including an appropriate methodology. The questionnaire was prepared by the CZSO in cooperation with the Czech National Bank. The CZSO has also provided training for the interviewers, concerning the content of the questionnaire as well as the methodology of the survey. Furthermore, the CZSO provided electronic questionnaire for data collection, conducted the data collection itself, applied case management system and processed the output data files in the structure required by the ECB with additional controls and data files validation according to the ECB validation files (CZSO, CNB, 2018).

The CNB has collaborated significantly with CZSO on the preparation of the questionnaire and the methodological guidelines, mainly concerning the correct financial terminology and derived variables compilation. It was also planned that the CNB will ensure statistical imputation of missing item nonresponse data according to the ESCB (European System of Central Banks) and the final validation of the datasets according to the HFCS rules and their transmission to the ECB and to the CZSO for joint sharing (CZSO, CNB, 2018). In practice, over the years it has become apparent that the CZSO has more experience with data processing, including validation procedures, and therefore currently carries out this part of the survey, too. Additionally, the CNB provides incentives for the respondents, as well as the material support for the project.

In the preparation phases, the CZSO had analysed the questions required for the survey. As a result, it was decided that the integration of FSD to the EU-SILC survey will take place, due to a high percentage of content overlap between the two surveys, mainly in the areas of income and consumption. This decision, as well as data collection methods and overall methodology, was implemented after it had been approved by the ECB.

Before the first wave took place, a pilot survey was conducted in 2019, mainly in order to test the data collection tool (i.e. questionnaire) and the organisation of the survey. During the pilot survey, approximately 150 households have participated from all Czech regions with a 100% response rate. In February of 2019, the questionnaires were distributed to the supervisors at the regional level and the collection of the data took place until June of the same year. After the data collection, its quality was measured by the type of dwelling and form of ownership, by the household type, by the social group and education level of the head of household, by the number of household members, municipality size and income level (Dvornáková, 2020). The first wave of FSD was conducted in 2020.

2 METHODOLOGY

The following part of the article briefly discusses the methodology of FSD. From 2020 onwards, the Czech version of HFCS has been conducted annually. As of spring/summer of 2024, the FSD has been conducted consecutively for the fourth year. The data, obtained by direct interviewing in the households, are then used at the national level by CNB, CZSO and other institutions and data users, as well as at the international level (CNB, CZSO, 2023).

2.1 Selection of households

The selection of interviewed households for the FSD is based on the four-year panel of the SILC survey. The unit of observation is a dwelling. During the first visit (i.e. on wave 1), households and their members with a habitual residence in the selected dwelling are interviewed. At follow-up visits, only the households with so-called panel persons, i.e. household members who had participated in wave 1, are interviewed. Households from all regions of the Czech Republic are contacted for the FSD. The survey unit is the household which was successfully interviewed in the SILC wave 4 (CNB, CZSO, 2023).

2.2 Fieldwork

Data collection is conducted as a face-to-face interview between the interviewer and the respondent. As a data collection instrument, interviewers use either an electronic questionnaire in a tablet, i.e. CAPI (Computer Assisted Personal Interviewing), or a paper questionnaire, i.e. PAPI (Pen and Paper Interviewing), which is then transcribed into an electronic version. The successfully interviewed households receive a set of commemorative coins as a gift provided by CNB. At the regional level, interviewers from CZSO ensure data collection, including primal data verification. At the CZSO headquarters, data from the questionnaires are then linked and subjected to a final 'super-check', editing including imputation of missing values, and centralised processing (CNB, CZSO, 2023).

Due to the aforementioned integration of the surveys, the FSD interview is usually conducted at the same time as a visit related to the SILC or HBS surveys, less often as a separate visit.

2.2.1 Survey content

The FSD questionnaire is divided into eight parts according to the surveyed topics. The first part of the interview focuses on whether the household has a mortgage or other property loan on their main dwelling, while the next part focuses on mortgage on other potential property. The third section includes questions on other loans and credit that the household may be repaying (e.g. consumer credit, hire purchase, financial leasing, credit card debt, etc.). The aim of this section is to map the different types and amount of debt (CZSO, 2023).

The fourth and fifth parts of the survey cover the ownership and estimated value of other assets such as cars and other vehicles, valuables (e.g. jewellery, paintings or antiques) and, where applicable, the value of a business or a share in a business. The sixth section of the questionnaire focuses on the use of financial products (bank accounts, building society accounts, pensions, investment funds, securities, etc.). This part maps not only whether members of the household use these financial products, but also their estimated value. In the next section, the questions concern the household's expectations vis-à-vis their financial situation in the next 12 months. Finally, the eight part of the interview focuses on the consumption expenditure of the household, mainly regarding groceries and leisure activities (CZSO, 2023).

Other information needed for the HFCS is obtained from the SILC survey, such as socio-economic data, data on income and economic activity of household members, housing etc. For more details on the integration of the HFCS and SILC surveys, see chapter 3.5.

2.3 Number of households in FSD

The results which were published for 2021 are based on a total of 3 122 households surveyed in 2020 and 2021, and the results for 2022 are based on a total of 3 155 households surveyed in 2021 and 2022 (CNB, CZSO, 2023). The results for 2023, which will be published in the autumn of 2024, are based on a total of 3 246 households surveyed in 2022 and 2023. The reason for this is to acquire a representative sample of households for the ensuing breakdown of the results. The numbers of surveyed households, as well as the overall response rate, are presented in Table 1.

Table 1 Number of households surveyed in FSD (2020–2023)

	2020	2021	2022	2023
Number of households sampled	2 062	2 031	2 064	2 062
Response rate (%)	76.2	76.4	77.7	79.6
Number of households surveyed	1 571	1 551	1 604	1 642

Source: CZSO

2.4 Converting data to population totals, corrections and calculations

The next part refers to the methodological aspects of the FSD results for 2021 and 2022. Due to the integration of the survey, i.e. the sample for the FSD being a subset of the SILC survey sample, the results of the SILC were used as the basis for the calculations. The resulting non-response rate, which was also influenced by SILC response rates from previous years, distorted the composition of the final dataset from which the FSD results were compiled. However, the limiting factor for the calculation method or the construction of conversion factors was the limited size of the household sample (CNB, CZSO, 2023).

The calculations consisted of the elimination of total non-response. It was necessary to add correction coefficients for individual households with respect to their representation in the FSD sample, relative to the SILC population estimates. An iterative weight calibration procedure was used to calculate the weights, using as a reference the SILC baseline estimates of aggregates (e.g. the number of households, persons, working household members, dependent children, inactive pensioners and unemployed persons), including selected characteristics used to classify households in published results:

- 4 groups of households according to the status of the head of the household (employees with lower/higher education, self-employed, pensioners);
- 4 groups of households according to the legal form of the dwelling use (owner-occupied house, owner-occupied apartment, cooperative dwelling, rented dwelling);
- 2 groups of households according to the number of dependent children (with or without children);
- 4 groups of households according to the municipality size (less than 2 thousand inhabitants, 2 000–9 999 inhabitants, 10 000–49 999 inhabitants, 50 thousand and more inhabitants);
- 5 groups of households by net monetary income per person (quintiles) (CNB, CZSO, 2023).

The above described procedure primarily corrects the social structure of households and at the same time eliminates the associated distortion in the distribution of income. In order to classify households by their income level (quintile), each household was assigned the net cash income according to the SILC survey (i.e. for the previous year) for the year in question. Thus, the FSD data for 2021 (2022) also included household data for 2020 (2021 respectively). House prices and car prices were recalculated accordingly using average (moving) annual indices of house prices and car prices respectively (CNB, CZSO, 2023).

2.5 Integration

As mentioned above, from 2020 onwards, the Czech version of HFCS is integrated into the EU-SILC household survey (for more information about the survey see for instance Linhartová Jiřičková, Dvornáková and Vopravil 2024). Thus, only the households that have been randomly selected and surveyed for SILC can be contacted and interviewed for FSD (CNB, CZSO, 2023).

As the CZSO's analysis has shown, some information required for HFCS is already obtained from the EU-SILC survey. EU-SILC, conducted in Czechia under the name Životní podmínky domácností (i.e. Living Conditions of Households) focuses mainly on the income of the household and its members, as well as on their living conditions. However, the survey maps a number of topics, some of which overlap with the HFCS/FSD focus, such as basic demographic information, but also information on income, employment and to some extent the financial situation of the household (Babecký and Dvornáková, 2024).

Selected households thus participate in several household surveys, which can be a greater burden for them. However, due to the integration of the surveys, some information does not have to be collected repeatedly and they are offered appropriate incentives. This leads to the reduction of both interviewers' and respondents' burden and to a higher response rate (Babecký and Dvornáková, 2024). Ideally, the combining of the three household surveys (including the Household Budget Survey) into one brings information on income (SILC), on consumption expenditure (HBS/SRÚ) and on financial and assets situation (HFCS/FSD) all from one household (Dvornáková, 2020).

The willingness to participate in the subsequent household survey is also much higher among households that have already participated in a previous survey and have prior experience with the Statistical office. Therefore, the CZSO prefers this integration with a relatively higher burden for an individual household, compared to separate surveys in households. The separate surveys, due to the relatively high non-response, would require relatively large samples which would have to be contacted and interviewed from the beginning with all questions.

Furthermore, the integration of the surveys facilitates the annual data collection for the FSD survey. The regularity of surveys leads to a distribution of interviewers' burden. It also enables the publishing of time series without brakes, with the regular results being more data user-friendly. What is more, the integration and subsequent more frequent data publication make it possible to determine the net wealth indicator more accurately (Dvornáková, 2020).

3 OUTCOMES

The primary outcome of the FSD survey is the Publication with the main results. The Publication is published annually. The latest available results are for years 2021 and 2022, and they were released at the end of 2023. The annual publication of the FSD survey results is possible due to the integration of the survey with the EU-SILC, which ensures annual data collection. On the other hand, European Central Bank carries out the survey in a three-year reference period (ECB, 2024).

In the Publication, the households are classified according to the number of household members, the municipality size, the type of dwelling and the form of ownership (CNB, CZSO, 2023). The resulting data are actively used by many institutions and other data users. Therefore, one of the important outcomes of the survey are also micro-data. For instance, at CNB, they are used in economic research to calibrate a model of household energy consumption and savings, and in financial stability research to evaluate the effectiveness of measures targeting debtors. According to Babecký (2024), results from FSD can be used in macroeconomics and macro-financial analyses, for example to assess the impact of changes in house prices or to better understand the consumption behaviour of indebted households and its impact on aggregate consumption (Babecký, 2024).

One of the key results of FSD is the *net wealth* indicator.

3.1 Net wealth

Household net wealth is the main outcome of the survey (i.e. of the integration of the surveys). In short, the indicator is calculated as *total assets minus total debts*.

3.1.1 Household assets and their components

According to an internationally comparable definition, household assets are composed of financial and non-financial assets. Non-financial (i.e. real) assets include real estate, vehicles, valuables and assets related to a business or other self-employment activity. Outcomes of the FSD survey offer the percentage shares of households by ownership of their main residence, ownership of at least one other property and ownership of at least one type of vehicle (CNB, CZSO, 2023).

The second component of household assets are financial assets. Financial assets include all forms of deposits and savings, the value of securities, funds invested in investment or capital life insurance and mutual and similar funds, long-term cash savings, alternative currencies, cryptocurrencies, etc. The category of deposits applies to households that have funds in current or savings bank accounts, building societies and time deposits or certificates of deposit (CNB, CZSO, 2023).

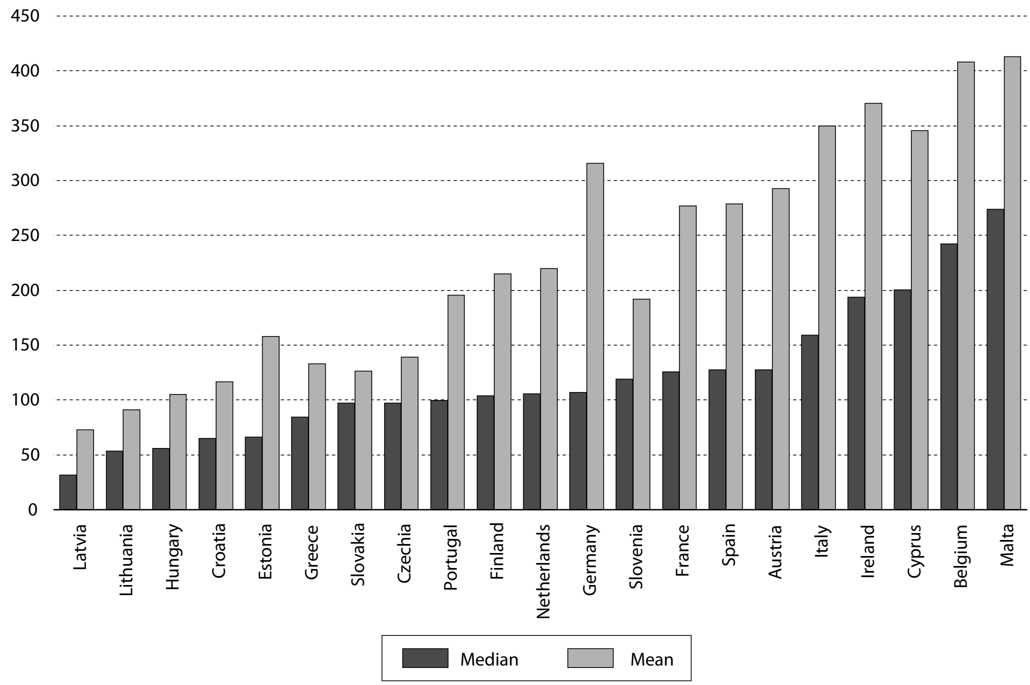
3.1.2 Debt and its components

In order to correctly estimate households' net wealth, it is important to determine the types and value of debt the households may have. Total household debt consists of mortgage debt on property and other types of debt. Mortgages also include other loans on property (e.g. building society loans). These are only mortgages or loans secured on the relevant property. Other debt consists of various types of bank loans (e.g. consumer loans), non-bank loans, outstanding balances on overdrafts, outstanding balances on credit cards, loans from relatives or friends and other loans (CNB, CZSO, 2023).

3.1.3 Net wealth in Czechia

The previous waves of the FSD survey have shown that the net wealth of Czech households is primarily determined by the ownership of the apartment or house in which the household lives. This is particularly interesting in comparison to the situation in some other countries where net wealth is calculated according to the same methodology. However, the amount of household net wealth also differs significantly according to the municipality size and according to the level of households' income (CNB, CZSO, 2023). Figure 1 shows a comparison of the median and mean net wealth of households in different EU countries in 2021. Luxembourg has been deliberately left out from the chart due to the country's very high values.

Figure 1 Median and mean net wealth of EU households in 2021 (EUR thousands)



Source: ECB (2023)

CONCLUSION

The Finanční situace domácností (FSD) survey is realized in Czechia as the national module of the Household Finance and Consumption Survey (HFCS), which is coordinated by the European Central Bank. The Czech Statistical Office carries out the survey in cooperation with the Czech National Bank. The cooperation is beneficial for a number of reasons, the Statistical Office can offer experience with data collection and processing, which can lead to financial savings, while the National Bank offers other expertise.

The FSD survey enables the two institutions in charge, as well as other data users, to acquire otherwise not accessible information on financial situation of Czech households. In relation to other related household surveys, primarily the EU-SILC, i.e. European Union – Statistics on Income and Living Conditions, and HBS, i.e. Household Budget Survey, this survey supplements information on households' *income* (available from EU-SILC, i.e. the Czech module of the survey) and *expenditure* (available from HBS) with data on their *financial situation*, mainly in terms of assets and debts. From these data, the net wealth indicator is calculated, which contributes to the assessment of households' financial situation. The survey results are primarily used by the Czech National Bank as a supporting component for the assessment of financial stability and monetary policy (Babecký, 2024). In general, this data will be used for the Income, Consumption and Wealth (ICW) concept. The Czech Statistical Office and the Czech National Bank have decided to integrate the FSD survey into the already well-established SILC survey. This results in an overall lower burden for respondents and brings further financial savings.

What is more, this integration allows the FSD survey to be carried out annually, contrary to the three-year reference period set by the European Central Bank. The CZSO and the CNB have agreed to conduct the survey on a smaller sample, but annually, which is beneficial for the users of the data.

References

BABECKÝ, J. (2024). V Česku je velmi nízká majetková nerovnost (Czechia has very low wealth inequality) [online]. *Statistika & My*, 2024/3. [cit. 10.7.2024]. <https://statistikaamy.csu.gov.cz/soubory/statistika-a-my/2024/03/18042403.pdf>.

BABECKÝ, J., DVORNÁKOVÁ, T. (2024). Jaké je čisté jmění českých domácností? (What is the net wealth of Czech households?) [online]. *Statistika & My*, 2024/3. [cit. 11.7.2024]. <https://statistikaamy.csu.gov.cz/jake-je-ciste-jmeni-ceskych-domacnosti>.

BRANDOLINI, A. (2023). Household wealth surveys and the HFCS. *Presentation presented at the Household Finance and Consumption Network Meeting in the Bank of Italy*, Rome on 7 June 2023.

CNB, CZSO. (2023). *Financial Situation of Households in 2021 and 2022* [online]. Prague: Czech Statistical Office. [cit. 4.7.2024]. <https://csu.gov.cz/docs/107508/c4942441-673a-faab-f49b-7c5671036282/160077-23_en.pdf>.

CZSO, CNB. (2018). *Dohoda o spolupráci pří provádění statistického šetření o finanční situací domácností*, evid. č. ČSÚ: 279-2018-S (Agreement on cooperation in conducting a statistical survey on the financial situation of households, CZSO No.: 279-2018-S) [online]. Registr smluv. [cit. 10.7.2024]. <https://smlouvy.gov.cz/smlouva/8526775?backlink=twrsx>.

CZSO. (2023). *Finanční situace domácností 2023* (Financial Situation of Households in 2023) [online]. Prague: Czech Statistical Office. [cit. 8.7.2024]. <https://csu.gov.cz/docs/107516/0f76d91a-8400-d638-a400-435f3fd22227/FSD_2023_dotaznik.pdf>.

DVORNÁKOVÁ, T. (2020). *Integrace šetření u domácností* (Household surveys integration). Presentation presented for the CZSO management on 27 January 2020.

ECB. (2023). *The Household Finance and Consumption Survey. Wave 2021. Statistical Tables* [online]. European Central Bank. [cit. 10.7.2024]. <https://www.ecb.europa.eu/home/pdf/research/hfcn/HFCS_Statistical_Tables_Wave_2021_July2023.pdf?0515e108613e1e4d0e839e816e9f07b8>.

ECB. (2024). *Household Finance and Consumption Survey (HFCS)* [online]. European Central Bank. [cit. 4.7.2024]. <https://www.ecb.europa.eu/stats/ecb_surveys/hfcs/html/index.en.html>.

LINHARTOVÁ JIŘIČKOVÁ, B., DVORNÁKOVÁ, T., VOPRAVIL, J. (2024). Statistics on Income and Living Conditions (SILC) Survey in the Czech Republic: Methodology and History [online]. *Statistika: Statistics and Economy Journal*, 104(2): 229–242. <http://doi.org/10.54694/stat.2023.43>

A Must Read for Those Who Wish to Learn More about Economic Analysis (Not Only) – Book Review

Ondřej Šíma[1] | *Prague University of Economics and Business, Prague, Czechia*

VEJMĚLEK, J., ŽĎÁREK, V. et al. *Makroekonomická analýza – od teorie k aplikaci* (Macroeconomic Analysis – from Theory to Application). Prague: Grada Publishing, 2025. ISBN 978-80-271-3735-0

In addition to the usual celebrations associated with the beginning of the New Year, another reason to rejoice was the publication of a new book *Macroeconomic Analysis – from Theory to Application*, whose main authors are Jan Vejmělek and Václav Žďárek and other co-authors are Lubomír Chaloupka and Marek Rojíček.

Attentive readers are going to look for an addendum such as 2nd Edition, revised edition, etc. to the title of the new book, as they remember the publication of a book with a similar focus in 2016 entitled *Macroeconomic Analysis – Theory and Practice*. Readers with an even broader knowledge will look for a connection with the book published in 2012 entitled *Macroeconomic Analysis*, and look for an addendum such as 3rd Edition, etc. However, it should be stressed at the outset that although these books have a similar focus, i.e. macroeconomic analysis, it is certainly not appropriate to regard them as merely a series of updated or expanded editions of the same "primer". The opposite is true.

The present book should rather be considered as a text that was compiled practically from scratch, and thus it cannot be spoken of as an expanded version of the book published in 2016 (or earlier).

[1] Department of Monetary Theory and Policy, Faculty of Finance and Accounting, Prague University of Economics and Business, W. Churchill Sq. 1938/4, 130 67 Prague 3, Czechia. Also Office of the Czech Fiscal Council, Holečkova 103/31, 150 00 Prague 5, Czechia. E-mail: ondrej.sima@vse.cz or ondrej.sima@unrr.cz. ORCID: <https://orcid.org/0000-0002-3756-4470>. The text contains the author's personal opinions and views and does not reflect the official position of the Prague University of Economics and Business or the (Office of the) Czech Fiscal Council.

On the one hand, there has been a partial renewal of the author team and a change in the leadership of this team. After the late V. Spěváček, J. Vejmělek became the head of the authors' collective. On the other hand, furthermore, the scope of the new publication speaks for itself, both in terms of the number of pages (888) and, above all, in terms of the considerably new content, which, in addition to current theoretical contributions, reflects the period of the COVID-19 pandemic and, in part, the energy crisis of 2022–2023.

The book consists of eight extensive chapters. The first chapter, *Introduction to Macroeconomic Analysis*, explains the nature of macroeconomic analysis and forecasting and emphasises the need to observe not only the technical terminology but also the need to interpret the results in a comprehensible way. Two boxes stand out in this chapter. The first box offers an informed theoretical information and an overview of time series in terms of their use in economic analysis and forecasting. Although the box does not contain explicit "how-to" guides on how to use various econometric approaches to test relationships between time series, etc., it does provide a valuable source of information that guides the reader to a method that is appropriate for a particular time series analysis. The second box turns its attention to the role of high-frequency data in economic analysis. This type of data became particularly important during the COVID-19 pandemic, and the authors clearly demonstrate this phenomenon in numerous graphs.

The second chapter, *Basic Relationships in Economics*, introduces the reader to the basic concepts and relationships that are necessary to know in order to produce an erudite macroeconomic analysis. The authors present the frequently used concepts of gross domestic product, etc., and place them in a broader context. It is therefore not a "boring" list of definitions and identities. The authors consistently – as throughout the book – supplement the text with illustrative graphs and tables showing the development of the economic variables discussed, mainly using the Czech economy as an example. Since the Czech Republic is a so-called small open economy, the authors do not neglect to discuss the primary and secondary redistribution of income. The second chapter concludes with a box linking internal and external macroeconomic equilibrium.

The third chapter, *National Accounts*, deals with the still often neglected issue of national accounts. National accounts are not a popular topic – perhaps because of its difficulty, less tractability or apparent opacity. The authors, however, keep these concerns of the readers to a minimum and once again balance the theoretical and practical view of the phenomenon of national accounts, supporting this extensive chapter with numerous charts and tables with realities from the Czech economy. The authors' use of actual data makes this difficult subject seem less "detached from reality" to readers, as hopefully everyone will realize that national accounts are nothing more than a mere mirror of the economic reality that surrounds us all. Moreover, the authors do not neglect to analyse the institutional sectors of the national economy (the emphasis on non-residents is given in the next chapter), which makes it even easier to visualise 'reality' by examining in detail the behaviour of non-financial corporations or households, for example.

External Economic Relations is the fourth chapter of the book. This part of the book does not only focus on the balance of payments, but also looks at detailed statistics on foreign trade, terms of trade, (institutional sector of) rest of the world, and the exchange rate. Although the chapter "stands alone", it clearly demonstrates that, especially for the small open Czech economy, it is not sufficient to focus only on internal economic relations (i.e. on the institutional sectors in the national economy), but it is also necessary to monitor external (im)balances, as already mentioned in the second chapter.

Economic Growth and the Supply Side of the Economy are the subject of Chapter 5. The title of the chapter evokes a purely theoretical focus of the text, but the opposite is true. The theoretical concepts of economic growth are used here as a "basis" for data analysis of economic growth of the Czech economy in particular and subsequent comparisons with other economies, mainly from the EU. Although some economic growth concepts – especially endogenous growth models – may seem "impractical" at first glance, a recently published CNB study[2] on the lack of innovation, investment and weak institutional framework in the Czech Republic makes the case that these more difficult-to-measure variables play crucial role in long-term economic growth.

The sixth chapter deals with *the Labour Market*. In this chapter in particular, readers will appreciate the strong link between theory and empirics (statistics). Here again, the authors draw attention to the need to distinguish consistently between the different terms, which are often confused in practice. The section focusing on the Czech labour market is excellent (in one box the authors present, among other things, the LUCI index used by the Czech National Bank). The chapter is well complemented by subchapters on labour productivity and labour costs.

The comprehensive seventh chapter, *Public Finance, Fiscal Policy and Sustainability*, as the title suggests, looks in detail at a major part of economic policy that has attracted more attention than ever in the context of the COVID-19 pandemic and the energy crisis. The authors start by defining the basic concepts which, as in the previous chapter on the labour market, are often confused or even mangled in practice. The subsection on debt dynamics is very valuable, tracing the reasons for the change in the ratio of general government debt to GDP over time. The topicality of the information the authors of the book have worked with is fully demonstrated in the subchapter on fiscal rules in the EU, where the current reform of the EU's fiscal framework from spring 2024 is not omitted. An integral part of the fiscal framework are the so-called independent fiscal institutions, which the authors also address, not forgetting to mention the Czech Fiscal Council. The chapter concludes with a section on the long-term sustainability of public finances.

The final eighth chapter, *Price Developments and Monetary Policy,* focuses on the "media stars" (not only) in the Czech airwaves from about mid-2021 until almost the present day (February 2025). Due to the high inflation rate in Czechia, there has been an unprecedented number of (in)expert articles on this topic explaining the origin of this phenomenon, followed by articles blaming the (un)responsive Czech National Bank, the (previous) government and who knows who or what else for the rapid rise in the price level. The authors of this book have once again managed to "stand up" to this often media-biased description of reality and with neutrality and a positive (i.e. not normative) analysis of their own, they present the measurement of the price level, the reasons for and effects of price movements, and insights into monetary policy issues. In the context of price developments, especially during and after the COVID-19 pandemic, it is worth highlighting the box where the authors present the often humorously, but very aptly, named price changes (see Box 8.2).

It is obvious that a few pages of this review cannot, even with the best efforts, summarize all the essential information contained in this extensive and valuable publication. However, it is hopefully clear from the review that a book of a considerably higher standard is entering our book market in many respects.

1) The composition of the team of authors led by J. Vejmělek suggests that the professionalism and high quality of the book's processing is guaranteed.
2) It is clear from the review that the authors have not "rested on their laurels" and offer readers very up-to-date data and information in their chapters.
3) The book contains a superior number of graphs, tables and charts to help readers understand macroeconomic analysis.
4) The authors have avoided the usual shortcomings often seen in other publications. The first shortcoming is the inclusion of only theory unsupported by evidence, and the opposite one is the inclusion of graphs, tables, etc. without an "economic story".

[2] See: <*https://www.cnb.cz/en/monetary-policy/monetary-policy-reports/boxes-and-articles/The-slowdown-in-the-long-term-potential-growth-of-the-Czech-economy*>.

I am quite sure that every reader will add to this list of positive qualities their own superlatives associated with this book. It is also my sincere hope that the book will become an integral and essential part of the library of every student of economics, economic analyst and commentator, and that it will stimulate interest in the subject among a wider range of people than has been the case to date. This book certainly has the prerequisites.

Impact of EU Membership on Statistical Business Register

Michal Čigáš[1] | *Czech Statistical Office, Prague, Czechia*

On 1st May 2004 the Czech Republic became a member of the European Union. The articles summarizes the main benefits and obligations of EU entering on the Statistical Business Register.

The entering of the Czech Republic into the European Union meant finishing of long term process during which the access consultations were led in order to secure harmonisation of Statistical Business Register and statistical units with relevant acquis communautaire. In this field the primary objective was to ensure that the Statistical Business Register – which was primarily used for national purposes at this time – will be able to prepare sample frame for statistical surveys in line with the other EU member countries. To achieve this objective it was necessary to implement significant changes in the content of Statistical Business Register and to delineate the new statistical units. In addition it was important to get familiar with EU decision procedures in the area of Statistical Business Registers and to arrange contacts with colleagues from other EU member countries.

The obligation that arose from EU membership is the necessity to implement and keep EU regulations in the field of Statistical Business Registers and provide data from Statistical Business Register for business demography statistics and the EuroGroups Register. The compatibility with EU regulations is monitored by Eurostat on yearly basis and in case it is not adequate the Eurostat begins consultations with the country. The Czech Statistical Business Register belongs to the top European business registers since for a long time it is more than 95% compatible with relevant EU regulations.

There are no doubts, that the EU membership brings not only the obligations but also benefits. Several examples of significant benefits for respondents and users can be also found in the field of Statistical Business Registers. One of them are European business demography statistics. Originally it was voluntary project focused on collection of harmonised data on business demography according the methodology produced by Eurostat and OECD. The collection of these data did not have impact on response burden as it was produced exclusively from Statistical Business Registers. The outputs of this project got popular among the statistical users because they provide information about the condition and development of entrepreneurial environment or labour market. The Eurostat contributed significantly to this success as it was able to use best practices from member states by preparing the methodology and provided finances for this project. At present the production of European business demography statistics is produced in all EU member states according to EU regulation and it extends its indicators to other areas of interest. Some of its outputs are available at the CZSO web page.[2]

Another evidence of the benefits from EU membership is development of EuroGroups Register (EGR), which collects basic data on multinational enterprises (MNEs) that operate in several EU countries. This register is managed by Eurostat and EU countries provide it information about MNEs activities in their national territories on regular basis. From these data the whole picture of MNEs is completed

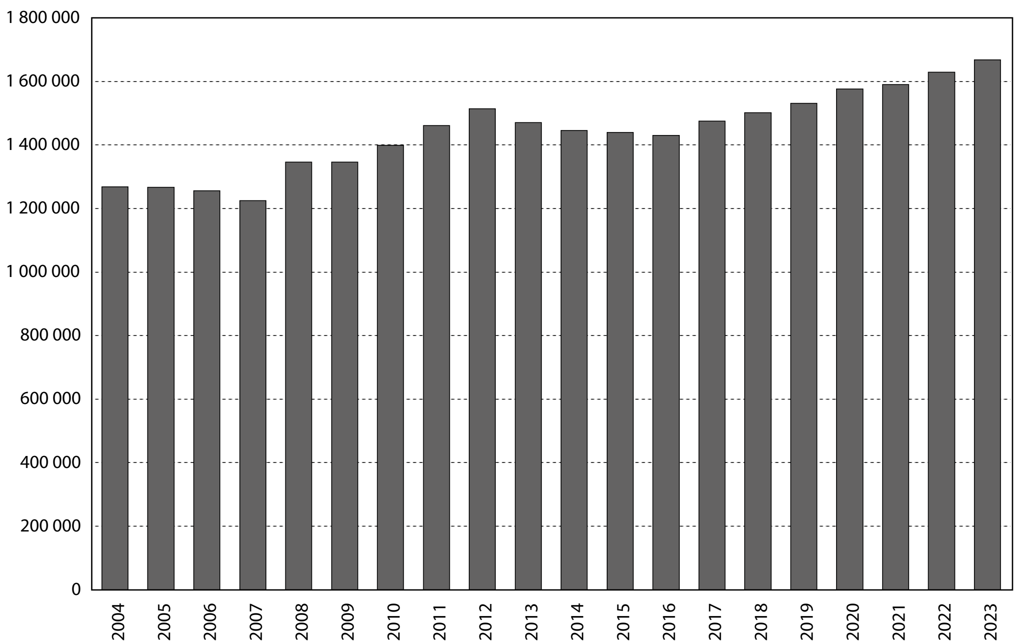
[1] Statistical Registers Department, Czech Statistical Office, Prague, Na Padesátém 3268/81, 100 82 Prague 10, Czechia. E-mail: michal.cigas@csu.gov.cz.

[2] *<https://csu.gov.cz/business-demography-statistics?pocet=10&start=0&podskupiny=142&razeni=-datumVydani>*.

in EGR by Eurostat and consequently it serves to member states as an additional source for creating FATS and other globalisation statistics. Even though EGR still has some problems with the data quality there is no doubt that it would be financially and personally unbearable if each member state developed and managed such register for itself. In this field the mutual international cooperation is crucial as it provides identification of Enterprise groups structures in abroad.

Due to EU grants the Czech Statistical Office (CZSO) is able to implement financially demanding projects in the field of Statistical Business Register that would be hardly possible to finance from national sources. Thus it was able to secure development of the statistical registers framework at CZSO and link it to the basic administrative registers. During this action – that was carried out from the year 2011 till 2014 – the farm register, register for tourism statistics and register of enterprise groups were added to the statistical registers framework. It was co-financed from the Structural European Funds. At present the CZSO is using the financial contribution from the European Recovery and Resilience Facility to implement National recovery plan's projects called SIS 5.0 and ODS that should – among other things – completely modernise the technical solution of statistical register framework, introduce digital services of Statistical Business Register and produce open geographical data from the Register of Census Districts and Buildings. Moreover, the CZSO uses the Eurostat's short term grants on regular basis to implement best practices or improve quality of Statistical Business Register. For instance, at present we are using these grants to implement NACE Rev. 2.1 into Statistical Business register as well as to carry out European profiling of multinational enterprise groups.

Figure 1 Number of enterprises in Czech SBR at the end of the year



Source: CZSO

Furthermore, the European Funds significantly helped CZSO to build the Basic Register of Legal and Natural Persons (ROS), which is currently one of the main pillars of digitalisation and E-government in the Czech Republic as it allows data sharing within the public administration.

Indisputable benefit of EU membership is cross boarder exchange of experience. During the preparing stage for EU membership it was possible to gain experience on creation and harmonisation of Statistical Business Registers from the experts of other EU countries. In this field the CZSO cooperated mainly with the experts from the French Statistical Office (INSEE). Their assistance contributed considerably to the improving professional knowledge and also language skills of the CZSO's staff responsible for the management of statistical registers framework. After entering EU this experience was used by CZSO experts to help other EU accessing or partnering countries. For example in the field of statistical business register the technical assistance was provided to North Macedonia, Azerbaijan or Turkey.

As regards the management of Statistical Business Register the big advantage is also regular participation at the Eurostat's working group on statistical units and business register which deals with various register issues (e.g. legislation, data sources, quality, variables, delineation of statistical units etc.). Important benefit is the possibility to participate on statistical business register legislation or recommendations and suggest solutions based on experience or needs of the CZSO.

Monitoring Sustainable Development through Indicators

Lucie Legnerová[1] | *Czech Statistical Office, Prague, Czechia*
Pavlína Mildnerová[2] | *Czech Statistical Office, Prague, Czechia*

INTRODUCTION

Describing the complex issue of sustainable development is almost as challenging as achieving it. Sustainability permeates our lives perhaps more than one may realize. Sustainable development concerns society as a whole, as well as each individual. It is crucial for the current generation as well as for the future ones. Monitoring sustainable development through indicators that reflect current societal trends is essential for future planning.

1 SUSTAINABILITY

Sustainability has been a frequently discussed concept for several decades. The effort to describe what sustainable behavior entails began to enter the public consciousness in the late 1980s. The questions that arose in connection with the topic were addressed by the international definition of sustainable development, first published in 1987. Sustainable development is development that meets the needs of the present generation without compromising the ability of future generations to meet their own needs. This means we cannot satisfy our current needs at the expense of our children's and grandchildren's needs. It is not ethical to satisfy our hunger knowing that our children will starve. This explicit example conveys an undeniable message. None of us (with few exceptions) would wish to participate in such a scenario.

Satisfying hunger is a physiological need that stands at the very base of Maslow's hierarchy of needs. Physiological needs are undeniable and easily perceived. However, as we move up this hierarchy, we uncover topics that are not as clearly visible but are directly related to sustainable development. In the context of sustainable development, we balance concepts such as "justice," "tolerability," and "viability," which evoke emotions in us. Would any of us want to live on the edge of tolerability? Do we prioritize viability over justice?

Economics involves decision-making about the use of resources and the distribution of goods and services for current and future consumption. The price of goods is determined by supply and demand, but there are goods that do not pass through the traditional market. Their price is not clearly quantified, but they have immense value for the population. What is the price of a certain level of air cleanliness? What is the aesthetic value of the landscape? How much do we value our own health? We cannot precisely quantify these prices, but through studies and research, we can at least approximate them.

[1] Trade, Transport, Services, Tourism, and Environmental Statistics Department, Czech Statistical Office, Na Padesátém 3268/81, 100 82 Prague 10, Czechia. E-mail: lucie.legnerova@csu.gov.cz.
[2] Trade, Transport, Services, Tourism, and Environmental Statistics Department, Czech Statistical Office, Na Padesátém 3268/81, 100 82 Prague 10, Czechia. E-mail: pavlina.mildnerova@csu.gov.cz.

The topic of sustainable development concerns society as a whole and intersects various areas of each individual's life. It affects today's generation as well as the future ones. However, future generations are disadvantaged because we largely decide their potential development at this moment. For decision-making and the creation of coherent public policies (strategies, laws, financial instruments), it is necessary to have relevant information based on facts. Sustainable development is not just a matter for one country but for our entire planet. Therefore, the information obtained is shared with other countries.

Since society is the driving force behind sustainable development, it is necessary to raise public awareness of this issue. Monitoring various indicators and using appropriate communication tools serve to disseminate information about sustainable development. Important prerequisites for strategic planning and management are a participatory approach and regular communication with the (expert) public. Through communication with the (expert) public, it is possible to gather and interpret relevant data, consider user experiences, and explain the intentions of policymakers and limitations, if any.

In 2015, the UN adopted the action plan "Transforming our world: the 2030 Agenda for Sustainable Development," known as Agenda 2030. It contains 17 goals to be achieved by 2030. The aim is to put the world on a sustainable path while leaving no one behind. The goals are indivisible and balance all three pillars of sustainable development: social, economic, and environmental.

The goals have a global scope and are evaluated using an indicator set that currently includes 248 indicators. At the national level, in response to Agenda 2030, the Strategic Framework Czech Republic 2030 was created, transforming the 17 goals of Agenda 2030 into the conditions of Czechia. This document focuses on goals relevant to our country. The goals are divided into 6 key areas with 27 strategic and 104 specific goals. Czech Republic 2030 represents the fundamental document of the state administration for sustainable development and improving the quality of life of the population, outlining areas that are crucial for Czechia in the long term. Achieving the objectives of Czech Republic 2030 should guide our country towards development that is sustainable socially, economically, and environmentally. The Strategic Framework Czech Republic 2030 is evaluated objectively using the Czech Republic 2030 indicator set, which currently contains 170 indicators. A subjective perspective is provided by the Quality of Life indicator set, which includes 39 indicators.

2 MONITORING SUSTAINABLE DEVELOPMENT

Without quality monitoring, we would be unable to identify if we are on the right path and if the measures taken are yielding the desired results. Monitoring sustainable development addresses various aspects of human society's existence. Capturing such a diverse set of data that reflects sustainability requires the collaboration of many experts from ministries, the Office of the Government of the Czech Republic, academia, and other institutions. The entire sustainable development agenda is overseen by the Ministry of the Environment in cooperation with the Ministry of Regional Development. The Government Council for Sustainable Development coordinates the activities of individual participants, under which the Committee for Sustainable Development Indicators, led by the Czech Statistical Office, operates. The Czech Statistical Office is also a crucial data provider.

Data collection is not a one-time affair. As the name suggests, sustainable development is a continuous process. Information that evolves and changes needs to be updated and supplemented to reflect reality as accurately as possible. Based on monitoring progress in sustainable development, measures and recommendations important for the further development of our society are created. Monitoring also helps identify areas where active intervention is most needed.

Coordinating data collection for all indicators of the global strategy Agenda 2030 and the national strategy Czech Republic 2030 is a complex process. In 2024, the Czech Statistical Office collected data for a total of 464 indicators. Data collection involved 36 institutions, and communication took place with 102 data providers.

For effective and quality work on sustainable development indicators, the aforementioned Committee for Sustainable Development Indicators (VPI) was established, which addresses tasks related to indicators together with the Internal Working Group of the Czech Statistical Office for Sustainable Development Indicators (IPS). Members of the VPI and IPS are experts from the statistical sphere and main data providers. Regular meetings have been held twice to three times a year since 2016. In October 2024, the 30th meeting took place. Members of the VPI and IPS discuss indicators and other activities related to sustainable development, are informed about the use of indicators, their changes, the progress of data collection, and the creation of new indicators. All members can also comment on the outputs in which the indicators are used.

Achieving the ideal state, where indicator sets are complete and become a comprehensive tool for evaluating strategies, will take some time. This is especially true for Agenda 2030, for which some data is currently unavailable, and any changes to indicators must be discussed at the international level.

2.1 Agenda 2030

Agenda 2030 is based on the action plan "Transforming our world: the 2030 Agenda for Sustainable Development," adopted in 2015 by the United Nations (UN). This plan aims to achieve 17 global Sustainable Development Goals (SDGs) by 2030 through collective effort. The implementation involves all UN member states, making the character of Agenda 2030 very diverse and including goals primarily focused on developing countries. The goals are further divided into 169 sub-targets.

In 2017, the Inter-agency and Expert Group on SDG Indicators (IAEG-SDGs), established within the UN to manage the indicators of Agenda 2030, created a global indicator framework comprising 248 indicators. This set is used to monitor progress in achieving the SDGs.

The IAEG-SDGs annually submits minor adjustments to the indicator framework to the Statistical Commission (SC) for approval. These changes mainly involve refining the wording of indicators. Every five years, a comprehensive review of the entire indicator set takes place, allowing member states to participate in the feedback process. This review can lead to significant adjustments, such as adding or removing indicators or revising methodologies. The revised indicators are then submitted to the SC for approval. The next review is currently underway, with the SC meeting scheduled for March 2025.

Under the auspices of the UN, the High-level Group for Partnership, Coordination, and Capacity-Building for statistics for the 2030 Agenda for Sustainable Development (HLG-PCCB) was also established within Agenda 2030. This group aims to create a global partnership for sustainable development data.

Although the entire set has 248 indicators, some are repeated across different sub-targets, resulting in 231 unique indicators. Each indicator has detailed metadata specifying how it should be populated with data. The indicators are designed to be internationally comparable in most cases. They are also classified into two levels, known as Tier 1 and Tier 2, based on the level of methodological development and data availability at the global level.

In Czechia, the Czech Statistical Office is responsible for collecting data for the Agenda 2030 indicators. Data collection occurred in two main stages. The first stage took place in 2019, successfully gathering data for 110 indicators. The second stage began in 2021 and continued until mid-2023, during which data for an additional 46 indicators were collected from the remaining 138. In 2023, 43 indicators were deemed irrelevant as they pertained to sub-targets focused on developing or coastal countries. Another 53 indicators were classified as unavailable in 2023 because, although relevant, the required data could not be obtained, and no data custodian was found. The previously collected 153 indicators were continuously updated, with significant changes in some cases, including data and indicator concepts. Some indicators were reclassified as unavailable or irrelevant. As of 2024, the Agenda 2030 set includes 146 available indicators.

The indicators have been continuously used in evaluation reports. The first report, published in January 2021, was the "Report on the Implementation of Agenda 2030 for Sustainable Development in the Czech Republic," utilizing 110 previously collected indicators. This report formed the basis for the second Voluntary National Review (VNR) of Agenda 2030 in Czechia, published in the same year. The VNR assesses the progress of Czechia in achieving the SDGs over the past four years. The VNR can be presented at the High-Level Political Forum on Sustainable Development under the auspices of the UN General Assembly to other UN member states. Czechia first participated in this activity in 2017, again in 2021, and will submit its third VNR in 2025.

The UN has established custodian agencies overseeing specific areas related to the Sustainable Development Goals. These agencies are responsible for the individual indicators of Agenda 2030. They regularly collect data for the Agenda 2030 indicators directly from UN member countries and then compare them internationally.

These agencies contact relevant data providers through questionnaires, mostly online, requesting data submission, feedback on topics, registration, comments, data verification, or updates. The agencies also provide countries with advice and technical support for monitoring Agenda 2030 through consultations, webinars, and conferences.

Communication between the state and agencies is crucial for effectively monitoring progress in achieving the goals of Agenda 2030 and, ultimately, for maintaining sustainable development. The Czech Statistical Office plays a key role in providing this data. While the Czech Statistical Office is not directly a custodian agency (CA), it supplies all necessary data and information to the agencies on behalf of Czechia. CAs typically contact specific focal points (contact persons) provided by the country upon request.

A global database is used to monitor progress in achieving the sustainable development goals defined in Agenda 2030, collecting internationally comparable data for the entire set of Agenda 2030 indicators from all UN member states.

2.2 Strategic Framework Czech Republic 2030

In 2017, the Czech government approved the Strategic Framework Czech Republic 2030, which sets the main priorities and development goals for the country for the specified period. The updated Strategic Framework Czech Republic 2030, with a vision extending to 2050, was approved by the Czech government in November 2024. This framework includes the Czech Republic 2030 indicator set, which is used to evaluate the achievement of goals.

In 2020, the first data collection for 192 Czech Republic 2030 indicators took place, and in the same year, data collection began for the Quality of Life indicator set, which reached a total of 140 indicators. Both indicator sets were intended to serve as one of the bases for evaluating the Strategic Framework Czech Republic 2030, specifically for the First Report on Quality of Life and its Sustainability.

Based on the experience from the first data collection and report writing, revisions were made to both sets. The main change was to set the objective role of the Strategic Framework Czech Republic 2030 to focus more on society as a whole, while the Quality of Life set began to take a subjective view of individual quality of life.

Other changes included removing indicators with insufficient explanatory value, eliminating duplicate or very similar indicators with Agenda 2030, and methodologically refining many indicators. Additionally, several new indicators were added to the set based on the needs of updating the Strategic Framework Czech Republic 2030, which will help better assess the current state of the strategy.

All these changes were made based on joint consultations between the Ministry of the Environment, the Ministry of Regional Development, and the Czech Statistical Office. Experts from the Czech Statistical Office evaluated the indicators and sought suitable data providers from whom they collected metadata and data. They were also responsible for the relevance of the data. A significant advantage of the national indicator set is its creation based on national needs, unlike the Agenda 2030 indicators, whose creation is determined by the UN, making changes to the set very difficult. Currently, the Czech Republic 2030 indicator set contains 170 indicators, and the Quality of Life indicator set contains 39 indicators.

3 PUBLICATION OF INDICATORS ON WEBSITES

The indicators website is one of the main outputs of the PUDR project. It was created as a platform for sharing sustainable development indicators. On the website sdg-data.cz, users can find indicators from the indicator sets – Agenda 2030, Czech Republic 2030, and Quality of Life. All indicators are presented in the form of visualizations supplemented with informative metadata, making the information easily understandable and useful for communication or promotional purposes.

The creation and administration of the website are managed by the Czech Environmental Information Agency which also processed all pre-prepared data from the Czech Statistical Office into the final interactive visualizations. The creation of visualizations was carried out in close cooperation with other partners, including the Czech Statistical Office, whose task was to check each visualization before its publication. During this process, a significant amount of data and metadata was supplemented and adjusted. The website and visualizations are available in both Czech and English, and the displayed data can be downloaded in CSV format.

On the homepage, users can find icons and a brief introduction to the two main indicator sets, namely the global Agenda 2030 set and the national Czech Republic 2030 set. The Quality of Life indicator set falls under the Czech Republic 2030 set, serving as a subjective complement to this objective set. Many indicators from the Agenda 2030 set are not relevant for Czechia, and some are unavailable due to specific data collection methodologies set by the UN. These irrelevant and unavailable indicators are appropriately marked on the <*sdg-data.cz*> website.

By clicking on the introductory icons, users can access the selection of individual goals and key areas. Moving one level further displays all sub-goals or strategic and specific goals and indicators falling under the given goal or key area.

Indicators can be displayed in two basic arrangements: hierarchical and catalog. In the hierarchical arrangement, all indicators are sorted precisely according to their classification under individual goals, key areas, sub-goals, or strategic and specific goals, and numerically in ascending order. In the Czech Republic 2030 set, it is possible to access the relevant Quality of Life indicators under each key area. The catalog overview, on the other hand, offers the option to display all indicators in a list without any classification. These indicators can then be filtered according to certain parameters, allowing users to select those of interest. Filters can be applied based on agenda, goal, key area, data availability, or specific thematic tags. Filters can be combined. Additionally, the indicators website offers links to similar or related indicators on individual visualization pages.

In addition to visualizations, some indicators from Agenda 2030 have been processed into clear infographics by CENIA, which were used in the Report on the Implementation of Agenda 2030 for Sustainable Development in Czechia. These infographics also help familiarize the general public with sustainable development indicators.

4 SUSTAINABLE DEVELOPMENT PROJECTS

From 2019 to 2023, activities were carried out under the project "Mechanisms for Promoting Sustainable Development Principles" (PUDR), reg. no. CZ.03.4.74/0.0/0.0/15_019/00014042, co-financed by the EU through the European Social Fund. The main partners of the project were the Czech Statistical Office, the Ministry of the Environment, the Ministry of Regional Development, and the Czech Environmental

Information Agency. One of the main outputs of the project was the collection of data for the indicators of Agenda 2030, Czech Republic 2030, and Quality of Life, including the creation of the website for these indicators: <*sdg-data.cz*>.

Currently, the Czech Statistical Office, the Czech Environmental Information Agency, the Ministry of the Environment, the Ministry of Regional Development, and the Ministry of Finance work jointly on the project "Policy Coherence for Sustainable Development (Improving Strategic Planning to Achieve Sustainability Goals)," reg. no. CZ.31.3.0/0.0/0.0/23_099/0010365, financed by the EU through the EU Recovery and Resilience Facility. One of the outputs of this project will be the semi-automation of updating data and metadata in the Agenda 2030, Czech Republic 2030, and Quality of Life sets. The main goal is to create a semi-automated system for data updates. The system will notify data providers of the need to update data, allowing them to access the indicators they are responsible for through a user-friendly interface. The purpose of semi-automation is to simplify and increase the independence of data providers in updating existing indicators.

NÁRODNÍ PLÁN OBNOVY

Financováno Evropskou unií
NextGenerationEU

CONCLUSION

Given the nature of sustainable development, it is clear that its continuous monitoring is essential for our society. Sustainable development indicators, which show the development of various trends in society, are the link between the past, present, and future planning. Activities related to the creation of new indicators, revisions of existing ones, and their updates are processes that will need to be maintained throughout the existence of human society.

References

UNITED NATIONS. (2024). *Sustainability* [online]. <https://www.un.org/en/academic-impact/sustainability>.

UNITED NATIONS. (2024). *Transforming our world: the 2030 Agenda for Sustainable Development* [online]. <https://sdgs.un.org/2030agenda >.

MINISTRY OF THE ENVIRONMENT. (2024). *Updated strategic framework Czech Republic 2030 with a view to 2050* [online]. <https://www.cr2030.cz/strategie/wp-content/uploads/sites/2/2024/11/CR-2030-s-vyhledem-do-2050.pdf>.

MINISTRY OF THE ENVIRONMENT. (2022). *Website for sustainable development indicators* [online]. <https://www.sdg-data.cz>.

Recent Events

Conferences

The ***27th Applications of Mathematics and Statistics in Economics Conference (AMSE 2025)*** will take place **from 27th to 31st August 2025 in Hradec Králové, Czechia**. The conference is based on scientific knowledge, mathematical and statistical methods and theoretical and practical problems of economics. The conference gives a unique chance to present research results achieved in the field where mathematics and statistics with economics. More at: <*https://www.amse-conference.eu*>.

The ***33nd Interdisciplinary Information Management Talks (IDIMT 2025)*** will be held **during 3–5 September 2025 in Hradec Králové, Czechia.** Over 30 years of history have established IDIMT-conferences as interdisciplinary international forum for the exchange of concepts and visions in the area of software intensive systems, management and engineering of information and knowledge, social media, business engineering, and related topics. IDIMT involves a multi-national, multidisciplinary audience in discussing up-to-date and evolving topics. More at: <*https://idimt.org*>.

The ***43rd International Conference on Mathematical Methods in Economics (MME 2025)*** will be held **from 3rd to 5th September 2025 in Zlín, Czechia.** The conference is a traditional meeting of professionals from universities and businesses interested in the theory and applications of operations research and econometrics. More at: <*https://www.ifors.org/43rd-international-conference-on-mathematical-methods-in-economics-mme-2025*>.

The ***19th International Days of Statistics and Economics (MSED 2025)*** will take place **during 4–5 September 2025 in Prague, Czechia.** The aim of the conference is to present and discuss current problems of statistics, demography, economics and management and their mutual interconnection. More at: <*http://msed.vse.cz*>.

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 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. Discussion 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. Reviews shall have up to 600 words or 1–2 1.5-spaced pages.

Information section contains informative (descriptive) texts, latest publications, or recent and upcoming scientific conferences. Recommended range 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 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) 111 222 333.

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 underline in the text. Paper parts numbering: 1, 1.1, 1.2, etc.

Headings
1 FIRST-LEVEL HEADING (Times New Roman 12, capitals bold)
1.1 Second-level heading (Times New Roman 12, bold)
1.1.1 Third-level heading (Times New Roman 12, bold italic)
1.1.1.1 Fourth-level heading (Times New Roman 12, 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 Effect 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 and numbers
Provide each table on a separate page. Indicate position of the table by placing in the text "insert Table 1 about here". Number tables in the order of their appearance in the text: 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. Specify the sources below all tables.
Separate thousands with a space (e.g. 1 528 000) and decimal points with a dot (e.g. 1.0).

Figures and pictures/maps
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 "insert Figure 1 about here". Number figures in the order of their appearance in the text: 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 data sources. Specify the sources below all figures.
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.

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@csu.gov.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 placed in the text (for final graphical typesetting must be supplied separately as specified in the instructions above).
Authors are required to read our **Publication Ethics** rules, i.e. *Authors responsibilities* (published at: http://www.csu.gov.cz/statistika journal) and respect them.

Managing Editor: Jiří Novotný
Phone: (+420) 274 054 299
E-mail: statistika.journal@csu.gov.cz | **web:** www.csu.gov.cz/statistika_journal
Address: Czech Statistical Office | Na padesátém 81 | 100 82 Prague 10 | Czechia

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 | Czechia

Subscriptions and orders
Czech Statistical Office | Information Services
Na padesátém 81 | 100 82 Prague 10 | Czechia
Phone: (+420) 274 052 733, (+420) 274 052 783
E-mail: objednavky@csu.gov.cz

Design: Toman Design
Layout: Ondřej Pazdera
Typesetting: Václav Adam
Print: Czech Statistical Office

105th year of the series of professional statistics and economy journals of the State Statistical Service in Czechia: ***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