STATISTIKA

STATISTICS AND ECONOMY JOURNAL

VOL. **106** (2) 2026

CZECH STATISTICAL OFFICE

Dear Readers,

It has been three years since the publication of the last special joint Czech-Hungarian Statistika issue (No. 3/2023), four since the Czech-Polish edition (No. 3/2022) and seven years since the publication of the first joint Czech-Slovakian journal issue (No. 4/2019).

In 2026, *Statistika: Statistics and Economy Journal* celebrates its 106th Volume, while it follows the tradition of the previous state statistics journals published in Czechoslovakia since in 1920, nowadays cited in Scopus and Web of Science scientific databases and has an Impact Factor.

We would like follow up the international cooperation through publication of this new joint Czech-Slovenian special issue of our professional quarterly: to commemorate and remind events, developments and current state and quality of research in official statistics, to strengthen the further developing and very successful cooperation between our national statistical offices as well as current quality and state of research in official statistics in our two countries: therefore, it is composed of articles from the Czech and Slovenian authors only.

We believe that papers published in this special issue will be interesting and beneficial for all its readers. We are looking for further cooperation (not only) with authors and reviewers from our two countries and wish all our colleagues, partners, and collaborators plenty of creative thoughts, professional success, and satisfaction.

Marek Rojíček
President of the Czech Statistical Office

Apolonija Oblak Flander
Director-General of the Statistical Office of Slovenia

CONTENTS

ANALYSES

DISCUSSION

INFORMATION

About Statistika

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

Statistika: Statistics and Economy Journal is professional double-blind peer reviewed open access journal included e.g. in the *Web of Science Emerging Sources Citation Index* (since 2016) of the **Web of Science Core Collection** database (**Impact Factor**), 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

Modeling Industrial Production in the EU Countries: Autoregressive Models Versus Sentiment--Enhanced Regression

Darja Števančec | *University of Ljubljana, Ljubljana, Slovenija*
Patricija Jaklič | *University of Ljubljana, Ljubljana, Slovenija*
Aleša Lotrič Dolinar | *University of Ljubljana, Ljubljana, Slovenija*
Mojca Bavdaž[1] | *University of Ljubljana, Ljubljana, Slovenija*

Received 8.7.2025, **Accepted** (reviewed) 22.8.2025, **Published** 12.6.2026

Abstract

Motivated by rising demand for timely economic insights, we explore the predictive power of the Industrial Confidence Indicator (ICI) for forecasting the Industrial Production Index (IPI) across EU Member States. Our purpose is to assess whether business sentiment data can serve as a real-time leading indicator for industrial production and explore possible structural patterns across EU countries and over time. The study analyses monthly IPI and ICI data for 27 EU Member States from 2008 to 2024 (while also considering only the pre-pandemic part of the time series) using correlation analysis, ARIMA/ARIMAX forecasting methods (with expanding and rolling window techniques) and clustering. Our analysis results in rather weak linear correlation between ICI and IPI, very limited forecasting dominance of ARIMAX models over ARIMA models (especially during volatile periods), and identification of contextually fairly meaningful clusters. Using all three methods the pre-pandemic data turn out to better reflect expected relationships compared to the whole time series.

Keywords	***DOI***	***JEL code***
Business tendencies, time-series, timeliness, real-time estimates, nowcasting, clustering	*https://doi.org/10.54694/stat.2025.29*	*C10, C13, C22, E23*

[1] University of Ljubljana, School of Economics and Business (SEB), Kardeljeva pl. 17, 1000 Ljubljana, Slovenia. Corresponding author: e-mail: mojca.bavdaz@ef.uni-lj.si, phone: (+386)15892630. ORCID: <https://orcid.org/0000-0003-1924-7516>.

INTRODUCTION

Timeliness, one of the five principal quality dimensions of official statistics besides relevance, accuracy and reliability, coherence and comparability, and accessibility and clarity (Eurostat, 2018), has always been particularly relevant for short-term statistics. However, the perception of what constitutes "timely" has evolved significantly, and with it, expectations regarding statistical timeliness. A series of systemic shocks in the 21st century, from the financial crisis to the pandemic and recent wars, has intensified user demand for immediate economic insights. The heightened demand has also been enabled by the proliferation of novel data streams that permit faster, and in some cases, real-time statistical generation. Consequently, we observe a surge in academic and institutional interest in early estimates.

Short-term statistics are the fastest and most frequent official economic statistics, providing the empirical foundation for effective economic governance and analysis. For key stakeholders, including governments and financial institutions, they are fundamental to shaping monetary and fiscal policy, evaluating economic conditions, performing business cycle analysis, and forecasting. In the analysis of the real sector, two short-term indicators, the Gross Domestic Product (GDP) and the Industrial Production Index (IPI), are widely recognized for their extensive time series, broad usage, relative methodological stability, and long-standing tradition. These indicators are crucial as the GDP encompasses a wide range of economic activities and the IPI covers an economic sector that is sensitive to economic fluctuations, thus suitable for the identification of turning points, and has strong linkages to other sectors. While forecasting and nowcasting GDP and IPI have spurred research and academic literature, the GDP has attracted much more attention (Stundziene et al., 2024). We therefore turn our attention to the less researched IPI that is also interesting because of its monthly frequency (as compared to typical quarterly for the GDP), shorter publication lags and time-series properties (Moody et al., 1993). In this paper, we work from the assumption that the accuracy of any early economic estimate is inherently dependent on the timeliness and accuracy of its input data. Namely, our primary objective is to examine the predictive capacity of the Industrial Confidence Indicator (ICI), one of the earliest monthly indicators available (if not the earliest), on the future trajectory of the IPI.

The remainder of this paper is structured as follows: Section 1 provides the literature review and articulates the research questions. Section 2 details the data sources, along with some background information on both studied indicators, the IPI and the ICI, and our methodological approach. Section 3 presents empirical results, followed by a discussion of their implications in Section 4 and the concluding section.

1 LITERATURE REVIEW

As already mentioned, general interest for timelier indicators primarily comes from turbulent environment and timelier new data sources. But the pursuit of accurate early estimates and robust forecasts of the IPI has long been seen in research covering major economies, such as the United States, Germany, the United Kingdom, France, and Italy (e.g. Bruno and Lupi, 2004; Hassani et al., 2009). Such interest is also driven by institutional changes, as was the creation of the Euro area (Bodo et al., 2000; Čižmešija et al., 2011).

Besides testing various methods, research has examined the predictive power of qualitative business surveys, also known as business tendency surveys, that provide qualitative or "soft" data for concurrent indicators (e.g. sales evolvement) and leading indicators (e.g. evolvement of orders or investment). Among the more recent studies, Boshnakov (2018) observed anticipated favourable effect of the business climate, as perceived by the managers in the industrial sector, on the production volume shifts at lag of one month for Bulgaria. Chipeva and Chavdarov (2019) similarly found that selected business indicators help explain the IPI variation in Bulgaria. Ptáčková and Fischer (2020) concluded for Czechia that the individual confidence indicator in the industry well predicts the monthly industrial production index and that the results are significant for the following month and the predictions two months ahead.

Reviewing studies assessing the forecasting power of the ifo Business Survey conducted by the German Ifo Institute, Lehmann (2023) concluded that several ifo indicators have been proven in the literature to be good leading indicators for industrial production despite some studies with contrary results.

However, as systemic shocks jeopardise the performance of forecasting and nowcasting in general, they also trigger debates about the relationship between the qualitative and quantitative data in such periods (see e.g. Bruno et al., 2019; and Sorić et al., 2022; for the 2008 financial crisis). While the effect of the COVID-19 pandemic period on predicting the industrial production has been researched, especially in the context of new data sources, the ability of traditional qualitative survey data to capture major shocks has been modestly investigated. For example, Furukawa and colleagues (2024) concluded that qualitative forecasts are unlikely to capture well changes in Japan industrial production, and Lehmann and Möhrle (2024) showed that high-frequency electricity consumption data beat qualitative ifo indicators for the German state of Bavaria. To the best of our knowledge, the predicting power of qualitative data for industrial production forecasting and nowcasting of the EU Member States has not yet been examined for the periods including the COVID-19 pandemic period. Focusing on the Industrial Production (Volume) Index (IPI) as a quantitative measure of industrial production and the Industrial Confidence Indicator (ICI) as a qualitative measure of business tendencies in industry, we formulated the following research questions (RQs):

RQ1: How do the IPI and the ICI correlate in the period 2008–2024 vs. 2008–2019?

RQ2: How does the inclusion of the ICI impact the accuracy of IPI predictions in the period 2008–2024 vs. 2008–2019?

RQ3: How do countries cluster with respect to their IPI and ICI movement in the period 2008–2024 vs. 2008–2019?

2 DATA AND METHODS

2.1 Data

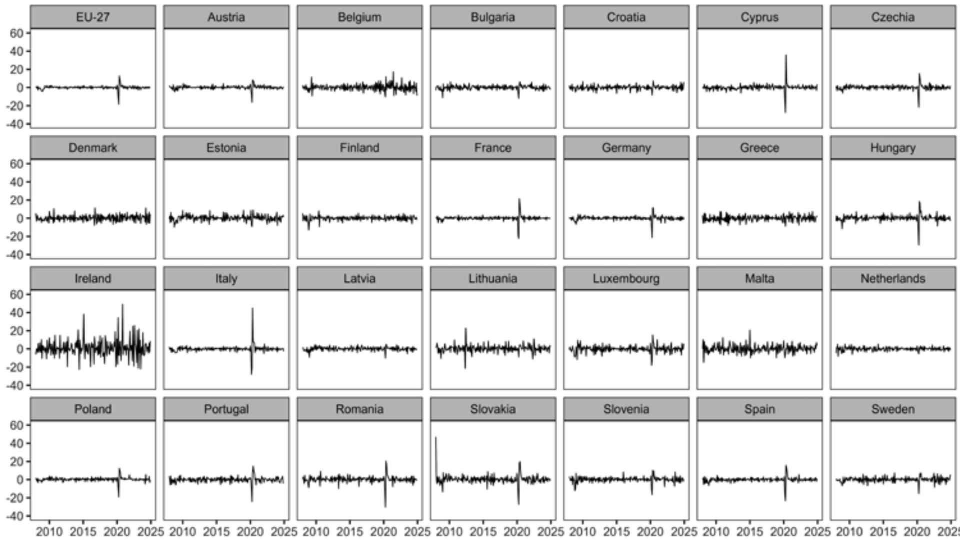
The analysis covers 27 EU Member States in the period between January 2008 and December 2024. The IPI measures monthly changes in the price-adjusted output of industry, with the purpose of tracking the development of value added. The index is calculated in the form of a Laspeyres type index and published with 40-day lag. Although countries may use different data sources (e.g. surveys and administrative data) and different calculation procedures, the geographical and temporal comparability is considered good (Eurostat, 2025c).

The ICI comes from the Joint Harmonised EU Programme of Business and Consumer Surveys that the European Commission launched in 1961 as a timely complement to official statistics (European Commission, 2025). The data are derived from (mostly voluntary) harmonised national surveys based on harmonised questionnaires and a common timetable. A high degree of representativeness is sought in national samples that altogether constitute a nominal sample of 38 000 units and an effective sample of 30 000 units. Answers obtained from the surveys are aggregated in the form of "balances". Balances are constructed as the difference between the percentages of respondents giving positive and negative replies. The ICI is the arithmetic average of the balances (in percentage points) of the answers to the survey questions on production expectations, order books and stocks of finished products (the last with inverted sign) addressed to representatives of the industry (Eurostat, 2024). Fieldwork for the monthly surveys is generally performed in the first two to three weeks of each month, so that the results are sent by email to Directorate-General for Economic and Financial Affairs (DG ECFIN) at least five working days before the end of the reference month (Eurostat, 2025c).

The time-series for the IPI and the ICI were downloaded from the Eurostat website. For the IPI (see Figure 1), we used seasonally and calendar adjusted data in the form of percentage change on previous period (Eurostat, 2025b). For the ICI (see Figure 2), we used seasonally adjusted monthly balances

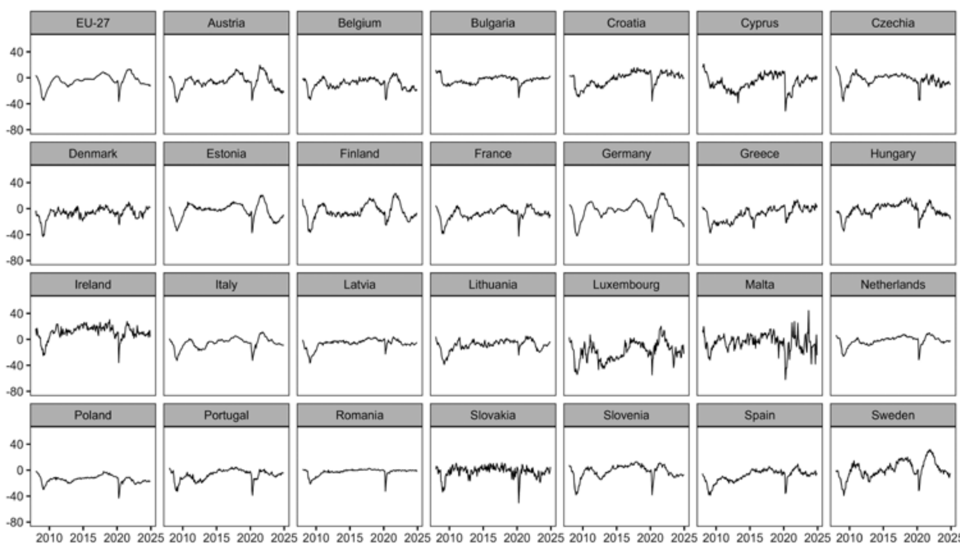
(Eurostat, 2025a). Missing values between January 2008 and April 2008 for Croatia were imputed using the average of May 2008 and December 2008 (all months before the crisis). The missing value for Italy for April 2020 was imputed using the average of March 2020 and May 2020. We can clearly observe the largest disruption in the industrial production at the outburst of the COVID-19 pandemic.

Figure 1 Industrial Production (Volume) Index (IPI) (percentage change on previous period, seasonally and calendar adjusted) by country, January 2008–December 2024



Source: Eurostat (sts_inpr_m, 2025)

Figure 2 Industrial Confidence Indicator (ICI) (seasonally adjusted monthly balances) by country, January 2008–December 2024



Source: Eurostat (ei_bsin_m_r2, 2025)

2.2 Methods

2.2.1 Correlation

To evaluate the relationship between the current business sentiment and the future production values (RQ1), we calculate Pearson's correlation coefficients between the IPI and the ICI. To analyse the forecasting power of business tendency data, we calculate correlations using the leads of 0 to 3 months in the industrial production data. We calculate the correlations for each EU Member State as well as the EU-27 aggregate. The calculations are carried out on the entire period 2008–2024, and separately for the pre-pandemic period 2008–2019. The analysis was performed using the open-source statistical software environment R using the `cor()` function.

2.2.2 Autoregressive integrated moving average (ARIMA) and autoregressive integrated moving average with exogenous inputs (ARIMAX)

To assess the forecasting power of business tendencies in the context of industrial production (RQ2), we compare forecast accuracy for models with and without the ICI as an exogenous variable. For each EU member state, as well as the EU-27 aggregate, we build an autoregressive integrated moving average model (ARIMA) (Box and Jenkins, 1970) on the IPI series and an autoregressive integrated moving average model with exogenous variables (ARIMAX) (Box and Tiao, 1975) that includes the ICI as the exogenous variable. Both models are constructed using the `auto.arima()` function in R with the `stepwise=F` option to improve the model selection by comparing all possible combinations of AR and MA terms (up to order 5).

We use both models independently to forecast industrial production for horizons of 1, 2 and 3 months ahead. Forecast accuracy is evaluated using Root Mean Squared Error (RMSE) as the criterium. Lower RMSE values indicate more accurate models and forecasts.

2.2.3 Expanding (EWF) and rolling window forecasts (RWF)

The principle that more data always leads to better results does not necessarily hold because of volatile data, reflecting financial crises, booms, the pandemic etc. Therefore, we forecast production using both Expanding Window Forecasts (EWF) and Rolling Window Forecasts (RWF) for each country. When EWF outperforms RWF, historic data provides valuable insights into the industrial production forecasts, while when RWF outperforms EWF, past data disturbs the current forecasting accuracy, possibly due to structural changes in the phenomenon.

For the EWF, the starting point remains fixed in January 2008, while the training window is increased by one month of data in each iteration. In each iteration, we fit both ARIMA and ARIMAX models independently, and use the models to generate forecasts for 1, 2 and 3 months ahead. We then use the resulting forecasts for the RMSE calculation. We exclude the initial three years from the RMSE calculation due to the volatile nature of the forecasts attributed to a small data size. The forecasts are prepared for the entire period 2008–2024, and separately for the pre-pandemic period 2008–2019.

While the EWF approach continuously extends its training window, the RWF employs a fixed window size. We choose the window size of 100 months to ensure a sufficient amount of data for reliable model estimation. For both methods, the training window endpoint is moved forward by one month in each iteration. The subsequent processes of model fitting, forecasting, and evaluation are conducted identically for both the EWF and RWF approaches.

2.2.4 Clustering

Clustering (RQ3) is performed on the IPI and ICI standardized data series, using the `scale()` function in R, ensuring that all variables contribute equally to distance calculations regardless of their original scale. Distances are calculated using the Euclidean method using the `dist()` function in R (option `method="euclidean")`. We first apply Ward's method of hierarchical clustering (Ward, 1963) to determine the initial clusters using the `hclust()` function in R (`method="ward.D2"`) for hierarchical clustering and the `cuttree()` function for determining initial groups. We use those clusters to calculate the centres of each cluster that are then used as seeds for *k*-means clustering (MacQueen, 1967) that assigns countries to their final group using the `kmeans()` function.

The clustering is conducted on the entire period 2008–2024 and also on the pre-pandemic period 2008–2019. Highly volatile data for Ireland (see Figures 1 and 2) led to its consistent classification into a separate group, therefore it is excluded from the cluster analysis.

3 RESULTS

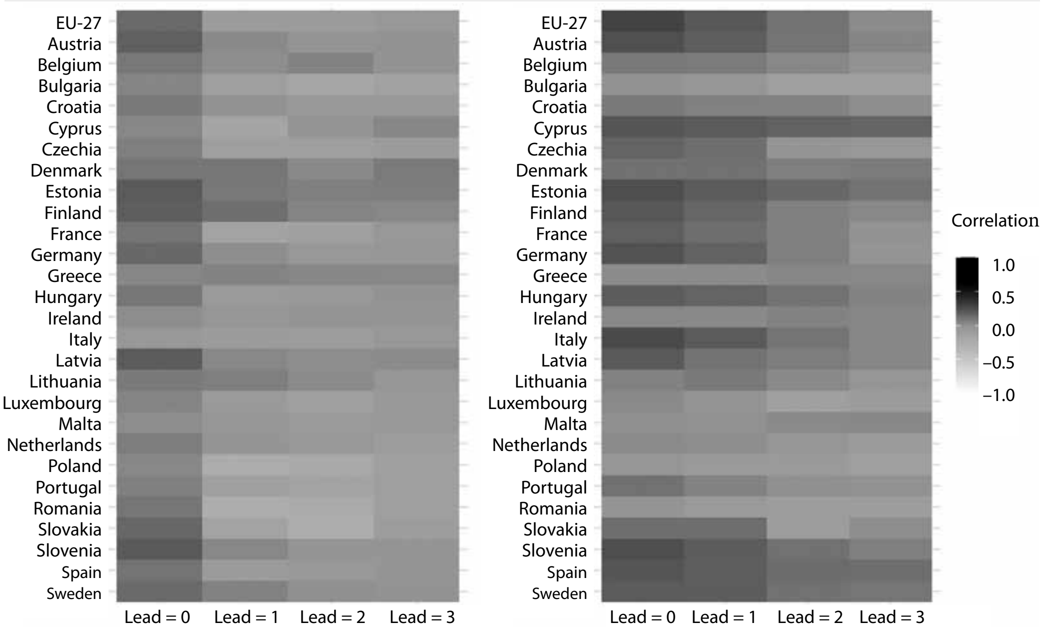
3.1 Correlations

With respect to RQ1, we expect to observe positive correlations between the ICI and IPI. For the contemporaneous relationship, this means that when businesses are more confident, the growth of industrial production tends to be higher, and vice-versa. From the forecasting perspective, this indicates that optimistic business sentiment tends to lead to a higher growth of industrial production, and vice versa. We also expect the strength of relationship to diminish with the length of forecast horizon.

The results for RQ1 show that none of the linear correlations between the IPI and the ICI is considered moderate, given the highest correlation coefficient of 0.29 (EU-27 and lead = 0). Out of 142 positive correlation coefficients 89 have values between 0 and 0.1, and further 53 out of 72 negative correlation coefficients have values between –0.1 and 0, meaning that nearly two thirds exhibit very weak or negligible linear correlation (see Figure 3 and detailed data in Table A1 in the Appendix).

We notice different patterns between countries. Countries with negative correlations for all leads, be it for the entire period or the pre-pandemic period, are Bulgaria, Luxembourg, Poland, Romania and Slovakia. Some other countries (Croatia, Cyprus, Czechia, France, Hungary, Italy, Malta, Portugal

Figure 3 Correlations between the Industrial Production (Volume) Index (IPI) and the Industrial Confidence Indicator (ICI) by country and lead, the entire period (left, 2008–2024) and the pre-pandemic period (right, 2008–2019)



Source: Own calculations based on Eurostat (sts_inpr_m, 2025; ei_bsin_m_r2, 2025) and Spain) show negative correlations only when the entire period is observed, indicating a possible major bias in assessing the effects of the Covid pandemic on industrial production. At least some positive correlations between 0.2 and 0.3 are observed in Austria, Cyprus, Estonia, Finland, Germany, Italy, Latvia, Slovenia and Spain. Generally, we observe the highest correlations when we use the contemporaneous values of the IPI and the ICI, with some exceptions (e.g. Slovakia with the strongest correlation of 0.26 at lead1).

The EU-27 aggregate reflects our previous finding that when analysing the entire period 2008–2024 this correlation is negative for all leads, but when we analyse only the pre-pandemic period 2008–2019, we see a positive correlation that diminishes with higher leads.

3.2 ARIMA and ARIMAX models

3.2.1 Models with an expanding window

To address RQ2, we first present models with an expanding window and then models with a rolling window. For the EFW setting for the entire period 2008–2024 (see Table 1), we find that ARIMAX models fit the data better in almost all cases (except for Bulgaria, Greece, and Slovakia), but the forecasting does not

Table 1 Root Mean Squared Error (RMSE) for ARIMA and ARIMAX models with an expanding window (entire period 2008–2024)

Country	ARIMA				ARIMAX			
	Model	h = 1	h = 2	h = 3	Model	h = 1	h = 2	h = 3
EU-27	1.436	**3.091**	**2.859**	**2.667**	**1.397**	3.283	3.391	2.900
Austria	2.811	**2.799**	2.311	2.252	**2.764**	2.817	**2.257**	**2.217**
Belgium	**2.354**	**3.857**	**3.876**	**3.897**	2.360	3.988	3.893	3.943
Bulgaria	2.176	**2.878**	**2.577**	**2.277**	**2.133**	3.084	2.925	2.319
Croatia	2.714	**2.968**	2.533	2.528	**2.518**	3.117	**2.422**	**2.447**
Cyprus	2.007	**4.769**	**4.357**	**4.227**	**1.964**	5.241	4.684	4.528
Czechia	3.122	**3.225**	3.234	**3.050**	**3.037**	3.284	**3.204**	3.127
Denmark	7.126	**5.230**	4.006	**3.727**	**7.091**	5.584	**3.973**	3.939
Estonia	**2.550**	3.908	**3.233**	**3.176**	2.589	**3.872**	3.287	3.208
Finland	2.289	**2.529**	**2.508**	**2.082**	**2.225**	2.790	2.926	2.519
France	1.975	**4.211**	**3.939**	**3.611**	**1.942**	4.546	4.500	4.028
Germany	2.141	**2.852**	**2.792**	2.773	**2.111**	2.875	2.797	**2.773**
Greece	2.619	**4.954**	**4.524**	**4.088**	**2.573**	5.718	5.531	5.007
Hungary	2.547	**5.682**	**4.533**	**4.045**	**2.523**	6.070	4.947	4.301
Ireland	2.097	10.403	10.381	10.383	**2.024**	**10.334**	**10.372**	**10.329**
Italy	4.140	**4.915**	**4.861**	**4.839**	**4.100**	4.982	4.872	4.844
Latvia	3.641	**2.461**	**2.242**	**2.243**	**3.632**	2.492	2.271	2.246
Lithuania	3.070	**4.614**	**4.628**	**4.378**	**3.011**	4.871	4.849	4.437
Luxembourg	4.229	**4.197**	3.889	3.988	**4.201**	4.305	**3.865**	**3.954**
Malta	2.290	**4.388**	**4.321**	**4.297**	**2.286**	4.392	4.325	4.301
Netherlands	1.748	**1.998**	**1.886**	**1.874**	**1.694**	2.148	2.016	2.000
Poland	1.754	**3.223**	**3.078**	**2.694**	**1.751**	3.596	3.447	2.939
Portugal	2.571	**3.629**	**3.546**	**3.384**	**2.537**	4.018	3.967	3.588
Romania	3.187	**4.411**	**4.322**	**4.214**	**3.182**	4.976	4.608	4.317
Slovakia	2.962	**4.502**	**4.441**	**4.384**	**2.828**	4.611	4.497	4.386
Slovenia	**5.805**	4.198	3.717	3.183	5.822	**4.051**	**3.676**	**3.067**
Spain	2.673	**3.063**	**3.144**	**3.122**	**2.529**	3.184	3.214	3.169
Sweden	2.367	2.956	3.049	2.992	**2.314**	**2.804**	**2.907**	**2.895**

Note: Bolded RMSE signals which model (ARIMA or ARIMAX) performs better in terms of a lower RMSE. For example, for EU-27, ARIMAX better describes the IPI of the same month, while ARIMA better predicts one (h = 1), two (h = 2) or three (h = 3) months ahead.
Source: Own calculations

reflect that as in most cases ARIMAX models produce less accurate forecasts reflected in higher RMSEs. The dominance of the ICI over univariate ARIMA models in terms of forecasting power can be observed for Belgium, Denmark (for h=1), Estonia, Croatia, Lithuania, Malta, Slovenia (all for h = 1 and h = 2), Ireland (for h = 2), Greece (for h = 2 and h = 3), and Slovakia and Sweden (for h = 1, h = 2 and h = 3).

The results for the pre-pandemic period 2008–2019 are similar (see Table 2). ARIMA models performed worse in the sense of model fitting, with exceptions again being Bulgaria, Greece and Slovakia; however, the forecasting accuracy does not show a clear pattern. Most countries, except for France, Italy, Romania and Finland, show at least one forecasting horizon in which the accuracy of ARIMAX models is higher compared to ARIMA models. But only in Greece, Cyprus, Malta, Austria, Portugal, Slovenia, Slovakia and Sweden some forecasting power of business tendencies can be observed regardless of the forecasting horizon.

Table 2 Root Mean Squared Error (RMSE) for ARIMA and ARIMAX models with an expanding window (pre-pandemic period 2008–2019)

Country	ARIMA				ARIMAX			
	Model	h = 1	h = 2	h = 3	Model	h = 1	h = 2	h = 3
EU-27	1.061	**0.931**	**0.919**	**0.863**	**1.016**	0.934	0.924	0.880
Austria	2.592	**1.548**	**1.381**	**1.344**	**2.536**	1.612	1.420	1.365
Belgium	2.347	**2.806**	2.812	2.769	**2.347**	2.809	**2.793**	**2.763**
Bulgaria	1.819	1.971	1.895	1.857	**1.763**	**1.969**	**1.878**	**1.856**
Croatia	2.641	**2.906**	2.529	2.498	**2.406**	3.046	**2.396**	**2.383**
Cyprus	1.699	2.179	2.191	2.167	**1.654**	**2.141**	**2.142**	**2.114**
Czechia	3.093	**1.994**	**1.871**	**1.748**	**3.003**	2.074	1.957	1.919
Denmark	6.318	**4.850**	3.487	**3.481**	**6.267**	5.260	**3.339**	3.859
Estonia	**2.488**	3.603	3.015	2.900	2.543	**3.541**	**2.997**	**2.872**
Finland	1.901	1.856	**1.880**	**1.857**	**1.843**	**1.856**	1.892	1.873
France	1.469	**1.393**	**1.438**	**1.357**	**1.435**	1.403	1.457	1.397
Germany	2.111	1.660	1.442	**1.431**	**2.075**	**1.636**	**1.434**	1.448
Greece	1.731	**2.752**	**2.778**	**2.781**	**1.670**	2.779	2.790	2.817
Hungary	1.909	**2.542**	**2.276**	**2.202**	**1.876**	2.601	2.279	2.227
Ireland	2.090	8.469	8.661	8.629	**1.985**	**8.397**	**8.637**	**8.570**
Italy	4.274	**2.178**	**1.531**	**1.403**	**4.239**	2.219	1.552	1.439
Latvia	3.596	**1.878**	**1.852**	**1.881**	**3.587**	1.917	1.892	1.895
Lithuania	2.676	4.539	**4.497**	4.491	**2.602**	**4.510**	4.499	**4.477**
Luxembourg	4.248	3.632	3.136	3.150	**4.224**	**3.585**	**2.929**	**2.990**
Malta	2.428	**4.720**	4.556	**4.499**	**2.426**	4.721	**4.556**	4.500
Netherlands	1.545	**1.779**	**1.744**	**1.759**	**1.493**	1.815	1.782	1.764
Poland	1.432	**1.308**	1.395	1.378	**1.424**	1.325	**1.389**	**1.378**
Portugal	2.282	2.114	2.287	2.221	**2.244**	**2.046**	**2.233**	**2.194**
Romania	2.829	**2.373**	**2.524**	**2.524**	**2.825**	2.375	2.528	2.536
Slovakia	2.933	**2.885**	**2.945**	**2.947**	**2.793**	3.006	2.981	2.977
Slovenia	**6.031**	2.637	2.355	2.217	6.042	**2.433**	**2.204**	**2.131**
Spain	2.864	**1.426**	**1.448**	**1.424**	**2.690**	1.621	1.562	1.499
Sweden	2.242	2.262	2.418	2.424	**2.197**	**2.178**	**2.362**	**2.374**

Note: Bolded RMSE signals which model (ARIMA or ARIMAX) performs better in terms of a lower RMSE. For example, for EU-27, ARIMAX better describes the IPI of the same month, while ARIMA better predicts one (h = 1), two (h = 2) or three (h = 3) months ahead.
Source: Own calculations

3.2.2 Models with a rolling window

For the RWF, setting the results are more conclusive. Here ARIMAX performs better in fitting the data in the entire period 2008–2024 (see Table A2 in the Appendix), with exceptions being Greece and France. But similarly to EWF, forecasts are predominantly more accurate using ARIMA models. Only in Estonia, Luxembourg and Finland, the ICI has some forecasting power for the IPI.

Results for the pre-pandemic period 2008–2019 (see Table A3 in the Appendix) are more in line with the expectations. Greece and Luxembourg are the only countries where business tendencies do not improve the model fit. Except for Ireland, Lithuania, Luxembourg and Netherlands, at least one forecasting horizon shows better forecasting accuracy when including business tendency data in the model. However, only for Belgium, Hungary, Slovenia, Slovakia and Sweden, there is noticeable forecasting power of business tendencies regardless of the forecasting horizon. This is not true for the EU-27 aggregate.

3.3 Clustering

This section presents the results of clustering, which was designed to answer RQ3. After excluding Ireland due to high volatility of its data, the cluster analysis result with three groups (see Table 3 and Figure A1 in the Appendix) turns out to be the most contextually meaningful, although one of the three groups is again a single country, Luxembourg.

Table 3 Three clusters based on the ICI and IPI

Cluster	Entire period 2008–2024			
	Mean ICI	Var ICI	Mean IPI	Var IPI
1: Austria, Bulgaria, Croatia, Czechia, Estonia, Finland, Germany, Hungary, Malta, Netherlands, Romania, Slovakia, Slovenia, Sweden	–2.7	108.8	0.110	10.643
2: Belgium, Cyprus, Denmark, France, Greece, Italy, Latvia, Lithuania, Poland, Portugal, Spain	–8.0	80.9	0.089	11.628
3: Luxembourg	–18.5	217.2	–0.055	14.627
Cluster	**Pre-pandemic period 2008–2019**			
	Mean ICI	**Var ICI**	**Mean IPI**	**Var IPI**
1: Austria, Bulgaria, Croatia, Czechia, Estonia, Finland, Germany, Hungary, Malta, Netherlands, Romania, Slovakia, Slovenia, Sweden	–2.4	91.8	0.095	8.007
2: Belgium, Cyprus, Denmark, France, Greece, Italy, Latvia, Lithuania, Poland, Portugal, Spain	–8.4	79.2	0.023	6.429
3: Luxembourg	–21.0	194.4	–0.128	10.175

Source: Own calculations

The differences between groups stem mainly from the differences in the ICI, however Luxembourg's IPI is negative as opposed to the other two groups. Performing the analysis taking only the pre-pandemic period into account doesn't change the allocation into clusters compared to the entire period, though. However, for the pre-pandemic period the mean of IPI is considerably larger in group 2 compared to group 1 with considerably lower mean of ICI, worse sentiment coupled with worse industrial production. The results based on the entire period show that the mean of ICI worsens for group 2 where the mean of IPI is slightly reduced while improves for group 1 where the mean of IPI considerably improves.

The relatively more stable pre-pandemic period reveals the relationship between the mean ICI and the mean IPI across the groups that can be expected from the contextual point of view: higher (less negative) mean ICI corresponds to higher mean IPI. Similar, but much less pronounced pattern can be observed also for the entire period.

DISCUSSION AND CONCLUSION

Persistently repetitive shocks in the business cycles, caused by various disruptions (financial crisis, COVID pandemic, recent wars), have intensified the need for immediate economic insights. This study aimed to evaluate the relationship between business tendencies, as captured by the ICI, one of the earliest and most frequent sentiment indicators, and actual industrial production, as captured by the IPI, across the EU Member States, using monthly data from the period 2008–2024. Besides the initial correlation analysis, we also investigated the utility of incorporating the ICI as an exogenous variable in ARIMA forecasting models for industrial production. In addition, we performed clustering analysis for the EU Member States.

Our analysis reveals generally weak linear correlations between the ICI and IPI across the EU Member States in the entire studied period 2008–2024 as in the pre-pandemic period 2008–2019. This finding is somewhat unexpected, as we initially anticipated a stronger positive correlation, given that business tendencies are in principle considered a leading indicator of economic activity.

The analysis of the time series using the classical univariate ARIMA models as a forecasting tool aimed at improving the forecasts by including the ICI as an exogenous explanatory variable within ARIMAX models. Consistent with the weak observed correlation, including the ICI as an exogenous variable in the ARIMAX forecasting models did not yield a convincing improvement in forecasting accuracy, particularly when evaluated over the entire period 2008–2024.

However, the cluster analysis results are more in line with the expected positive relationship between the ICI and the IPI. After excluding Ireland as an outlier (due to its considerably higher IPI volatility), we ended up with three clusters of EU Member States, where again one of the clusters consisted only of one country, i.e. Luxembourg, which is consistent with its distinctiveness in economic research due to the dominating financial sector (e.g. Böwer and Guillemineau, 2006; Lehwald, 2013). But for the other two clusters the relationship between the mean IPI and the mean ICI is rather clear: the higher the ICI, the higher the IPI. This holds for both studied periods but is much more pronounced for the pre-pandemic period 2008–2019.

The study acknowledges several limitations. Not only the entire period 2008–2024, also the pre-pandemic period was marked with turbulences (e.g. the global financial crisis and euro-zone debt crisis), which represent inherent challenges for identifying stable statistical relationships. The assumption of linearity in Pearson correlation and standard ARIMA models may also not capture the complex relationship between qualitative assessments and quantitative measures. Furthermore, while the ICI is a valuable indicator, it reflects sentiment and expectations, which might not always perfectly translate to real economic output. We recommend that official statistics institutions further assess the relationship between the ICI and IPI, explore alternative operationalisations of the business tendencies and produce other real-time indicators to supplement sentiment data, while policymakers rely on a broader set of evidence beyond a single sentiment indicator.

In conclusion, this study successfully achieved its primary goal of assessing the predictive power of the ICI as a leading indicator for the IPI across EU countries and over time. While business sentiment theoretically holds promise as a leading indicator for industrial production, its practical utility in a highly volatile period like 2008–2024 appears limited when assessed through traditional linear correlation and inclusion of an exogenous variable in ARIMA models. Future research should delve into the limitations of the current study to better understand the outcomes of this study.

ACKNOWLEDGMENT

The authors would like to express their gratitude to the Statistical Office of the Republic of Slovenia and the Czech Statistical Office for the opportunity to present this research. The authors also wish to acknowledge the assistance of several digital tools in the preparation of this paper. Specifically, Google

Scholar, Web of Science, Elicit and ClickUp were used to identify relevant literature, and Gemini and MS Copilot were utilized to enhance the clarity, precision and flow of the language.

References

BÖWER, U., GUILLEMINEAU, C. (2006). Determinants of Business Cycle Synchronisation across Euro Area Countries [online]. *ECB Working Paper No. 587*. <http://dx.doi.org/10.2139/ssrn.880429>.

BODO, G., GOLINELLI, R., PARIGI, G. (2000). Forecasting industrial production in the Euro area [online]. *Empirical Economics,* 25: 541–561. <https://doi.org/10.1007/s001810000032>.

BOSHNAKOV, V. (2018). Business Climate and Production Dynamics Nexus: Empirical Evidence for Bulgarian Manufacturing Sector [online]. *Hradec Economic Days*. <https://doi.org/10.36689/uhk/hed/2018-01-011>.

BOX, G. E. P., JENKINS, G. M. (1970). *Time series analysis; forecasting and control*. San Francisco: Holden-Day.

BOX, G. E. P., TIAO, G. C. (1975). Intervention Analysis with Applications to Economic and Environmental Problems [online]. *Journal of the American Statistical Association,* 70(349): 70–79. <https://doi.org/10.1080/01621459.1975.10480264>.

BRUNO, G., LUPI, C. (2004). Forecasting industrial production and the early detection of turning points [online]. *Empirical Economics,* 29: 647–671. <http://doi.org/10.1007/s00181-004-0203-y>.

BRUNO, G., CROSILLA, L., MARGANI, P. (2019). Inspecting the Relationship Between Business Confidence and Industrial Production: Evidence on Italian Survey Data [online]. *Journal of Business Cycle Research,* 15: 1–24. <https://doi.org/10.1007/s41549-018-00033-4>.

CHIPEVA, S., CHAVDAROV, V. (2019). One Opportunity for Forecasting the Total Industrial Production Index (IPI) based on Business Indicators in Industry [online]. *9th International conference on application of information and communication technology and statistics in economy and education (ICAICTSEE – 2019)*, Bulgaria. <https://www.researchgate.net/publication/359454554_ICAICTSEE-2019>.

ČIŽMEŠIJA, M., COPAK, M., KURNOGA ŽIVADINOVIĆ, N. (2011). Industrial confidence indicator and manufacturing industry in the Euro area [online]. *Proceedings of the 11th International Symposium on Operational Research SOR'11*, Slovenia. <https://drustvo-informatika.si/uploads/documents/d7bd350e-fcab-429a-8faf-ab884f15ecfd//SOR11Proceedings.pdf>.

EUROPEAN COMMISSION. (2025). *The Joint Harmonised EU Programme of Business and Consumer Surveys. User Guide* (updated February 2025) [online]. <https://ec.europa.eu/economy_finance/db_indicators/surveys/documents/methodological_guidelines/bcs_user_guide.pdf>.

EUROSTAT. (2018). *European Statistics Code of Practice – revised edition 2017* [online]. Publications Office of the European Union. <https://ec.europa.eu/eurostat/documents/4031688/8971242/KS-02-18-142-EN-N.pdf>.

EUROSTAT. (2024). *Business and consumer surveys (ei_bcs)* [online]. Reference Metadata in Euro SDMX Metadata Structure, October 29. [cit. 14.1.2025]. <https://ec.europa.eu/eurostat/cache/metadata/en/ei_bcs_esms.htm>.

EUROSTAT. (2025a). *Business Surveys – NACE Rev. 2 activity: Industry – monthly data (ei_bsin_m_r2)* [online]. January 8. [cit. 14.1.2025]. <https://doi.org/10.2908/EI_BSIN_M_R2>.

EUROSTAT. (2025b). *Production in industry – monthly data (sts_inpr_m)* [online]. January 14. [cit. 14.1.2025]. <https://doi.org/10.2908/STS_INPR_M>.

EUROSTAT. (2025c). *Short-term business statistics (sts)* [online]. Reference Metadata in Euro SDMX Metadata Structure, January 14. [cit. 14.1.2025]. <https://ec.europa.eu/eurostat/cache/metadata/en/sts_esms.htm>.

FURUKAWA, K., HISANO, R., MINOURA, Y., YAGI, T. (2024). A nowcasting model of industrial production using alternative data and machine learning approaches [online]. *Japan and the World Economy*, 71. <https://doi.org/10.1016/j.japwor.2024.101271>.

HASSANI, H., HERAVI, S., ZHIGLJAVSKY, A. (2009). Forecasting European industrial production with singular spectrum analysis [online]. *International Journal of Forecasting,* 25: 103–118. <https://doi.org/10.1016/j.ijforecast.2008.09.007>.

HASSANI, H., HERAVI, S., ZHIGLJAVSKY, A. (2013). Forecasting UK industrial production with multivariate singular spectrum analysis [online]. *Journal of Forecasting,* 32: 395–408. <http://doi.org/10.1002/for.2244>.

LEHMANN, R. (2023). The Forecasting Power of the Ifo Business Survey [online]. *Journal of Business Cycle Research,* 19: 43–94. <https://doi.org/10.1007/s41549-022-00079-5>.

LEHMANN, R., MÖHRLE, S. (2024). Forecasting regional industrial production with novel high-frequency electricity consumption data [online]. *Journal of Forecasting*, 43(6): 1918–1935. <https://doi.org/10.1002/for.3116>.

LEHWALD, S. (2013). Has the Euro changed business cycle synchronization? Evidence from the core and the periphery [online]. *Empirica*, 40: 655–684. <https://doi.org/10.1007/s10663-012-9205-8>.

MACQUEEN, J. (1967). Multivariate observations. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability,* 1: 281–297.

MOODY, J. et al. (1993). Predicting the U.S. Index of Industrial Production. Proceedings of PASE '93: Parallel Applications in Statistics and Economics. *Neural Network World*, 3(6): 791–794.

PTÁČKOVÁ, V., FISCHER, J. (2020). Can the Business Tendency Survey Predict the Economic Indicators in Czechia? [online]. *Statistika: Statistics and Economy Journal*, 100(3): 256–264. <https://csu.gov.cz/docs/107508/3c1658c3-6753-2b23-f66b-fdad3ca83c64/32019720q3_ptackova_analyses.pdf?version=1.0>.
SORIĆ, P., ŠKRABIĆ PERIĆ, B., MATOŠEC, M. (2022). Breaking new grounds: a fresh insight into the leading properties of business and consumer survey indicators [online]. *Quality & Quantity,* 56: 4511–4535. <https://doi.org/10.1007/s11135-021-01306-4>.
STUNDZIENE, A., PILINKIENE, V., BRUNECKIENE, J., GRYBAUSKAS, A., LUKAUSKAS, M., PEKARSKIENE, I. (2024). Future directions in nowcasting economic activity: a systematic literature review [online]. *Journal of Economic Surveys*, 38: 1199–1233. <https://doi.org/10.1111/joes.12579>.
WARD Jr., J. H. (1963). Hierarchical grouping to optimize an objective function [online]. *Journal of the American Statistical Association,* 58(301): 236-244. <https://doi.org/10.1080/01621459.1963.10500845>.

APPENDIX

Table A1 Correlations between the Industrial Production (Volume) Index (IPI) and the Industrial Confidence Indicator (ICI) by country and lead, entire period (2008–2024) and pre-pandemic period (2008–2019)

Country	Entire period				Pre-pandemic period			
	Lead = 0	Lead = 1	Lead = 2	Lead = 3	Lead = 0	Lead = 1	Lead = 2	Lead = 3
EU-27	0.14	**–0.07**	**–0.08**	**–0.03**	0.29	0.21	0.11	0.03
Austria	0.17	0.03	0.00	0.01	0.23	0.18	0.11	0.05
Belgium	0.09	0.02	0.06	0.00	0.09	0.08	0.04	0.00
Bulgaria	0.05	**–0.11**	**–0.19**	**–0.15**	0.01	**–0.01**	**–0.12**	**–0.11**
Croatia	0.09	0.01	**–0.04**	**–0.04**	0.09	0.07	0.06	0.02
Cyprus	0.04	**–0.17**	**–0.01**	0.04	0.21	0.20	0.17	0.15
Czechia	0.07	**–0.11**	**–0.11**	**–0.08**	0.16	0.12	0.00	**–0.03**
Denmark	0.10	0.10	0.03	0.08	0.11	0.12	0.07	0.08
Estonia	0.19	0.09	0.07	0.07	0.24	0.19	0.14	0.11
Finland	0.18	0.12	0.05	0.04	0.21	0.15	0.06	0.04
France	0.11	**–0.16**	**–0.12**	**–0.03**	0.17	0.13	0.07	0.00
Germany	0.14	0.02	**–0.03**	**–0.02**	0.23	0.17	0.07	**–0.01**
Greece	0.04	0.06	0.04	0.04	0.03	0.03	0.04	0.04
Hungary	0.10	**–0.07**	**–0.04**	0.01	0.18	0.16	0.11	0.06
Ireland	0.03	**–0.01**	0.00	0.00	0.04	0.03	0.06	0.04
Italy	–0.04	**–0.09**	**–0.10**	**–0.03**	0.25	0.19	0.11	0.04
Latvia	0.19	0.04	0.02	0.03	0.20	0.11	0.08	0.04
Lithuania	0.09	0.07	0.03	**–0.01**	0.06	0.08	0.04	**–0.02**
Luxembourg	0.05	**–0.06**	**–0.12**	**–0.05**	0.03	–0.01	–0.13	–0.08
Malta	0.02	**–0.02**	**–0.06**	**–0.03**	0.02	0.01	0.03	0.04
Netherlands	0.08	0.00	**–0.03**	**–0.10**	0.03	0.02	**–0.01**	**–0.08**
Poland	0.04	**–0.25**	**–0.20**	**–0.12**	**–0.02**	**–0.06**	**–0.07**	**–0.11**
Portugal	0.07	**–0.13**	**–0.16**	**–0.11**	0.11	0.05	0.02	0.01
Romania	0.10	**–0.27**	**–0.22**	**–0.13**	0.00	**–0.07**	**–0.09**	**–0.10**
Slovakia	0.15	**–0.16**	**–0.26**	**–0.10**	0.12	0.13	**–0.10**	0.02
Slovenia	0.20	0.04	0.00	**–0.01**	0.24	0.18	0.11	0.07
Spain	0.11	**–0.08**	**–0.05**	0.00	0.21	0.18	0.13	0.12
Sweden	0.14	0.06	0.01	0.00	0.19	0.18	0.12	0.10

Source: Own calculations

Table A2 Root Mean Squared Error (RMSE) for ARIMA and ARIMAX models with a rolling window (entire period 2008–2024)

Country	ARIMA				ARIMAX			
	Model	**h = 1**	**h = 2**	**h = 3**	**Model**	**h = 1**	**h = 2**	**h = 3**
EU-27	1.776	**4.240**	**3.590**	**3.270**	1.756	4.534	4.539	3.815
Austria	2.998	**3.540**	**2.657**	2.610	**2.982**	3.602	2.689	**2.610**
Belgium	2.065	**4.676**	**4.648**	**4.692**	**2.049**	4.924	4.710	4.738
Bulgaria	2.496	**3.511**	**3.120**	**2.490**	**2.468**	4.161	4.234	2.637
Croatia	2.786	3.266	2.523	2.576	**2.752**	**3.253**	**2.456**	**2.495**
Cyprus	2.165	**6.327**	**5.199**	**5.248**	**2.132**	6.778	5.765	5.343
Czechia	2.851	**3.608**	**3.719**	3.678	**2.818**	3.860	3.761	**3.678**
Denmark	8.983	**5.384**	**4.755**	4.097	**8.848**	5.970	4.941	**4.066**
Estonia	**2.517**	**3.793**	**3.358**	3.352	2.541	3.808	3.381	**3.345**
Finland	2.481	**3.411**	**2.896**	**2.660**	**2.454**	3.662	3.415	2.940
France	**2.451**	**5.898**	**4.993**	**4.708**	2.522	6.020	5.192	4.817
Germany	2.103	**3.320**	**3.357**	**3.327**	**2.067**	3.613	3.449	3.341
Greece	3.544	7.039	**4.998**	5.441	**3.472**	**7.006**	6.904	**5.268**
Hungary	3.342	**7.512**	**5.601**	**4.735**	**3.305**	7.610	6.563	5.493
Ireland	1.913	11.351	**11.369**	**11.287**	**1.906**	**11.331**	11.382	11.300
Italy	3.633	5.944	6.033	6.031	**3.606**	**5.882**	**6.019**	**6.030**
Latvia	3.144	3.505	2.917	2.600	**3.143**	**3.298**	**2.685**	**2.522**
Lithuania	3.338	**5.178**	**4.664**	**4.111**	**3.312**	5.639	5.511	4.664
Luxembourg	4.118	**4.203**	**3.968**	**4.206**	**4.094**	4.493	4.135	4.245
Malta	1.722	3.664	3.747	**3.772**	**1.707**	**3.655**	**3.740**	3.795
Netherlands	1.844	**2.461**	**2.044**	**1.984**	**1.803**	2.752	2.362	2.351
Poland	2.119	**4.210**	**3.744**	**3.315**	**2.097**	4.755	4.573	3.606
Portugal	2.795	**4.357**	**4.254**	**4.155**	**2.771**	5.435	5.272	4.584
Romania	3.498	**5.775**	**5.346**	**4.971**	**3.473**	7.402	6.872	5.050
Slovakia	2.554	**5.364**	**5.137**	**4.982**	**2.465**	5.959	5.664	5.140
Slovenia	3.795	**4.796**	**4.423**	3.695	**3.741**	5.040	5.094	**3.630**
Spain	1.829	**3.735**	**3.841**	**3.828**	**1.790**	3.795	3.863	3.845
Sweden	2.390	**3.230**	**3.175**	**3.188**	**2.361**	3.233	3.240	3.243

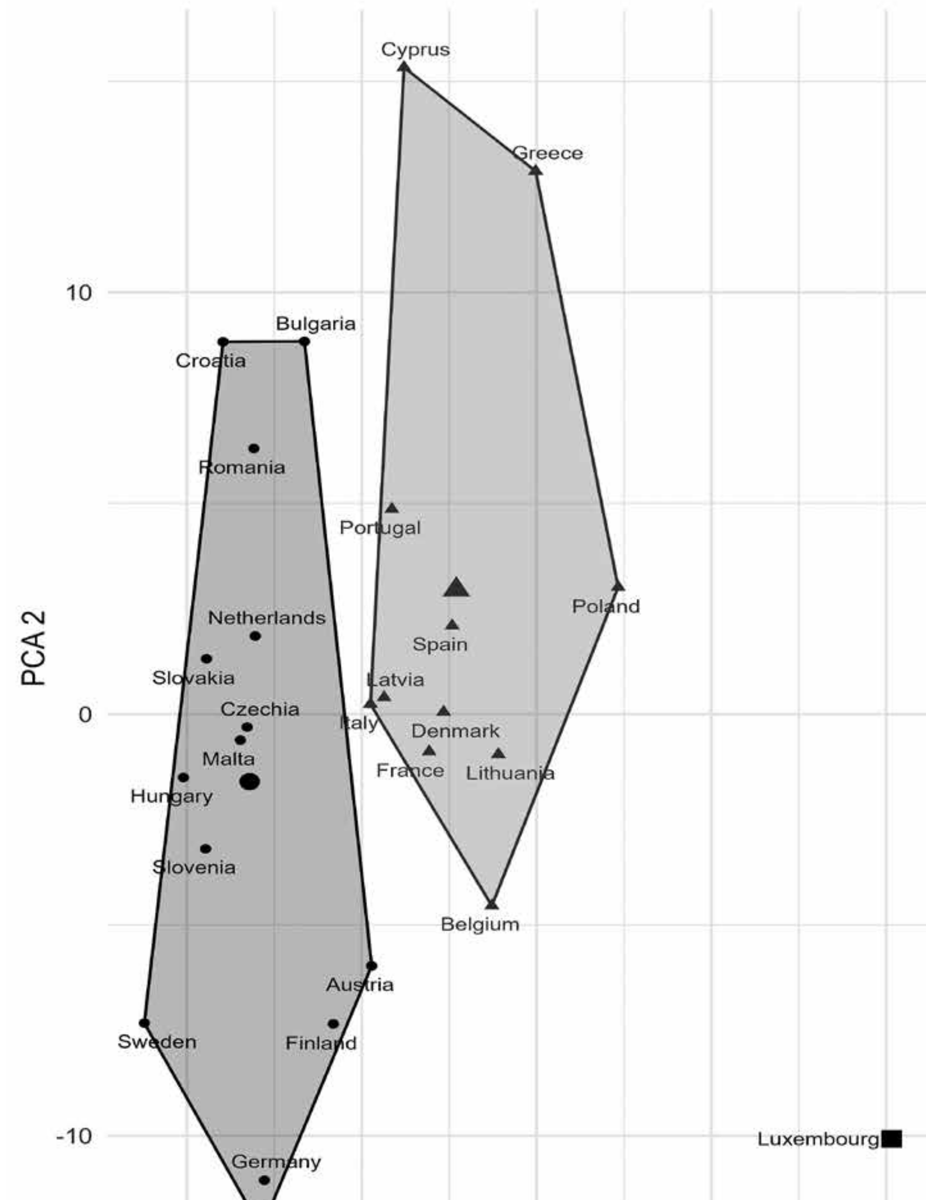
Note: Bolded RMSE signals which model (ARIMA or ARIMAX) performs better in terms of a lower RMSE. For example, for EU-27, ARIMAX better describes the IPI of the same month, while ARIMA better predicts one (h = 1), two (h = 2) or three (h = 3) months ahead.
Source: Own calculations

Table A3 Root Mean Squared Error (RMSE) for ARIMA and ARIMAX models with a rolling window (pre-pandemic period 2008–2019)

Country	ARIMA				ARIMAX			
	Model	h = 1	h = 2	h = 3	Model	h = 1	h = 2	h = 3
EU-27	0.819	**0.876**	**0.872**	**0.868**	**0.803**	0.908	0.881	0.880
Austria	2.230	**2.038**	1.301	1.267	**2.199**	2.118	**1.282**	**1.218**
Belgium	1.838	**3.538**	**3.481**	3.529	**1.833**	3.573	3.527	**3.518**
Bulgaria	1.558	**1.888**	1.836	1.772	**1.526**	1.906	**1.795**	**1.772**
Croatia	2.572	3.188	2.467	2.495	**2.515**	**3.155**	**2.299**	**2.303**
Cyprus	1.309	**1.962**	1.913	1.746	**1.293**	1.975	**1.907**	**1.723**
Czechia	2.643	2.171	**1.894**	1.932	**2.620**	**2.113**	1.937	**1.891**
Denmark	7.748	**4.817**	4.261	4.076	**7.654**	5.092	**4.129**	**3.971**
Estonia	**2.452**	**2.949**	**2.798**	**2.831**	2.526	3.080	2.832	2.833
Finland	1.425	**1.733**	1.871	**1.797**	**1.390**	1.776	**1.794**	1.829
France	1.239	**1.147**	1.313	**1.346**	**1.222**	1.189	**1.309**	1.348
Germany	2.047	**1.470**	1.439	**1.468**	**1.996**	1.743	**1.367**	1.477
Greece	1.395	**1.878**	2.013	**2.147**	**1.366**	1.913	**1.999**	2.157
Hungary	1.978	**2.104**	2.017	**1.948**	**1.936**	2.208	**1.996**	2.051
Ireland	1.826	8.068	**8.583**	8.485	**1.811**	**8.030**	8.609	**8.476**
Italy	3.953	**2.171**	1.446	1.523	**3.936**	2.270	**1.352**	**1.475**
Latvia	**2.585**	**2.142**	1.828	**1.783**	2.590	2.191	**1.788**	1.815
Lithuania	2.132	3.327	3.258	3.321	**2.103**	**3.302**	**3.244**	**3.304**
Luxembourg	4.282	2.657	2.296	2.267	**4.257**	**2.636**	**2.266**	**2.253**
Malta	1.870	**3.609**	3.635	**3.584**	**1.862**	3.621	**3.624**	3.586
Netherlands	1.226	**1.430**	**1.295**	1.357	**1.191**	1.493	1.345	**1.344**
Poland	1.285	**1.265**	**1.440**	**1.452**	**1.271**	1.287	1.458	1.488
Portugal	1.995	**2.076**	2.294	2.258	**1.962**	2.127	**2.288**	**2.249**
Romania	2.376	2.365	**2.117**	**2.104**	**2.371**	**2.343**	2.133	2.114
Slovakia	2.057	**2.643**	**2.686**	**2.680**	**1.968**	2.788	2.689	2.691
Slovenia	2.819	2.276	1.659	**1.502**	**2.788**	**2.268**	**1.606**	1.579
Spain	2.044	**1.591**	**1.458**	**1.448**	**1.999**	1.655	1.494	1.459
Sweden	2.040	1.895	2.170	2.187	**2.036**	**1.799**	**2.071**	**2.102**

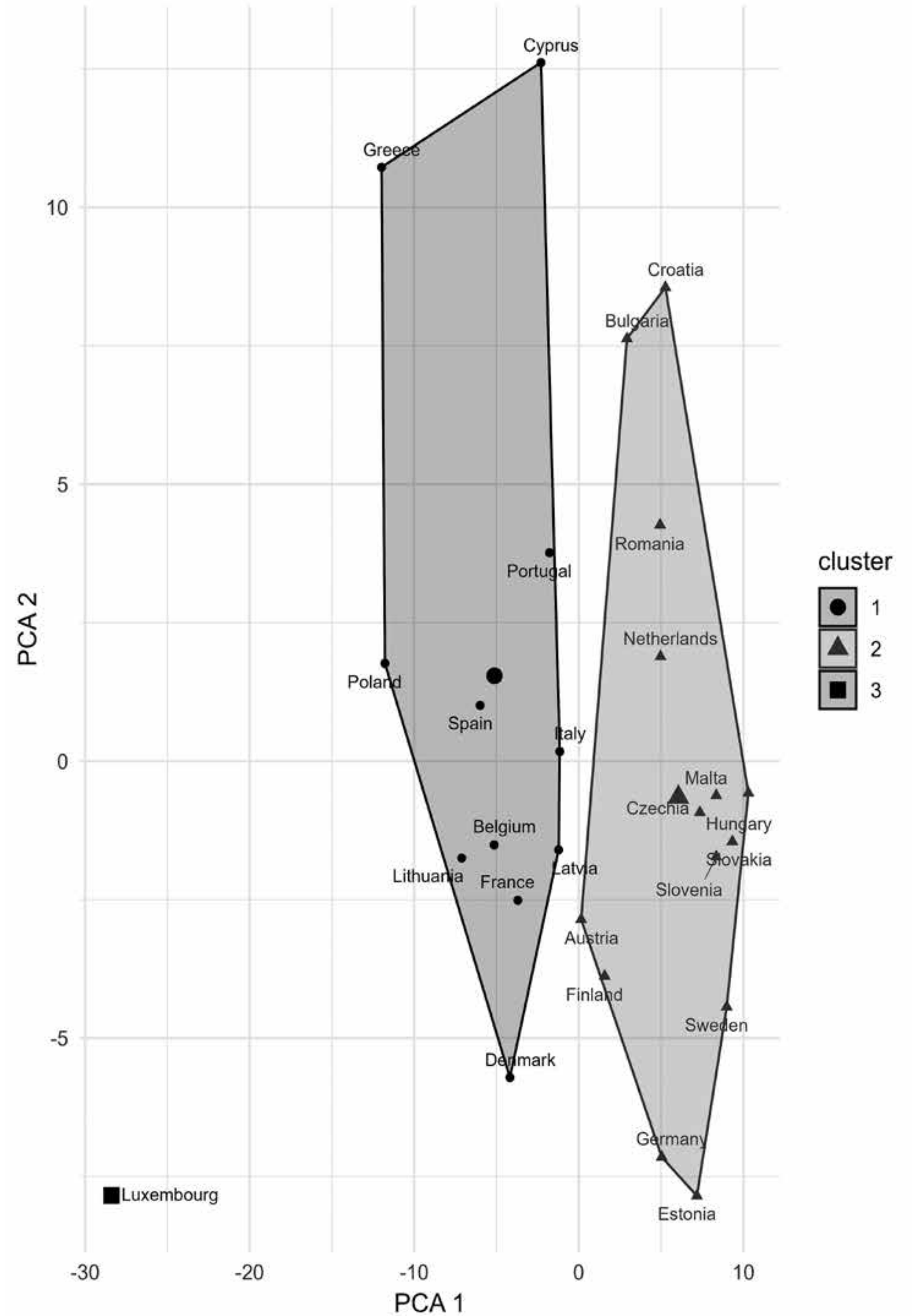
Note: Bolded RMSE signals which model (ARIMA or ARIMAX) performs better in terms of a lower RMSE. For example, for EU-27, ARIMAX better describes the IPI of the same month, while ARIMA better predicts one (h = 1), two (h = 2) or three (h = 3) months ahead.
Source: Own calculations

Figure A1 Three clusters based on the ICI and IPI (left: the pre-pandemic period 2008–2019; right: the entire period 2008–2024)



Note: The clusters 1 and 2 on the right are labelled 2 and 1 in Table 3.
Source: Own calculations

Figure A1 (continuation)



Note: The clusters 1 and 2 on the right are labelled 2 and 1 in Table 3.
Source: Own calculations

Optimization for Partitional Time-Series Clustering

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

Received 19.9.2025 (revision received 8.11.2025), **Accepted** (reviewed) 6.12.2025, **Published** 12.6.2026

Abstract

Time-series clustering is a convenient tool for analysing hidden structures in data. However, as is the case with clustering, it is possible to encounter a number of complications, especially with regards to the sensitivity to the initial algorithm conditions and the subjective choice of the number of groups. The aim of this article is to conduct an experiment using real life data of housing prices in the EU to suppress subjectivity, whether in terms of finding subgroups in the data or the validation of the result for partitional clustering. The proposed procedure is based on a modified bootstrapping principle, where the principle of stability via repetition is applied to the algorithm and its results. As such, this method is applied both to the group selection by monitoring the Calinski-Harabasz index and the final assignment of the resulting classification of clustered objects. The result of this process is a structure that has a better informative value about the relationships in the data.

Keywords	***DOI***	***JEL code***
Partitional clustering DTW distance, k-means algorithm, time-series	*https://doi.org/10.54694/stat.2025.47*	*C15, C18, C38*

INTRODUCTION

The necessity of working with large quantities of volatile high-dimensional datasets containing a wide range of key information needed for the smooth operation of established processes is an issue many fields are facing. Common examples of such data are time-series, the specific nature of which allows for versatile use beyond the basic definition of a sequence of values for a selected statistical feature over time (Liao, 2005). Given this diversity, many tools have been developed to extract said information from time-series, with time-series clustering being one of many different emerging fields offering a versatile method looking for hidden structures inside data. Especially useful is the ability for identification of different subgroups of time-series in a large set of subjects and the subsequent description of group characteristics, e.g. for revealing subphenotypes of patients with infection (Bhavani et al., 2023).

[1] Department of Statistics and Probability, Faculty of Informatics and Statistics, Prague University of Economics and Business, W. Churchill Sq. 1938/4, 130 67 Prague 3 – Žižkov, Czechia. Corresponding author: e-mail: karolina.bakuncova@vse.cz, phone: (+420)224095482. ORCID: <https://orcid.org/0009-0008-6300-243X>.
[2] Department of Statistics and Probability, Faculty of Informatics and Statistics, Prague University of Economics and Business, W. Churchill Sq. 1938/4, 130 67 Prague 3 – Žižkov, Czechia. Corresponding author: e-mail: marek@vse.cz, phone: (+420)224095410. ORCID: <https://orcid.org/0000-0003-4761-1936>.

Clustering analysis has been widely researched and is an attractive method from the point of view of statistical analysis. But many issues arise in its specifications, especially in the clustering of time-series. The core of the issue stems from the reality that time-series does not represent points of n-dimensional space, but rather an entire sequence. At the same time, the complication of time itself influences the monitoring of the development. As such, it is necessary to consider both the issue of possible dependence between values and their time shift. The need to solve these complications leads to a combination of the field of statistical analysis of time-series with recently developed options for information mining methods (Aghabozorgi et al., 2015).

Current knowledge about time-series clustering often adapts already established clustering algorithms with a new similarity measure, the application of which is suitable for dynamic data (Aghabozorgi et al., 2015). The classification of possible clustering methods for time-series clustering is thus not new, where a large part of the methods used are hierarchical or partitional algorithms, which comprise a significant portion of the most frequently used algorithms.

Partitional methods in the field of time-series are predominantly based either on a suitable prototype function or the choice of an appropriate group representative (Biehl et al., 2016; Petitjean et al., 2011). This builds the basis for the optimal solution search in an iterative manner. As is the case with conventional clustering, and especially in the case of time-series, where the situation is further complicated by the volatility of the time-series itself, the vast majority of partitional methods are sensitive to the choice of initial parameters. Suppression of the subjectivity for the final results thus represents a key problem where the usual validation of the clustering results is made more difficult (Ben-David et al., 2006).

The field of data mining offers tools for the application of some methods that allow for a much more accurate view of the real clustering result (e.g. bootstrap). These methods can be utilized in order to point out the shortcomings of the resulting solution or, in some cases, its disputability. The aim of this article is therefore to point out the possibilities of optimizing partitional clustering, specifically, the evaluation of subjectivity suppression of the results for partitional clustering methods applied to time-series and their later validation.

1 LITERATURE SURVEY

Literature about clustering analysis is a diverse set of sources that describe in detail various areas of clustering, such as hierarchical clustering (Murtagh and Contreras, 2017; Ran et al., 2023), partitional clustering (Celebi, 2016), the use of neural networks (Du, 2010) as well as specialized procedures (Barton et al., 2019). It is possible to find extensive overviews comparing individual methods (Ezugwu et al., 2022). Similarly, there exists a wide range of literature related to machine learning, particularly on the usage of clustering methods in the field of classification and prediction (Kotsiantis et al., 2006), or description of bootstrapping methods for validation (Schumacker, 2014).

Contrarily, the amount of literature dealing with the issue of time-series clustering is low. One of the most important publications dealing with time-series clustering is published by Aghabozorgi et al. (2015), which contains an exhaustive summary of methods and procedures implemented in the field of time-series clustering up until 2015. Since then, several textbooks for time-series clustering have been published, e.g. the textbook created by Maharaj et al. (2019), which summarizes the development of traditional and fuzzy clustering methods, including approaches based on observations, features, models and machine learning methods.

The main issue which arises during literature review for time-series clustering is that a high percentage of existing articles focus on the development of their own procedures for time-series clustering, applied only on simulated data, which leads to overfitting and other issues concerning the wider application of these methods and validation of their results (Bagnall et al., 2017).

Block bootstrapping in the context of time-series is a tool used to estimate standard errors, confidence intervals and other relevant statistics. It allows the enhancement of predictions by evaluating the forecasts' uncertainty in chosen econometric models. Conceptually, this method is a non-parametric resampling technique applied to a single time-series, where instead of resampling individual points, resampling results in several blocks of observations. There are two main variants of this application, one assuming the resulting blocks are non-overlapping or overlapping (Hall, 1985). However, block bootstrapping can be very fastidious in terms of real application, as the convergence of errors is rather slow (Härdle, 2003).

Despite the method's imperfections, most of the currently available tools trying to improve upon the results of the block bootstrapping do not achieve significantly better results. The issue worsens when one tries to apply this principle on time-series clustering, as there is little to no literature about its effectiveness nor results.

2 MATERIAL FRAMEWORK

Statistical clustering attempts to classify similar objects based on given characteristics under the assumption of homogeneity to their own group members and heterogeneity to other group members. Differences in definition of similarity used in clustering can lead to significant variability of the results, depending on the chosen approach. This applies both to the classic application of clustering and to its modification for time-series. Another difficulty from the point of view of time-series arises from the fact that the distance measure must not only comply with the chosen concept for the definition of similarity but also be able to deal with the issues of multidimensionality and volatility connected to the nature of working with time-series.

2.1 Methodology

For the purposes of this article, a standard combination of time-series clustering methods will be used, namely the Dynamic time warping (DTW) distance (Vintsyuk, 1968; Sakoe and Chiba, 1978) and the k-means algorithm. The DTW distance as a similarity measure is currently considered to be one of the most successful metrics for time-series clustering usage, as it is very suitable for working around time distortions.

Suppose that to calculate the distance of two time-series $X = (x_1, ..., x_T)$ and $Y = (y_1, ..., y_T)$ the Euclidean distance would be used, defined as:

$$Euclidean(X,Y) = \sqrt{\sum_{i=1}^{T}(x_i - y_i)^2}. \tag{1}$$

The given definition of distance for time-series brings with it two main complications:

– The time-series being compared must be of the same length,
– The individual events of the time-series being compared must occur at the same point in time, otherwise their resulting distance will be noticeably higher even though both time-series undergo similar development.

The DTW distance essentially provides an optimization of the Euclidean distance, the goal of which is to find such an assignment of points in a pair of time-series that would lead to the minimization of the resulting total distance. For the Euclidean distance, the optimization problem can be written as follows:

$$DTW(X,Y) = \min_{\pi \in A(X,Y)} \left(\sum_{(i,j)\in\pi} Euclidean(X_i, Y_j) \right), \tag{2}$$

where the optimal assignment π of length k is a sequence of k indexed pairs of time-series values X and Y. The set $A(X,Y)$ contains all possible assignment paths. In essence, the optimization problem can also be illustrated as finding the optimal path of the original distance matrix using the Euclidean distances. From the definition of the DTW distance, it is obvious that the main issue of its calculation is its computational complexity. Even with a small increase in the number of the time-series, the quantity of possible sorting combinations increases drastically. There are ways to simplify the calculation, but for the purposes of this article we will use the basic definition of the DTW distance.

The k-means algorithm is an unsupervised clustering algorithm based on the idea of the similarity of neighboring points (MacQueen, 1967). Its application to time-series is in practice the same as its common application to classic data sets. In other words, time-series clustering mainly differs in the method of calculating the distance, where it is possible to subsequently use common, well-known clustering algorithms. There are also specific clustering algorithms targeted at time-series, e.g. k-Shape (Paparrizos and Gravano, 2016).

2.2 Validation

Calinski-Harabasz (CH) index was chosen for the purpose of validation for both the intermediate results from the optimization process introduced in this article and the final clustering results themselves. In general, the CH index quantifies the ratio between-cluster separation to the within-cluster dispersion (Calinski and Harabasz, 1974). It is a popular metric commonly used to evaluate the quality of clustering results in unsupervised learning.

The core idea behind the CH index is to examine how distinct the clusters are (high inter-cluster variance) and how tightly grouped the data points are within each cluster (low intra-cluster variance). The mathematical formulation for this principle is as follows:

$$CH = \frac{\left(B_k / (k-1)\right)}{\left(W_k / (n-k)\right)}, \tag{3}$$

where B_k is the between-cluster separation and W_k is within-cluster dispersion, both normalised by the corresponding degrees of freedom.

A higher CH index indicates better-defined clusters, meaning the clusters are well separated and the objects within the cluster are similar. However, a lower score suggests overlapping or poorly defined clusters. Because the index depends on the number of clusters, it is often used to determine the optimal in clustering tasks. While the CH index performs well for convex, spherical clusters, it may not be as effective for complex, non-linear cluster shapes.

2.3 Optimization

When applying partitional clustering to time-series, there are many practical issues related to the selection of an appropriate number of clusters and their result. Since most real time-series data do not have a predetermined group assignment, it is necessary to consider a procedure that would allow for the effective detection of hidden structures in the data. However, partitional methods themselves assume that the number of clusters is predetermined, which makes the selection of an appropriate parameter a key step in the entire process.

One possible approach to automate and refine this process is to repeatedly apply the algorithm with different initial conditions. The results obtained this way can then be consolidated into a single, less subjective solution that better reflects consistent patterns in the data rather than their individual runs. This principle assumes that repeated application of clustering helps to suppress random effects and better captures persistent structures in the data.

To identify the structure, further processing of the results of repeated clustering was proposed. A set of 100 outputs for the selected optimal number of clusters was used to construct the co-association matrix (Zhong et al., 2019). This matrix of dimensions (corresponding to the number of time-series analyzed) captures the intensity with which individual pairs of time-series were assigned to the same cluster across individual runs of the algorithm. Specifically, if two time-series were assigned to the same cluster in any run, value 1 is added to the corresponding cell of the matrix (defined by the row and column corresponding to both time-series). The given matrix was further modified with the following notation:

$$x_{pairs_{ij}} = 1 - \frac{x_{ij}}{100} \ . \tag{4}$$

After processing all 100 repetitions, the matrix contains values in the range $\langle 0,1 \rangle$, which represents the degree of co-association of the given time-series in the same cluster, so that lower values signal a smaller distance between time-series and therefore significant similarity. This information provides a comprehensive view of the resulting structure in the data as well as the degree of similarity between individual pairs of time-series and serves as a basis for further consolidation of the results. Additionally, the given matrix can be interpreted as a distance matrix, which may be used to determine the resulting classification of time-series based on the result of 100 different runs of the algorithm.

2.3.1 Number of clusters

The final number of clusters was selected based on repeated partitional clustering for different values of the parameter *k*, specifically in the range from 2 to $T - 1$ (where T is the number of time-series). For each value of *k*, the clustering was repeated 100 times with different random initializations of the starting point, which resulted in a total of $(T - 1) \times 100$ runs of the algorithm.

Each individual solution was subsequently evaluated using the CH index. For each value of *k*, the average, maximum and median index values from the respective 100 runs were calculated. A higher index value indicates a better-quality clustering structure.

2.3.2 Bootstrapping

We can further modify the aforementioned method via the application of bootstrapping. The principle used in the article combines the general usage of bootstrapping with the idea proposed by block bootstrapping. Since the time-series data are notoriously challenging to analyze due to the issue of correlation and volatility, the proposed approach remains intentionally simple.

Given a dataset of n time-series of length T, n new datasets were generated, each created by omitting one of the original time-series. For each of these modified datasets, the clustering procedure proposed earlier was repeated. This resulted in n co-association matrices, one per modified dataset. Finally, by aggregating these matrices into a single, more robust co-association matrix, we hope to achieve a distance matrix that better captures the underlying structure of the data.

2.4 Data

Real estate prices are an important indicator of economic development and the standard of living for the population of any country. Eurostat regularly publishes detailed data on the development of residential real estate prices across the Member Countries and selected non-EU countries. This data allows not only the monitoring of long-term trends in housing, but also the comparison of different regions of the EU.

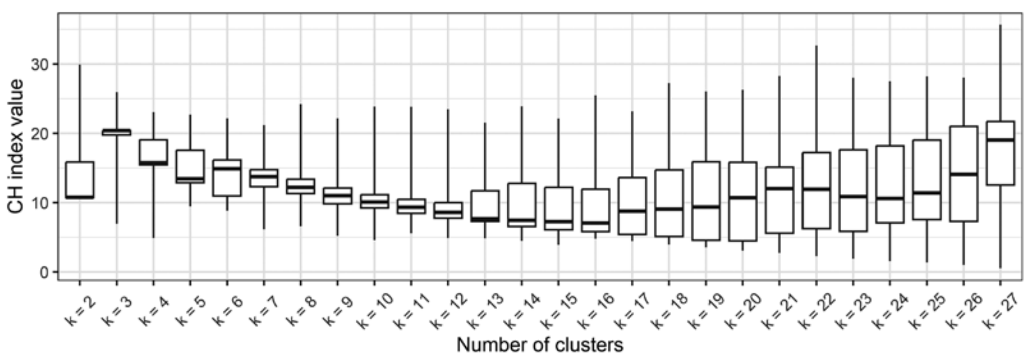
The dataset used in the article contains a total of 28 selected European and non-European countries, recording 59 observations of quarterly house price indices from the first quarterof 2010 to the third quarter of 2024 (Eurostat, 2025). Given that the main principle of time-series clustering is based on the comparison of shape and structure in the time-series development, many of the methods behave better when they are not applied directly on the raw data. One of the possible and simplest methods for data preprocessing used for time-series clustering is standardisation, where the transformation of a time-series means rescaling the values in a way where the mean is equal to zero and standard deviation is equal to one. This can be achieved by a simple subtraction of the sample mean from the observed values and division by the standard deviation. Rather than be bottled down by the differences in values,by rescaling the time-series to a common range, we can focus more on the actual comparison of shapes, something that often contributes to cleaner and more concise results for time-series clustering. However, this method introduces a certain degree of bias into the data which should be handled carefully.

Processing of the data and the results was done by using the R programming language (R Core Team, 2025) with several key libraries, specifically the package *caret* (Kuhn, 2024) for data-preprocessing, *TSclust* (Montero and Vilar, 2014) and *dtwclust* (Sardá-Espinosa, 2019) for time-series classification and *clusterCrit* (Desgraupes, 2024) for clustering validation.

3 RESULTS

The first step of the analysis consists of choosing an appropriate number of clusters. Based on the analysis of the resulting values, it was found that the first significant and stable value of the CH index occurs at point $k = 3$, which indicates the existence of three significant subgroups in the data (Figure 1). After this point, the index decreases and starts to increase again only when the parameter is greater than 18. From the point of view of clustering needs, the given index values are already too high, and the results cannot be interpreted as meaningful. For these reasons $k = 3$ was chosen as the optimal number of clusters.

Figure 1 Values of the CH index for values of *k* 2 to T-1



Source: Author's calculations

Based on the above procedure, a rough estimate of the optimal number of clusters was obtained, which serves as a starting point for further analysis. However, it is necessary to once again emphasize that the results of individual runs of partitional clustering can vary significantly depending on the initial conditions of the algorithm. This variability is evident with the given parameter $k = 3$, where it is possible to observe significantly different index values across repeated runs. At the same time, it is evident that some solutions are repeated multiple times based on the same index value. This implies the need to further analyse the clustering output in a way that achieves consistent and interpretable results.

Based on the resulting co-association matrix (Table 1), it is evident that some countries were often classified into the same cluster within the repeated runs of the algorithm, in some extreme cases the value being equal to zero, meaning that in all 100 iterations these countries were classified into the same cluster. This phenomenon points to the existence of a strong structural similarity of the development within the analyzed time-series.

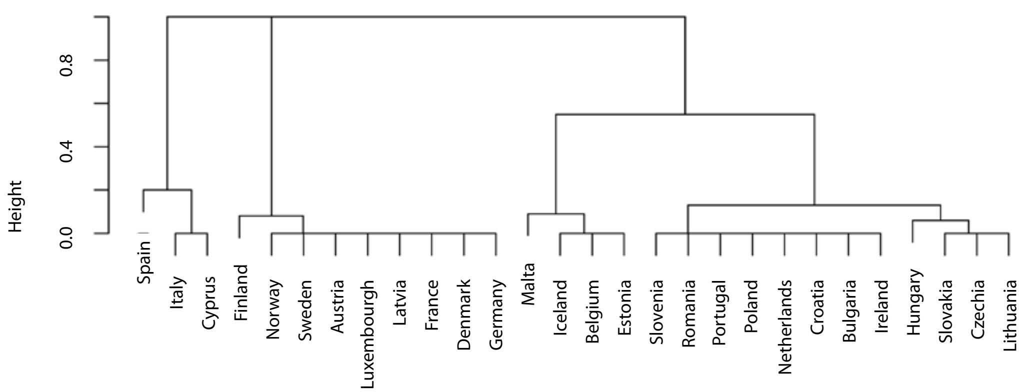
Table 1 Co-association matrix for the first five countries (full matrix in Table A1)

	Bulgaria	Czechia	Denmark	Germany
Belgium	0.55	0.42	0.44	0.44
Bulgaria		0.13	0.99	0.99
Czechia			0.86	0.86
Denmark				0.00

Source: Author's calculations

The algorithm of the farthest neighbor method was then used to search the matrix. Resulting classification can therefore be illustrated in Figure 2. Based on the image it can be argued that most analyzed countries display a strong tendency to be classified together and form smaller groups on the lowest level of the resulting dendrogram.

Figure 2 Initial resulting ranking of European and other selected countries

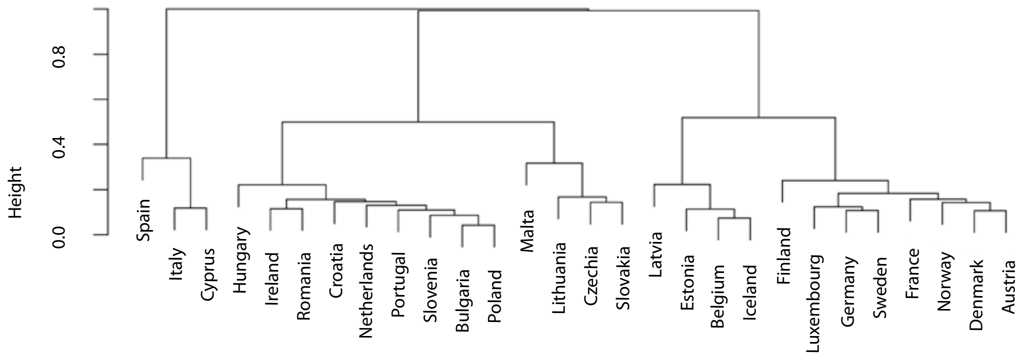


Source: Author's calculations

Further possible modification is in the application of machine learning validation procedures. For each variant, 100 iterations of partitional clustering assuming $k = 3$ have been performed, similarly to the previous cases. The resulting aggregated co-association matrix will thus capture not only the variability caused by the choice of the starting point of the algorithm, but also the sensitivity of the model to the presence of individual countries in the input data.

This modification led to a solution that remains largely stable in its basic layout (see Figure 3). Again, a smaller cluster consisting of Cyprus, Italy and Spain appears, followed by a cluster comprising mainly Western European and Scandinavian countries, and finally a diverse, heterogeneous cluster comprised of, among others, Croatia, Poland and Slovenia. The most significant difference from the initial result is the new assignment of countries such as Latvia, Estonia and Belgium. The assignment stayed otherwise consistent.

Figure 3 Modified result of time-series clustering



Source: Author's calculations

4 DISCUSSION

Approximately three main clusters can be identified (Figure 4). The smallest and most clearly defined cluster is the third cluster formed by Southern European countries, namely Spain, Italy and Cyprus, whose real estate price dynamics show a high degree of similarity in their development.

The second, larger cluster mainly consists of Northern and Western European countries, its first subgroup including countries such as Finland, Norway, Austria and Germany. These countries are characterized by relatively stable development and common economic features as well as high economic prosperity, which may also be reflected in the time-series of real estate prices. Second subgroup consists of countries

Figure 4 Assignment of housing prices for chosen countries based on the modified results

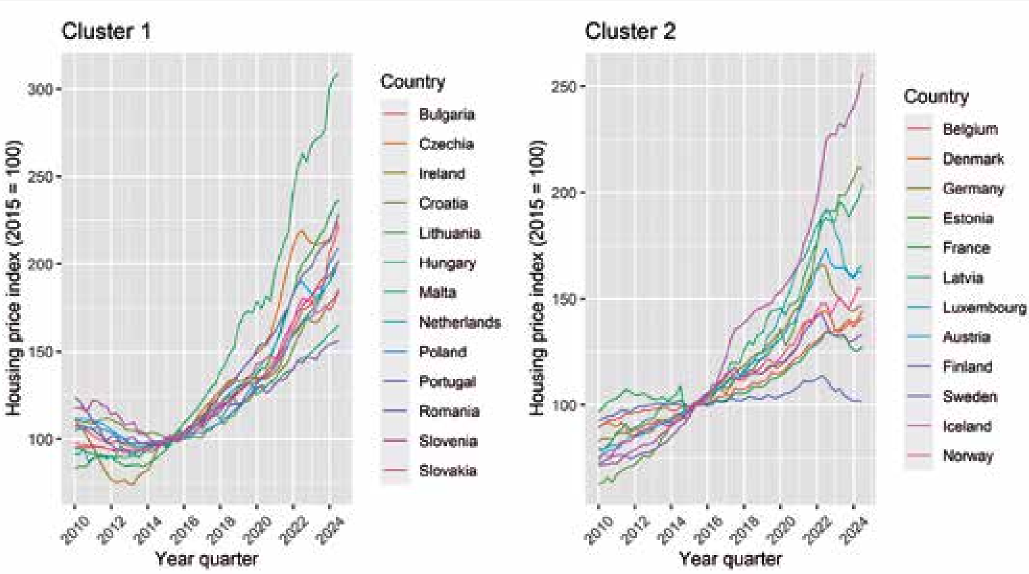
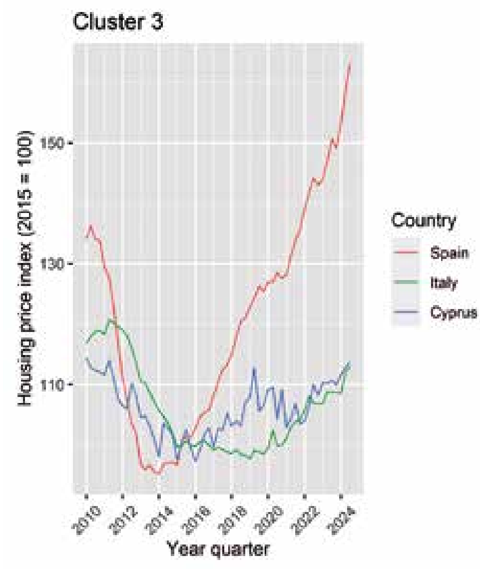


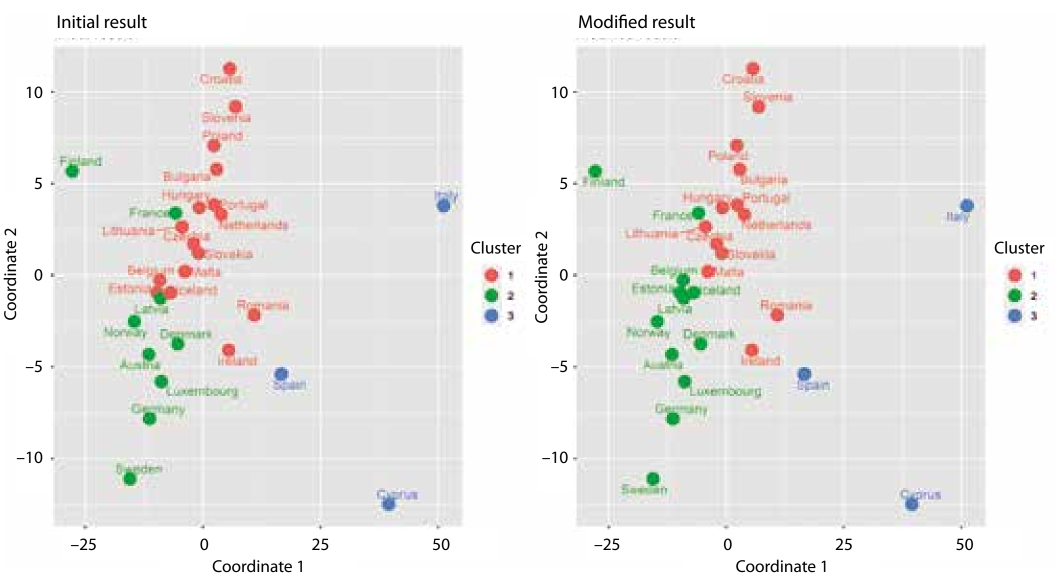
Figure 4 continuation



Source: Eurostat

Latvia, Estonia, Belgium and Iceland. These countries are all small, open, high-income European economies that depend on services, trade and external markets. It is important to note that Belgium, Estonia and Iceland are on a border between two of the main big clusters suggesting a shared structure with both of the groups (Figure 5).

Figure 5 Multidimensional scaling of the initial and modified solutions



Source: Author's calculations

Last cluster consists mainly of countries of Central and Eastern Europe, which could also be divided into two smaller subgroups. One group encapsulates Malta, Lithuania, Czechia and Slovakia, while the other subgroup focuses on countries like Poland, Croatia and Romania. Main feature of this cluster is a noticeably lower economic prosperity, especially compared to the countries of the other big final cluster.

CONCLUSION

Time-series clustering allows not only the assessment of individual development of time-series, but also the comparison of broader datasets and the uncovering of common features or tendencies that may indicate the existence of hidden influences. It is also possible to assume that these influences may act differently within identified groups of time-series and thus have a significant, albeit different impact on their dynamics. In this sense, clustering can be perceived as a means of exposing the hidden structure in the data, which can then be visualized and interpreted. At the same time, it is necessary to emphasize that time-series clustering contains a significant degree of subjectivity. The results are sensitive to the choice of algorithm, parameter settings, definition of the distance metric and the choice of initial conditions, as well as the interpretation itself. Although subjectivity can be partially limited by using different methods and comparing their outputs, e.g. bootstrapping, it cannot be completely eliminated. For this reason, it would be misleading to claim that the obtained results are absolutely accurate.

However, the analysis conducted in this study using real life data of housing prices in EU showed that despite the aforementioned limitations in most cases a somewhat consistent structure can be observed in the data. This fact supports the idea that the greatest benefit of time-series clustering lies in the discovery and description of these hidden structures rather than in a fixed and unambiguous classification. From a practical point of view, it is therefore appropriate to direct further research towards methods that will enable this latent structure to be revealed more clearly and made visually accessible. For example, it could be beneficial to link clustering with methods for visualizing high-dimensional data or with advanced dimensionality reduction techniques. Such an approach could lead to a deeper understanding of the internal relationships between individual time-series and at the same time provide tools for a more robust interpretation of the results.

Overall, it can be concluded that time-series clustering represents a valuable analytical framework, the strength of which does not lie in achieving complete unambiguousness of the results, but in the ability to reveal stable, recurring patterns and structures in the data. Their detailed investigation and description can bring new insights into the dynamics of the observed phenomena and open the way to their deeper interpretation and predictive modelling.

References

AGHABOZORGI, S., SHIRKHORSHIDI, A. S., WAH, T. Y. (2015). Time-series clustering–a decade review [online]. *Information systems*, 53: 16–38. <https://doi.org/10.1016/j.is.2015.04.007>.

BAGNALL, A., LINES, J., BOSTROM, A., LARGE, J., KEOGH, E. (2017). The great time-series classification bake off: a review and experimental evaluation of recent algorithmic advances [online]. *Data mining and knowledge discovery*, 31: 606–660. <https://doi.org/10.1007/s10618-016-0483-9>.

BARTON, T., BRUNA, T., KORDIK, P. (2019). Chameleon 2: an improved graph-based clustering algorithm [online]. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 13(1): 1–27. <https://doi.org/10.1145/3299876>.

BEN-DAVID, S., VON LUXBURG, U., PÁL, D. (2006). A sober look at clustering subjectivity [online]. *Lecture Notes in Computer Science*, 4005: 5–19. <https://doi.org/10.1007/11776420_4>.

BHAVANI, S. V., XIONG, L., PIUS, A., SEMLER, M., QIAN, E. T., VERHOEF, P. A., ROBICHAUX, C., COOPERSMITH, C. M., CHURPEK, M. M. (2023). Comparison of time-series clustering methods for identifying novel subphenotypes of patients with infection [online]. *Journal of the American Medical Informatics Association*, 30(6): 1158–1166. <https://doi.org/10.1093/jamia/ocad063>.

BIEHL, M., HAMMER, B., VILLMANN, T. (2016). Prototype-based models in machine learning [online]. *Wiley interdisciplinary reviews. Cognitive science,* 7(2): 92–111. <https://doi.org/10.1002/wcs.1378>.

CALIŃSKI, T., HARABASZ, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics* [online]. 3(1): 1–27. <https://doi.org/10.1080/03610927408827101>.

CELEBI, M. E. (eds.). (2016). *Partitional Clustering Algorithms*. Springer Cham. <https://doi.org/10.1007/978-3-319-09259-1>.

DESGRAUPES, B. (2024). *ClusterCrit: Clustering indices* (Version 1.3.0). [Computer software]. <https://CRAN.R-project.org/package=clusterCrit>.

DU, K. L. (2010). Clustering: A neural network approach [online]. *Neural networks*, 23(1): 89–107. <https://doi.org/10.1016/j.neunet.2009.08.007>.

EUROSTAT. (2025). *House price index (quarterly, PRC_HPI_Q)* [Dataset]. European Commission. [cit. 18.4.2025]. <https://doi.org/10.2908/PRC_HPI_Q>.

EZUGWU, A. E., IKOTUN, A. M., OYELADE, O. O., ABUALIGAH, L., AGUSHAKA, J. O., EKE, C. I., AKINYELU, A. A. (2022). A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects [online]. *Engineering Applications of Artificial Intelligence*, 110: 104743. <https://doi.org/10.1016/j.engappai.2022.104743>.

HALL, P. (1985). Resampling a coverage process [online]. *Stochastic Process Applications*, 19: 259–269. <https://doi.org/10.1016/0304-4149(85)90212-1>.

HÄRDLE, W., HOROWITZ, J., KREISS, J. P. (2003). Bootstrap methods for time series [online]. *International Statistical Review*, 71(2): 435–459. <https://doi.org/10.1111/j.1751-5823.2003.tb00485.x>.

KOTSIANTIS, S. B., ZAHARAKIS, I. D., PINTELAS, P. E. (2006). Machine learning: a review of classification and combining techniques [online]. *Artif Intell Rev*, 26: 159–190. <https://doi.org/10.1007/s10462-007-9052-3>.

KUHN, M. (2024). *Caret: Classification and regression training* (Version 6.0-94) [Computer software]. <https://CRAN.R-project.org/package=caret>.

LIAO, T. W. (2005). Clustering of time-series data – a survey [online]. *Pattern recognition*, 38(11): 1857–1874. <https://doi.org/10.1016/j.patcog.2005.01.025>.

MACQUEEN, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability,* 5(1): 281–298.

MAHARAJ, E. A., D'URSO, P., CAIADO, J. (2019*). Time-series clustering and classification*. Chapman and Hall/CRC.

MONTERO, P., VILAR, J. A. (2014). *TSclust: An R package for time series clustering* (Version 1.3.1) [Computer software]. *Journal of Statistical Software,* 62(1): 1–43. <https://doi.org/10.18637/jss.v062.i01>.

MURTAGH, F., CONTRERAS, P. (2017). Algorithms for hierarchical clustering: an overview, II [online]. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7(6): e1219. <https://doi.org/10.1002/widm.1219>.

PAPARRIZOS, J., GRAVANO, L. (2016). K-Shape: Efficient and Accurate Clustering of Time-series [online]. *SIGMOD Rec.,* 45(1): 69–76. <https://doi.org/10.1145/2949741.2949758>.

PETITJEAN, F., KETTERLIN, A., GANÇARSKI, P. (2011). A global averaging method for dynamic time warping, with applications to clustering [online]. *Pattern recognition*, 44(3): 678–693. <https://doi.org/10.1016/j.patcog.2010.09.013>.

R CORE TEAM (2025). *R: A Language and Environment for Statistical Computing* (Version 4.4.1) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org>.

RAN, X., XI, Y., LU, Y., WANG, X., LU, Z. (2023). Comprehensive survey on hierarchical clustering algorithms and the recent developments [online]. *Artificial Intelligence Review*, 56(8): 8219–8264. <https://doi.org/10.1007/s10462-022-10366-3>.

SAKOE, H., CHIBA, S. (1978). Dynamic programming algorithm optimization for spoken word recognition [online]. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1): 43–49. <https://doi.org/10.1109/tassp.1978.1163055>.

SARDÁ-ESPINOSA, A. (2019). *Dtwclust: Time-series clustering along with optimizations for dynamic time warping* (Version 5.5.12) [Computer software]. <https://CRAN.R-project.org/package=dtwclust>.

SCHUMACKER, R. E. (2014). *Bootstrapping: A nonparametric approach to statistical inference*. SAGE Publications.

VINTSYUK, T. K. (1968). Speech discrimination by Dynamic Programming [online]. *Cybernetics*, 4(1): 52–57. <https://doi.org/10.1007/bf01074755>.

ZHONG, C., HU, L., YUE, X., LUO, T., FU, Q., XU, H. (2019). Ensemble clustering based on evidence extracted from the co-association matrix [online]. *Pattern Recognition*, 92: 93–106. <https://doi.org/10.1016/j.patcog.2019.03.020>.

APPENDIX

Table A1 Full co-association matrix

COUNTRY	Bulgaria	Czechia	Denmark	Germany	Estonia	Ireland	Spain	France	Croatia	Italy	Cyprus	Latvia	Lithuania	Luxembourg	Hungary	Malta	Netherlands	Austria	Poland	Portugal	Romania	Slovenia	Slovakia	Finland	Sweden	Iceland	Norway
Belgium	0.55																										
Bulgaria	0.42	0.13																									
Czechia	0.44	0.86	1.00																								
Denmark	0.44	0.99	0.86	0.00																							
Germany	0.00	0.55	0.42	0.44	0.44																						
Estonia	0.55	0.00	0.13	0.99	0.99	0.55																					
Ireland	1.00	0.73	0.86	1.00	1.00	1.00	0.73																				
Spain	0.44	0.99	0.86	0.00	0.00	0.44	0.99	1.00																			
France	0.55	0.00	0.13	0.99	0.99	0.55	0.00	0.73	0.99																		
Croatia	1.00	0.93	1.00	1.00	1.00	1.00	0.93	0.20	1.00	0.93																	
Italy	1.00	0.93	1.00	1.00	1.00	1.00	0.93	0.20	1.00	0.93	0.00																
Cyprus	0.44	0.99	0.86	0.00	0.00	0.44	0.99	1.00	0.00	0.99	1.00	1.00															
Latvia	0.42	0.13	0.00	0.86	0.86	0.42	0.13	0.86	0.86	0.13	1.00	1.00	0.86														
Lithuania	0.44	0.99	0.86	0.00	0.00	0.44	0.99	1.00	0.00	0.99	1.00	1.00	0.00	0.86													
Luxembourg	0.48	0.07	0.06	0.92	0.92	0.48	0.07	0.80	0.92	0.07	1.00	1.00	0.92	0.06	0.92												
Hungary	0.09	0.46	0.33	0.53	0.53	0.09	0.46	1.00	0.53	0.46	1.00	1.00	0.53	0.33	0.53	0.39											
Malta	0.55	0.00	0.13	0.99	0.99	0.55	0.00	0.73	0.99	0.00	0.93	0.93	0.99	0.13	0.99	0.07	0.46										
Netherlands	0.44	0.99	0.86	0.00	0.00	0.44	0.99	1.00	0.00	0.99	1.00	1.00	0.00	0.86	0.00	0.92	0.53	0.99									
Austria	0.55	0.00	0.13	0.99	0.99	0.55	0.00	0.73	0.99	0.00	0.93	0.93	0.99	0.13	0.99	0.07	0.46	0.00	0.99								
Poland	0.55	0.00	0.13	0.99	0.99	0.55	0.00	0.73	0.99	0.00	0.93	0.93	0.99	0.13	0.99	0.07	0.46	0.00	0.99	0.00							
Portugal	0.55	0.00	0.13	0.99	0.99	0.55	0.00	0.73	0.99	0.00	0.93	0.93	0.99	0.13	0.99	0.07	0.46	0.00	0.99	0.00	0.00						
Romania	0.55	0.00	0.13	0.99	0.99	0.55	0.00	0.73	0.99	0.00	0.93	0.93	0.99	0.13	0.99	0.07	0.46	0.00	0.99	0.00	0.00	0.00					
Slovenia	0.42	0.13	0.00	0.86	0.86	0.42	0.13	0.86	0.86	0.13	1.00	1.00	0.86	0.00	0.86	0.06	0.33	0.13	0.86	0.13	0.13	0.13	0.13				
Slovakia	0.52	1.00	0.94	0.08	0.08	0.52	1.00	1.00	0.08	1.00	1.00	1.00	0.08	0.94	0.08	1.00	0.61	1.00	0.08	1.00	1.00	1.00	1.00	0.94			
Finland	0.44	0.99	0.86	0.00	0.00	0.44	0.99	1.00	0.00	0.99	1.00	1.00	0.00	0.86	0.00	0.92	0.53	0.99	0.00	0.99	0.99	0.99	0.99	0.86	0.08		
Sweden	0.00	0.55	0.42	0.44	0.44	0.00	0.55	1.00	0.44	0.55	1.00	1.00	0.44	0.42	0.44	0.48	0.09	0.55	0.44	0.55	0.55	0.55	0.55	0.42	0.52	0.44	
Iceland	0.44	0.99	0.86	0.00	0.00	0.44	0.99	1.00	0.00	0.99	1.00	1.00	0.00	0.86	0.00	0.92	0.53	0.99	0.00	0.99	0.99	0.99	0.99	0.86	0.08	0.00	0.44

Source: Author's calculations

Probability-Based Web Panels for Official Statistics: Basic Insights and Analysis of the Bias of Survey Estimates

Gregor Čehovin[1] | *University of Ljubljana, Ljubljana, Slovenia*
Vasja Vehovar[2] | *University of Ljubljana, Ljubljana, Slovenia*

Received 8.7.2025 (revision received 24.9.2025), **Accepted** (reviewed) 28.10.2025, **Published** 12.6.2026

Abstract

The past three decades of survey research have revealed a shift from traditional probability-based surveys to various types of web surveys. In official statistics, the probability-based web panels (PWPs), for which respondents are recruited once and then incentivized for repeated participation in web surveys, play a particularly important role in this process. This article provides insights into the use of PWPs for official statistics. First, a survey of European Union national statistics offices revealed that one-quarter had already implemented or were planning to implement PWPs; the main barriers to their implementation were lack of knowledge and expertise. In addition, we evaluated the quality of the estimates in the Slovenian PWP (1KA Panel) that replicated questions from 12 traditional probability-based surveys. The findings showed that 205 of 651 PWP estimates (31%) exhibited relative bias exceeding 10%. Biases varied substantially across survey topics, indicating the selective suitability of PWPs for official statistics.

Keywords	***DOI***	***JEL code***
Probability-based web panels, survey estimates, relative bias, nonresponse error, measurement error, coverage error, processing error	*https://doi.org/10.54694/stat.2025.28*	*C83, Y80*

INTRODUCTION

Since the 1970s, the evolution of digital technologies has steadily transformed the survey research process from *traditional survey modes* – i.e., face-to-face, telephone, and mail surveys – to web surveys. Survey science (Groves et al., 2009), however, was originally grounded on these traditional modes,

[1] Centre for Social Informatics, Faculty of Social Sciences, University of Ljubljana, Kardeljeva ploščad 5, 1000 Ljubljana, Slovenia. Corresponding author: e-mail: gregor.cehovin@fdv.uni-lj.si. ORCID: <https://orcid.org/0000-0002-9431-3694>
[2] Centre for Social Informatics, Faculty of Social Sciences, University of Ljubljana, Kardeljeva ploščad 5, 1000 Ljubljana, Slovenia. E-mail: vasja.vehovar@fdv.uni-lj.si. ORCID: <https://orcid.org/0000-0002-3253-7959>.

typically implemented using the design-based sampling theory (e.g., Cochran, 1977; Särndal et al., 1992). We hereafter refer to them as *traditional probability-based surveys* (TPSs), a term that also includes contemporary TPSs that combine traditional modes with web surveys (i.e., mixed mode).

In recent years, rising TPS costs and declining response rates have been addressed through inexpensive, flexible web surveys (Callegaro et al., 2015), especially nonprobability ones (Callegaro et al., 2014; Golini and Righi, 2024; Vehovar et al., 2016). *Probability-based web panels* (PWPs) have emerged as viable alternatives for official statistics grounded in probability principles. Respondents are recruited via probabilistic sampling and then invited to participate in recurring web surveys (e.g., monthly for two years), typically with incentives (Couper, 2017; Hays et al., 2015). PWPs reduce per-survey costs by requiring only one recruitment drive, enable faster data collection, and maintain probability principles.

Although PWPs involve lower costs and facilitate more flexible data collection, they generally yield lower response rates than TPSs (e.g., Kocar and Kaczmirek, 2024). This increases the risk of nonresponse bias – the difference between a survey estimate and its true (population) value caused by nonresponse. The issue is critical for official statistics, where unbiased estimates are essential. Whether the operational advantages of PWPs offset potential declines in estimate quality remains unclear. Empirical evidence concerning the role of PWPs in official surveys, particularly the effects of transitioning from TPSs to PWPs, is limited. To address these gaps, we pose the following research questions (RQs):

- RQ1: What are the practices, plans, and perceptions of PWPs for official statistics?
- RQ2: What is the extent of bias of the estimates derived from PWPs compared to those obtained from TPSs?

To address these questions, we surveyed *national statistical institutes* (NSIs) in the European Union (EU) for RQ1 and conducted a comprehensive study of Slovenian official statistics for RQ2 by comparing estimates from 12 TPSs with those from a parallel PWP.

1 PROBABILITY-BASED WEB PANELS

PWPs possess three essential characteristics: (a) long-term respondent participation across multiple studies, based on informed consent and incentives; (b) primary data collection via the web, supplemented by traditional modes for internet nonusers but not for recontacting web survey nonrespondents; and (c) probability-based sampling, requiring known, positive inclusion probabilities for all population units and a list of all units in the target population or a surrogate sample, such as an area sample. Unfortunately, constructing and exploiting such lists is resource-intensive. In general population surveys, the absence of email address lists necessitates traditional sampling methods, including population registers, address databases, telephone directories, or random digit dialing, typically via postal, telephone, or in-person approaches. Recruitment may also use the existing TPSs ("piggybacking"). PWPs thus combine the rigor of probability-based sampling with the cost-efficiency of web surveys.

Several PWPs have been established in EU countries, including the Longitudinal Internet Studies for the Social Sciences (LISS) panel in the Netherlands in 2007, followed by initiatives such as the German Internet Panel (GIP), GESIS Panel, Étude Longitudinale par Internet Pour les Sciences Sociales (ELIPSS), and Norwegian Citizen Panel (NCP) (Bottoni and Sommer, 2019; Kocar and Kaczmirek, 2024). The European Social Survey (ESS) has also conducted three pilot PWP studies (the CROss-National Online Survey [CRONOS] projects) and plans to integrate them into ESS surveys (European Social Survey, 2025).

In the United States (U.S.; Couper, 2017), PWPs began in the commercial sector in 1998 with KnowledgePanel (formerly KnowledgeNetworks), followed by academic and nonprofit initiatives such as the Face-to-Face Recruited Internet Survey Platform (FFRISP), NORC's AmeriSpeak, Pew Research Center's American Trends Panel, RAND Corporation's American Life Panel, and the Understanding America Study (UAS). Other commercial panels include the SSRS Opinion Panel and Gallup Panel.

The U.S. Census Bureau (2025) also established the Household Trends and Outlook Pulse Survey (HTOPS) for official statistics. Several PWPs have been established outside the EU and U.S. (Kocar and Kaczmirek, 2024).

PWPs differ in recruitment, operations, incentives, eligible respondent ages, and size (e.g., Blom et al., 2017; Bosnjak et al., 2018). Most comprise a few thousand participants and rarely exceed 10 000. To improve coverage, some PWPs, such as CRONOS (European Social Survey, 2023) and LISS (Centerdata, 2024), provide participation options for individuals without internet access through in-person, postal, or telephone surveys; however, the overall cost and time implications remain uncertain (e.g., Bach et al., 2024; Kocar and Biddle, 2023).

Comparisons of PWPs and TPSs are methodologically complex, as mode effects and other measurement errors must be distinguished from sampling, noncoverage, processing, and nonresponse errors (Berrens et al., 2003; Cornesse et al., 2020; Struminskaya et al., 2015). Thus, it is not surprising that findings are often inconclusive. Some studies show that PWPs approximate TPS accuracy, with measurement instruments performing similarly or better for sensitive questions, though concerns persist about nonresponse bias and coverage limitations (Cornesse et al., 2020; Pennay et al., 2018; Yeager et al., 2011). Similarly, Bosnjak et al. (2018) revealed minimal differences between the GESIS Panel and related TPSs. However, some studies identified notable discrepancies between PWPs and TPSs (e.g., Ivanovska et al., 2025; Mercer and Lau, 2023; Pforr and Dannwolf, 2017; Struminskaya et al., 2015).

These methodological challenges have limited the adoption of PWPs for official statistics, and caution has generally prevailed. Empirical applications remain rare, primarily concentrated in health statistics (e.g., Ivanovska et al., 2025; Lemcke et al., 2024; Peytchev, 2025), with additional studies on sociodemographic indicators (e.g., Seol et al., 2023), consumer expenditure (Graf et al., 2025), and civic, economic, and well-being measures (Kocar and Biddle, 2023). Many contributions, however, remain conceptual (Bethlehem, 2014; Svensson, 2014). Nevertheless, the suspension of in-person surveys during COVID-19 renewed interest in PWPs as an alternative to TPSs. Recent methodological innovations further support their implementation, including knock-to-nudge (K2N), in which interviewers visit households to encourage participation (Smith, 2022), and electronic questionnaire delivery (EQD), which provides customized tablets to simplify online survey completion (Fitzgerald, 2022).

1.1 Response rates

PWPs face significant challenges, including low response rates, mode effects, panel conditioning, and attrition, all of which may compromise data quality (Kennedy et al., 2016). Response rates – the proportion of individuals invited who participate (American Association for Public Opinion Research, 2023) – and corresponding nonresponse rates are a key concern. Cumulative response rates (i.e., *overall response rates* [ORRs]), which account for all recruitment stages, are a central PWP nonresponse indicator (Callegaro and DiSogra, 2008).

Historically, TPS response rates of 70–80% were considered benchmarks by institutions such as the ESS (Koch et al., 2010), the Organisation for Economic Co-operation and Development (OECD, 2014), and the United States Office of Management and Budget (OMB, 2016). Although these benchmarks have been largely abandoned (Vehovar and Beullens, 2018), many TPS, particularly for official statistics, still exceed 50% (Jabkowski and Kołczyńska, 2020; Vehovar and Beullens, 2018), though this is increasingly rare in developed countries.

By contrast, PWP response rates are typically less than half the rate observed in comparable face-to-face surveys. In Slovenia, face-to-face recruitment yields about 50%, but only half of these respondents join the corresponding PWP. The Slovenian CRONOS-1 pilot panel recorded a 54% initial response rate linked to the 2018 ESS face-to-face survey, yet only 43% participated in the PWP, resulting in an ORR of 23% (Bottoni and Sommer, 2019). A similar pattern appeared in CRONOS-2 (Bottoni, 2023; Maslovskaya and Lugtig, 2022). Postal recruitment without supplemental sampling of internet nonusers generally yields ORRs below 20%, as in the CRONOS surveys (1KA Panel, 2025).

A study of 23 PWPs in 15 countries found that all but four reported ORRs below 20% (Kocar and Kaczmirek, 2024). In the early years of PWP development, panels such as the LISS and FFRISP panels, which used in-person recruitment and support for internet nonusers, achieved ORRs near 40% (Sakshaug et al., 2009; Scherpenzeel and Bethlehem, 2011), though such levels are now unlikely in developed countries. Commercial PWPs typically report even lower, often single-digit, response rates (Kennedy et al., 2016; Mercer et al., 2018; Olson et al., 2021; Pasek, 2016). Persistently low response rates therefore remain a key barrier to PWP use in official statistics.

1.2 Bias of estimates

Within the Total Survey Error (TSE) framework, survey error is classified into sampling and non-sampling components (Groves et al., 2009). Sampling error arises from estimating population parameters based on a sample and is typically quantified using variance measures. Non-sampling error refers to various deviations from true population values, with nonresponse and measurement errors typically most notable, along with coverage and processing errors. The present study does not estimate these error components separately, but their combined effect.

Formally, the bias of an estimate can be formulated as follows:

$$\mathrm{Bias}(\bar{y}) = \mathrm{E}\left[\bar{y}\right] - \bar{Y}, \tag{1}$$

where $\bar{y}$ is the estimate, $E[\bar{y}]$ is the expected value of this estimate, and $\bar{Y}$ is the true value (Groves et al., 2009). Typically, the *absolute value* of this bias is applied, which also holds for the corresponding estimate of the bias($\bar{y}$), i.e.,

$$\mathrm{bias}(\bar{y}) = \bar{y} - \bar{Y}. \tag{2}$$

For example, a Pew Research Center analysis comparing 28 demographic and lifestyle benchmarks across three U.S. PWPs revealed that the estimates of absolute bias ranged from 2.3 to 3.0 percentage points, with an overall average of 2.6 points (Mercer and Lau, 2023). An Australian PWP study reported average absolute bias of 3.6 points for substantive and 5.9 for demographic items (Pennay et al., 2018).

Comparative studies often assess *relative bias* (*RB*), typically expressed as a proportion. When *RB* is estimated from a single survey, we denote this estimate as *RB*' (Groves et al., 2009):

$$\mathrm{RB}' = \frac{\mathrm{bias}(\bar{y})}{\bar{Y}}. \tag{3}$$

RB' is defined only when $\bar{Y} > 0$. The estimate $\bar{y}$ typically derives from the PWP, while the reference value $\bar{Y}$ in the numerator is taken from an external benchmark (e.g., a census or administrative source) or its estimate, such as a high-quality estimate from TPS, which we denote $\bar{Y}'$. Unlike absolute bias, *RB* reflects the size effect: an identical absolute bias of 2 percentage points corresponds to an *RB* of 40% when the population share is 5% (estimate 7%) but only 4% when the share is 50% (estimate 52%).

A meta-analysis of 137 health-related survey items from 14 studies reported a median *RB*' of 12.7% for PWP estimates, with *RB*' ranging from 2.2% for doctor treatment to 23.6% for disability-related estimates (Ivanovska et al., 2025). An analysis of six waves of the Life in Australia PWP across 18 benchmarked items (e.g., health status, home ownership, and life satisfaction) reported median *RB*' values ranging from 4.7% to 6.4% (Kocar and Biddle, 2023). Recalculation of *RB*' for the above-mentioned Pew Research Center study (Mercer and Lau, 2023) showed a median *RB*' of 13.0% and a range of 0%–800% (Ivanovska et al., 2025).

Researchers often overemphasize response rates while underrepresenting other quality indicators such as nonresponse bias, sampling variance, response quality, and costs. Relationships among these indicators are complex, particularly between nonresponse rates and bias (Groves and Peytcheva, 2008), and further complicated by cost considerations (Callegaro et al., 2015; Vehovar and Beullens, 2018). In this context, a consensus statement by survey methodologists (Maslovskaya et al., 2025) recommends prioritizing sample composition and representativeness over raw response rates. Nonetheless, response rates remain the most visible quality parameter when the credibility of national statistics is challenged, as in the United Kingdom Labor Force Survey (LFS), for which response rates are critically low (Casey, 2024; Office for National Statistics, 2025).

2 PROBABILITY-BASED WEB PANELS FOR OFFICIAL STATISTICS

2.1 Methods

To assess PWP use in official statistics, we surveyed NSIs in EU member states and those engaged with Eurostat processes. The survey, sponsored by the University of Ljubljana (UL) and Statistical Office of the Republic of Slovenia (SORS), was conducted from October 11, 2022, to January 25, 2023. SORS provided a contact list. Invitations were sent to methodology section heads in 31 NSIs (27 EU states plus Iceland, Kosovo, Switzerland, and Norway). Reminders followed in October and December. Anonymity was ensured by omitting identifying information and internet protocol addresses.

Nineteen NSIs responded (61% response rate). The questionnaire addressed current practices, plans for PWPs, and perceptions of PWP-related factors. The results offer descriptive insights into the PWP landscape in the EU. Further technical details are available in Vehovar et al. (2023).

2.2 Results

Of the 19 NSIs, five (26%) reported engagement with PWPs. Two had implemented PWPs: one since 2000 using an internal panel recruited face-to-face or by telephone for 12 household and person surveys (response rate 6%–10%, no incentives, personal interviews for non-internet users); and one since 2001 using an external panel of 5 000–10 000 participants recruited by post or face-to-face, with devices and internet access provided to non-internet users (response rate 11%–20%, with incentives). Three NSIs were developing or considering PWPs: one in conceptual development targeting 10 000–20 000 members with a 21%–30% response rate, and two exploring implementation, including one planning an internal PWP.

Table 1 Perceived barriers to PWP adoption among NSIs not yet engaged in PWP implementation (n = 11; 1 = strongly disagree to 7 = strongly agree)

Statement (perceived barriers)	Mean	Standard deviation
Lack of knowledge and expertise regarding this topic	5.6	1.03
Low sample representativeness (e.g., noncoverage)	5.1	1.64
Too little evidence in the literature that this works for official statistics	5.0	1.41
Low response quality (e.g., validity and reliability)	4.6	1.44
Potential surveys too specific to be included in such panels	4.4	1.50
Low response rates	3.9	1.38
Managerial complexity	3.8	1.40
Relatively little cost savings	3.6	1.04

Source: Original data and analysis by the authors

For the above five NSIs that had engaged with PWPs, perceived advantages and disadvantages compared with TPSs were assessed on a 7-point scale (1 = strongly disagree, 7 = strongly agree); the values reported below are means. Advantages included ease of repeated observations (6.0), simplified data processing (5.8), faster fieldwork (5.6), simplified management (5.6), lower costs (5.4), easier sampling (5.2), and better response quality (5.2). Disadvantages were panel conditioning (5.4), panel attrition (5.4), lower representativeness (4.8), coverage problems for non-internet users (4.8), nonresponse error (4.6), lower response quality (4.6), survey topics being too specific (4.2), and lower overall response rates (4.0).

Among the 14 NSIs (74%) not engaged with PWPs, 11 assessed barriers to adoption, also on the same 7-point scale (Table 1). The most cited were lack of knowledge and expertise (5.6), low representativeness (5.1), insufficient evidence of suitability for official statistics (5.0), and low response quality (4.6).

3 BIAS IN PROBABILITY-BASED WEB SURVEYS

To address potential bias in the PWP estimates, we conducted a comprehensive study using estimates from TPSs in the context of Slovenian official statistics.

3.1 Sample and recruitment strategies

We analyzed data collected from October 2022 to July 2023, during which a Slovenian PWP (i.e., the 1KA Panel,[3] run by UL's Faculty of Social Sciences) operated concurrently with multiple TPSs. Data were collected across five waves of the PWP-0 (December 5, 2022–February 17, 2023), 1 (February 10, 2023–April 5, 2023), 2 (April 3, 2023–May 14, 2023), 3 (May 19, 2023–July 3, 2023), and 4 (June 14, 2023–July 23, 2023) – as well as during a parallel recruitment Wave 0' (October 26, 2022–January 23, 2023) to facilitate estimate comparisons. Notably, no surveys pertinent to this analysis were conducted during Wave 3, which focused exclusively on topics outside the scope of official statistics.

The panel included 1 628 Slovenian residents aged 18 or older, recruited via two TPS surveys (i.e., piggybacking) – the Comparative Study of Electoral Systems (CSES) and the Slovene Public Opinion Survey (SJM) 2022/2 – as well as through fresh recruitment from the Central Population Register. All invitations for the fresh sample were sent by post; no web pages or email addresses were used for frame construction or contact. All recruitment sources drew exclusively from population registers. A random systematic sample with implicit stratification by 12 statistical regions and five settlement types ensured equal selection probabilities across all three samples. The fresh sample contributed 717 panelists (20%) from a 3 600-unit sample. Piggybacking added 413 panelists from a 3 000-unit CSES sample and 498 from a 2 500-unit SJM 2022/2 sample. Wave 1 included all 1 628 consenting panelists. Waves 2–4 targeted the same cohort, adjusted for 15 opt-outs.

All waves employed push-to-web methods, with initial wave recruitment via postal invitations offering conditional €10 gift cards for Waves 0 and 0' and unconditional €5 cards for Waves 1–4; email was used solely for reminders. Wave 0 used two postal reminders and a conditional €10 gift card for 2 800 units, 1 540 of which were assigned to incentive experiments. Wave 0' used three postal reminders and a conditional €10 gift card for 2 500 units. Waves 1–4 offered unconditional €5 gift cards to all panelists, including prior nonrespondents. These methods yielded an 18% overall recruitment rate, with register-based recruitment alone achieving 20% and per-wave response rates ranging from 87% to 93%. Wave 1 achieved a 17% ORR with 1% breakoff; across Waves 1–4, the ORR was 16%. In Wave 4, 1,439 respondents participated, representing 88% of recruited panelists – within the 70–90% retention typical of high-quality panels – and 16% of the initial gross sample. Attrition remained low: only 1% of the initial sample (6% of panelists) dropped out between Waves 1 and 4. Completion rates were high: 81% completed all four waves, 9% three, 4% two, 4% one, and 3% none.

---

[3] More information about the 1KA Panel is available at: <*https://panel.1ka.si/?lang_id=2*>.

We compared estimates from various TPSs with PWP Waves 0–4 (i.e., the 1KA Panel) conducted from October 2022 to July 2023. The PWP relied exclusively on web surveys, whereas the TPSs typically employed mixed-mode surveys, combining web with face-to-face, telephone, or mail surveys. In TPSs, the second mode was applied only as a follow-up for nonrespondents. Each TPS drew its initial sample from a single frame based on population registers; no web pages or email addresses were used for frame construction, so each unit appeared only once in the sample. We compared 12 survey sets: 9 for SORS and 3 for the National Institute of Public Health (NIPH) (see Table 2). SORS and NIPH decided to include only selected estimates from these TPSs in the PWP survey; the number of variables selected appears in the final column of Table 2. The corresponding PWP questionnaires were developed using the 1KA[4] software.

Table 2 Overview of the 12 survey sets implemented for the TPSs and replicated in the PWP (the 1KA Panel)

Survey set acronym: full survey set name	TPS provider	TPS mode*	PWP wave	Variable count
Consumers: Consumer opinion	SORS	web, CATI	0′	8
Less salt: Monitoring excreted sodium, potassium, and iodine	NIPH	CAPI	0′	26
COVID: Work and living conditions	SORS	web, CATI	0	58
Adult learning: Adult learning and education	SORS	web, CAPI	0	143
SILC: Living conditions survey (persons)	SORS	CAPI, CATI	1	57
SILC (HH): Living conditions survey (households)	SORS	CAPI, CATI	1	54
Tourism: Tourist trips – domestic population	SORS	web, CATI	2	24
Tourism (HH): Tourist trips – domestic population (households)	SORS	web, CATI	2	16
Drugs: National survey on tobacco, alcohol, and other drugs	NIPH	web, CAPI	2	141
ICT: ICT use in households and by individuals	SORS	web, CAPI	2	39
Activity: Active and inactive population (LFS)	SORS	CAPI, CATI	4	17
Mental health: National survey on attitudes toward mental health	NIPH	web, PAPI	4	68

Note: * the second TPS mode was used only for nonrespondents; no additional sampling frames were introduced. SORS = Statistical Office of the Republic of Slovenia, NIPH = National Institute of Public Health of Slovenia, CAPI = Computer-Assisted Personal Interviewing, CATI = Computer-Assisted Telephone Interviewing, PAPI = Pen-and-Paper Personal Interview, ICT = Information and Communications Technology.
Source: Original data and analysis by the authors

3.2 Bias calculations

In this study, TPS estimates were used as the benchmark under the assumption that they are unbiased, as they represent the best available TPSs in Slovenia. We note that this approach has limitations, as TPS estimates may include some error and may also reflect mode or measurement effects. For the bias analysis, we evaluated the bias for all PWP estimates using Formula (2), assuming $\bar{Y}'$ from TPS as an unbiased estimate of the population value $\bar{Y}$. We then calculated RB' with Formula (3), expressing RB' values as percentages, and considered bias($\bar{y}$) statistically significant at $p < .05$. Analyses were conducted with weighted data in IBM SPSS Statistics and R (R Core Team, 2024). Significance was tested with Welch's *t*-test (i.e., comparing means from two independent samples), and standard errors were calculated using the Taylor linearization method for the ratio estimator, as weights were applied.

[4] *<https://www.1ka.si/d/en>.*

We used the Survey Weighting GUI application (Štrlekar and Vehovar, 2025) for weighting via the raking method, implemented with R's *anesrake* package (Pasek, 2022). Auxiliary variables included gender, age, region, settlement type, and education, with population distributions obtained from SORS. We trimmed weights above 5 to stabilize standard errors, normalized them to a mean of 1 and applied identical weighting to both PWP and TPS data.

3.3 Results

Throughout the remainder of this paper, we only interpret *RB*' that were related to statistically significant biases ($p < .05$). Table 3 shows, for 12 survey sets, the proportion of variables with *RB*' exceeding thresholds of 5%, 10%, and 20%. For subsequent analyses, we focused on *RB*' > 10% to identify substantial bias variation across survey sets. The proportion of variables exceeding this threshold ranged from 0% to 69%. Overall, 205 of 651 variables (31%) had *RB*' > 10%. The median *RB*' across all 651 variables was 16%.

Table 3 The 12 survey sets showing variable counts, median RB', and proportions of variables with RB' with corresponding biases statistically significant at p < .05 and exceeding thresholds (5%, 10%, 20%), listed in descending order of the proportion with RB' > 10%.

Survey set	Variable count	Median *RB'*	*RB'* > 5%		*RB'* > 10%		*RB'* > 20%	
		Proportion	Count	Proportion	Count	Proportion	Count	Proportion
ICT	39	22%	29	74%	27	69%	20	51%
SILC	57	23%	43	75%	35	61%	28	49%
SILC (HH)	54	14%	35	65%	31	57%	19	35%
Activity	17	39%	11	65%	9	53%	7	41%
Tourism (HH)	16	29%	7	44%	7	44%	6	38%
Less salt	26	21%	13	50%	10	38%	8	31%
Adult learning	143	14%	35	24%	33	23%	24	17%
Drugs	141	20%	33	23%	29	21%	21	15%
Mental health	68	10%	15	22%	14	21%	9	13%
COVID	58	11%	10	17%	8	14%	6	10%
Tourism	24	7%	2	8%	2	8%	1	4%
Consumers	8	4%	4	50%	0	0%	0	0%
All survey sets	651	16%	237	36%	205	31%	149	23%

Note: Refer to Table 2 for full names and providers of the survey sets.
Source: Original data and analysis by the authors

Variation in the proportion of estimates with *RB*' exceeding 10%, as well as in median *RB*' values within survey sets, was primarily associated with the survey topic. One survey set contained no such estimates (Consumers), while Tourism (individuals) and COVID constituted only 8% and 14% of the variables where *RB*' exceeded 10%. In 9 out of 12 sets, more than 20% of the estimates had *RB*' > 10%. For survey sets, more than 50% of estimates exceeded this threshold: Activity (53%), SILC (households 57%, persons 61%), and ICT (69%).

Variation in the proportion of estimates with *RB'* values exceeding 10% could be attributed to nonresponse bias, but also to other sources of errors (e.g., noncoverage, processing error, measurement error). Estimates from the PWPs were most comparable to the TPS surveys conducted using push-to-web approaches (see Table 2), with *RB'* > 10% for Tourism (8% of the estimates), COVID (14%), Drugs (21%), Mental health (21%), and Adult learning (23%) (Table 3). However, the ICT survey set, using a web survey plus Computer-Assisted Personal Interviewing (CAPI) follow-ups for nonrespondents in the TPS implementation, had the highest proportion of variables with *RB'* > 10% (69%). High *RB'* values were also found in TPS survey sets administered in person, particularly for SILC (61%), SILC (HH) (57%), Activity (53%), and Less salt (38%), where in-person administration may have amplified response differences, mainly via higher response rates.

A summary comparison of the estimates indicated that relative to TPS respondents, PWP respondents represented a distinct demographic—potentially younger and more digitally engaged – with a higher proportion residing in urban areas. These differences may have also contributed to the observed biases. Additionally, PWP respondents, on average, reported lower levels of satisfaction with life, relationships, and social trust. Their health-related differences included reduced health awareness, lower fruit and vegetable intake, poorer self-perceived health, longer-term medical conditions, but fewer disabilities. They had higher internet usage, higher employment, and lower retirement rates. PWP respondents also showed lower home ownership, more rental accommodation, and more single-person households. Educational attainment and interest in further learning, especially informal or online learning, were higher. Financial hardship was more common and was characterized by material deprivation and delayed payments. Travel was more frequent, longer, and more expensive.

4 DISCUSSION

In this section, we provide a summary evaluation of the RQs and highlight the limitations of the study and future research opportunities.

4.1 The use of probability-based panels for official statistics

In relation to RQ1, the survey of NSIs in the EU indicated that the web survey mode was only partially used for traditional longitudinal panels, demonstrating that comprehensive adoption across household and person surveys continues to present a significant challenge. In this context, PWPs constitute an even more advanced application of web-based data collection. Although adoption remains limited, approximately one-quarter of NSIs reported that they had implemented or initiated the implementation of PWPs. Nonetheless, substantial limitations and perceived barriers persist, as outlined in Section 2.2. The five NSIs that commenced the PWP adoption identified several operational advantages (e.g., ease of repeated observations, simplified survey procedures, increased speed), while the primary disadvantages were directly or indirectly associated with the risks of introducing bias into estimates due to panel conditioning, attrition, compromised representativeness, noncoverage, nonresponse, and reduced response quality. The NSIs that had not yet adopted PWPs expressed caution due to similar concerns; however, the most frequently cited impediments were a lack of technical expertise and insufficient scientific literature addressing these issues.

In addition, both adopters and non-adopters of PWPs recognized that certain official surveys were unsuitable for the PWP integration. Although this was ranked among the less serious concerns, it remains important for specific surveys and necessitates thorough methodological consideration. Often, for example, PWPs have sample sizes that are too small, or complex data collection is required (e.g., biological samples; Couper, 2017).

The issues identified above partially reflect the general findings in the literature on PWPs (Cornesse et al., 2020; Mercer et al., 2018). In particular, repeated participation can lead to panel conditioning, altering respondent behavior, attrition, and undermining longitudinal consistency (Kennedy et al., 2016; Struminskaya and Bosnjak, 2021). Low response rates heighten the risk of nonresponse bias; however, response rate alone is a weak quality indicator, necessitating alternative assessments (e.g., benchmark comparisons, R-indicators). Very low response rates may reduce confidence in survey results unless accompanied by transparent evidence of representativeness. For the PWP or web-first approaches in official statistics, providing nonresponse-bias diagnostics is recommended to support stakeholders' confidence in the results (Cornesse and Bosnjak, 2018; Maslovskaya et al., 2025).

The appropriateness of PWPs varies by topic and mode, requiring methodological refinements to reduce bias without escalating costs (Bethlehem, 2014; Callegaro et al., 2015; Cornesse et al., 2020). However, limited expertise in advanced statistical methods and a scarcity of robust, peer-reviewed research on mitigating self-selection and noncoverage biases have hindered their wider adoption by NSIs (Svensson, 2014). Although PWPs are obviously less suitable when they involve high nonresponse or coverage errors, emerging approaches, such as adaptive survey designs and dual-frame methodologies, offer promising avenues for balancing costs and errors, emphasizing the need for comprehensive optimization beyond merely improving response rates or reducing bias (Schouten et al., 2017; Tourangeau et al., 2017; Vannieuwenhuyze, 2014).

As described in Section 2, PWPs have been implemented in multiple countries, although their integration into official statistics remains limited and uneven. A recent meta-analysis (Kocar and Kaczmirek, 2024) documents global adoption patterns and challenges. From a broader perspective, PWPs are unlikely to fully replace TPSs in the foreseeable future. Instead, official statistics increasingly adopt web-first mixed-mode designs, in which web surveys are supplemented by targeted interviewer follow-ups (e.g., CATI or CAPI) among nonrespondents or offline populations. This strategy preserves most cost and speed advantages of web surveys while reducing coverage and nonresponse issues, leaving TPSs essential for topics in which PWPs alone remain inadequate (Vannieuwenhuyze, 2014).

4.2 Biases of the estimates in probability-based web surveys

With respect to RQ2, concerning the extent of the bias in estimates derived from PWPs compared to those from TPSs, we conducted a comprehensive empirical study of Slovenian official statistics, and the results indicated a mixed pattern.

Several TPS survey sets exhibited a high proportion of estimates with *RB'* exceeding 10%, particularly for ICT (69%), SILC (61%), SILC (HH) (57%), and Activity (53%). These findings suggest that surveys involving complex socioeconomic or behavioral indicators are more susceptible to bias. This observation is consistent with Cornesse et al. (2020), who reported that PWPs tend to yield estimates comparable to TPSs for general attitudinal measures but display greater bias for behavioral or domain-specific variables, such as the ICT usage and socioeconomic indicators. Notably, the ICT survey demonstrated the highest proportion of estimates with *RB'* > 10% (69%), primarily attributable to the exclusion of internet nonusers from PWPs, as well as to the overrepresentation of individuals with higher levels of technological access or proficiency. The bias in ICT estimates thus reflects pronounced nonresponse and undercoverage issues (e.g., Groves and Peytcheva, 2008), as well as selection bias (e.g., Scherpenzeel and Bethlehem, 2011). Similar discrepancies were observed by Šoštarič (2020) for the CRONOS-1 panel, for which variables related to internet use exhibited high deviations compared to TPS estimates.

In contrast, certain survey sets showed comparatively low proportions of *RB'* > 10%, particularly for Consumers (0%), Tourism (8%), and COVID (14%). This pattern indicates that surveys involving less sensitive or cognitively simpler topics (e.g., consumer opinions) are less affected by PWP specifics. Conversely, surveys addressing sensitive domains (e.g., health, wealth, or life satisfaction) tend to encounter biases, warranting careful methodological consideration. However, mode-related effects are generally small (Tourangeau et al., 2013) and limited to socially desirable variables, such as those related to sensitive topics (e.g., health or income), for which PWPs may even yield more accurate responses due to reduced social desirability bias. The relatively low bias observed for Tourism was somewhat unexpected, although it may be attributed to the specific nature of the selected variables because some key variables (i.e., questions on domestic travel) within this survey demonstrated substantial discrepancies (see Bučar, 2022).

Overall, 205 of the 651 variables (31%) in the PWP had *RB'* values exceeding 10%, while the overall median *RB'* was 16%. This median is relatively high compared to the other evaluations presented in Section 2.2, where the median values ranged from 4.7% to 13%. The primary explanation for this discrepancy is probably the specific characteristics of the TPSs used in this study. Unlike most other surveys, which primarily covered health or general social indicators, the TPS sets we analyzed focused on core behavioral variables regarding official statistics.

The results also revealed that PWP respondents constituted a specific segment of the population characterized by lower levels of life satisfaction, social trust, health-related awareness, poorer health (but with fewer disabilities), more rental accommodation, less home ownership, and generally lower social status, although more travel was observed. However, this segment has higher internet usage, digital literacy, education involvement, and employment levels.

4.3 Limitations and future research

There are certain technical limitations worth highlighting with respect to the survey of NSIs. Namely, 19 of 31 NSIs responded (61%), which could have introduced a potential bias if nonresponding NSIs differed systematically regarding the PWP engagement. NSIs with fewer resources or lower PWP involvement may have chosen not to respond. In addition, although the sample included a comprehensive set of Eurostat-affiliated NSIs, generalizability to non-Eurostat NSIs remains limited. Future researchers should consider expanding the scope to encompass additional international statistical institutions. Moreover, incorporating more targeted items into the questionnaire may facilitate the identification of a wider range of technical, legal, operational, or cultural barriers to PWP implementation.

Several technical limitations also pertain to the analysis of *RB'* in estimates taken from PWPs. First, the PWPs were limited by 1KA panel characteristics, particularly the exclusion of internet nonusers. Integrating PWPs with mixed-mode approaches may enhance response rates (Callegaro et al., 2015; Scherpenzeel and Bethlehem, 2011; Vannieuwenhuyze, 2014) and reduce coverage bias (Eckman, 2016), especially among older or less technologically adept individuals. Second, recruitment relied solely on mailed invitations, which generally yield lower response rates than personal contact (Blom et al., 2017; Maslovskaya and Lugtig, 2022). Additionally, consent procedures were not optimized or experimentally varied (Bottoni and Sommer, 2019; Sakshaug et al., 2020). Future researchers should further investigate recruitment and consent strategies (Callegaro et al., 2015; Vehovar and Beullens, 2018). Third, the relatively small sample size in this PWP study (i.e., 1 628 panelists) may have reduced the quality of the estimates and compromised representativeness. Fourth, the omnibus structure of the PWP survey may have shaped the respondent context differently from that of corresponding single-topic TPS surveys, thereby constraining topic-specific generalizability. Fifth, the weighting model employed may have insufficiently accounted for unobserved confounders, such as attitudinal or behavioral traits, that influence PWP nonresponse or coverage bias (Mercer et al., 2018). In addition, the trimming of the weights above a threshold of 5 may have resulted in many extreme values. Future researchers should assess more advanced weighting strategies, including propensity score adjustments (Vehovar and Beullens, 2018) and adaptive survey designs that incorporate real-time error diagnostics (Schouten et al., 2017). Integrating topic-relevant auxiliary variables may further improve the accuracy of weighting procedures for future PWP applications (Callegaro et al., 2014). Sixth, transitioning from mailed vouchers to digital incentives may enhance cost efficiency. Finally, although attrition across the five PWP waves was minimal (one percentage point in each wave; see Section 4.1), both panel conditioning and attrition may have introduced data quality concerns (Kennedy et al., 2016).

In addition to technical issues, several conceptual aspects warrant discussion. Although we identified topic-specific variations in *RB'* (e.g., high *RB'* for ICT but low for Consumers), we did not assess the possible causal mechanisms of these differences. In this context, the high variation in the proportion of estimates with *RB'* > 10% (range: 0%–69% across 12 survey sets) indicates the need for a more granular understanding of the substantive and methodological factors affecting *RB'* levels. Separate treatment of nonresponse, noncoverage, and measurement errors would also provide additional insight (e.g., Bosnjak et al., 2018; Couper, 2011; Groves, 2006; Struminskaya et al., 2015; Vehovar and Beullens, 2018). Furthermore, we limited the *RB'* comparisons to cross-sectional estimates across PWP waves without assessing longitudinal trends or measurement equivalence (Cernat and Revilla, 2021; Cornesse and Blom, 2020). Future researchers should incorporate item-level diagnostics, response quality indicators, and error decomposition frameworks (e.g., Pennay et al., 2018; Unangst et al., 2020).

Furthermore, the bias-related findings are limited to a single PWP in Slovenia and may not be generalizable to contexts with different demographic, societal, or technological characteristics, although Slovenia is often considered a typical EU country, with median levels of socioeconomic indicators, including an internet penetration rate of 91% among individuals aged 16–74 years (Jevnikar, 2025). Similarly, the 651 variables we analyzed were purposively selected from 12 survey sets, although many more variables were considered in the corresponding TPSs. Although these selected variables were often essential for or typical of the corresponding TPSs, they may not be the most consistently representative. Ideally, a PWP would fully represent an entire TPS survey.

Nevertheless, despite these constraints, the 1KA Panel demonstrated favorable recruitment and retention outcomes, effectively functioning as a prototype for web probability panels in Slovenia. The cumulative response rate across all three recruitment channels and five PWP waves (2022–2023), based on a 9 100-unit sample, was 16%, aligning with the results reported by Villar et al. (2018) and Kocar and Kaczmirek (2024). This study thus represents a typical scenario in which PWPs (rather than TPSs) are used for official statistics. Increased resource allocation to improve 1KA Panel operations and procedures could potentially raise ORRs (e.g., above 20%) and reduce certain biases; however, it is likely that the overall pattern of the biases will remain largely unchanged (i.e., a considerable number of topics in official statistics seem to involve unacceptable biases when included in PWPs).

CONCLUSION

This article addresses a notable gap in the literature regarding the use, experience, and prospects of PWPs for official statistics. In this context, a survey conducted among NSIs in the EU provided a specific insight into the adoption of PWPs for official statistics. The related findings indicated that the adoption process is well underway, with one-quarter of NSIs already engaged in the PWP implementation. The results further revealed that, in addition to various methodological challenges that may increase the bias of estimates (e.g., attrition, conditioning, nonresponse, noncoverage, and measurement), a primary concern among NSIs was the lack of expertise and established good practices related to PWPs.

The other focus of this article was the bias of the PWP estimates. We employed a case study of Slovenian official statistics using the 1KA Panel to compare the PWP estimates with those from concurrently conducted TPS sets based on 651 variables taken from 12 official statistics surveys. The results revealed substantial discrepancies, with the PWP estimates exhibiting considerably more bias than TPS estimates. The median *RB'* across corresponding estimates was 16%, while the median proportion of variables with *RB'* exceeding 10% across the 12 survey sets was 31%. The PWP estimates demonstrated higher *RB'* for behavioral and factual variables, particularly for ICT usage (69% of estimates with *RB'* > 10%) and living conditions (SILC, 61%; Activity, 53%). In contrast, attitudinal variables exhibited lower *RB'*, such as for

Consumers (0%). In general, the biases identified for 651 variables suggest that PWP respondents have specific characteristics: more active lifestyles and internet usage but also lower social statuses.

Identifying bias patterns across survey topics allowed us to propose a framework for strategically allocating the use of TPSs to surveys with low or moderate *RB'* values. The findings suggest that TPSs may be unnecessary for surveys with consistently low *RB'* values, such as those focusing on attitudinal measures. In particular, our results suggest that PWPs may be suitable for producing certain estimates in surveys focused on general social attitudes, where relatively low bias levels were observed (e.g., in the Consumers and COVID survey sets). However, the potential cost savings of larger PWP samples may be insufficient to offset the higher *RB'* values observed in behavioral and factual domains. Therefore, despite the cost- and time-efficiency gains offered by PWPs, the extent of bias they involve remains a substantial limitation.

Thus, this article offers a general overview of PWP adoption for official statistics and provides a case study illustrating which survey topics are appropriate for PWPs as potential replacements for TPSs. Future researchers should address the limitations of empirical studies and refine their methodological approaches.

ACKNOWLEDGMENT

The Slovenian Research and Innovation Agency supported this research (grant numbers P5-0399, J5-3100, V5-2157, and J5-50159).

References

1KA PANEL. (2025). *1KA Panel Methodology* [online]. Ljubljana: Centre for Social Informatics. [cit. 22.9.2025]. <https://panel.1ka.si/?a=metodologija&lang_id=2>.

AMERICAN ASSOCIATION FOR PUBLIC OPINION RESEARCH. (2023). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys* [online]. Alexandria, Virginia: AAPOR. [cit. 22.9.2025]. <https://aapor.org/wp-content/uploads/2023/05/Standards-Definitions-10th-edition.pdf>.

BACH, R. L., CORNESSE, C., DAIKELER, J. (2024). Equipping the Offline Population with Internet Access in an Online Panel: Does It Make a Difference? [online]. *Journal of Survey Statistics and Methodology*, 12(1): 80–93. <http://doi.org/10.1093/jssam/smad003>.

BERRENS, R. P., BOHARA, A. K., JENKINS-SMITH, H., SILVA, C., WEIMER, D. L. (2003). The Advent of Internet Surveys for Political Research: a Comparison of Telephone and Internet Samples [online]. *Political Analysis*, 11(1): 1–22. <http://doi.org/10.1093/pan/11.1.1>.

BETHLEHEM, J. (2014). Web Surveys in Official Statistics. In: ENGEL, U., JANN, B., LYNN, P., SCHERPENZEEL, A., STURGIS, P. (eds.). *Improving Survey Methods*, New York: Routledge.

BLOM, A. G., HERZING, J. M. E., CORNESSE, C., SAKSHAUG, J. W., KRIEGER, U., BOSSERT, D. (2017). Does the Recruitment of Offline Households Increase the Sample Representativeness of Probability-Based Online Panels? Evidence From the German Internet Panel [online]. *Social Science Computer Review*, 35(4): 498–520. <http://doi.org/10.1177/0894439316651584>.

BOSNJAK, M., DANNWOLF, T., ENDERLE, T., SCHAURER, I., STRUMINSKAYA, B., TANNER, A., WEYANDT, K. W. (2018). Establishing an Open Probability-Based Mixed-Mode Panel of the General Population in Germany: The GESIS Panel [online]. *Social Science Computer Review*, 36(1): 103–115. <http://doi.org/10.1177/0894439317697949>.

BOTTONI, G. (2023). *CROss-National Online Survey (CRONOS-2) Panel: Data and Documentation User Guide* [online]. London: ESS ERIC. [cit. 22.9.2025]. <https://www.europeansocialsurvey.org/sites/default/files/2023-07/CRONOS2_user_guide_e01_0.pdf>.

BOTTONI, G., SOMMER, E. (2019). *Best Strategies to Recruit and Maintain Web Panels* [online]. London: Synergies for Europe's Research Infrastructures in the Social Sciences. [cit. 22.9.2025]. <https://seriss.eu/resources/deliverables/>.

BUČAR, N. (2022). *Stroški in Napake v Anketi Turistična Potovanja Domačega Prebivalstva* [online]. Ljubljana: University of Ljubljana. [cit. 22.9.2025]. <https://repozitorij.uni-lj.si/IzpisGradiva.php?id=143434>.

CALLEGARO, M., BAKER, R. P., BETHLEHEM, J., GÖRITZ, A. S., KROSNICK, J. A., LAVRAKAS, P. J. (2014). (eds.). *Online Panel Research: A Data Quality Perspective*. Chichester: Wiley.

CALLEGARO, M., DISOGRA, C. (2008). Computing Response Metrics for Online Panels [online]. *Public Opinion Quarterly*, 72(5): 1008–1032. <https://doi.org/10.1093/poq/nfn065>.

CALLEGARO, M., LOZAR MANFREDA, K., VEHOVAR, V. (2015). *Web Survey Methodology*. London: SAGE.

CASEY, B. (2024). *Can We Trust the UK Labour Force Survey?* [online]. London: UK in a Changing Europe. [cit. 22.9.2025]. <https://ukandeu.ac.uk/can-we-trust-the-uk-labour-force-survey/>.

CENTERDATA. (2024). *LISS Panel* [online]. Tilburg: LISS Panel. [cit. 22.9.2025]. <https://www.lissdata.nl/>.

CERNAT, A., REVILLA, M. (2021). Moving From Face-to-Face to a Web Panel: Impacts on Measurement Quality [online]. *Journal of Survey Statistics and Methodology*, 9(4): 745–763. <http://doi.org/10.1093/jssam/smaa007>.

COCHRAN, W. G. (1977). *Sampling Techniques.* 3rd Ed. New York: Wiley.

CORNESSE, C., BLOM, A. G. (2020). Response Quality in Nonprobability and Probability-Based Online Panels [online]. *Sociological Methods & Research*. <http://doi.org/10.1177/0049124120914940>.

CORNESSE, C., BLOM, A. G., DUTWIN, D., KROSNICK, J. A., DE LEEUW, E. D., LEGLEYE, S., PASEK, J., PENNAY, D., PHILLIPS, B., SAKSHAUG, J. W., STRUMINSKAYA, B., WENZ, A. (2020). A Review of Conceptual Approaches and Empirical Evidence on Probability and Nonprobability Sample Survey Research [online]. *Journal of Survey Statistics and Methodology*, 8(1): 4–36. <http://doi.org/10.1093/jssam/smz041>.

CORNESSE, C., BOSNJAK, M. (2018). Is There an Association Between Survey Characteristics and Representativeness? A Meta-Analysis [online]. *Survey Research Methods*, 12(1): 1–13. <http://doi.org/10.18148/srm/2018.v12i1.7205>.

COUPER, M. P. (2011). The Future of Modes of Data Collection [online]. *Public Opinion Quarterly*, 75(5): 889–908. <http://doi.org/10.1093/poq/nfr046>.

COUPER, M. P. (2017). New Developments in Survey Data Collection [online]. *Annual Review of Sociology*, 43(1): 121–145. <http://doi.org/10.1146/annurev-soc-060116-053613>.

ECKMAN, S. (2016). Does the Inclusion of Non-Internet Households in a Web Panel Reduce Coverage Bias? [online]. *Social Science Computer Review*, 34(1): 41–58. <http://doi.org/10.1177/0894439315572985>.

EUROPEAN SOCIAL SURVEY. (2023). *CROss-National Online Survey (CRONOS-2) Panel* [online]. London: ESS ERIC. [cit. 22.9.2025]. <https://www.europeansocialsurvey.org/methodology/methodological-research/modes-data-collection/cronos>.

EUROPEAN SOCIAL SURVEY. (2025). *New Panel Survey Data Now Available* [online]. London: ESS ERIC. [cit. 22.9.2025]. <https://www.europeansocialsurvey.org/news/article/new-panel-survey-data-now-available>.

FITZGERALD, R. (2022). *ESS Electronic Questionnaire Device* [online]. London: ESS ERIC. [cit. 22.9.2025]. <https://www.ncrm.ac.uk/documents/EQD-presentation_rory%20fitzgerald.pdf>.

GOLINI, N., RIGHI, P. (2024). Integrating Probability and Big Non-Probability Samples Data to Produce Official Statistics [online]. *Statistical Methods & Applications*, 33(2): 555–580. <http://doi.org/10.1007/s10260-023-00740-y>.

GRAF, N. L., JONES, G., MILLER, T. (2025). Testing Self-Administration of the Consumer Expenditure Diary Survey in a Probability-Based Online Panel [online]. Los Angeles: *CIPHER 2025 Conference*. [cit. 22.9.2025]. <https://dornsife.usc.edu/cesr/wp-content/uploads/sites/54/2025/03/GRAF-main.pdf>.

GROVES, R. M. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys [online]. *Public Opinion Quarterly*, 70(5): 646–675. <http://doi.org/10.1093/poq/nfl033>.

GROVES, R. M., FOWLER, F. J., COUPER, M. P., LEPKOWSKI, J. M., SINGER, E., TOURANGEAU, R. (2009). *Survey Methodology*. 2nd Ed. New York: Wiley.

GROVES, R. M., PEYTCHEVA, E. (2008). The Impact of Nonresponse Rates on Nonresponse Bias: a Meta-Analysis [online]. *Public Opinion Quarterly*, 72(2): 167–189. <http://doi.org/10.1093/poq/nfn011>.

HAYS, R. D., LIU, H., KAPTEYN, A. (2015). Use of Internet Panels to Conduct Surveys [online]. *Behavior Research Methods*, 47(3): 685–690. <http://doi.org/10.3758/s13428-015-0617-9>.

IVANOVSKA, A., BOSNJAK, M., VEHOVAR, V. (2025). Relative Bias in Health Estimates From Probability-Based Online Panels: Systematic Review and Meta-Analysis [online]. *Journal of Health and Rehabilitation Sciences*, 4(1): 10–23. <http://doi.org/10.33700/jhrs.4.1.137>.

JABKOWSKI, P., KOŁCZYŃSKA, M. (2020). Sampling and Fieldwork Practices in Europe: Analysis of Methodological Documentation From 1,537 Surveys in Five Cross-National Projects, 1981–2017 [online]. *Methodology*, 16(3): 186–207. <http://doi.org/10.5964/meth.2795>.

JEVNIKAR, Ž. (2025). *Usage of Internet in Households and by Individuals* [online]. Ljubljana: Statistical Office of the Republic of Slovenia. [cit. 22.9.2025]. <https://www.stat.si/StatWeb/en/News/Index/13336>.

KENNEDY, C., MERCER, A. W., KEETER, S., HATLEY, N., MCGEENEY, K., GIMENEZ, A. (2016). *Evaluating Online Nonprobability Surveys* [online]. Washington, DC: Pew Research Center. [cit. 22.9.2025]. <https://www.pewresearch.org/wp-content/uploads/2016/04/Nonprobability-report-May-2016-FINAL.pdf>.

KOCAR, S., BIDDLE, N. (2023). Do We Have to Mix Modes in Probability-Based Online Panel Research to Obtain More Accurate Results? [online]. *Methods, Data, Analyses*, 17(1): 93–120. <http://doi.org/10.12758/mda.2022.11>.

KOCAR, S., KACZMIREK, L. (2024). A Meta-Analysis of Worldwide Recruitment Rates in 23 Probability-Based Online Panels, Between 2007 and 2019 [online]. *International Journal of Social Research Methodology*, 27(5): 589–604. <http://doi.org/10.1080/13645579.2023.2242202>.

KOCH, A., FITZGERALD, R., STOOP, I., WIDDOP, S. (2010). *Field Procedures in the European Social Survey Round 5: Enhancing Response Rates* [online]. London: ESS ERIC. [cit. 22.9.2025]. <https://www.academia.edu/88218012/Field_Procedures_in_the_European_Social_Survey_Round_5_Enhancing_Response_Rates>.

LEMCKE, J., LOSS, J., ALLEN, J., ÖZTÜRK, I., HINTZE, M., DAMEROW, S., KUTTIG, T., WETZSTEIN, M., HÖVENER, C., HAPKE, U., ZIESE, T., SCHEIDT-NAVE, C., SCHMICH, P. (2024). Health in Germany: Establishment of a Population-Based Health Panel [online]. *Journal of Health Monitoring*, 9(S2): 1–21. <http://doi.org/10.25646/11992.2>.

MASLOVSKAYA, O., LUGTIG, P. (2022). Representativeness in Six Waves of CROss-National Online Survey (CRONOS) Panel [online]. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. <http://doi.org/10.1111/rssa.12801>.

MASLOVSKAYA, O., LYNN, P., CALDERWOOD, L., DURRANT, G., FITZGERALD, R., NICOLAAS, G., WILLIAMS, J. (2025). *Response Rates Position Statement* [online]. Colchester: Survey Futures. [cit. 22.9.2025]. <https://surveyfutures.net/publications/response-rates-position-statement/>.

MERCER, A., LAU, A. (2023). *Comparing Two Types of Online Survey Samples* [online]. Washington, DC: Pew Research Center. [cit. 22.9.2025]. <https://www.pewresearch.org/methods/2023/09/07/comparing-two-types-of-online-survey-samples>.

MERCER, A. W., LAU, A., KENNEDY, C. (2018). *For Weighting Online Opt-In Samples, What Matters Most?* [online]. Washington, DC: Pew Research Center. [cit. 22.9.2025]. <https://www.pewresearch.org/methods/2018/01/26/for-weighting-online-opt-in-samples-what-matters-most/>.

OECD. (2014). *PIAAC Technical Standards and Guidelines* [online]. Paris: OECD. [cit. 22.9.2025]. <https://www.oecd.org/en/about/programmes/piaac/piaac-data.html>.

OFFICE FOR NATIONAL STATISTICS. (2025). *Labour Force Survey Performance and Quality Monitoring Reports* [online]. Newport: Office for National Statistics. [cit. 22.9.2025]. <https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyperformanceandqualitymonitoringreports>.

OFFICE OF MANAGEMENT AND BUDGET. (2016). *Questions and Answers When Designing Surveys for Information Collections* [online]. Washington, DC: Office of Management and Budget. [cit. 22.9.2025]. <https://obamawhitehouse.archives.gov/sites/default/omb/inforeg/pmc_survey_guidance_2006.pdf>.

OLSON, K., WAGNER, J., ANDERSON, R. (2021). Survey Costs: where Are We and what Is the Way forward? [online]. *Journal of Survey Statistics and Methodology*, 9(5): 921–942. <http://doi.org/10.1093/jssam/smaa014>.

PASEK, J. (2016). When Will Nonprobability Surveys Mirror Probability Surveys? Considering Types of Inference and Weighting Strategies as Criteria for Correspondence [online]. *International Journal of Public Opinion Research*, 28(2): 269–291. <http://doi.org/10.1093/ijpor/edv016>.

PASEK, J. (2022). *Anesrake: ANES Raking Implementation* [online]. R Foundation for Statistical Computing. <http://doi.org/10.32614/CRAN.package.anesrake>.

PENNAY, D., NEIGER, D., LAVRAKAS, P. J., BORG, K. (2018). *The Online Panels Benchmarking Study: a Total Survey Error Comparison of Findings from Probability-Based Surveys and Non-Probability Online Panel Surveys in Australia* [online]. Canberra: ANU Centre for Social Research and Methods. [cit. 22.9.2025]. <https://csrm.cass.anu.edu.au/research/publications/online-panels-benchmarking-study-total-survey-error-comparison-findings>.

PEYTCHEV, A. (2025). *Evaluating Error From the Addition of Web Panel Data to a National Probability-Based Interviewer-Administered Survey* [online]. Washington, D.C.: CIPHER 2025 Conference. [cit. 22.9.2025]. <https://dornsife.usc.edu/cesr/wp-content/uploads/sites/54/2025/03/3-Peytchev.pdf>.

PFORR, K., DANNWOLF, T. (2017). What Do We Lose with Online-Only Surveys? Estimating the Bias in Selected Political Variables Due to Online Mode Restriction [online]. *Statistics, Politics and Policy*, 8(1): 105–120. <http://doi.org/10.1515/spp-2016-0004>.

R CORE TEAM. (2024). *The R Project for Statistical Computing* [online]. R Foundation for Statistical Computing. [cit. 22.9.2025]. <https://www.R-project.org>.

SAKSHAUG, J. W., HÜLLE, S., SCHMUCKER, A., LIEBIG, S. (2020). Panel Survey Recruitment with or without Interviewers? Implications for Nonresponse, Panel Consent, and Total Recruitment Bias [online]. *Journal of Survey Statistics and Methodology*, 8(3): 540–565. <http://doi.org/10.1093/jssam/smz012>.

SAKSHAUG, J. W., KROSNICK, J. A., ACKERMANN, A., MALKA, A., DEBELL, M., TURAKHIA, C. (2009). Dispositions and Outcome Rates in the 'Face-to-Face Recruited Internet Survey Platform' (the FFRISP). *AAPOR 64th Annual Conference,* Hollywood, Florida.

SÄRNDAL, C.-E., SWENSSON, B., WRETMAN, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

SCHERPENZEEL, A., BETHLEHEM, J. (2011). How Representative Are Online Panels? Problems of Coverage and Selection and Possible Solutions. In: DAS, M., ESTER, P., KACZMIREK, L. (eds.). *Social and Behavioral Research and the Internet*, New York: Routledge, pp. 105–132.

SCHOUTEN, B., PEYTCHEV, A., WAGNER, J. (2017). *Adaptive Survey Design*. New York: CRC Press.

SEOL, D.-H., JANG, D.-H., LOCASCIO, S. P. (2023). RDD with Follow-Up Texting: a New Attempt to Build a Probability-Based Online Panel in South Korea [online]. *Asian Journal for Public Opinion Research*, 11(3): 257–273. <http://doi.org/10.15206/ajpor.2023.11.3.257>.

SMITH, P. (2022). *What Is Knock-to-Nudge and Does It Have a Future?* [online]. London: Ipsos. [cit. 22.9.2025]. <https://www.ncrm.ac.uk/documents/4_Patten%20Smith_Data%20collection%20network%202_2_22.pdf>.

ŠOŠTARIČ, T. (2020). *Izzivi Spletnih Verjetnostnih Panelov: Primer Panela CRONOS* [online]. Ljubljana: University of Ljubljana. [cit. 22.9.2025]. <https://repozitorij.uni-lj.si/Dokument.php?id=135272&lang=slv>.

ŠTRLEKAR, L., VEHOVAR, V. (2025). *SurveyWeightingGUI* [online]. Ljubljana: 1KA Panel. [cit. 22.9.2025]. <https://enklikanketa.shinyapps.io/SurveyWeightingGUI/>.

STRUMINSKAYA, B., BOSNJAK, M. (2021). Panel Conditioning: Types, Causes, and Empirical Evidence of what We Know So Far [online]. In: LYNN, P. (eds). *Advances in Longitudinal Survey Methodology*, New York: Wiley, pp. 272–301. <http://doi.org/10.1002/9781119376965.ch12>.

STRUMINSKAYA, B., DE LEEUW, E., KACZMIREK, L. (2015). Mode System Effects in an Online Panel Study: Comparing a Probability-Based Online Panel with Two Face-to-Face Reference Surveys [online]. *Methods, Data, Analyses*, 9(1): 3–56. <http://doi.org/10.12758/mda.2015.001>.

SVENSSON, J. (2014). Web Panel Surveys – a Challenge for Official Statistics [online]. *Proceedings of Statistics Canada Symposium*, Ottawa: Statistics Canada. [cit. 22.9.2025]. <https://www.statcan.gc.ca/eng/conferences/symposium2014/program/14280-eng.pdf>.

TOURANGEAU, R., BRICK, J. M., LOHR, S., LI, J. (2017). Adaptive and Responsive Survey Designs: a Review and Assessment [online]. *Journal of the Royal Statistical Society*, 180(1): 203–223. <http://doi.org/10.1111/rssa.12186>.

TOURANGEAU, R., CONRAD, F., COUPER, M. (2013). *The Science of Web Surveys*. New York: Oxford University Press.

UNANGST, J., AMAYA, A. E., SANDERS, H. L., HOWARD, J., FERRELL, A., KARON, S., DEVER, J. A. (2020). A Process for Decomposing Total Survey Error in Probability and Nonprobability Surveys: a Case Study Comparing Health Statistics in US Internet Panels [online]. *Journal of Survey Statistics and Methodology*, 8(1): 62–88. <http://doi.org/10.1093/jssam/smz040>.

U.S. CENSUS BUREAU. (2025). Agency Information Collection Activities; Submission to the Office of Management and Budget (OMB) for Review and Approval; Comment Request; High-Frequency Surveys Program/Household Trends and Outlook Pulse Survey (HTOPS) [online]. Washington, DC, *Federal Register*, 90(39): 10879. [cit. 22.9.2025]. <https://www.federalregister.gov/documents/2025/02/28/2025-03194/agency-information-collection-activities-submission-to-the-office-of-management-and-budget-omb-for>.

VANNIEUWENHUYZE, J. (2014). On the Relative Advantage of Mixed-Mode Versus Single-Mode Surveys [online]. *Survey Research Methods*, 8(1): 31–42. <http://doi.org/10.18148/srm/2014.v8i1.5500>.

VEHOVAR, V., BEULLENS, K. (2018). Cross-National Issues in Response Rates [online]. In: VANNETTE, D. L., KROSNICK, J. A. (eds.). *The Palgrave Handbook of Survey Research*, Cham: Springer International Publishing, pp. 29–42. <http://doi.org/10.1007/978-3-319-54395-6_6>.

VEHOVAR, V., ČEHOVIN, G., PRAČEK, A. (2023). *The Use of Probability Web Panels in National Statistical Institutes* [online]. Ljubljana: Centre for Social Informatics. [cit. 22.9.2025]. <https://1ka.cdi.si/uploadi/editor/doc/1676613266NSIReport-final.pdf>.

VEHOVAR, V., TOEPOEL, V., STEINMETZ, S. (2016). Non-Probability Sampling. In: WOLF, C., JOYE, D., SMITH, T. W., FU, Y. (eds.). *The SAGE Handbook of Survey Methodology*, London: SAGE, pp. 329–346.

VILLAR, A., SOMMER, E., FINNØY, D., GAIA, A., BERZELAK, N., BOTTONI, G. (2018). *Cross-National Online Survey (CRONOS) Panel: Data and Documentation User Guide* [online]. London: ESS ERIC. [cit. 22.9.2025]. <https://openaccess.city.ac.uk/id/eprint/30024/1/CRONOS_user_guide_e01_1.pdf>.

YEAGER, D. S., KROSNICK, J. A., CHANG, L., JAVITZ, H. S., LEVENDUSKY, M. S., SIMPSER, A., WANG, R. (2011). Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples [online]. *Public Opinion Quarterly*, 75(4): 709–747. <http://doi.org/10.1093/poq/nfr020>.

APPENDIX

The dataset corresponding to the 651 estimates we analyzed is available at:
<https://doi.org/10.5281/zenodo.15829184>.

Domestic Competition and Export Performance in the Beer Industry: Evidence from the EU

Miriam Brellíková[1] | *Prague University of Economics and Business, Prague, Czechia*

Received 15.7.2025 (revision received 30.11.2025), **Accepted** (reviewed) 10.1.2026, **Published** 12.6.2026

Abstract

This paper investigates how domestic market competition influences beer export performance in the European Union between 2014 and 2023. Employing a gravity model framework and Poisson Pseudo-Maximum Likelihood estimation, it incorporates standard trade variables alongside two measures of domestic competition: the Herfindahl-Hirschman Index and the ratio of microbreweries to total breweries. The empirical results, based on 1 277 observations, indicate that lower market concentration and a higher share of microbreweries are significantly associated with greater export volumes. Conceptually, the paper extends heterogeneous-firm trade to a mature, differentiated consumer industry and identifies both an efficiency/selection channel and a non-price differentiation channel. Empirically, it offers new sector-level, multi-country evidence that links domestic market dynamism and firm diversity to external competitiveness, and informs policies on competition and SME support.

Keywords	***DOI***	***JEL code***
Competition, export performance, international trade, beer industry, microbrewery	*https://doi.org/10.54694/stat.2025.34*	*D43, F14, L11, L66, O24*

INTRODUCTION

The European beer industry has undergone significant changes over the last decade, especially with the increasing number of microbreweries, defined as breweries producing up to 10 000 hectoliters annually, across the continent. As of 2023, craft breweries accounted for approximately 20% of total beer production in the United Kingdom (Brewers of Europe, 2025; Society of Independent Brewers, 2024) and 7–8% in France (Businesscoot, 2025), illustrating their substantial presence in two of Europe's

[1] Faculty of International Relations, Prague University of Economics and Business, W. Churchill Sq. 1938/4, 130 67 Prague 3, Czechia. Phone: (+420) 224 09 5297, e-mail: miriam.brellikova@vse.cz. ORCID: <https://orcid.org/0009-0009-3355-7335>.

most active microbrewery markets. The Czech Republic, consistently the world's top beer consumer, mirrors this trend. Surveys reveal that 81% of Czech consumers are open to trying craft beers, and 45% are willing to pay a premium for locally produced options (Ipsos, 2017), indicating a rising demand for distinctive, high-quality beer products that challenge the dominance of traditional brewing firms.

This shift is shaped by broader cultural and generational transformations. Spáčil and Teichmannová (2016) document how beer consumption among Generation Y (Millennials) follows a moderate, normally distributed pattern, where moderate drinking is most common, while fewer individuals drink either very frequently or rarely. In contrast, the distribution in Generation X has a fatter tail, meaning that many drink beer several times per week, but only a small percentage do not drink at all. On the other hand, Generation Z (today's youngest alcohol consumers) exhibit markedly lower consumption levels, reflecting growing health consciousness and lifestyle reorientation (Ipsos, 2024).

These patterns underpin what researchers have called the "Drink Less, But Better" trend (Lerro et al., 2020; Mastanjević et al., 2019), emphasizing premiumization and responsible drinking. Culturally, this resonates with the rise of neolocalism, a revalorization of local identity and production (Maier, 2016). Neolocalist dynamics are particularly evident in the beer sector, where small breweries embed place-based narratives, regional dialects, and historical references into branding strategies (Materna, 2022), offering a form of cultural resistance to the global standardization of tastes.

This trend extends across Europe. In Italy, Garavaglia (2022) finds that post-war concentration gave way to a surge of microbrewery entry, driven by consumer demand for differentiated, locally rooted products. The number of breweries across Europe more than doubled between 2010 and 2020 (Brewers of Europe, 2025), signaling a continent-wide reconfiguration of industrial structure. Figure 1 shows how the number of microbreweries rose sharply between 2014 and 2023, contributing to a more fragmented domestic competitive environment. This evolution reflects both shifting consumer preferences and a structural transformation of industry dynamics, wherein large incumbents now coexist with a growing cohort of adaptive small-scale producers.

Figure 1 Comparison of number of microbreweries in EU between 2014 and 2023

Source: Own elaboration based on Brewers of Europe (2025)

These developments raise an important question about how domestic market conditions translate into international competitiveness. According to heterogeneous-firm trade theory, stronger domestic rivalry reallocates output toward more productive firms, thereby enhancing export capacity (Melitz, 2003; Melitz and Ottaviano, 2008). Empirical evidence supports this link between competition and exports in manufacturing and technology-intensive sectors, where efficiency and scale dominate (e.g., Clougherty and Zhang, 2009; Goodwin and Pierola, 2015). In contrast, the mechanisms connecting domestic competition to export success in mature consumer industries, such as beer, where differentiation, quality, and local identity play a central role, remain quantitatively underexplored.

This study extends heterogeneous-firm trade mechanisms to such differentiated consumer markets by combining the efficiency/selection channel with a non-price differentiation channel proxied by producer diversity. It addresses this gap by examining whether changes in domestic market structure influence export performance in the European beer industry.

Employing a reduced-form gravity model and panel data on EU beer trade between 2014 and 2023, it tests whether stronger rivalry (lower market concentration) and greater producer diversity (higher share of small breweries) are associated with higher export volumes. The central research question is: *How does domestic competition affect beer export performance across EU member states?*

The paper is structured as follows: Section 1 provides the literature review, Sections 2 and 3 outline the data sources and methodology, Section 4 presents the empirical findings and discusses the implications. Finally, the paper concludes with recommendations for industry stakeholders and policymakers.

1 LITERATURE REVIEW

This paper investigates how domestic market structure affects export performance, with a specific focus on the European Union's beer industry. While much of the traditional trade literature emphasizes cross--country determinants of export volumes (e.g., Anderson and van Wincoop, 2003; Tinbergen, 1962; Yotov, 2016), an expanding body of research highlights the critical role of within-country factors, including domestic competition, market structure and industry dynamics, in shaping export volumes (e.g., Clougherty and Zhang, 2009; Goodwin and Pierola, 2015; Melitz, 2003; Melitz and Ottaviano, 2008). This section reviews relevant theoretical and empirical literature across the field and provides the theoretical basis for the empirical model developed in this paper.

1.1 Domestic competition and export performance

The link between market concentration and firm performance is well established in industrial organization and trade literature. Concentrated markets tend to reduce competitive pressure, leading to lower efficiency and innovation (Boone, 2000; Vives, 2008), while lower concentration typically fosters rivalry, reallocation, and better industry outcomes (Melitz and Ottaviano, 2008). Syverson (2011) emphasizes that firm-level productivity differences significantly shape aggregate outcomes, with market structure evolving through competitive dynamics. Empirical studies confirm that less concentrated industries achieve higher allocative efficiency and productivity. Loualiche et al. (2021) show that increased entry erodes incumbents' power, improving efficiency.

The post-Melitz (2003) literature highlights how domestic competition enhances export performance. Greater rivalry raises industry productivity, enabling more firms to cover the fixed costs of exporting - a self-selection process favoring the most efficient firms (Bernard et al., 2012; Wagner, 2007). Several studies validate this. Goodwin and Pierola (2015) link pro-competition reforms to stronger SME exports. Clougherty and Zhang (2009) show that domestic airline competition raises export intensity. Chen et al. (2009) find similar effects from EU reforms. Argentesi et al. (2024) show exporters report gains in efficiency, innovation, and competitiveness from domestic rivalry. Freund and Pierola (2015) warn of the risks of export concentration, noting top exporters often dominate trade. Broader participation, enabled by competition and lower barriers, enhances resilience and performance.

Competition also drives non-price competitiveness. The Alchian-Allen effect (Hummels and Skiba, 2004) predicts that high trade costs lead exporters to specialize in quality. Competitive pressure motivates quality upgrading (Amiti and Khandelwal, 2013; Iacovone and Javorcik, 2012), and De Loecker (2011) finds that exporters raise prices due to quality gains rather than cost increases. Recent studies further emphasize non-price differentiation: Keil (2024) finds that branding, design, and certification intensity raise export shares in consumer industries, while Reimer and Langpap (2022) show that sustainability and regional identity strengthen export competitiveness in European food and beverage sectors. However, although the latter study extends the analysis to the beverage sector, it relies primarily on survey-based evidence rather than quantitative trade data, leaving room for a sector-level empirical assessment that this paper provides.

Chaney's (2008) gravity model shows that trade depends on both export volumes and the number of exporting firms: lower barriers expand the extensive margin by allowing smaller firms to enter export markets. In sectors like brewing, where entry barriers have fallen, this suggests that increased domestic competition enhances exports through broader firm participation.

1.2 Industry evolution and microbrewery dynamics in European beer markets

The European beer industry has transformed in recent decades, with declining concentration and a surge in microbreweries. Between 2010 and 2020, the number of breweries in the EU more than doubled, driven largely by small-scale entrants (Brewers of Europe, 2025).

Literature highlights the impact of this shift. In Italy, craft breweries increased variety and catered to local tastes (Fastigi et al., 2018). UK microbreweries proved more agile in adapting to demand (Ellis and Bosworth, 2015), while Germany and CEE saw declining concentration and greater price dispersion (Materna et al., 2022; Pokrivčák et al., 2019).

Innovation and branding are central to microbrewery strategies. They emphasize authenticity, quality, and locality (Aquilani et al., 2015; Lerro et al., 2020), often prompting strategic responses from larger firms, such as product diversification and quality upgrades. In Slovenia, small breweries have fostered both innovation and internationalization (Faganel and Rižnar, 2023), while Czech producers are quick to adopt niche trends like non-alcoholic or locally inspired beers (Maier, 2016). Storytelling tied to ingredients, techniques, and local identity enhances consumer loyalty and supports price premiums despite limited scale (Ellis and Bosworth, 2015; Lerro et al., 2020).

Garavaglia (2022) documents the rise of microbreweries in Italy and explores how entry patterns vary by institutional and demand-side factors, revealing regional disparities. More competitive regions support the persistence of efficient producers, a dynamic consistent with Chaney's (2008) model of competitive sorting and firm reallocation.

Evidence from Belgium shows that microbreweries can succeed internationally by leveraging uniqueness and quality, with exports reaching up to 80% of output (Poelmans and Ostyn, 2020). These cases illustrate how domestic competition strengthens differentiation and increases export potential.

1.3 Gravity in practice: empirical approaches to beer trade patterns

The gravity model, first introduced by Tinbergen (1962), explains bilateral trade flows based on the economic size of trading partners and the distance between them. Drawing from Newton's law of gravitation, it posits that trade increases with GDP and decreases with distance. In its log-linear form, it typically includes GDPs of both countries, distance, and trade resistance factors (Anderson and van Wincoop, 2003; Yotov, 2016). Its predictive accuracy and flexibility make it a core tool in empirical trade analysis.

Recent studies apply the gravity model to beer trade, confirming the continued relevance of GDP, distance, and shared borders while incorporating sector-specific dynamics. Dreyer et al. (2017) show that German beer exports are shaped by economic size, EU membership, and exchange-rate volatility, where higher exporter GDP and institutional integration increase exports, while distance and currency fluctuations reduce them. Similarly, Bieleková and Pokrivčák (2020) extend the analysis to a broader panel of European beer exporters and find that GDP and cultural proximity (common borders, language, or colonial ties) stimulate trade, whereas distance, landlockedness, and importer population exert negative effects.

Olper et al. (2012) explore home bias in EU beer consumption, finding that proximity and production scale significantly affect trade. Their gravity estimates indicate stronger home market effects for beer than for wine, especially in countries with smaller export sectors.

Although not using a gravity model directly, Garavaglia and Swinnen (2018) emphasize the need to include firm heterogeneity, distribution, and product differentiation when modeling beer trade, reinforcing the value of sector-specific controls.

Building on this literature, the present study extends the gravity approach by incorporating indicators of domestic market structure - specifically the Herfindahl-Hirschman Index (HHI) and the microbrewery ratio - to test how rivalry and producer diversity affect export performance in the EU beer sector. Drawing on heterogeneous-firm trade theory (Melitz, 2003; Melitz and Ottaviano, 2008) and recent evidence on non-price competitiveness (Keil, 2024; Reimer and Langpap, 2022), the analysis tests two hypotheses:

Hypothesis 1: Lower market concentration increases beer export volumes.

Hypothesis 2: A higher share of small producers enhances beer export performance.

2 DATA

Data for the empirical analysis is obtained from several sources. Country and year selection depends on data availability and comparability. The panel data set covers 1 277 bilateral observations for 28 EU countries (including the United Kingdom) over 2014–2023. Trading partners were selected each year until their cumulative share reached at least 70% of the exporting country's total beer export volumes. In this way, the coverage of the largest volumes of exports from each country is ensured. The reason for choosing these countries is that EU national brewers associations provide sufficient data quality, however, extensive time series are not available. Therefore, a shorter time period that is associated with the notable surge in the popularity of specialty beers, craft beers, and microbrewing in the EU is chosen (e.g., Lerro et al., 2020).

A detailed summary of all variables, including descriptions and data sources is provided in Appendix Table A1. The dependent variable is total beer export volumes from exporter to importer (*tot_exp_eur*), in millions of EUR, capturing bilateral trade flows (e.g., Tinbergen, 1962). Explanatory variables follow the standard gravity framework and include economic size and demand indicators, GDP per capita and population of exporter and importer, capturing income and market potential (Anderson and van Wincoop, 2003; Bieleková and Pokrivčák, 2020). Geographic distance, measured as the shortest air route between capitals, proxies trade costs that have declined with lower barriers and improved logistics. Industry-specific variables include beer production (Bieleková and Pokrivčák, 2020; Olper et al., 2012), consumption (Faganel and Rižnar, 2023) (both per capita), and average nominal price in the exporting country (Garavaglia and Swinnen, 2018). These reflect supply, demand, and price competitiveness. The key competitiveness measure is the Herfindahl-Hirschman Index (*HHI*) of the exporting country's beer market. Additional controls include dummy variables for advertising restrictions (*after10*) (Anderson et al., 2009), shared borders (*contig*) and common currency (*curr*) (Bieleková and Pokrivčák, 2020; Dreyer et al., 2017; Olper et al., 2012), whether the importer is an island (*island*) (Armstrong and Read, 2006), and COVID-19 period (*covid*) (Baldwin and Tomiura, 2020), accounting for institutional and temporal effects.

This analysis employs the HHI as an indicator of competitiveness within the beer industry. The HHI considers the full distribution of market shares, squaring each firm's share before summing the values.

This gives greater weight to larger firms while still reflecting the influence of smaller players. Unlike the Concentration Ratio or the Gini coefficient, the HHI is sensitive to structural changes across the entire market, making it especially valuable in industries like brewing, where market consolidation and the entry of new firms can occur simultaneously (Peleckis, 2022). Because the HHI is highly sensitive to shifts in market share, it is widely used to assess market structure and its impact on competition, pricing, and efficiency.

Mathematically, the HHI is calculated by summing the squares of each firm's market share:

$$HHI = \sum_{i=1}^{N} \left(MS_i \right)^2, \tag{1}$$

where MS_i is the market share of firm i in the market and N is the number of firms on the market. When expressed on a normalized 0–1 scale, an HHI of 1 represents a pure monopoly (one firm controlling the entire market), while an HHI close to zero signals a highly competitive market with many firms holding similar shares. In the context of the EU beer industry, the HHI provides a powerful tool for assessing how market structure has evolved over time. Traditionally, a handful of multinational brewers have held dominant positions in most European markets. At the same time, the past decade has seen a surge in the number of microbreweries, introducing greater diversity in beer styles, branding, and consumer preferences (Ellis and Bosworth, 2015; Fastigi et al., 2018). This dual dynamic, consolidation among the largest players and fragmentation due to an influx of small-scale producers, makes the beer industry a particularly interesting case for studying market concentration.

Table 1 presents an overview of the descriptive statistics for the variables utilized in the model. The number of observations of the baseline model is 1 277, although it varies slightly across models, which results from missing data for certain countries or years in some indicators, leading to the exclusion of these observations when constructing the extended model specifications. For instance, data necessary

Table 1 Descriptive statistics

Variable	N	Mean	SD	Median	Min	Max	VIF
tot_exp_eur	1 373	37.695	82.513	11.892	0.034	789.999	-
GDP_exp	1 373	27 318.080	15 413.545	24 440.000	6 120.000	97 300.000	3.899
GDP_imp	1 349	28 792.321	17 373.178	28 250.000	6 72.000	82 780.000	1.335
dist	1 373	2 147.092	2 961.849	845.000	54.880	16 366.000	2.878
pop_imp	1 362	148.646	350.654	38.400	0.076	1 425.893	2.221
prod_exp	1 360	0.884	0.530	0.805	0.070	3.717	2.438
prod_imp	1 246	0.740	0.819	0.608	0.000	21.977	2.952
cons_exp	1 354	72.536	26.750	71.000	28.000	150.308	2.694
cons_imp	1 157	62.061	24.105	66.625	2.500	150.308	2.851
microbrew_brew	1 131	0.771	0.190	0.833	0.318	1.000	-
contig	1 373	0.404	0.491	0.000	0.000	1.000	1.482
curr	1 373	0.320	0.467	0.000	0.000	1.000	1.439
island	1 373	0.110	0.313	0.000	0.000	1.000	1.185
after10	1 373	0.283	0.450	0.000	0.000	1.000	1.530
price	864	2.152	0.979	2.005	0.770	5.725	3.451
HHI	1 299	0.232	0.081	0.226	0.082	0.494	1.404
covid	1 373	0.210	0.408	0.000	0.000	1.000	1.033

Note: The last column reports the Variance Inflation Factor values for Model (6) in Table 3, i.e., the specification that includes all explanatory variables.
Source: Own calculations

Table 2 Correlation matrix of the model

	tot_exp_eur	GDP_exp	GDP_imp	dist	pop_imp	prod_exp	prod_imp	cons_exp	cons_imp	microb_rew_brew	contig	curr	island	after10	price	HHI	covid
tot_exp_eur	1.000***	0.377***	0.171***	0.051	0.095*	0.194***	−0.057	0.034	−0.125**	−0.030	0.037	0.040	0.165***	−0.253***	0.011	−0.300***	−0.039
GDP_exp		1.000***	0.390***	0.098*	−0.004	0.141***	0.020	−0.113**	−0.067	0.072	−0.033	0.068	0.108**	−0.203***	0.695***	−0.298***	−0.024
GDP_imp			1.000***	0.040	−0.228***	−0.008	0.125**	−0.096*	0.013	0.088*	−0.123**	−0.023	0.239***	−0.091*	0.265***	−0.139***	−0.060
dist				1.000***	0.687***	−0.090*	−0.264***	−0.150***	−0.335***	0.089*	−0.423***	−0.283***	0.024	−0.142***	0.050	−0.168***	0.037
pop_imp					1.000***	−0.103*	−0.288***	−0.123**	−0.394***	0.063	−0.268***	−0.219***	−0.088*	−0.025	−0.047	−0.118**	0.023
prod_exp						1.000***	−0.072	0.680***	0.037	−0.168***	0.071	−0.072	−0.054	−0.325***	−0.155***	0.087*	−0.028
prod_imp							1.000***	−0.001	0.691***	0.031	0.143***	0.167***	−0.048	0.096*	0.042	−0.064	−0.034
cons_exp								1.000***	0.114**	−0.337***	0.107*	−0.139***	−0.119**	−0.220***	−0.354***	0.031	−0.015
cons_imp									1.000***	−0.040	0.176***	0.024	0.018	0.109**	0.028	0.078	−0.088*
microbrew_brew										1.000***	−0.154***	−0.199***	0.090*	0.151***	0.347***	−0.038	0.052
contig											1.000***	0.168***	−0.100*	0.074	−0.084*	0.091*	−0.026
curr												1.000***	−0.167***	0.195***	0.075	0.005	−0.051
island													1.000***	−0.123**	0.121**	−0.114**	−0.008
after10														1.000***	0.130**	0.256***	0.030
price															1.000***	−0.097*	−0.042
HHI																1.000***	−0.034
covid																	1.000***

Note: *p<0.1; **p<0.05; ***p<0.01.
Source: Own calculations

for the calculation of the Herfindahl–Hirschman Index are not available for Malta, Cyprus, and Luxembourg, while information on the number of microbreweries is missing for Belgium, the Netherlands, and Poland. High variability in exporter GDP and distance reflects the dominance of large exporters such as Germany, Belgium, and the Netherlands and the inclusion of both intra-EU and extra-EU destinations. Winsorization mitigates these outliers while preserving cross-country heterogeneity essential for gravity estimations. Given the model's nature, all variables (excluding the dummy variables) are transformed into logarithmic form. Following the application of logarithms to the variables, there is a notable decrease in the standard deviations of these variables.

The exporter country population variable was initially included but omitted owing to high correlation with GDP to ensure the unbiasedness of the estimators. Consequently, Table 2 shows the correlation matrix of the model, expressing the weak correlation between the variables under study.

3 RESEARCH DESIGN

To investigate the impact of domestic competition on beer export volumes within the European Union, this study employs the gravity model of international trade. The gravity model posits that trade volume between two countries is proportional to their economic size and inversely related to distance:

$$X_{ij} = G\left(\frac{Y_i^{\alpha} Y_j^{\beta}}{D_{ij}^{\gamma}}\right), \tag{2}$$

where X_{ij} represents the trade flow from country i to country j; Y_i and Y_j are the economic sizes (GDP) of countries i and j, respectively; D_{ij} denotes the distance between the two countries; G is a constant of proportionality; α, β and γ are parameters to be estimated. By taking the natural logarithm of both sides, the equation becomes linear and suitable for econometric analysis:

$$lnX_{ij} = lnG + \alpha lnY_i + \beta lnY_j - \gamma lnD_{ij} + \varepsilon_{ij}, \tag{3}$$

where ε_{ij} is the error term capturing unobserved factors affecting trade flows.

The model includes standard gravity variables, economic size and distance, augmented by sector--specific and policy-related factors. By integrating a measure of domestic competition based on theoretical concepts, this study extends the existing literature.

Considering the gravity model structure, the proposed model for the empirical analysis is as follows:[2]

$$\begin{aligned} tot\_exp\_eur_{ijt} = {} & \beta_0 + \beta_1 \ln\left(GDP_{it}\right) + \beta_2 \ln\left(GDP_{jt}\right) + \beta_3 \ln\left(dist_{ij}\right) + \boldsymbol{\gamma}' \ln\left(\boldsymbol{X}_{it}\right) + \boldsymbol{\delta}' \ln\left(\boldsymbol{M}_{jt}\right) \\ & + \boldsymbol{\theta}' \ln\left(\boldsymbol{I}_{it}\right) + \boldsymbol{\lambda}' \boldsymbol{C}_{ij} + \epsilon_{ijt}, \end{aligned} \tag{4}$$

where $\boldsymbol{X}_{it}$ represents a vector of macroeconomic variables for the exporting country:

$$\boldsymbol{X}_{it} = \gamma_1 pop_{it} + \gamma_2 prod_{it} + \gamma_3 cons_{it}, \tag{5}$$

$\boldsymbol{M}_{jt}$ includes macroeconomic variables for the importing country:

$$\boldsymbol{M}_{jt} = \delta_1 pop_{jt} + \delta_2 prod_{jt} + \delta_3 cons_{jt}, \tag{6}$$

$\boldsymbol{I}_{it}$ denotes industry-specific variables as:

$$\boldsymbol{I}_{it} = \theta_1 price_{it} + \theta_2 HHI_{it}, \tag{7}$$

[2] For variables description please see Table A1 in Appendix.

$\boldsymbol{C}_{ij}$ consists of control dummy variables such as:

$$\boldsymbol{C}_{ij} = \lambda_1 after10_{it} + \lambda_2 contig_{ij} + \lambda_3 curr_{ij} + \lambda_4 island_j + \lambda_5 covid \,. \tag{8}$$

Two key variables representing competitiveness are included in the model: the Herfindahl–Hirschman index (HHI_{it}) in the exporting country and the average nominal price of beer in the exporting country ($price_{it}$).

The Herfindahl–Hirschman index in the exporting country is used in this study as a proxy for domestic competition intensity within the beer industry. Based on the theoretical framework, a more concentrated market is expected to have a negative impact on export volumes. The average nominal price of beer in the exporting country is included to assess the price competitiveness of exporters. Higher domestic beer prices may render export volumes less competitive in international markets due to increased production costs and reduced-price attractiveness to foreign consumers (Dreyer et al., 2017).

The gravity model is estimated using the Poisson Pseudo-Maximum Likelihood (PPML) estimator, which is well-suited for trade data as it handles heteroskedasticity and zero trade flows effectively (Silva and Tenreyro, 2006). Unlike OLS on log-linear models, which can yield biased estimates due to $E[\ln y] \neq \ln E[y]$ (Silva and Tenreyro, 2006) and the exclusion of zero flows, PPML estimates the model in levels, preserving zero values and delivering consistent results (Silva and Tenreyro, 2011). It also downweights large trade flows, supporting robustness. Given these advantages, PPML is now standard in gravity model applications (e.g., Yotov et al., 2016).

To assess the impact of domestic competition on beer export volumes, a stepwise estimation strategy is followed. The baseline model includes key gravity variables, GDPs, distance, and the HHI, to identify the direct effect of competition. Covariates are then incrementally added to test the robustness of HHI's influence and examine how other factors affect model fit and coefficient stability. Year fixed effects are included in selected model specifications to control for global shocks. Their inclusion serves as a robustness check to ensure that the estimated relationships are not merely driven by time--specific influences, allowing identification of country-specific effects like domestic competition, which would be absorbed by more restrictive fixed effect structures (Baldwin and Taglioni, 2006; Yotov et al., 2016).

4 RESULTS AND DISCUSSION

This section presents the empirical findings from the gravity model estimations assessing the impact of domestic competition on beer export volumes within the European Union. The results are summarized in Table 3 (core specifications including country-, industry-specific, and institutional controls) and Table 4 (robustness analysis employing an alternative competition indicator). The key variable of interest is the Herfindahl–Hirschman index in the exporting country (HHI_{it}), which serves as a proxy for domestic competition intensity. Consistent with theoretical expectations (Bosma et al., 2011; Melitz and Ottaviano, 2008), stronger domestic competition enhances market efficiency and firm-level productivity. All models were subjected to standard diagnostic procedures for multicollinearity and heteroskedasticity. Variance Inflation Factor values were well below conventional thresholds (<5) – for VIF values of Model (6), Table 3, the one that includes all the explanatory variables, see Table 1. The Breusch–Pagan test indicated the presence of heteroskedasticity at $\alpha = 5\%$ significance level; therefore, heteroskedasticity-robust standard errors are reported throughout.

In the baseline model (Model 1, Table 3), which includes the fundamental gravity variables and the domestic competition measure, the coefficient for HHI_{it} is negative and highly significant ($B = -0.414$, $p < 0.05$). The negative coefficient on HHI confirms that higher market concentration, indicating weaker domestic competition, reduces export volumes. The negative effect remains consistent and significant across all subsequent models, even as additional variables are introduced. This robustness underscores the central role of domestic competition in enhancing export performance in the beer industry, thereby supporting Hypothesis 1.

The estimated coefficients for standard gravity variables are consistent with theoretical priors (Anderson and van Wincoop, 2003; Tinbergen, 1962; Yotov, 2016). When macroeconomic variables are added (Model 3), the exporting country's GDP exhibits positive and significant effect on beer export volumes, suggesting that larger economies with greater production capacity tend to export more. The positive coefficients for both exporting and importing countries' populations in Model 3–7 suggest that larger markets facilitate greater trade flows, a principle that is consistent with the market size effect described in economic literature (e.g., Anderson and van Wincoop, 2003; Tinbergen, 1962; Yotov, 2016). This effect suggests that countries with larger populations tend to trade more because they offer larger markets for goods and services, as well as greater diversity in production capabilities. Distance, as expected (e.g., Anderson and van Wincoop, 2003; Tinbergen, 1962; Yotov, 2016), has a negative and significant impact on export volumes from Model (3) onwards, reinforcing the importance of trade costs.

Industry-specific variables provide further insights. Exporter-side beer production per capita is positively associated with export volumes, while importer-side production is insignificant, consistent with home-market absorption effects and the "home bias" in beer consumption (Olper et al., 2012). Notably, the exporting country per capita consumption effect is large and significantly negative across Models (4) to (7), indicating that strong domestic demand may crowd out export potential. This mirrors findings from Dreyer et al. (2017) and Wieczorek and Czupryna (2021), who highlight the importance of internal consumption dynamics. The coefficient on the average domestic beer price is negative and highly significant, confirming the relevance of cost competitiveness even in a differentiated consumer industry (Garavaglia and Swinnen, 2018).

Control variables added in Models 5–7 do not show (based on the dataset) statistically significant effects on beer export volumes. The statistically insignificant impact of the Covid-19 pandemic on beer export volumes (Model 6) challenges initial presumptions (Baldwin and Tomiura, 2020) about global trade disruptions. This resilience might be attributed to the essential nature of food and beverage products, combined with the adaptability of supply chains and trade networks.

The inclusion of time fixed effects supports the robustness of the findings by accounting for global shocks and trends. The fact that the key coefficients remain significant and consistent when time fixed effects are included indicates that the observed relationships are not driven by temporal factors but are inherent to the dynamics of the beer industry and trade patterns within the EU. Overall, the adjusted R^2 values increase progressively with the inclusion of additional variables, indicating improved model fit and explanatory power.

As a robustness check, the key variable of interest, market concentration measured by the Herfindahl-Hirschman Index, was substituted with an alternative indicator of domestic competition: the ratio of microbreweries to total breweries in the exporting country. The results in Table 4 support the main findings. The coefficient on the microbrewery ratio is positive and highly significant across all specifications, indicating that a higher share of microbreweries, interpreted as a more competitive and dynamic domestic industry, is associated with increased export volumes. This confirms Hypothesis 2, suggesting that non-price competitiveness factors enhance export volumes. While Keil (2024) and Reimer and Langpap (2022) develop these mechanisms in general models of trade and firm behavior, this research extends their insights to a mature consumer industry. It demonstrates empirically that rivalry and diversity within domestic markets can translate differentiation advantages into stronger export performance. Other variables behave similarly to those in Table 3: exporter GDP remains a key determinant; exporter consumption retains its negative effect; and, in the robustness models, contiguity and island status are positive and significant, reflecting reduced trade frictions for neighbors and higher import dependence for islands, consistent with Bieleková and Pokrivčák (2020), and Dreyer et al. (2017).

Table 3 Estimation output

	Dependent variable: tot_exp_eur						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log(HHI)	−0.414*** (0.100)	−0.411*** (0.101)	−0.745*** (0.096)	−0.799*** (0.118)	−0.775*** (0.122)	−0.783*** (0.122)	−0.799*** (0.122)
log(GDP_exp)	1.630*** (0.067)	1.627*** (0.067)	1.296*** (0.064)	1.493*** (0.129)	1.446*** (0.132)	1.448*** (0.132)	1.501*** (0.135)
log(GDP_imp)	0.127*** (0.046)	0.128*** (0.046)	0.166*** (0.045)	0.158*** (0.059)	0.155** (0.061)	0.153** (0.061)	0.132** (0.062)
log(dist)	−0.003 (0.029)	−0.003 (0.029)	−0.237*** (0.039)	−0.344*** (0.046)	−0.302*** (0.053)	−0.301*** (0.053)	−0.288*** (0.053)
log(pop_imp)			0.267*** (0.026)	0.262*** (0.032)	0.261*** (0.032)	0.260*** (0.032)	0.256*** (0.032)
log(prod_exp)			0.826*** (0.055)	1.273*** (0.099)	1.252*** (0.105)	1.260*** (0.106)	1.284*** (0.107)
log(prod_imp)			-0.029 (0.032)	0.031 (0.059)	0.028 (0.060)	0.030 (0.060)	0.033 (0.060)
log(cons_exp)				−1.499*** (0.147)	−1.458*** (0.153)	−1.469*** (0.153)	−1.500*** (0.153)
log(cons_imp)				−0.218* (0.132)	-0.220 (0.135)	-0.228* (0.136)	−0.215 (0.135)
log(price)				−1.049*** (0.150)	−0.996*** (0.154)	−1.004*** (0.155)	−1.093*** (0.160)
after10					−0.061 (0.098)	−0.054 (0.098)	−0.009 (0.099)
contig					0.154* (0.090)	0.153* (0.090)	0.147* (0.089)
curr					0.060 (0.093)	0.054 (0.093)	−0.039 (0.093)
island					0.101 (0.128)	0.101 (0.128)	0.101 (0.128)
covid						−0.066 (0.077)	
Constant	−15.700*** (0.664)	−15.580*** (0.674)	−12.364*** (0.657)	−5.608*** (1.523)	−5.632*** (1.544)	−5.519*** (1.550)	−5.893*** (1.573)
Observations	1,277	1,277	1,277	703	703	703	703
Adjusted R^2	0.403	0.401	0.546	0.591	0.591	0.591	0.596
Time F.E.	N	Y	N	N	N	N	Y

Note: *p<0.1; **p<0.05; ***p<0.01; heteroskedasticity consistent standard errors are reported in parentheses.
Source: Own calculations

Overall, the empirical evidence from both competition indicators, market concentration and producer diversity, demonstrates that domestic market structure plays a decisive role in determining export performance in the European beer industry. The results complement qualitative studies of microbrewery dynamics (Bieleková and Pokrivčák, 2020; Garavaglia, 2022; Pokrivčák et al., 2019) and align with gravity-based research emphasizing the importance of economic scale, proximity, and institutional integration in beverage trade (Bieleková and Pokrivčák, 2020; Dreyer, 2017; Olper et al., 2012). From a policy perspective, these findings suggest that competition-enhancing regulations, reduced entry barriers, and targeted support for small and innovative producers can generate indirect trade benefits by strengthening both efficiency-based and differentiation-based competitiveness within the internal market.

Table 4 Estimation output – robustness check

	Dependent variable: tot_exp_eur						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log(microbrew_brew)	0.816*** (0.135)	0.858*** (0.137)	1.096*** (0.130)	1.380*** (0.171)	1.478*** (0.179)	1.477*** (0.180)	1.540*** (0.178)
log(GDP_exp)	1.479*** (0.074)	1.486*** (0.074)	1.236*** (0.073)	2.084*** (0.135)	2.063*** (0.138)	2.066*** (0.138)	2.189*** (0.140)
log(GDP_imp)	0.037 (0.051)	0.040 (0.051)	0.092* (0.053)	0.047 (0.070)	0.022 (0.072)	0.019 (0.073)	−0.039 (0.073)
log(dist)	−0.044 (0.032)	−0.042 (0.032)	−0.170*** (0.047)	−0.262*** (0.056)	−0.149** (0.064)	−0.149** (0.064)	−0.120* (0.064)
log(pop_imp)			0.194*** (0.031)	0.189*** (0.038)	0.184*** (0.038)	0.183*** (0.038)	0.180*** (0.038)
log(prod_exp)			0.867*** (0.068)	1.123*** (0.128)	1.131*** (0.128)	1.129*** (0.128)	1.067*** (0.129)
log(prod_imp)			−0.015 (0.037)	−0.082 (0.067)	−0.080 (0.068)	−0.078 (0.068)	−0.089 (0.068)
log(cons_exp)				−1.292*** (0.180)	−1.183*** (0.184)	−1.184*** (0.184)	−1.096*** (0.184)
log(cons_imp)				0.203 (0.154)	0.198 (0.159)	0.191 (0.159)	0.235 (0.158)
log(price)				−1.887*** (0.166)	−1.879*** (0.168)	−1.884*** (0.169)	−2.062*** (0.173)
after10					0.153 (0.103)	0.155 (0.103)	0.192* (0.102)
contig					0.337*** (0.106)	0.336*** (0.106)	0.321*** (0.104)
curr					0.139 (0.108)	0.135 (0.109)	0.151 (0.108)
island					0.373** (0.153)	0.375** (0.153)	0.397*** (0.152)
covid						−0.046 (0.089)	
Constant	−12.571*** (0.738)	−12.411*** (0.748)	−9.995*** (0.760)	−11.483*** (1.750)	−12.455*** (1.781)	−12.416*** (1.783)	−13.866*** (1.807)
Observations	1,185	1,185	1,021	616	616	616	616
Adjusted Pseudo R^2	0.315	0.316	0.433	0.516	0.527	0.526	0.538
Time F.E.	N	Y	N	N	N	N	Y

Note: *p<0.1; **p<0.05; ***p<0.01; heteroskedasticity consistent standard errors are reported in parentheses.
Source: Own calculations

CONCLUSION

This paper examined how domestic market structure shapes export performance in a mature, differentiated consumer industry. Using a gravity framework estimated by PPML on EU beer trade between 2014 and 2023, it incorporated two complementary measures of domestic competition, market concentration (HHI) and producer diversity (microbrewery share). The estimates indicate that lower concentration and greater producer diversity are robustly associated with higher export volumes; the magnitudes are economically meaningful, with a 10% decline in HHI corresponding roughly to a 4% increase in exports, and results remain stable across specifications.

Theoretically, the present study extends heterogeneous-firm trade mechanisms to a context where non-price differentiation plays a central role, showing that domestic rivalry operates through both an efficiency/selection channel and a differentiation channel that is proxied by producer diversity. Empirically, it provides new sector-level evidence for the European beer industry using a multi-country panel linking domestic market transformation to external competitiveness within a gravity framework.

The findings have clear implications for policy and practice. Competition-enhancing regulation, timely entry and scale-up support for smaller producers, and instruments that lower the fixed costs of quality upgrading (such as certification facilitation, origin and sustainability labeling, and coordinated export promotion) are likely to yield indirect trade gains even in established consumer sectors. For incumbent firms, coexistence with dynamic small producers appears complementary: variety expansion, quality upgrading, and collaborative production can help translate identity-driven advantages into export performance.

Limitations remain. Reverse causality cannot be ruled out, and industry-level data obscure firm-level heterogeneity. Future research could exploit external shocks (mergers, advertising rules, tax reforms) or link brewery registries with customs data to analyze extensive and intensive export margins, quality premia, and reallocation dynamics, and to measure differentiation more directly.

Overall, the evidence indicates that export performance in mature consumer industries depends not only on macroeconomic scale and trade costs, but also on the internal organization of domestic markets. Policies that preserve rivalry and foster diverse producer ecosystems strengthen both productivity-based and differentiation-based competitiveness, with measurable gains in external performance.

ACKNOWLEDGMENT

Financial support from the Prague University of Economics and Business (grant IG14/2025) is gratefully acknowledged.

References

ANDERSON, J. E., VAN WINCOOP, E. (2003). Gravity With Gravitas: a Solution to the Border Puzzle [online]. *American Economic Review*, 93: 170–192. <https://doi.org/10.1257/000282803321455214>.

AMITI, M., KHANDELWAL, A.K. (2013). Import Competition and Quality Upgrading [online]. *Review of Economics and Statistics*, 95: 476–490. <https://doi.org/10.1162/REST_a_00271>.

ANDERSON, P., DE BRUIJN, A., ANGUS, K., GORDON, R., HASTINGS, G. (2009). Impact of Alcohol Advertising and Media Exposure on Adolescent Alcohol Use: a Systematic Review of Longitudinal Studies [online]. *Alcohol and Alcoholism*, 44: 229–243. <https://doi.org/10.1093/alcalc/agn115>.

AQUILANI, B., LAURETI, T., POPONI, S., SECONDI, L. (2015). Beer Choice and Consumption Determinants When Craft Beers are Tasted: an Exploratory Study of Consumer Preferences [online]. *Food Quality and Preference*, 41: 214–224. <https://doi.org/10.1016/j.foodqual.2014.12.005>.

ARGENTESI, E., DE SIMONE, L., PAETZ, S., SCRUTINIO, V. (2024). *Does Domestic Competition Help EU Firms Compete Abroad?* [online]. Luxembourg: Publications Office of the European Union. <https://doi.org/10.2763/012974>.

ARMSTRONG, H., READ, R. (2006). Geographical 'Handicaps' and Small States: Some Implications for the Pacific from a Global Perspective [online]. *Asia Pacific Viewpoint*, 47(1): 79–92. <https://doi.org/10.1111/j.1467-8373.2006.00294.x>.

BALDWIN, R., TAGLIONI, D. (2006). *Gravity for Dummies and Dummies for Gravity Equations* [online]. Cambridge, MA: National Bureau of Economic Research. <https://doi.org/10.3386/w12516>.

BALDWIN, R., TOMIURA, E. (2020). Thinking Ahead About the Trade Impact of COVID-19. In: BALDWIN, R., WEDER DI MAURO, B. (eds.). *Economics in the Time of COVID-19*, London: CEPR Press, 59–71.

BERNARD, A. B., GRAZZI, M., TOMASI, C. (2012). Intermediaries in International Trade: Products and Destinations [online]. *Review of Economics and Statistics*, 94: 356–369. <https://doi.org/10.1162/REST_a_00495>.

BIELEKOVÁ, E., POKRIVČÁK, J. (2020). Determinants of International Beer Export [online]. *AGRIS On-Line Papers in Economics and Informatics*, 12(2): 17–27. <https://doi.org/10.22004/ag.econ.303941>.

BOONE, J. (2000). Competitive Pressure: The Effects on Investments in Product and Process Innovation [online]. *The RAND Journal of Economics*, 31: 549–569. <https://doi.org/10.2307/2601000>.

BOSMA, N., STAM, E., SCHUTJENS, V. (2011). Creative Destruction and Regional Productivity Growth: Evidence from the Dutch Manufacturing and Services Industries [online]. *Small Business Economics*, 36: 401–418. <https://doi.org/10.1007/s11187-009-9257-8>.

BREWERS OF EUROPE. (2025). *European beer trends* [online]. [cit. 6.6.2025]. <https://brewersofeurope.eu>.

BUSINESSCOOT. (2025). *The Craft Beer Market – France* [online]. [cit. 6.6.2025]. <https://www.businesscoot.com/en>.

CHANEY, T. (2008). Distorted Gravity: The Intensive and Extensive Margins of International Trade [online]. *American Economic Review*, 98: 1707–1721. <https://doi.org/10.1257/aer.98.4.1707>.

CLOUGHERTY, J. A., ZHANG, A. (2009). Domestic Rivalry and Export Performance: Theory and Evidence from International Airline Markets [online]. *Canadian Journal of Economics*, 42: 440–468. <https://doi.org/10.1111/j.1540-5982.2009.01515.x>.

DE LOECKER, J. (2011). Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity [online]. *Econometrica*, 79: 1407–1451. <https://doi.org/10.3982/ECTA7617>.

DREYER, H., FEDOSEEVA, S., HERRMANN, R. (2017). Gravity Meets Pricing to Market: What a Combined-Method Approach Tells us on German Beer Exports and Pricing [online]. *Journal of Economics and Statistics (Jahrbücher für Nationalökonomie und Statistik)*, 237: 295–328. <https://doi.org/10.1515/jbnst-2017-0106>.

ELLIS, V., BOSWORTH, G. (2015). Supporting Rural Entrepreneurship in the UK Microbrewery Sector [online]. *British Food Journal*, 117: 2724–2738. <https://doi.org/10.1108/BFJ-12-2014-0412>.

EUROMONITOR INTERNATIONAL. (2025). *Passport: Beer Industry Company Shares in the EU* [online]. [cit. 6.6.2025]. <https://www.euromonitor.com>.

FAGANEL, A., RIŽNAR, I. (2023). The Growth in Demand for Craft Beer and the Proliferation of Microbreweries in Slovenia [online]. *Beverages*, 9: 86. <https://doi.org/10.3390/beverages9040086>.

FASTIGI, M., VIGANO, E., ESPOSTI, R. (2018). The Italian Microbrewing Experience: Features and Perspectives [online]. *Bio-Based and Applied Economics*, 7: 59–86. <https://doi.org/10.22004/ag.econ.301225>.

FREUND, C., PIEROLA, M. D. (2015). Export Superstars [online]. *Review of Economics and Statistics*, 97: 1023–1032. <https://doi.org/10.1162/REST_a_00511>.

GARAVAGLIA, C. (2022). Industry Evolution: Evidence from the Italian Brewing Industry [online]. *Competition & Change*, 26: 75–95. <https://doi.org/10.1177/102452942110074>.

GARAVAGLIA, C., SWINNEN, J. F. M. (2018). The Craft Beer Revolution: an International Perspective. In: GARAVAGLIA, C., SWINNEN, J.F.M. (eds.). *Economic Perspectives on Craft Beer: a Revolution in the Global Beer Industry*, Cham: Palgrave Macmillan, 3–52.

GOODWIN, T. K., PIEROLA CASTRO, M. D. (2015). *Export Competitiveness: Why Domestic Market Competition Matters*. Public Policy for the Private Sector Note 348, Washington, DC: World Bank Group.

HUMMELS, D., SKIBA, A. (2004). Shipping the Good Apples Out? An Empirical Confirmation of the Alchian – Allen Conjecture [online]. *Journal of Political Economy*, 112: 1384–1402. <https://doi.org/10.1086/422562>.

IACOVONE, L., JAVORCIK, B. S. (2012). Getting Ready: Preparation for Exporting. *International Economic Review*, 53: 997–1027.

IPSOS. (2017). *Jak vnímá česká populace malé pivovary?* [online]. [cit. 20.8.2025]. <https://www.ipsos.com/sites/default/files/ct/news/documents/2017-12/tz_2017_08_17_ipsos_minipivovary_pro_amsp_media_bez_metody_sberu.pdf>.

IPSOS. (2024). *Budoucnost piva: nové příchutě, nealko varianty a hektolitry inovací* [online]. [cit. 30.10.2024]. <https://www.ipsos.com/cs-cz/budoucnost-piva-nove-prichute-nealko-varianty-hektolitry-inovaci>.

KEIL, S. (2024). Assessing the Role of Non-Price Factors: Shedding New Light on the European Competitiveness Puzzle [online]. *International Economics,* 178(C). <https://doi.org/10.1016/j.inteco.2024.100495>.

LERRO, M., MAROTTA, G., NAZZARO, C. (2020). Measuring Consumers' Preferences for Craft Beer Attributes Through Best-Worst Scaling [online]. *Agricultural and Food Economics*, 8: 1–13. <https://doi.org/10.1186/s40100-019-0138-4>.

LOUALICHE, E. et al. (2021). *Asset Pricing with Entry and Imperfect Competition* [online]. Working paper. [cit. 30.10.2024]. <https://www.loualiche.com/docs/AP_entry.pdf>.

MAIER, T. (2016). Sources of Microbrewery Competitiveness in the Czech Republic [online]. *Agris On-Line Papers in Economics and Informatics*, 8: 97–110. <https://doi.org/10.22004/ag.econ.254029>.

MASTANJEVIĆ, K., KRSTANOVIĆ, V., LUKINAC, J., JUKIĆ, M., LUČAN, M., MASTANJEVIĆ, K. (2019). Craft Brewing – Is it Really about the Sensory Revolution? [online]. *Kvasný průmysl*, 65: 13–16.<https://doi.org/10.18832/kp2019.65.13>.

MATERNA, K., BERNHÄUSEROVÁ, V., HASMAN, J., HÁNA, D. (2022). How Microbreweries Flooded Europe: Mapping a New Phenomenon in the Beer Industry [online]. *Journal of Maps*, 18: 18–25. <https://doi.org/10.1080/17445647.2021.2012536>.

MELITZ, M.J. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity [online]. *Econometrica*, 71: 1695–1725. <https://doi.org/10.1111/1468-0262.00467>.

MELITZ, M. J., OTTAVIANO, G. I. (2008). Market Size, Trade, and Productivity [online]. *The Review of Economic Studies*, 75: 295–316. <https://doi.org/10.1111/j.1467-937X.2007.00463.x>.

OLPER, A., CURZI, D., FRISIO, D. G., RAIMONDI, V. (2012). Home Bias in Consumption: a Comparison between Wine and Beer [online]. *German Journal of Agricultural Economics (GJAE)*, 61(4): 223–234. <https://doi.org/10.52825/gjae.v61i4.1883>.

PELECKIS, K. (2022). Determining the Level of Market Concentration in the Construction Sector – Case of Application of the HHI Index [online]. *Sustainability*, 14: 779. <https://doi.org/10.3390/su14020779>.

POELMANS, E., OSTYN, T. P. (2020). On the Existence of Belgian Craft Breweries: Explorative Research at the Microlevel. In: PATTERSON, M. W., HOALST-PULLEN, N. (eds.). *The Geography of Beer: Culture and Economics*, Cham: Springer, 179–200.

POKRIVČÁK, J., SUPEKOVÁ, S.C., LANČARIČ, D., SAVOV, R., TÓTH, M., VAŠINA, R. (2019). Development of Beer Industry and Craft Beer Expansion. *Journal of Food & Nutrition Research*, 58: 1–9.

REIMER, J. J., LANGPAP, C. (2022). New Developments in Trade Theory and the Food and Beverage Industry [online]. *Journal of the Agricultural and Applied Economics Association*, 1(1): 4–17. <https://doi.org/10.1002/jaa2.5>.

SILVA, J. S., TENREYRO, S. (2006). The log of Gravity. *The Review of Economics and Statistics*, 88: 641–658.

SILVA, J. S., TENREYRO, S. (2011). Further Simulation Evidence on the Performance of the Poisson Pseudo-Maximum Likelihood Estimator [online]. *Economics Letters*, 112: 220–222. <https://doi.org/10.1016/j.econlet.2011.05.008>.

SOCIETY OF INDEPENDENT BREWERS. (2024). *SIBA Independent British Craft Beer Report 2024* [online]. [cit. 20.1.2025]. < https://toolbox-siba.co.uk/documents/Facts%20&%20Figures/SIBA%20British%20Craft%20Beer%20Report/SIBA%20independent%20Beer%20Report%202024.pdf>.

SPÁČIL, V., TEICHMANNOVÁ, A. (2016). Intergenerational Analysis of Consumer Behaviour on the Beer Market [online]. *Procedia – Social and Behavioral Sciences*, 220: 487–495. <https://doi.org/10.1016/j.sbspro.2016.05.524>.

SYVERSON, C. (2011). What Determines Productivity? [online]. *Journal of Economic Literature*, 49: 326–365. <https://doi.org/10.1257/jel.49.2.326>.

TINBERGEN, J. (1962). *Shaping the World Economy: Suggestions for an International Economic Policy*. New York: Twentieth Century Fund.

VIVES, X. (2008). Innovation and Competitive Pressure [online]. *The Journal of Industrial Economics*, 56: 419–469. <https://doi.org/10.1111/j.1467-6451.2008.00356.x>.

WAGNER, J. (2007). Exports and Productivity: a Survey of the Evidence from Firm-Level Data [online]. *World Economy*, 30: 60–82. <https://doi.org/10.1111/j.1467-9701.2007.00872.x>.

WIECZOREK, L., CZUPRYNA, M. (2021). What Drives the Development of Craft Beer Production in Poland? Insights From a Hedonic Regression Analysis of the Polish Beer Market [online]. *Eastern European Economics*, 59: 597–610. <https://doi.org/10.1080/00128775.2021.1976062>.

YOTOV, Y.V., PIERMARTINI, R., LARCH, M. et al. (2016). *An Advanced Guide to Trade Policy Analysis: the Structural Gravity Model*. Geneva: WTO.

APPENDIX

Table 1 Overview of variables used

Variable	Description	Source
tot_exp_eur	Volume of total beer exports from exporting to importing country (in millions of EUR)	The Observatory of Economic Complexity
GDP_exp, GDP_imp	Real Gross Domestic Product per Capita of exporting and importing country (in EUR)	Eurostat
dist	Shortest air distance between the capital of exporting and importing country (in kilometers)	CEPII gravity database
pop_exp, pop_imp	Number of habitants of exporting and importing country (in millions)	CEPII gravity database
prod_exp, prod_imp	Beer production per capita of exporting and importing country (in hectoliters)	Brewers of Europe
cons_exp, cons_imp	Beer consumption per capita of exporting and importing country (in liters)	Brewers of Europe
HHI	Herfindahl–Hirschman index of the beer industry in the exporting country	Own calculation based on Euromonitor International
microbrew_brew	Ratio of the number of microbreweries to the number of breweries of the exporting country	Own calculation based on Brewers of Europe
price	Average nominal price of beer per unit in the exporting country (in EUR)	Statista Market Insights
tax_revenue	Government revenues generated from excise duties on the sale of beer of the exporting country per capita (in EUR)	Brewers of Europe
after10	Dummy variable, if the country allows TV advertisement of alcohol only after 10pm = 1 (in the exporting country)	European Centre for Monitoring Alcohol Marketing, European Commission
contig	Dummy variable, if countries share borders = 1	CEPII gravity database
curr	Dummy variable, if countries have the same currency = 1	CEPII gravity database
island	Dummy variable, if the importing country is an island = 1	Own compilation
covid	Dummy variable, if t = 2020 and 2021, = 1	Own compilation

Source: XXX

Sustainable Dairy Farming in the Visegrad Group Countries' Regions: Linking Eco-Efficiency and Competitiveness

Eva Richterová[1] | *Slovak University of Agriculture, Nitra, Slovakia*
Martin Richter[2] | *Slovak University of Agriculture, Nitra, Slovakia*

Received 23.9.2025 (revision received 10.12.2025), **Accepted** (reviewed) 10.2.2026, **Published** 12.6.2026

Abstract

This study evaluates regional eco-efficiency and market competitiveness in the dairy sector of the Visegrad Four (V4) regions – Czechia, Slovakia, Hungary, and Poland – for 2015 and 2022. Eco-efficiency was assessed using input-oriented Data Envelopment Analysis (DEA) with an undesirable output, assuming constant returns to scale, while competitiveness was measured with a composite Dairy Competitiveness Index based on economic and sectoral indicators. Results indicate that high environmental performance does not consistently align with market competitiveness, with only the Polish region PL92: Mazowiecki regionalny excelling in both dimensions. The research hypothesis – The eco-efficiency of a region ensures a higher level of outputs for given inputs, thereby increasing its competitiveness – was rejected. Four regional groups were identified – Leaders, Market-driven, Eco-driven, and Laggards – highlighting persistent structural differences across the V4. The findings provide evidence for designing region-specific policies that support sustainable, competitive, and resilient dairy systems.

Keywords	***DOI***	***JEL code***
dairy eco-efficiency, competitiveness, V4 regions, interrelationship	*https://doi.org/10.54694/stat.2025.51*	*C14; C43; D24; O13; O44; R11*

INTRODUCTION

The dairy sector represents one of the most important areas of agricultural production in the European Union. Its significance lies not only in the volume of production or its contribution to food security, but

[1] Institute of Statistics, Operations Research and Mathematics, Faculty of Economics and Management, Slovak University of Agriculture in Nitra, Slovakia. E-mail: eva.richterova@uniag.sk. ORCID: <https://orcid.org/0000-0001-7265-7413>.
[2] Institute of Economics and Management, Faculty of Economics and Management, Slovak University of Agriculture in Nitra, Slovakia. E-mail: martin.richter@uniag.sk. ORCID: <https://orcid.org/0000-0001-6577-9287>.

it also has a broad influence on rural economies, employment opportunities, and trade performance of the EU member states. The dairy sector is closely linked to the agricultural sustainability, regional economic growth and cohesion, and the stability of rural areas particularly in areas with limited alternative economic opportunities.

Following the removal of milk quotas in 2015, the EU dairy sector expanded as producers gained greater flexibility to adjust output to market conditions. Since then, milk production has continued to grow, stimulated by the intensification of dairy farming, the use of high-quality concentrated feeds, advances in breeding programmes, improvements in herd and feed management, better livestock housing, and the increasing adoption of technologies such as automated milking systems. This upward trend has been reinforced over the last decades by the Common Agricultural Policy (CAP), which in its 2023–2027 framework prioritises strengthening the competitiveness of companies, regions, and countries in response to rising global demand and the need to support world dairy markets (European Commission, 2023). However, while productivity has accelerated through the shift toward larger and more automated facilities, these developments have also raised concerns regarding environmental impacts, animal welfare, and the overall sustainability of the dairy sector (Heise & Theuvsen, 2018; Autio et al., 2023). In particular, the dairy industry represents one of the primary sources of agricultural greenhouse gas (GHG) emissions--mainly methane and nitrous oxide – resulting from enteric fermentation, manure management, and feed production (FAO, 2001). Therefore, sustainability and environmental concerns have become central themes in public debate and scientific discourse and are now integral to government agendas. The global dairy industry is increasingly facing pressure to address these challenges (Bhat, 2025). Achieving economic sustainability in the dairy sector depends on establishing a durable competitive position in the market, where the adoption of advanced technologies, precision farming practices, and value-added product innovations can enhance productivity and boost market competitiveness (Agostinho et al., 2019). Moreover, agricultural production in general is confronted with a range of long-term sustainability risks, including the need to increase output while simultaneously mitigating the negative environmental consequences associated with agricultural activities (Arru et al., 2024). In this context, advancing towards more sustainable dairy systems requires a shift to practices that integrate environmental protection with improved resource-use efficiency and long-term sector resilience (Brito et al., 2021). Therefore, assessing eco-efficiency becomes essential, as it provides a comprehensive framework for evaluating how effectively the dairy sector can balance economic performance with environmental responsibility. Most existing studies examine dairy eco-efficiency at the farm level within a single country (Urdiales et al., 2016; Grassauer et al., 2022; Martinsson & Hansson, 2021; Baležentis et al., 2022; Novaković et al., 2025). However, from the perspective of the common EU cohesion policy, it is necessary to analyse the dairy sector also for larger territorial units and make international comparisons. The dairy sectors vary among regions within each nation. European regions are characterised by different environmental conditions, market conditions, consumer behaviour, and policy environment, which inevitably affect the structure of their dairy farming (Bianchi et al., 2020, Thorsøe et al., 2020). Thus, evaluating dairy eco-efficiency regionally is important, yet studies at this level remain limited.

At the same time, it is necessary to monitor the sector's market position from the perspective of competitiveness, to capture the potential synergies between economic viability and ecological sustainability. Aligning economic incentives with environmental objectives allows dairy producers to implement practices that ensure both profitability and ecological performance (Gardašević et al., 2024). While Latruffe (2010) notes that efficiency is often cited as an indicator of competitiveness, the potential ecological dimension is not considered. According to Pietrzak et al. (2023), a significant research gap exists regarding the integration of competitiveness and eco-efficiency analyses in dairy cooperatives. Few studies address the link between eco-efficiency and competitiveness in the dairy sector; however, these are primarily limited to farm-level analyses within individual countries (Menning and Szigeti, 2025; Arru, et al., 2024;

Pietrzak et al., 2023; Zhu et al, 2023; Grzelak et al., 2022; Lioy et al., 2022). Therefore, our paper focuses on assessing regional dairy eco-efficiency together with competitiveness in the dairy market, to simultaneously analyse the relationship between them. Based on Lioy et al.' argument (2022) that, at the farm level, high environmental performance is both an advantage in terms of economic competitiveness and a prerequisite for optimal economic outcomes, we formulate the following research hypothesis: *The eco--efficiency of a region ensures a higher level of outputs for given inputs, thereby increasing its competitiveness.*

Eco-efficiency and competitiveness are evaluated for two years, 2015 and 2022, to compare changes over time. The year 2015 was chosen as the starting point of the analysis, as it marks the abolition of milk quotas and allows capturing their impacts. The year 2022 was selected because it is the most recent year for which the analysed data are available. The analysis focuses on the NUTS2 regions within the Visegrad Four, i.e. Czechia, Slovakia, Hungary, and Poland. Svatoš et al. (2013) highlight that comparing the agricultural sectors of the Visegrad countries is meaningful, as the economic structures of the V4 members have undergone substantial changes over time, with agriculture being among the most affected sectors. Agricultural production in the V4 experienced two major shocks: first in the 1990s, due to the transition from centrally planned to market economies, and later through integration into the European Union.

Dairy eco-efficiency is assessed using an input-oriented Data Envelopment Analysis (DEA) model under the assumption of constant returns to scale.

The paper is structured as follows. First, the concepts of eco-efficiency and competitiveness are introduced, along with a review of relevant literature. Second, the data and methodologies used to assess dairy eco-efficiency and competitiveness are described, including the CCR DEA model, the super-efficiency DEA model, and the construction of a composite index. In the third section, the results are presented and compared with findings from other studies. Finally, the article discusses key findings, limitations, and offers suggestions for future research.

1 LITERATURE SURVEY

The European Commission (EC) wants European food production to be a global example of sustainability through several initiatives, such as the EU Green Deal, the Farm to Fork (F2F) and Biodiversity Strategies, Horizon Europe, Next Generation EU, and the Common Agricultural Policy (CAP). These initiatives aim to support a fair transition (by promoting cooperation between different sustainability goals and reducing conflicts among them. For example, the CAP 2023–2027 focuses on fighting climate change, protecting the environment, and preserving landscapes and biodiversity, while also supporting rural communities, ensuring fair incomes for farmers, improving competitiveness, and strengthening farmers' role in the food chain (European Commission, 2021, 2023). Similarly, the Farm to Fork (F2F) strategy seeks to take advantage of the economic benefits of moving towards more sustainable practices and to create win-win opportunities that enhance both sustainable development and farm profitability, reflecting the synergy between environmental and economic aspects of farming (Mowlds, 2020). To support the successful transformation of the dairy sector, it is essential to develop methods that capture the multiple dimensions of sustainability and consider the interactions between them, providing a practical and usable framework for decision-making at all levels (Dumanski et al., 1998). One of the most used approaches for assessing the environmental dimension of farming is the evaluation of production eco-efficiency, which serves as a basis for linking sustainability with economic performance (UNESCAP, 2009).

The concept of eco-efficiency was first introduced by Schaltegger & Sturm (1990) as a „business link to sustainable development". Eco-efficiency, often described as the integration of economic and ecological objectives, became popular in the 1990s to attain long-term agricultural growth (Gołaś et al., 2020). The concept of eco-efficiency is applicable to different products and sectors, as well as at different levels (Caiado et al., 2017).

Measuring the efficiency of the dairy sector is important not only for improving dairy cow yield performance, but also for enhancing the efficiency of other resources and increasing the incomes of individual dairy farms (Kovács & Szücs, 2020). There are several methods used to quantify eco--efficiency performance, with the ratio approach and the frontier approach in the foreground (Liu et al., 2020; Bianchi et al., 2020). Due to the need to consider multiple input and output variables, as well as to account for the negative aspects of agriculture as undesirable outputs, Data Envelopment Analysis is often preferred (Dyckhoff & Allen, 2001).

A number of studies have applied various DEA approaches (Adenuga et al., 2018; Cecchini et al., 2018; Graussauer et al., 2022; Baležentis, 2022; Novaković et al., 2025; Cortés et al., 2021; Martinsson & Hansson, 2021; Soteriades et al., 2020) and parametric stochastic frontier analysis (Alem, 2023a, 2023b; Le et al., 2020) to evaluate the environmental efficiency of dairy production at the farm level within a single country. Efficiency analysis can also be used to assess only the economic dimension of dairy production; however, in this case it refers solely to economic sustainability, without considering environmental aspects (Špička & Smutka, 2014; Špička, 2015; Špička & Machek, 2015; Náglová & Rudinskava, 2021; Žáková Kroupová et al., 2020; Kovács & Szücs, 2020). In addition, insights from efficiency measurement may be linked with market competitiveness analyses, as both perspectives offer complementary information on performance and resource use within the dairy sector, even though they capture different aspects of sectoral dynamics. Fostering competitiveness of the dairy sector at the national, regional and farm levels is one of the objectives of the CAP 2023–2027. Competitiveness represents the ability of a specific geographical area to stimulate economic growth and enhance social prosperity through the efficient use of resources and improvements in the quality of life of its inhabitants (Judrupa, 2021). A few studies have assessed the competitiveness of the dairy sector at the macro level, applying individual indicators, aggregated indicators and composite indices (Bojnec & Fertő, 2014; Irz & Jansik, 2015; Simo et al., 2016; Nagy & Jámbor, 2019; Parzonko et al., 2024).

Achieving synergy between economic performance and ecological sustainability represents both a challenge and an opportunity for the dairy sector (Britt et al., 2018). Aligning economic incentives with environmental goals allows producers to adopt models that support long-term profitability while reducing environmental impacts (Gardašević et al., 2024). However, ensuring both dimensions of sustainability remains a complex, multidimensional task requiring integrated strategies, stakeholder cooperation, and innovation (Brkić & Puvača, 2024). Studies emphasise the need to improve resource efficiency, minimise environmental impacts, and maintain economic viability (Bi et al., 2024), while also addressing challenges related to technology adoption and diverse regional conditions across the EU (Arvidsson Segerkvist et al., 2020; Penker, 2024). These pressures place eco-efficiency and competitiveness at the forefront of priorities for both producers and policymakers.

Despite growing interest in both eco-efficiency and competitiveness, current research rarely examines these two dimensions simultaneously at the regional level. Many existing studies rely on national averages or evaluate environmental and economic outcomes separately, overlooking substantial regional differences in production conditions, market dynamics, and policy impacts. This leaves a gap in understanding how regions position themselves in terms of both environmental performance and economic strength. By analysing these aspects within a unified regional framework, this study offers evidence that can support more targeted, region-specific policy measures and improve the alignment between sustainability and competitiveness objectives.

2 METHODOLOGY

Five input variables are selected: dairy cows per thousand inhabitants, live bovine per thousand inhabitants, employment in agriculture per thousand inhabitants, gross fixed capital formation in agriculture in € per capita, and compensation of employees in agriculture in € per capita. The output side includes two desirable outputs: milk production in tons per cow and gross value added from agriculture in € per capita, and one undesirable output: greenhouse gas emissions in ton CO_2eq per 1 ton of milk. The selected input and output variables were chosen because they comprehensively capture the key determinants of dairy sector performance at the regional level. Inputs such as dairy cows, live bovine population, employment, capital formation, and labour compensation reflect the scale, labour, and capital intensity of dairy production, which are critical drivers of both productivity and economic output. The desirable outputs-milk production per cow and agricultural gross value added represent the economic efficiency and value creation of the sector, while the undesirable output, greenhouse gas emissions per ton of milk, accounts for its environmental impact. Together, these indicators allow for a multidimensional assessment of eco-efficiency, linking resource use and economic outcomes with sustainability, and are consistent with previous studies on agricultural and dairy sector efficiency (Song et al., 2012; Piao et al., 2019).

Data were obtained from EUROSTAT for the years 2015 and 2022, selected to allow a comparison over time. The year 2015 was chosen as the starting point of the analysis because it marks the abolition of milk quotas and allows capturing their impacts. The year 2022 was selected as it is the most recent year for which the analysed data are available. To ensure comparability under similar geographical conditions, the analysis focuses on the NUTS2 regions within the Visegrad Four: 8 Czech regions, 8 Hungarian regions, 17 Polish regions, and 4 Slovak regions.

The following research hypothesis is established: *The eco-efficiency of a region ensures a higher level of outputs for given inputs, thereby increasing its competitiveness.*

Descriptive statistics for the output variables are computed to enable a multidimensional comparison of the selected V4 regions (Coluccia *et al*., 2020; Cui *et al*., 2024).

A non-parametric frontier method originally proposed by Charnes, Cooper, and Rhodes in 1978 (Charnes *et al*., 1978), input-oriented Data Envelopment Analysis (DEA) assuming constant returns to scale, is used to assess dairy eco-efficiency in V4 regions. DEA applies linear programming to evaluate the relative efficiency or inefficiency of decision-making units (DMUs) that produce multiple outputs using multiple inputs. In this paper, DMUs correspond to the chosen V4 regions. DEA is a useful method to compile a complex efficiency indicator because no explicit weights are required (Dyckhoff & Allen, 2001). Nevertheless, the method has certain limitations – its outcomes are highly influenced by the selection of input and output variables, and the number of efficient DMUs on the frontier often rises as more variables are included (Laurinavičius & Rimkavienė, 2017).

Let's assume that we have *n* independent homogeneous decision-making units, denoted by *DMUj (j = 1, 2, …, n)*. For given *p* nondiscretionary inputs $Z_j = \left(z_{1j}, z_{2j}, \ldots, z_{pj}\right)^T$, each DMU consumes *m* discretionary inputs $X_j = \left(x_{1j}, x_{2j}, \ldots, x_{mj}\right)^T$ to produce *s* outputs $Y_j = \left(y_{1j}, y_{2j}, \ldots, y_{sj}\right)^T$ (Hua et al., 2007). Since greenhouse gas emissions represent an undesirable output, they should be converted into a positive one using an appropriate vector (w) (Seiford & Zhu, 2002).

$$y_j^{-b} = -y_j^b + w > 0 \tag{1}$$

$$y_j^{-b} = -y_j^b + \max\left(y_j^b\right) + 1 \tag{2}$$

The standard linear input-oriented CCR model with a constant return to scale and with undesirable output could be written as the following linear programming problem:

$$Min\,\theta_q$$

$$\sum_{j=1}^{n} y_{rj}^g \lambda_j \geq Y_{rq}^g, \quad r \in G\left(desirable\right)$$

$$\sum_{j=1}^{n} y_{tj}^{-b}\lambda_j \geq Y_{tq}^{-b},\ \ t \in B(undesirable) \tag{3}$$

$$\sum_{j=1}^{n} x_{ij}\lambda_j \leq \theta X_{iq},\ \ i = 1,2,\ldots,m$$

$$\lambda_j \geq 0,\ \ j = 1,2,\ldots,n$$

Where θ_q stands for the technical efficiency of the DMUq and λj represents the weight assigned to the *DMUj (j = 1, 2, …, n)*.
To identifying the most efficient units among those considered efficient by standard DEA, Super-efficiency DEA model is used, and one additional condition is included in the original model: Min θ_q

$$\sum_{j=1}^{n} y_{rj}^{g}\lambda_j \geq Y_{rq}^{g},\ \ r \in G(desirable)$$

$$\sum_{j=1}^{n} y_{tj}^{-b}\lambda_j \geq Y_{tq}^{-b},\ \ t \in B(undesirable)$$

$$\sum_{j=1}^{n} x_{ij}\lambda_j \leq \theta X_{iq},\ \ i = 1,2,\ldots,m \tag{4}$$

$$\lambda_0 = 0$$

$$\lambda j \geq 0,\ \ j = 1,2, ..., m$$

To rank the V4 regions based on their competitiveness in the dairy sector, a Dairy Competitiveness Index is constructed for the years 2015 and 2022. The compilation of this composite index follows several steps, based on the *Manual on the Compilation of Composite Indicators* (OECD, 2008):

1. Adjustment to ensure consistent direction of development – undesirable outputs are transformed to reflect a positive orientation.
2. Standardization of variables – due to differing units of measurement, min-max standardization is used to standardize the variables to a common scale.
3. Determination of indicator weights – the weights of individual indicators are assigned based on Factor analysis. The weight of each indicator (variable) within its specific factor is based on its squared factor loading. Squaring the loadings avoids negative weights and represents the proportion of the variable's variance explained by that specific factor.

$$weight_{ij} = \frac{(Factor\ Loading_{ij})^2}{\sum_{k=1}^{n_i}(Factor\ Loading_{ik})^2} \tag{5}$$

where: $weight_{ij}$ – the calculated weight for the *j-th* variable within the *i-th* factor,
$Factor\ Loading_{ij}$ – the loading of the j-th variable on the i-th factor from rotated factor matrix with Varimax method,
$\sum_{k=1}^{n_i}(Factor\ Loading_{ik})^2$ – the sum of the squared factor loadings for all variables that load highly on the *i-th* factor/dimension.

4. Aggregation of indicators – the indicators are combined using the additive method into a single composite index according to the following formula:

$$I_i^t = \sum_{j=1}^{k} v_j^t U_{ij}^t \quad (6)$$

where: I_i^t – composite index value for region$_i$ at time$_t$,
v_j^t– weight of indicator$_j$ at time$_t$,
U_{ij}^t– standardized value of indicator$_j$ for region$_i$ at time$_t$.
The results of the leave-one-variable-out robustness test show that removing any single indicator does not lead to substantial changes in regional rankings. Only minor variations were observed, while the overall structure and relative positions of regions remained stable. This confirms that the composite indicator is not driven by any single variable and can be considered methodologically robust (Alqararah, 2023).

Pearson's correlation coefficient (r) is used to verify the existence and strength of the relationship between eco-efficiency and competitiveness in the dairy sector and to either accept or reject the established research hypothesis. The coefficient is calculated using the following formula:

$$r = \frac{Cov(x,y)}{s_{xs_y}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (7)$$

where: r – Pearson's correlation coefficient,
x_i and y_i - individual values of the variables *eco-efficiency* (X) and *Dairy Competitiveness Index* (Y) for each V4 region,
$\bar{x}$ and $\bar{y}$ - mean values of *eco-efficiency* (X) and *Dairy Competitiveness Index* (Y) across all V4 regions,
n - number of regions compared.

To assess whether the observed relationship is statistically significant, a hypothesis test is conducted with the following null and alternative hypotheses:
H_0: There is no relationship between the variables ($r = 0$),
H_1: There is a relationship between the variables ($r \neq 0$).
The resulting p-value is compared with the significance level $\alpha = 0.05$. If p-value < α, the null hypothesis is rejected, indicating a statistically significant relationship (Al-Hameed, 2022).

3 RESULTS

At the beginning of the results section, a multidimensional comparison of the V4 regions in terms of key agricultural output indicators, namely milk production, gross value added (GVA) from agriculture and greenhouse gas emissions (GHG) from agriculture, in two time periods 2015 and 2022. Basic descriptive statistics for key agricultural output indicators are presented in Table 1.

On average, between 2015 and 2022, milk production per cow and agricultural gross value added per capita increased in all V4 countries, reflecting overall improvements in productivity and economic performance. However, regional disparities widened, particularly in Hungary and Poland, where growth was uneven and driven by a few high-performing areas, while some regions experienced declines, especially in milk production. Even though overall agricultural greenhouse gas emissions in the V4 increased slightly, emissions in Czechia and Slovakia declined (except for the Slovak region SK02: Západné Slovensko). This overall increase was largely driven by rising emissions in most Hungarian and Polish regions. Hungary and Poland generally have a larger and more intensive livestock sector, which contributes to higher methane emissions. Differences in agricultural policy dynamics also play a role, as environmental performance has improved more rapidly in Czechia and Slovakia, whereas production-driven growth has dominated in many Hungarian and Polish regions. These trends underscore the need for more balanced regional development and environmentally sustainable practices in the V4 dairy sector.

Table 1 Descriptive statistics for key output agricultural indicators

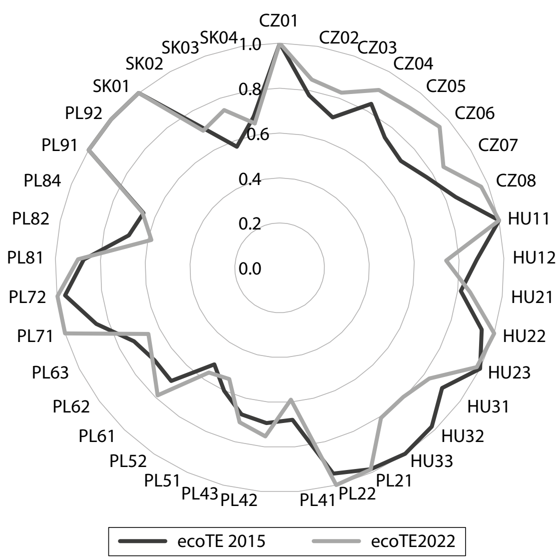
Variable	MEAN	MEDIAN	STDEV	MIN	REGION 1/2	MAX	REGION 2/2
Milk prod. 15	6.99	7.17	1.27	4.80	PL21	8.84	HU21
Milk prod. 22	7.73	7.51	2.32	3.00	PL43	16.20	HU11
GVA 15	346.00	308.89	182.03	36.87	HU11	784.35	HU33
GVA 22	545.58	527.00	311.32	59.22	HU11	1 557.26	PL92
GHG 15	0.031	0.025	0.024	0.003	PL84	0.130	PL42
GHG 22	0.032	0.026	0.026	0.004	PL84	0.134	HU11

Source: Own calculations, data obtained from Eurostat for 2015, 2022

Monitoring the achievement of not only performance goals, but also sustainability goals in the dairy sector is possible based on an assessment of the eco-efficiency of this sector. Eco-efficiency reflects the ability to generate greater desirable output from given inputs while reducing undesirable outputs such as greenhouse gas emissions.

The dairy sector`s eco-efficiency scores of V4 regions in 2015 and 2022 are analysed in the second part and are shown in Figure 1.

Figure 1 Eco-efficiency of V4 regional dairy sector in 2015 and 2022



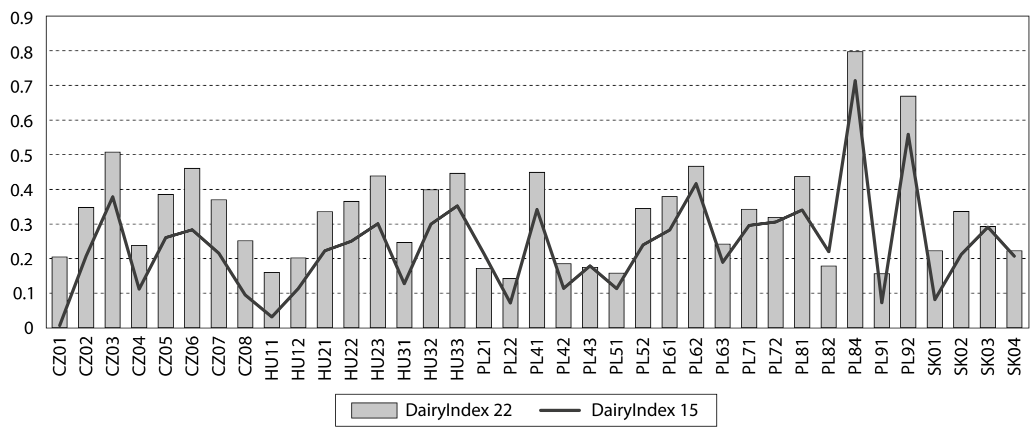
Source: Own calculations, data obtained from Eurostat for 2015, 2022

In 2015, only seven V4 regions achieved eco-efficiency in the dairy sector (ecoTE = 1.0 in Figure 1): These included the capital regions of each V4 country – CZ01: Praha, HU11: Budapest, PL91: Warszawski Stołeczny, and SK01: Bratislavský kraj – along with one other Polish region, PL92:

Mazowiecki regionalny, and two Hungarian regions, HU33: Dél-Alföld, and HU23: Dél-Dunántúl. In 2022, all capital regions and PL92: Mazowiecki regionalny remained eco-efficient. Their strong human capital, infrastructure, research capacity and access to financial and policy support continue to facilitate innovation and adoption of eco-efficient dairy practices. Three additional regions became eco-efficient in 2022 (HU22: Nyugat-Dunántúl, PL22: Śląskie and PL71: Łódzkie). Conversely, the previously eco-efficient Hungarian regions HU33: Dél-Alföld and HU23: Dél-Dunántúl became eco-inefficient. Altogether, eight V4 regions were eco-efficient in 2022. Eco-efficiency declined in most Hungarian regions except Pest HU11: Pest, HU21: Közép-Dunántúl, and HU22: Nyugat-Dunántúl. Several Polish (PL41: Wielkopolskie, PL51: Dolnośląskie, PL6 All other V4 regions were eco-inefficient in both years.3: Pomorskie, PL82: Podkarpackie) and Slovak regions (SK02: Západné Slovensko, and SK04: Východné Slovensko) also recorded decreases, driven by rising input costs, farm structural changes, labour shortages, limited investment capacity, and insufficient environmental incentives.

To complement the eco-efficiency results, regional performance was also assessed through a Dairy Competitiveness Index for 2015 and 2022 (DairyIndex 15 and DairyIndex 22 in Figure 2). The index ranges from 0 to 0.8, and V4 regions were ranked from the most (1st) to the least competitive (37th). In both years, the top four positions were held by the same regions, reflecting their strong and stable role in the V4 dairy production. The most competitive region was the Polish PL84: Podlaskie, supported by the presence of major dairy processors (SM Mlekpol, SM Mlekovita, SM Piątnica). Second was PL92: Mazowiecki regionalny, also home to leading companies. In 2015, PL62: Warmińsko-Mazurskie ranked third and CZ03: Jihozápad fourth; these two exchanged places by 2022. The Hungarian region HU33: Dél-Alföld ranked fifth in 2015 but dropped to seventh by 2022, while the Czech region CZ06: Jihovýchod improved markedly from 13th to 5th, reflecting its strong processing capacity. Competitiveness improved in most V4 regions between 2015 and 2022. Only three Polish regions (PL82: Podkarpackie, PL21: Małopolskie, and PL43: Lubuskie) – declined substantially due to lower productivity, structural constraints, and rising GHG emissions per unit of milk. In 2015, the least competitive region was CZ01: Praha, but it rose to 28th place by 2022 due to its low initial agricultural base. By contrast, PL22: Śląskie became the least competitive in 2022, largely due to limited dairy production, small farm sizes, and the absence of major processors. Capital-city regions in general (Praha, Budapest, Warsaw, Bratislava) consistently ranked near the bottom, reflecting their predominantly urban and service-oriented economic structure rather than dairy production.

Figure 2 Dairy Competitiveness Index of V4 regions in 2015 and 2022



Source: Own calculations, data obtained from Eurostat for 2015, 2022

To rank the V4 regions based on the achieved eco-efficiency values, the original input-oriented DEA model with the assumption of constant returns to scale is modified to the Super-efficiency DEA model. Based on the super eco-efficiency values, the V4 regions are ranked in Table 2 (SuperTE15 and SuperTE22). To assess the relationship between eco-efficiency and competitiveness, the table also presents the ranking of the V4 regions based on the constructed Dairy Competitiveness Index. The rankings are based on the relative position of each region in the year 2015 and 2022, enabling an assessment of both performance and dynamics over time.

Table 2 Ranking of V4 regions based on super TE and Dairy Competitiveness Index in 2015, 2022

NUTS2	superTE15	superTE22	Dairy Index15	Dairy Index22	NUTS2	superTE15	superTE22	Dairy Index15	Dairy Index22
HU11	2	1	36	34	HU31	13	20	27	23
CZ01	1	2	37	28	CZ03	26	21	4	3
PL91	4	3	34	36	HU33	5	22	5	7
PL92	3	4	2	2	HU32	9	23	10	10
HU22	12	5	16	14	PL61	27	24	14	12
SK01	6	6	33	27	PL42	28	25	28	30
PL22	11	7	35	37	SK03	36	26	12	21
PL71	16	8	11	17	HU12	14	27	30	29
PL72	10	9	8	20	PL43	33	28	26	32
PL21	8	10	21	33	PL62	29	29	3	4
HU23	7	11	9	8	SK02	25	30	22	18
CZ08	17	12	32	22	PL84	34	31	1	1
CZ06	24	13	13	5	PL63	23	32	25	24
CZ05	22	14	15	11	SK04	31	33	23	26
CZ04	18	15	31	25	PL41	32	34	6	6
PL81	15	16	7	9	PL82	30	35	19	31
CZ07	21	17	20	13	PL52	37	36	17	16
HU21	19	18	18	19	PL51	35	37	29	35
CZ02	20	19	24	15					

Source: Own calculations, data obtained from Eurostat for 2015, 2022

Almost all the most competitive V4 regions in dairy production (PL84: Podlaskie, CZ03: Jihozápad, PL62: Warmińsko-mazurskie, and PL41: Wielkopolskie) ranked relatively low in terms of eco-efficiency, except for PL92: Mazowiecki regionalny. These regions are key dairy producers and prioritize high output levels to meet domestic and export demand. However, achieving high volumes does not necessarily mean that resources (e.g., feed, water, energy) are used efficiently relative to output. While large farms benefit from economies of scale, they can also be less flexible in adopting new eco-friendly technologies if those require significant capital investment or operational change. In contrast, regions with smaller, innovative farms sometimes adopt eco-efficient practices faster. These highly dairy competitive regions could further enhance their competitiveness by improving eco-efficiency and achieving higher outputs from a given level of input. On the other hand, the most eco-efficient regions, such as HU11: Budapest, CZ01: Praha, PL91: Warszawski Stołeczny and SK01: Bratislavský kraj, ranked among the lowest in terms of dairy competitiveness. This suggests that while eco-efficiency can contribute to better competitiveness, it may not be sufficient in regions where the dairy sector is economically less significant or small in scale. Only Polish region PL92: Mazowiecki regionalny was both eco-efficient and highly competitive. The presence of large, modern dairy farms and processing facilities in this region contributes to higher productivity. This Polish region is located near Warsaw, and, therefore, the region enjoys excellent infrastructure, including transport, logistics, and access to research institutions and extension services. Unlike in many other regions where these two aspects diverge, PL92: Mazowiecki regionalny shows that economic performance and environmental responsibility can align, particularly when supported by smart investment and innovation policies. Hungarian region HU33: Dél-Alföld was highly competitive in both years (6^{th} in 2015 and 8th in 2022) but lost its eco-efficiency status in 2022. This highlights that a region can maintain competitiveness even as eco-efficiency declines, particularly if scale, market access, or economic output remain strong.

The relationship between dairy eco-efficiency levels, based on the super-DEA model, and the Dairy Competitiveness Index was examined using Pearson's correlation coefficient. The coefficient was –0.32 in 2015 and –0,19 in 2022, with p-value above α (0.05) in both analysed years, indicating a very weak and statistically insignificant relationship between these two variables (Table 3).

Table 3 Pearson Correlation Coefficient in 2015 and 2022

Pearson Correlation Coefficients, N = 37 Prob > |r| under H0: Rho=0	
	TE15
IDC15	–0.32456
	0.0500
Pearson Correlation Coefficients, N = 37 Prob > |r| under H0: Rho=0	
	TE22
IDC22	–0.19391
	0.2502

Source: Own calculations, data obtained from Eurostat for 2015, 2022

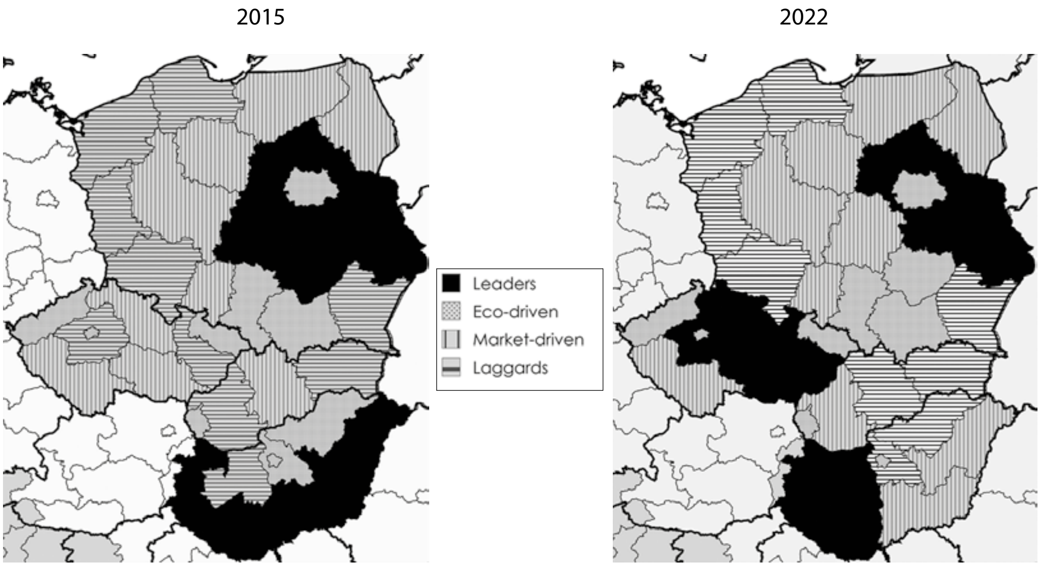
The proposed research hypothesis – *The eco-efficiency of a region ensures a higher level of outputs for given inputs, thereby increasing its competitiveness.* – was rejected because the link between eco-efficiency and competitiveness in the dairy sector was not statistically confirmed.

In both years, based on dairy eco-efficiency values and Dairy Competitiveness Index values, the V4 regions are classified into four groups: Leaders – regions achieving above-average values in both eco-efficiency and the Dairy Competitiveness Index (highlighted in black on the maps), Market-driven regions – regions achieving above-average values in the Dairy Competitiveness Index but below-average values in eco-efficiency (highlighted with checker pattern on the maps), Eco-driven regions – regions achieving above-average eco-efficiency but below-average values in the Dairy Competitiveness Index (highlighted in white on the maps) and Laggards – regions with below-average values in both the Dairy Competitiveness Index and eco-efficiency (highlighted with horizontal stripes on the maps).

In 2015, the Leaders group included eight regions: four Hungarian (HU22: Nyugat-Dunántúl, HU23: Dél-Dunántúl, HU32: Észak-Alföld, and HU33: Dél-Alföld); and four Polish (PL71: Łódzkie, PL72: Świętokrzyskie, PL81: Lubelskie and PL92: Mazowiecki regionalny) (Figure 3). The Market-driven group consisted of nine regions, mainly strong dairy producers from the Czech Republic (CZ03: Jihozápad, CZ05:

Severovýchod and CZ06: Jihovýchod), Poland (PL41: Wielkopolskie, PL52: Opolskie, PL61: Kujawsko-pomorskie, PL62: Warmińsko-mazurskie and PL84: Podlaskie), and Slovakia (SK03: Stredné Slovensko). The Eco-driven group included ten regions, comprising the capital regions CZ01: Praha, HU11: Budapest, PL91: Warszawski stołeczny, and SK01: Bratislavský kraj, as well as other environmentally strong regions: CZ04: Severozápad, CZ08: Moravskoslezsko, HU12: Pest, HU31: Észak-Magyarország, PL21: Małopolskie, and PL22: Śląskie. The Laggards group also included ten regions from all V4 countries, generally those with below-average eco-efficiency and competitiveness, namely: The Czech regions CZ02: Střední Čechy and CZ07: Střední Morava; the Hungarian regions HU21: Közép-Dunántúl; the Polish regions PL42: Zachodniopomorskie, PL43: Lubuskie, PL51: Dolnośląskie, PL63: Pomorskie and PL82: Podkarpackie; and the Slovak regions SK02: Západné Slovensko and SK04: Východné Slovensko (Figure 3).

Figure 3 Map of V4 groups in 2015 and 2022



Source: Own calculations, data obtained from Eurostat for 2015, 2022

By 2022, the Leaders expanded to nine regions (Figure 3): HU22: Nyugat-Dunántúl, HU23: Dél-Dunántúl, PL81: Lubelskie, PL92: Mazowiecki regionalny (maintained); CZ05: Severovýchod, CZ06: Jihovýchod (improved eco-efficiency); HU21: Közép-Dunántúl, CZ02: Střední Čechy, CZ07: Střední Morava (improved both).

The Market-driven group included ten regions: CZ03: Jihozápad, PL41: Wielkopolskie, PL52: Opolskie, PL61: Kujawsko-pomorskie, PL62: Warmińsko-mazurskie, PL84: Podlaskie (maintained); HU3: Észak-Alföld, HU33: Dél-Alföld, PL71: Łódzkie (downgraded from Leaders); SK02: Západné Slovensko (improved competitiveness).

The Eco-driven group included nine regions: CZ01: Praha, CZ04: Severozápad, CZ08: Moravskoslezsko, HU11: Budapest, PL21: Małopolskie, PL22: Śląskie, PL91: Warszawski stołeczny, SK01: Bratislavský kraj (maintained), and PL72 Świętokrzyskie (downgraded from Leaders).

The Laggards group included nine regions: PL42: Zachodniopomorskie, PL43: Lubuskie, PL51: Dolnośląskie, PL63: Pomorskie, PL82: Podkarpackie, SK04: Východné Slovensko (maintained); HU12: Pest, HU31: Észak-Magyarország (downgraded from Eco-driven); SK03: Stredné Slovensko (downgraded from Market-driven).

These classifications highlight persistent disparities in eco-efficiency and competitiveness. Czech regions have improved the most, with several moving into the Leaders by 2022 and no remaining Laggards by 2022. Polish regions largely maintained positions, Hungarian regions stagnated or declined, and Slovak regions remained the weakest. The coexistence of Eco-driven and Market-driven regions in both years shows that eco-efficiency and competitiveness do not necessarily align, confirming limited convergence in the V4 dairy sector and the need for tailored, region-specific strategies.

4 DISCUSSION

Based on the results, except for regions containing a capital city, the remaining eco-efficient regions were from Poland and Hungary. According to Náglová & Rudinskaya (2021) and Žáková Kroupová et al. (2020), Poland had the most efficient dairy sector among the V4 countries. In line with our focus, Rybaczewska-Błazejowska & Masternuk-Janus (2018) argued that ecological efficiency should be analysed within specific agricultural sectors, rather than agriculture. As their analysis suggests, Polish regions PL21: Malopolskie and PL41: Wielkopolskie were overall agriculturally eco-efficient. However, similarly as in this study, they were eco-inefficient in the case of dairy farming, despite having a significant number of cows. They produce relatively less milk due to a combination of factors, including smaller herds, less specialized farming, and potentially less focus on maximizing the milk yield of individual cows.

Consistent with the findings of this study, the Hungarian region HU22: Nyugat-Dunántúl was also recognized as fully efficient by Špička & Smutka (2014) and demonstrated the highest Malmquist index value among the V4 regions in the analysis by Špička & Machek (2015). On the other hand, over time, the eco-efficiency decline was observed in almost all Hungarian regions. According to Századvég (2022), this decline in eco-efficiency can be attributed to a combination of economic, structural, and technological challenges affecting the Hungarian dairy sector. These regions face a significant shortage of labour, especially skilled workers essential for modern dairy operations, which hinders the adoption of efficient practices and advanced technology management. Moreover, rising energy costs have increased operating expenses, making it financially difficult for dairy farms to maintain or improve environmentally efficient practices.

When examining dairy competitiveness, the best positions were achieved predominantly by eco--inefficient Polish regions, except for PL92: Mazowiecki regionalny. Parzonko (2024) and Nagy & Jámbor (2019) also recognized Poland as the most competitive V4 country in the dairy sector, thanks to its key dairy regions that prioritize high output levels to meet both domestic and export demand.

Based on the eco-efficiency and competitiveness rankings, as well as the computed Pearson's correlation coefficient, which indicated a very weak and statistically insignificant relationship between eco-efficiency and competitiveness, the research hypothesis was rejected. At the regional level, high environmental performance does not play a key role in supporting economic competitiveness, unlike at the farm level (Lioy et al., 2021). Similarly, Arru et al. (2024) reported that economic performance and environmental outcomes in all livestock subsectors developed in different directions, suggesting that improvements in one do not necessarily translate into gains in the other. From both the ecological and economic standpoint, Czechia recorded the greatest progress between 2015 and 2022. This positive trend reflects successful modernization processes and a gradual movement toward the EU average. Consistent with these findings, Bórawski et al. (2020) also reported that dairy farms in Czechia exhibited the highest level of economic sustainability among the V4 countries.

CONCLUSION

This study provides a multidimensional assessment of the dairy sector in the Visegrad four (V4) regions, analysing its economic output, environmental performance, eco-efficiency and competitiveness over the period 2015 and 2022.

Overall, milk production per cow and agricultural gross value added in € per capita increased in all V4 countries, reflecting improvements in productivity and economic performance. However, this growth was uneven across regions, with some Hungarian and Polish regions even experiencing a decline. Rising standard deviations for both indicators suggest a widening gap between high-performing and lagging regions, particularly in Hungary and Poland. Overall agricultural greenhouse gas emissions in the V4 also increased slightly, mainly driven by most Hungarian and Polish regions, where larger and more intensive livestock production, combined with more production-oriented agricultural policies, contributed to the rise. These trends highlight the importance of promoting balanced regional development and environmentally sustainable practices in the V4 dairy sector.

The eco-efficiency analysis further revealed that while some regions improved between 2015 and 2022, others, such as almost all Hungarian regions, four Polish regions, and two Slovak regions, experienced a decline. Structural challenges – including labour shortages, rising production costs, and limited technological adoption – have eroded environmental performance in several areas. In 2015, all capital regions of the V4 countries, along with the Polish region PL92: Mazowiecki regionalny and two Hungarian regions HU33: Dél-Alföld and HU23: Dél-Dunántúl, had eco-efficient dairy sectors (a total of seven V4 regions). By 2022, this number increased to eight. Each of the capital regions, as well as PL92: Mazowiecki regionalny near Warsaw, maintained their eco-efficiency. These regions benefit from well-developed infrastructure, including transport, logistics, and access to research institutions and advisory services. Additionally, the Hungarian region HU22: Nyugat-Dunántúl and the Polish regions PL22: Śląskie and PL71: Łódzkie improved their eco-efficiency and became eco-efficient by 2022. Excluding the capital regions, most eco-efficient dairy regions in 2022 were in Poland.

The Dairy Competitiveness Index adds another layer of insight. The most competitive regions remained consistent over time. In both years examined, the top four positions were maintained by the same V4 regions. The most competitive dairy sector among the V4 regions had the Polish region PL84: Podlaskie, home to three leading Polish dairy companies. In 2015, the second position was held by another Polish region, PL92: Mazowiecki regionalny, followed by PL62: Warmińsko-Mazurskie in third place and the Czech region CZ03: Jihozápad in fourth. In 2022, the latter two regions swapped positions, with PL62: Warmińsko-Mazurskie taking fourth and CZ03: Jihozápad moving up to third place. All mentioned regions host major dairy companies and exhibit strong production capacity and market orientation. Conversely, capital city regions, which had eco-efficient dairy sector, ranked among the least competitive due to the relatively low scale and economic relevance of dairy production in urban settings.

A notable finding is the very weak connection between eco-efficiency and competitiveness. Based on computed Pearson's correlation coefficient, the proposed research hypothesis – *The eco-efficiency of a region ensures a higher level of outputs for given inputs, thereby increasing its competitiveness.* – was rejected.

Some regions perform well environmentally but lag in market competitiveness, and vice versa. For example, HU11: Budapest, CZ01: Praha, PL91: Warszawski Stołeczny, and SK01: Bratislavský kraj were the most eco-efficient in 2022 but among the least competitive. Conversely, key dairy-producing regions prioritizing high output often ranked low in eco-efficiency, highlighting that strong environmental performance does not guarantee market success, nor does competitiveness ensure sustainability. This underscores the need for integrated strategies addressing productivity, market strength, and eco-efficiency simultaneously.

The classification of V4 regions into Leaders, Market-driven, Eco-driven, and Laggards revealed persistent structural disparities:

1. Czechia improved most significantly. Four regions moved up: two from Market-driven to Leaders via improved eco-efficiency, and two from Laggards to Leaders via gains in both eco-efficiency and competitiveness. By 2022, no Czech region remained in the Laggards, reflecting structural modernization and convergence with the EU average.

2. Poland largely maintained positions, with minor shifts: one region dropped from Leaders to Market--driven, and another from Leaders to Eco-driven.
3. Hungary saw several regions decline, moving from Leaders to Market-driven or from Eco-driven to Laggards, due to declining milk yields, rising GHG emissions, aging farm infrastructure, and slower adoption of low-emission technologies. Only HU21 Közép-Dunántúl improved into the Leaders.
4. Slovakia remained the weakest performer, with no regions in Leaders and most in Laggards. Only minor improvements were observed (e.g., SK02 Západné Slovensko moving to Market-driven).

The coexistence of Eco-driven and Market-driven groups in both years confirms that eco-efficiency and competitiveness do not automatically align. Regions face different combinations of environmental and economic challenges, requiring differentiated policy interventions: Eco-driven regions need support to boost productivity; Market-driven regions require investment in environmental technologies; Laggards need comprehensive support; Leaders can act as hubs for knowledge transfer and innovation.

These findings align closely with the CAP 2023–2027, especially eco-schemes aiming to reduce environmental pressures while maintaining productivity. Measures such as precision nutrient management, improved grassland practices, low-emission livestock strategies, and enhanced animal welfare can help regions reduce GHG emissions and improve eco-efficiency without compromising output. Eco-driven but less competitive regions can benefit from incentives for sustainable practices, while high output but environmentally weaker regions require measures to reduce emissions intensity. Overall, the CAP framework is well-suited to address structural disparities and promote climate-smart innovations in the V4 dairy sector.

The study is limited using regional-level data, which may mask farm-level heterogeneity in productivity, technology adoption, and environmental performance. Only a limited set of inputs and outputs was included, focusing on milk production, agricultural GVA, and GHG emissions, which may not capture all dimensions of eco-efficiency and competitiveness. The DEA model assumes constant returns to scale and considers only GHG as an undesirable output, potentially oversimplifying production realities and environmental impacts. Finally, the analysis compares only two years (2015 and 2022), limiting the ability to observe longer-term trends or causal relationships. Future research could address these limitations by incorporating farm-level data, additional environmental and economic indicators, and longer time series to better understand the drivers of eco-efficiency and competitiveness.

References

ADENUGA, A. H., DAVIS, J., HUTCHINSON, G., DONNELLAN, T., & PATTON M. (2018). Modelling regional environmental efficiency differentials of dairy farms on the island of Ireland [online]. *Ecological Indicators*, 95(1): 851–861. <https://doi.org/10.1016/j.ecolind.2018.08.040>.

AGOSTINHO, F., OLIVEIRA, M.W., PULSELLI, F.M., ALMEIDA, C.M.V.B., & GIANNETTI, B.F. (2019). Emergy Accounting as a Support for a Strategic Planning towards a Regional Sustainable Milk Production [online]. *Agricultural Systems*, 176(2019): 102647. <https://doi.org/10.1016/j.agsy.2019.102647>.

ALEM, H. (2023a). Measuring dynamic and static eco-efficiency in Norwegian dairy farms: a parametric approach [online]. *Front. Environ. Econ., Sec. Agricultural Economics*, 2(2023): 1182236. <https://doi.org/10.3389/frevc.2023.1182236>.

ALEM, H. (2023b). A parametric analysis of eco-efficiency and its determinants: Evidence from Norwegian dairy farms [online]. *Agricultural Economics – Czech*, *69*(7), 284–290. <https://doi.org/10.17221/149/2023-AGRICECON>.

AL-HAMEED, K. A. A. (2022). Spearman's correlation coefficient in statistical analysis [online]. *Int. J. Nonlinear Anal. Appl.*, 13(1). 3249–3255. <https://doi.org/10.22075/ijnaa.2022.6079>.

ALQARARAH, K. (2023). Assessing the robustness of composite indicators: the case of the Global Innovation Index [online]. *Journal of Innovation and Entrepreneurship*, 12(61), 1–22. <https://doi.org/10.1186/s13731-023-00332-w>.

ARRU, B., CISILINO, F., SAU, P., FURESI, R., PULINA, P., & MADAU, F. A. (2024). The economic and environmental sustainability dimensions of agriculture: a trade-off analysis of Italian farms [online]. *Front. Sustain. Food Syst.*, 8(2024), 1474903. <https://doi.org/10.3389/fsufs.2024.1474903>.

ARVIDSSON SEGERKVIST, K., HANSSON, H, SONESSON, U., & GUNNARSSON, S. (2020). Research on environmental, economic, and social sustainability in dairy farming: A systematic mapping of current literature [online]. *Sustainability,* 12(14), 5502. <https://doi.org/10.3390/su12145502>.

AUTIO, M., SEKKI, S., AUTIO, J., PELTONEN, K., & NIVA, M. (2023). Towards de-dairyfication of the diet?—Consumers downshifting milk, yet justifying their dairy pleasures [online]. *Front. Sustain*., 4(2023), 975679. <https://doi.org/10.3389/frsus.2023.975679>.

BALAŽENTIS, T., DABKIENĖ, V., & ŠTREIMIKIENĖ, D. (2022). Eco-efficiency and shadow price of greenhouse gas emissions in Lithuanian dairy farms: An application of the slacks-based measure [online]. *Journal of Cleaner Production*, 356(2022), 131857. <https://doi.org/10.1016/j.jclepro.2022.131857>.

BHAT, R. & INFASCELLI, F. (2025). The path to Sustainable Dairy Industry: Addressing Challenges and Embracing Opportunities [online]. *Sustainability*, 1*7*(2025), 3766. <https://doi.org/10.3390/su17093766>.

BIANCHI, M., DEL VALLE, I., & TAPIA, C. (2020). Measuring eco-efficiency in European regions: evidence from a territorial perspective [online]. *Journal of Cleaner Production*, 276(2020), 123246. <https://doi.org/10.1016/j.jclepro.2020.123246>.

BI, A. Z., UMESH, K. B., ABDUL, B. MD., SIVAKUMA, D., & SRIKANTH, P. (2024). Economic and environmental sustainability of agriculture production at the crop level [online]. *Global J. Environ. Sci. Manage*, 10(3), 1433–1456. <https://doi.org/10.22034/gjesm.2024.03.29>.

BOJNEC, Š., & FERTÖ, I. (2014). Export competitiveness of dairy products on global markets: The case of the European Union countries [online]. *Journal of Dairy Science*, 97(10), 6151–6163. <https://doi.org/10.3168/jds.2013-7711>.

BÓRAWSKI, P., PAWLEWICZ, A., MICKIEWICZ, B., PAWLEWICZ, K., BEŁDYCKA-BÓRAWSKA, A., HOLDEN, L., & BRELIK, A. (2020). Economic Sustainability of Dairy Farms in the EU [online]. *European Research Studies Journal 1,* XXIII(1), 955–978. <https://doi.org/10.35808/ersj/1864>.

BRITO, L. F., BEDERE, N., DOUHARD, F., OLIVEIRA, H.R., ARNAL, M., PEÑAGARICANO, F., SCHINCKEL, A. P., BAES, C.F., & MIGLIOR, F. (2021). Genetic Selection of High-Yielding Dairy Cattle toward Sustainable Farming Systems in a Rapidly Changing World [online]. *Animal,* 15(2021), 100292. <https://doi.org/10.1016/j.animal.2021.100292>.

BRITT, J. H., CUSHMAN, R. A., DECHOW, C. D., DOBSON, H., HUMBLOT, P., HUTJENS, M. F., JONES, G. A., RUEGG, P. S., SHELDON, I. M., & STEVENSON, J. S. (2018). Learning from the Future—A Vision for Dairy Farms and Cows in 2067 [online]. *Journal of Dairy Science*, 101(2018), 3722–3741. <https://doi.org/10.3168/jds.2017-14025>.

BRKIĆ, I. & PAVUČA, N. (2024). Economic and Ecological Sustainability of Dairy Production [online]. *Agron Technol Eng Manag*, 7(3), 1088–1104. <https://doi.org/10.55817/TQXU2115>.

CAIADO, R. G. G., DIAS, R. F., MATTOS, L. V., QUELHAS, O. L. G., & FILHO, W. L. (2017). Towards sustainable development through the perspective of ecoefficiency – a systematic literature review [online]. *Journal of Cleaner Production*, *165*(2017), 890–904. <https://doi.org/10.1016/j.jclepro.2017.07.166>.

CECCHINI, L., VENANZI, S., PIERRI, A., & CHIORRI, M. (2018). Environmental efficiency analysis and estimation of CO2 abatement costs in dairy cattle farms in Umbria (Italy): A SBM-DEA model with undesirable output [online]. *Journal of Cleaner Production*, *197*(1), 895–907. <https://doi.org/1016/j.jclepro.2018.06.165>.

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>.

COLUCCIA, B., VALENTE, D., FUSCO, G., DE LEO, F., & PORRINI, D. (2020). Assessing agricultural eco-efficiency in Italian Regions [online]. *Ecological Indicators* 116(4), 10648, 7–10. <https://doi.org/10.1016/j.ecolind.2020.106483>.

CORTÉS, A., FEIJOO, G., FERNÁNDEZ, M., & MOREIRA, M. T. (2021). Pursuing the route to eco-efficiency in dairy production: The case of Galician area [online]. *Journal of Cleaner Production*, 285(2021), 124861. <https://doi.org/10.1016/j.jclepro.2020.124861>.

CUI, X., XIONG, J., & LIU, Y. (2024). Agricultural Eco-Efficiency: Progress, Challenges and Prospects. *Journal of Resources and Ecology*, 15(5), 1358–1367. <https://doi.org/ 10.5814/j.issn.1674-764x.2024.05.022>.

DUMANSKI, J., TERRY, E., BYERLEE, D., and PIERI, C. 1998. Performance indicators for sustainable agriculture. Washington: The World Bank, 115–1124.

DYCKHOFF, H., & ALLEN, K. (2001). Measuring ecological efficiency with data envelopment analysis [online]. *European Journal of Operational Research*, 132(2), 312–325. <https://doi.org/10.1016/S0377-2217(00)00154-5>.

EUROPEAN COMMISSION (2021). A long-term vision for the EU's rural areas—Towards stronger, connected, resilient and prosperous rural areas by 2040. [online]. European Commission. [cit. 10.6.2025]. <https://www.fao.org/family-farming/detail/en/c/1414177/>.

EUROPEAN COMMISSION (2023). Key policy objectives of the CAP 2023-27. [online]. European Commission. [cit. 16.5.2025]. <https://agriculture.ec.europa.eu/common-agricultural-policy/cap-overview/cap-2023-27/key-policy-objectives-cap-2023-27_en#documents>.

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS (FAO) (2001). Agriculture and the environment: changing pressures, solutions and trade-offs. [online]. Food and Agriculture Organization of the United Nations. [cit. 12.6.2025]. <https://www.fao.org/3/y4252e/y4252e12.pdf>.

GARDAŠEVIĆ, J., BRKIĆ, I., & KOVAČEVIĆ, M. (2024). Management of the Entrepreneurial Ecosystem [online]. *J Agron Technol Eng Manag,* 7(2024), 1067–1073. <https://doi.org/10.55817/KUNO3911>.

GOŁAŚ, M., SULEWSKI, P., WAS, A., KLOCZKO-GAJEWSKA, A., & JUREK, K. (2020). On the Way to Sustainable Agriculture—Eco-Efficiency of Polish Commercial Farms [online]. *Agriculture*, 10 (10), 438. <https://doi.org/10.3390/agriculture10100438>.

GRASSAUER, F., HERNDL, M., NEMECEK, T., FRITZ, Ch., GUGGENBERGER, T., STEINWIDDER, A., & ZOLLITSCH, W. (2022). Assessing and improving eco-efficiency of multifunctional dairy farming: The need to address farms' diversity [online]. *Journal of Cleaner Production*, 338(5), 130627. <https://doi.org/10.1016/j.jclepro.2022.130627>.

GRZELAK, A., BORYCHOWSKY, M., & STANISZEWSKI, J. (2022). Economic, environmental, and social dimensions of farming sustainability – trade-off or synergy? [online]. *Technological and Economic Development of Economy*, 28(3), 1–21. <https://doi.org/10.3846/tede.2022.16463>.

HEISE, H., & THEUVSEN, L. (2018). German dairy farmers' attitudes toward farm animal welfare and their willingness to participate in animal welfare programs: A cluster analysis [online]. *Int. Food Agribus. Manag. Rev.*, 21(2018), 1121–1136. <https://doi.org/10.22434/IFAMR2017.0066>.

HUA, Z., BIAN, Y., & LIANG, L. (2007). Eco-efficiency analysis of paper mills along the Huai river: an extended DEA approach [online]. *Omega*, 35(5), 578–587. <https://doi.org/10.1016/j.omega.200511.001>.

IRZ, X., & JANSIK, C. (2015). Competitiveness of dairy farms in northern Europe: A cross-country analysis [online]. *Agricultural and Food Science,* 24(3), 206–218. <https://doi.org/10.23986/afsci.50881>.

JUDRUPA, ILZE. (2021). Regional Competitiveness as an Aspect Promoting Sustainability of Latvia [online]. *European Journal of Sustainable Development*, 10(1), 2021, 1., pp. 650. <https://doi.org/10.14207/ejsd.2021.v10n1p650>.

KOVÁCS, K., & SZÜCS, I. (2020). Explorign efficiency reserves in Hungarian milk production [online]. *Studies in Agricultural Economics*, 122(2020), 37–43. <https://doi.org/10.7896/j.1919>.

LATRUFFE, L. (2010). Competitiveness, Productivity and Efficiency in the Agricultural and Agri-Food Sectors [online]. *Agriculture and Fisheries Working Papers, No. 30,* OECD Publishing, Paris. <https://doi.org/10.1787/5km91nkdt6d6-en>.

LAURINAVIČIUS, E., & RIMKAVIENĖ, D. (2017). The Comparative Efficiency Analysis of EU Members Agriculture Sectors [online]. *Rural Sustainability Research*, 37(332), 14–18. <https://doi.org/10.1515/plua-2017-0002>.

LE, S., JEFFREY, S. & AN, H. (2020). Greenhouse Gas Emissions and Technical Efficiency in Alberta Dairy Production: What Are the Trade-Offs? [online]. *Journal of Agricultural and Applied Economics*, 52(2020), 177–193. <https://doi.org/10.1017/aae.2019.41>.

LIOY, R., BATTHEU-NOIRFALISE, C., LEHNEN, A., REDING, R., & DUSSELDORF, T. (2021). Is Environmental Efficiency Compatible with Economic Competitiveness in Dairy Farming? A Case Study of 80 Luxembourgish Farms [online]. In book: Towards a Sustainable Future - Life Cycle Management, Challenges and Prospects (pp.49–60). <https://doi.org/10.1007/978-3-030-77127-0_5>.

LIU, Y., ZOU L., & WANG, Y. (2020). Spatial-temporal characteristics and influencing factors of agricultural eco-efficiency in recent 40 years [online]. *Land Use Policy*, 97(2020), 104794. <https://doi.org/10.1016/j.landusepol.2020.104794>.

MARTINSSON, E., & HANSSON, H. (2021). Adjusting eco-efficiency to greenhouse gas emissions targets at farm level – The case of Swedish dairy farms [online]. *Journal of Environmental Management*, 287(1), 112313. <https://doi.org/10.1016/j.jenvman.2021.112313>.

MENNING, P. & SZIGETI, Z. (2025). Synergies and trade-offs between environmental impacts and farm profitability: The case of pasture-based dairy production systems [online]. *Journal of Agriculture and Food Research*, 21(2025), 101798. <https://doi.org/10.1016/j.jafr.2025.101798>.

MOWLDS, S. (2020). The EU's farm to fork strategy: missing links for transformation [online]. *Acta Innovations*, 3(2020), 17–30. <https://doi.org/10.32933/ActaInnovations.36.2>.

NAGY, J., & JÁMBOR, Z. (2019). Competitiveness in Dairy Trade – the Case of EU and the Visegrad Group Countries [online]. *AGRIS on-line Papers in Economics and Informatics*, 11(4), 61–74. ISSN 1804-1930. <https://doi.org/7160/aol.2019.110406>.

NÁGLOVÁ, Z., & RUDINSKAYA, T. (2021). Factors Influencing Technical Efficiency in the EU Dairy Farms [online]. *Agriculture*, 11(11), 1–14. <https://doi.org/10.3390/agriculture11111114>.

NOVAKOVIĆ, T., NOVAKOVIĆ, D., MILIĆ, D., SIMIN, M. T., RADIŠIĆ, M., RADIŠIĆ, M., NIKOLIĆ, S., & MIHAJLOVIĆ, M. (2025). Assessment of Technical and Eco-Efficiency of Dairy Farms in the Republic of Serbia: Towards the Implementation of a Circular Economy [online]. *Agriculture*, 15(8), 899. < https://doi.org/10.3390/agriculture15080899>.

OECD (2008). Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD PUBLICATIONS, France, 2008, 158 s. ISBN 978-92-64-04345-9.

PARZONKO, A., PARZONKO, A. J., & BÓRAWSKI, P. (2024). Economic Competitiveness of Dairy Farms from the Top Milk-Producing Countries in the EU: Assessment in 2014–2021 [online]. *Agriculture*, 14(1), 123. <https://doi.org/10.3390/agriculture14010123>.

PENKER, M. 2024. Emerging policies and contradictions in the EU: A fair, healthy and environmentally friendly food system by 2030 [online]. *J. Rural Probl.*, 60(2024), 41–48. <https://doi.org/10.7310/arfe.60.41>.

PIAO, S. R., LI, J., & TING, CH. J. (2019). Assessing regional environmental efficiency in China with distinguishing weak and strong disposability of undesirable outputs [online]. *Journal of Cleaner Production,* 227(1), 748–759. <https://doi.org/10.1016/j.jclepro.2019.04.207>.

PIETRZAK, M., DOMAGAŁA, J., & CHLEBICKA, A. (2023). Competitiveness, productivity, and efficiency of farmer' cooperatives: evidence from the Polish dairy industry [online]. *Problems of Agricultural Economics,* 376(3), 1–25. <https://doi.org/10.30858/zer/170891>.

RYBACZEWSKA-BŁAZEJOWSKA, M., & MASTERNAK-JANUS, A. (2018). Eco-efficiency assessment of Polish regions: joint application of life cycle assessment and data envelopment analysis [online]. *Journal of Cleaner Production,* 172(2018), 1180–1192. <https://doi.org/10.1016/j.jclepro.2017.10.204>.

SCHALTEGGER, S., & STURM, A. (1990). Ökologische Rationalität: Ansatzpunkte zur Ausgestaltung von ökologieorientierten Managementinstrumenten. Die Unternehmung, 44(4), 273–290.

SEIFORD, M. L., & ZHU, J. (2002). Modeling undesirable factors in efficiency evaluation [online]. *European Journal of Operational Research*, 142(1), 16–20. <https://doi.org/10.1016/S0377-2217(01)00293-4>.

SIMO, D., MURA, L., & BULECA, J. (2016). Assessment of milk production competitiveness of the Slovak Republic within the EU-27 countries [online]. *Agric. Econ. - Czech*, 62(10), 482–492. <https://doi.org/10.17221/270/2015-AGRICECON>.

SONG, M., QINGXIAN, A., ZHANG, W., WANG, Z., & WU, J. (2012). Environmental efficiency evaluation based on data envelopment analysis: A review [online]. *Renewable and Sustainable Energy Reviews*, 16(7), 4465–4469. <https://doi.org/10.1016/j.rser.2012.04.052>.

SOTERIADES, A.D., FOSKOLOS, A., STYLES, D., & GIBBONS, J.M. (2020). Maintaining production while reducing local and global environmental emissions in dairy farming [online]. *Journal of Environmental Management*, 272(2020), 111054. <https://doi.org/10.1016/j.jenvman.2020.111054>.

STANÍČKOVÁ, M., & MALECKÝ, L. (2012). Assessment of efficiency in Visegrad countries and regions using DEA models [online]. *Central European review of Economic*, 15(2012), 145–156. <https://doi.org/10.7327/cerei.2012.09.02>.

SVATOŠ, M., SMUTKA, L., ELSHIBANI, B. A. M., & MOUSBAH, S. A. (2013). Development of Visegrad countries' agricultural trade in relation to agricultural production development [online]. *Agris on-line Papers in Economics and Informatics*, 5(1), 1–11. <https://doi.org/10.22004/ag.econ.148104>.

SZÁZADVÉG (2022). Challenges and outlooks – Future of the Hungarian dairy sector. [online]. Századvég. [cit. 12.6.2025]. <https://szazadveg.hu/en/cikkek/challenges-and-outlooks-future-of-the-hungarian-dairy-sector>.

ŠPIČKA, J., & SMUTKA, L. (2014). The technical efficiency of specialised milk farms: a regional view [online]. *The Scientific World Journal*, *2014*, 985149, 1–13. <https://doi.org/10.1155/2014/985149>.

ŠPIČKA, J. (2015). The Efficiency Improvement of Central European Corporate Milk Processors in 2008 – 2013 [online]. *AGRIS on-line Papers in Economics and Informatics*, 7(4), 175–188, ISSN 1804–1930.

ŠPIČKA, J., & MACHEK, O. (2015). Change in the production efficiency of European specialized milk farming [online]. *Agric.Econ – Czech*, 61(1), 1–13. <https://doi.org/10.17221/112/2014-AGRICECON>.

THORSØE, M., NOE, E., MAYE, D., VIGANI, M., KIRWAN, J., CHISWELL, H., GRIVINS, M., ADAMSONE-FISKOVICA, A., TISENKOPFS, T., & TSAKALOU, E. (2020). Responding to Change: Farming System Resilience in a Liberalized and Volatile European Dairy Market [online]. *Land Use Policy*, 99(2020), 105029. <https://doi.org/10.1016/j.landusepol.2020.105029>.

UN.ESCAP (2009). Eco-efficiency indicators: measuring resource-use efficiency and the impact of economic activities on the environment. Greening of economic growth series. [online]. Bangkok: UN.ESCAP. [cit. 12.6.2025]. <https://sustainabledevelopment.un.org/co ntent/documents/785eco.pdf>.

URDIALES, M., P., LANSIK, A., O., & WALL, A. (2016). Eco-efficiency Among Dairy Farmers: The Importance of Socio-economic Characteristics and Farmer Attitudes [online]. *Environ Resource Econ*, 64(2016), 559–579. <https://doi.org/10.1007/s10640-015-9885-1>.

ZHU, L., K. SCHNEIDER, & A. OUDE LANSINK. 2023. Economic, environmental, and social inefficiency assessment of Dutch dairy farms based on the dynamic by-production model [online]. *Eur. J. Oper. Res.*, 311(2023), 1134–1145. <https://doi.org/10.1016/j.ejor.2023.05.032>.

ŽÁKOVÁ KROUPOVÁ, Z., HÁLOVÁ, P., & RUMÁNKOVÁ, L. (2020). Productivity of Czech Milk Production in European Comparison [online]. *Agris on-line Papers in Economics and Informatics,* 12(3), 115–127. <https://doi.org/10.7160/aol.2020.120310>.

Predicting Young Bovine Slaughter Numbers Using Statistical Modelling

Nikolina Rizanovska[1] | *Statistical Office of the Republic of Slovenia, Ljubljana, Slovenia*
Aleš Stele | *Statistical Office of the Republic of Slovenia, Ljubljana, Slovenia*
Andreja Smukavec | *Statistical Office of the Republic of Slovenia, Ljubljana, Slovenia*

Received 14.7.2025 (revisions received 1.11.2025, 2.2.2026), **Accepted** (reviewed) 12.2.2026, **Published** 12.6.2026

Abstract

The Statistical Office of the Republic of Slovenia (SURS) developed a predictive model to estimate the intended slaughter or breeding of young bovine animals using administrative data from the *Central Register of Bovine Animals (CRB)*. A binomial regression model with a logit link was employed to forecast slaughter rates, replacing the traditional, resource-intensive survey-based approach. Internal bootstrap validation and external calibration confirmed the model's robustness, ensuring that predictions align with real-world occurrences and are suitable for future forecasting. The model demonstrated a significant improvement in predictive accuracy, with a difference of around 2% between the model's estimates and the survey results, equating to approximately 3 000 animals per year. The model is now closely aligned with observed values, demonstrating that administrative data can effectively replace costly telephone surveys. This shift promises both cost savings and enhanced accuracy in official agricultural statistics, with potential for broader application in other agricultural sectors or regions.

Keywords	***DOI***	***JEL code***
Central Register of Bovine Animals, logistic regression, prediction, calibration, validation	*https://doi.org/10.54694/stat.2025.37*	*CQ18, C25, C53*

INTRODUCTION

The European Union (EU) has a longstanding commitment to agricultural statistics, with regulations dating back to 2004. Building on this foundation, the EU adopted the *Statistics on agricultural input and output (SAIO)* regulation, which aims to provide reliable, comparable, and timely data to support the *Common Agricultural Policy (CAP)* as well as other EU policies related to agriculture, food security, and rural development.

Regulation (EU) 2022/2379, adopted on 23 November 2022 and applicable from 1 January 2025, establishes a unified framework for collecting and reporting statistics on agricultural inputs and outputs

[1] Methodology section, Statistical Office of the Republic of Slovenia, Litostrojska 54, 1000 Ljubljana, Slovenia. E-mail: nikolina.rizanovska@gov.si. ORCID: <https://orcid.org/0009-0007-0540-8223>.

across the European Union. The latest requirements for livestock statistics, discussed in this article, are closely linked to Commission Implementing Regulation (EU) 2023/2745, adopted on 8 December 2023. This regulation imposes rules for implementing Regulation (EU) 2022/2379, specifically concerning animal production statistics, and establishes clear deadlines for reporting livestock statistics. Slovenia is one of the countries obligated to report livestock data annually, with the reference date set as 1 December. According to the reporting schedule, the deadline for submitting the forecast of bovine animals is 15 May of the following year.

Regarding bovine animals, previous versions of the regulation required Member States to conduct livestock surveys. The new Regulation (EU) 2022/2379 introduces the possibility for significant methodological changes, aimed at increasing flexibility and reducing the administrative burden on respondents, national statistical institutes (NSIs), and other national authorities. In addition to statistical surveys, Member States are now allowed to use a variety of data sources and methods, including administrative records, and scientifically sound techniques such as imputation, estimation, and modelling. Nevertheless, the quality of the resulting statistics – particularly regarding accuracy, timeliness, and comparability – must be rigorously maintained.

The main bovine categories relevant to these analyses include bovine animals less than 1 year old (hereafter referred to as young bovines). This category is further disaggregated based on their intended purpose (i.e., for slaughter or not for slaughter). To determine the number of young bovine animals in each category, the previous methodology at the Statistical Office of Slovenia (SURS) – prior to the introduction of the model prediction – relied on the December livestock survey.

The general objective of this project was to develop a model to predict the number of slaughtered young bovines using administrative data from the CRB, thereby replacing the previous survey-based data acquisition method. Our goal was to build a parsimonious model that is likely to perform well on new datasets, particularly those to be used annually for forecasting slaughter numbers. To achieve this, we developed a predictive model based on the approach suggested by Hosmer et al. (2013), purposely selecting covariates by combining statistical methods, experience, and common sense. A detailed explanation of the model follows later.

1 LITERATURE SURVEY

1.1 December livestock survey vs Central Register of Bovine Animals (CRB)

In Slovenia, the *December Livestock Survey (KME-DEC)* is conducted annually using 1 December as the reference date. A sample is drawn from all farmers who own bovine animals on that date. These selected farmers are then contacted to report the status of their animals, including their final intentions, specifically, how many animals they plan to *slaughter* and how many they intend to *retain for breeding (not for slaughter)*.

This survey-based approach has advantages and disadvantages. One notable advantage is the ability to exercise full control over the collected variables. However, the survey incurs logistical and financial costs, including expenses related to employing and training enumerators. There is also a risk of miscommunication or misinterpretation between enumerators and respondents, which could affect the consistency and quality of the collected data.

On the other hand, *The Central Register of Bovine Animals* in Slovenia is a national database that monitors the entire bovine population in the country. It includes data on individual animals, such as births, imports, movements between farms (with separate records for departures and arrivals), deaths, slaughter, exports to the EU, and animal losses. This data set ensures that detailed information about each animal is always accessible. Animal owners are required by regulations to report data either directly through online applications or through a network of institutes.

According to the SAIO legislation, information on the purpose of breeding for each cohort group, specifically whether the animals are intended for slaughter or further breeding, must be reported in advance, no later than 15 May. This deadline is 165 days after the reference date and 200 days before the final classification of young bovines.

Since approximately 15 days are required for data processing and publication, we considered the status of the cohort group on the agreed date of 30 April. On that date, approximately 50% of the young bovines required prediction, while the rest were already classified. A prediction model is developed to classify animals not yet categorized as of 30 April each year.

2 METHODS

2.1 Overview of the data sets and timeline of events

In this section, the timeline of events is explained in detail. The study covers the entire population of young bovines in Slovenia, thus including both female and male bovines under one year of age, registered on Slovenian agricultural holdings as of the reference date. These animals are grouped into mutually exclusive cohorts, each corresponding to a specific annual reference date. For this project, data from two cohorts, registered on 1 December 2022 (N = 149 262) and 1 December 2023 (N = 144 655) were analysed. The model is trained using the 2022 cohort and subsequently applied to generate predictions for the 2023 cohort. In general, model calibration for each reference year relies on data from the immediately preceding year.

Cohort 2022 and Cohort 2023 were monitored longitudinally over a 365-day period from their respective reference dates, with the status of each animal recorded at two time points: 150 days and 365 days after the reference date. At the first time point (150 days after the reference date), approximately 50% of the animals had been classified, and by the second time point (365 days after the reference date), classification of all animals was complete.

As illustrated in Table 1, at the 150-day mark, animals within each cohort were classified into one of five categories. Those that had reached an age greater than one year by this time were classified as *not slaughtered*, indicating they had continued breeding (2022: 43.70%; 2023: 44.40%). Animals slaughtered during this period were classified as *slaughtered* (2022: 5.08%; 2023: 4.29%), whilst the animals that died or were exported were categorised as *losses* (2022: 1.12%; 2023: 1.19%) and *exports* (2022: 0.10%; 2023: 0.04%), respectively. Finally, animals still under one year of age and not slaughtered, dead, or exported

Table 1 Classification of young bovine animals 150 days after the reference date

		1 Dec 2022 (N = 149 262)				1 Dec 2023 (N = 144 655)			
		N	%	N	%	n	%	N	%
Classified				74 646	50.00			71 709	49.57
	Not slaughtered	65 232	43.70			63 713	44.40		
	Slaughtered	7 586	5.08			6,208	4.29		
	Losses	1 673	1.12			1,723	1.19		
	Exports	155	0.10			65	0.04		
Not classified				74 616	50.0			72 946	50.43

Source: Authors' estimations

were classified as *unclassified* (2022: 50.00%; 2023: 50.43%). At this point, about half of the animals had been assigned to one of the four definitive outcome categories (these being slaughtered, not for slaughter, losses, or exported), while others remained unclassified.

The primary objective of this project is to classify the animals that remained unclassified at the 150-day time point. In our example for the year 2023, 50.43% of the animals remained unclassified. Additionally, for the same year, two other groups present classification challenges: young bovines that died (1.19%) or were exported (0.04%) before the 150-day time point. These animals must also be classified. However, the absence of a system to monitor the status of exported young bovines presents a challenge. In practice, the same model will be applied to both the defined groups (i.e., exported or lost) and the unclassified animals. This approach involves predicting how many of these animals would have been slaughtered under a counterfactual scenario in which those that died or were exported had remained alive and present in Slovenia. The predictions will consider each animal's age at the time monitoring ceased, that is the date of death or export.

Table 2 presents the status of the cohort at the second time point, 365 days after the reference date, when all animals within each cohort were fully classified into one of four categories. Animals that had reached an age exceeding one year were classified as *not slaughtered* (2022: 91.90%; 2023: 92.93%;), indicating continued breeding. Those that were slaughtered were classified as *slaughtered* (2022: 6.62%, 2023: 5.60%), while animals that died or were exported were categorised as *losses* (2022: 1.35%, 2023: 1.40%) and *exports* (2022: 0.13%, 2023: 0.07%), respectively. At this stage, the exact number of animals in each category was determined, resulting in the complete classification of all individuals within the cohort. At the time of developing the model, the status of Cohort 2023 was not known in advance and had to be predicted. It is now known, and we are presenting the results.

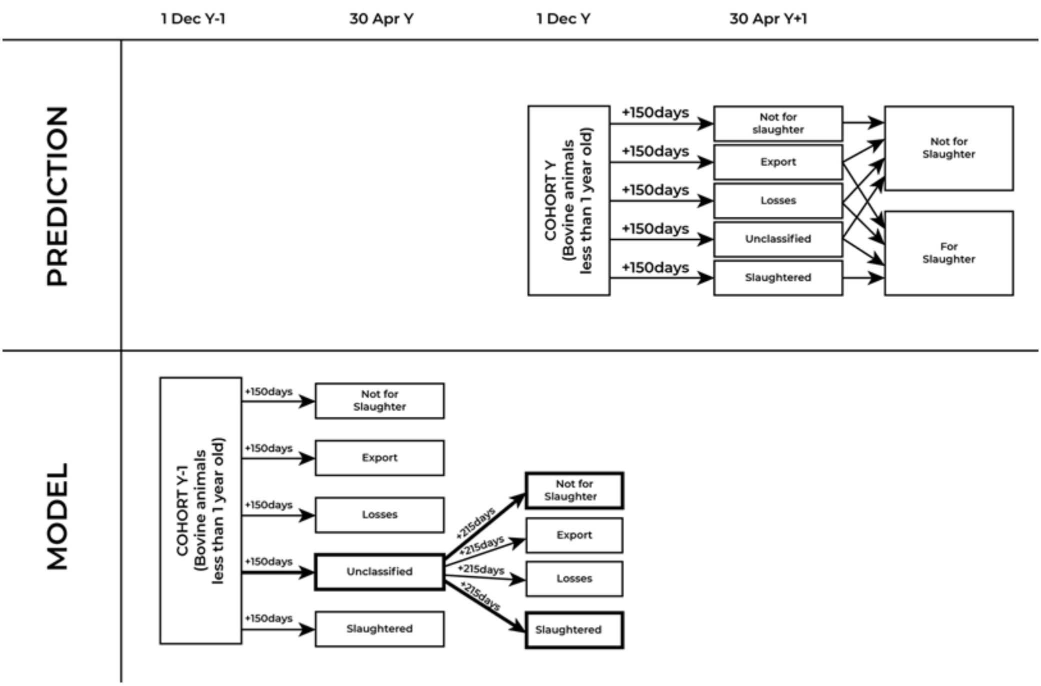
Table 2 Classification of young bovine animals 365 days after the reference date

		1 Dec 2022 (N = 149 262)				1 Dec 2023 (N = 144 655)			
		N	%	N	%	n	%	N	%
Classified				149 262	100.00%			144 655	100.00%
	Not slaughtered	137 165	91.90			134 433	92.93		
	Slaughtered	9 882	6.62			8 092	5.60		
	Losses	2 019	1.35			2 026	1.40		
	Exports	196	0.13			104	0.07		
Not classified				0.0%				0.0%	

Source: Authors' estimations

To estimate the coefficients of our model and predict the number of slaughtered and non-slaughtered animals, we trained the model using subgroup of animals from the previous year. The subgroup is visually represented and highlighted at the lower part of Figure 1. As an example, consider the cohort group with a reference date of 1 December 2022. To be included in the training data, bovine animals had to meet the following conditions: they had to be less than one year old on that date, registered with Slovenian agricultural holdings as of 1 December 2022, remain unclassified regarding slaughter status until 30 April 2023, and be definitively classified as either slaughtered or not slaughtered by 1 December 2023. This final classification served as the dependent variable (y), where a value of 1 was assigned to *slaughtered* animals and 0 to those *not slaughtered*.

Figure 1 Timeline of young bovines' categorization



Source: Authors' visualisation

2.2 Building the model

In this section, we outline the methodology used to select the variables that produce the "best" predictive model and describe how we assessed its adequacy, both for individual variables and overall performance. Our primary goal is to develop a model using a pseudo-out-of-sample evaluation, which will be used annually to estimate the probability of being slaughtered for the observed population. Estimating the effects of influencing factors (animal characteristics) is of secondary importance. The model was developed using *R Statistical Software* (R Core Team, 2025) with which the *Least Absolute Shrinkage and Selection Operator (LASSO)* method was applied by calling the glmnet package (Friedman et al., 2025), whilst the validation and calibration were conducted using the rms package (Harrell, 2025).

When we began building the model, we considered basic descriptive statistics about the animals, including their *age* on the reference date, *sex* (male = 1, female = 0), *region* (western = 1, eastern = 0), and *breed*. Our focus was on the four most common breeds in Slovenia: simmental (breed1), holstein (breed2), brown (breed3), and beef (breed4). All other 517 breeds were grouped under the category "other" (breed5). As Hosmer et al. (2013) noted, it is inappropriate to model a nominal variable as if it were an interval variable. To address this, we created a set of design variables to represent the different breed categories. Specifically, we introduced four dichotomous variables, each comparing one breed (i.e., holstein, brown, beef, and other) to the reference group, i.e., simmental, the most common and frequent breed found in Slovenia.

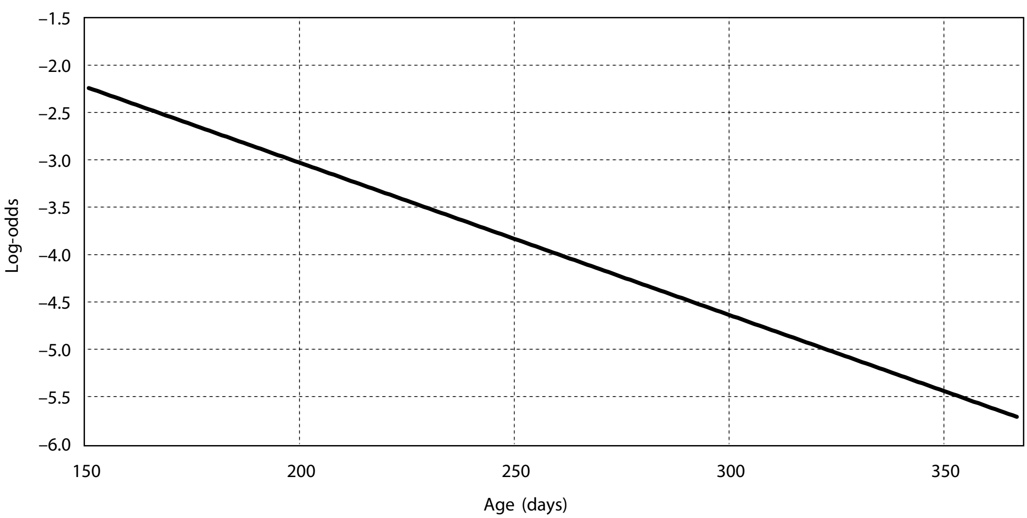
Additionally, we had information about the farm on which each animal was raised. Based on this data, a content specialist recommended creating a new variable that could be important for our predictions: the *type of farm*. To classify the farm type, we checked on which farm each animal was located as of 1 December 2022. We considered two key factors: the total number and the number of milking cows on the farm. If the proportion of milking cows exceeded 30%, the farm was classified as a dairy farm; otherwise, it was categorised as a beef farm (dairy = 0, beef = 1).

The dataset used to build the model consisted of data from 74 616 cattle, of which 2 296 had been slaughtered. Since the response variable is binary, the appropriate method for modelling the relationship is binomial logistic regression using logit link. Coefficients were estimated using maximum likelihood estimation. The covariates are as follows: *age* ranges from 150 to 365 days ($\bar{x}$ = 263.3, SD = 63.78), where *sex* is nearly equally distributed, with 51.13% female and 48.87% male. For *type of farm*, 45.59% are dairy and 54.41% are beef. *Region* distribution is 70.70% eastern and 29.30% western. *Breed* includes 40.65% simmental, 18.70% holstein, 3.60% brown, 15.48% beef, and 21.57% other.

With only five covariates and a large sample size, we followed Hosmer et al. (2013) recommendation to begin by fitting an initial model that included all covariates. The Wald test for all coefficients was significant ($p < 0.001$). Thus, our preliminary main effect model ("model_initial") contains eight covariates: *age*, *breed2*, *breed3*, *breed4*, *breed5*, *sex*, *type_farm*, *region*.

Furthermore, we assessed the assumption that the model is linear in the logit for the continuous variable age (days). To check this, we used a smoothed scatterplot, as suggested by Crawley (2012), with LOWESS smoothing providing a nonparametric description of the relationship between the logit (log-odds) and age. Visual inspection of the Figure 2 indicated that the logit decreases linearly as a function of age, confirming a clear linear relationship.

Figure 2 Lowess smooth on the log – odds scale of outcome slaughter, versus the covariate AGE, n = 74 229



Source: Authors' visualisation

The next step in the purposeful selection procedure is to explore possible interactions among the main effects. We fit models that individually added each of the 22 possible interactions to the main effect model. At this stage, the main effects are considered fixed and could not be removed. The results are summarized in Table 3.

We identified 15 interaction terms that were statistically significant at the p=0.10 level. To obtain a parsimonious model, we applied the Least Absolute Shrinkage and Selection Operator (LASSO) (Friedman et al., 2010) in R, further reducing the number of significant interaction terms identified in the previous step. Predictor variables included main effects and selected two-way interactions constructed

Table 3 Log-Likelihood, Likelihood Ratio (G, df = 1), and p-value for the addition of the interaction to the main effects model

Interaction	Log-likelihood	G	p
Main Effects Model			
age*breed2	−8,449.170	99.699	< 0.001*
age*breed3	−8,498.710	0.609	0.435
age*breed4	−8,497.195	3.651	0.056*
age*breed5	−8,495.108	7.823	0.005*
age*sex	−8,470.572	56.896	< .001*
age*type_farm	−8,474.738	48.565	< .001*
age*region	−8,498.923	0.194	0.659
sex*breed2	−8,154.567	688.905	< 0.001*
sex*breed3	−8,494.297	9.447	0.002*
sex*breed4	−8,434.204	129.632	< 0.001*
sex*breed5	−8,421.576	154.889	< 0.001*
sex*type_farm	−8,192.901	612.238	< 0.001*
sex*region	−8,494.017	10.007	0.002*
type_farm*breed2	−8,498.003	2.033	0.153
type_farm*breed3	−8,498.830	0.380	0.537
type_farm*breed4	−8,498.515	1.010	0.314
type_farm*breed5	−8,497.218	3.604	0.058*
type_farm*region	−8,479.428	39.184	< 0.001*
region*breed2	−8,495.943	6.154	0.013*
region*breed3	−8,497.846	2.347	0.126
region*breed4	−8,497.787	0.465	0.495
region*breed5	−8,497.028	3.983	0.046*

Source: Authors' estimations

with *model.matrix*. The penalty parameter (λ) was chosen by 10-fold cross-validation using *cv.glmnet*. We selected λ_min as the optimal value. All predictors were standardised before estimation. As stated by Kuhn and Johnson (2013), the practical implications of using this technique are significant. While the regression coefficients are shrunk towards 0, a consequence of penalising the absolute values is that some parameters are actually set to 0 for some values of *λ*. Thus, the LASSO yields models that simultaneously use regularisation to improve the model and to conduct feature selection.

The results of the LASSO regularisation revealed that 7 interactions had coefficients of zero, suggesting minimal contribution to the model. Interactions *age*breed2, age*breed4, age*breed5, age*sex, age*type_farm, type_farm*breed5, region*breed5* were therefore excluded from the final model. One interaction *sex*breed2,* had a coefficient of 0.19, suggesting a moderate contribution, and was retained in the model. The remaining 7 interactions showed small contributions, with coefficients ranging from 0.01 to 0.1. After consulting with a content expert and based on these coefficients, three additional interactions *sex*breed3, sex*type_farm,* and *region*type_farm* were included in the initial model.

The final model consists of one continuous variable, *age*, along with seven categorical dummy variables and four significant interactions: *sex*breed2, sex*breed3, sex*type_farm,* and *region*type_farm*. According to Hosmer et al. (2013), when an interaction term is statistically significant, but the corresponding main effect is not, both the main effect and the interaction should be retained in the model to ensure accurate calculation of the odds ratio. Therefore, the selected model includes a nonsignificant covariate (*breed3*).

The four-degree-of-freedom likelihood ratio test comparing the interactions model to the model consisting only of the main effects (eight covariates) resulted in a G value of 863.71 with p < 0.001, indicating that the model with interactions statistically reduces the deviance. Thus, in aggregate, the interactions contribute significantly to the model. The final model is presented in the Table 4.

Table 4 Results of fitting the Final multivariable model fit, n = 74,229

Variable	Coeff.	Std. Err.	*Z*	p	95% CI		OR
(Intercept)	–1.847	0.137	–13.480	< 0.001	–2.118	–1.581	0.158
age	–0.015	0.001	–34.069	< 0.001	–0.016	–0.014	0.985
breed2	0.437	0.141	3.101	< 0.001	0.159	0,711	1.548
breed3	–0.066	0.330	–0.199	.842	–0.778	0.529	0.937
breed4	0.688	0.090	7.674	< 0.001	0.511	0.863	1.989
breed5	1.177	0.071	16.613	< 0.001	1.038	1.316	3.243
sex	0.885	0.106	8.386	< 0.001	0.681	1.095	2.424
type_farm	1.267	0.111	11.448	< 0.001	1.052	1.487	3.550
Region	0.658	0.066	10.018	< 0.001	0.529	0.786	1.930
sex*breed2	2.300	0.147	15.631	< 0.001	2.013	2.590	9.972
sex*breed3	1.509	0.353	4.271	< 0.001	0.860	2.259	4.521
sex*type_farm	–1.140	0.119	–9.624	< 0.001	–1,374	–0,910	0.320
Region*type_farm	–0.636	0.095	–6.698	< 0.001	–0,823	–0,451	0.529

Source: Authors' estimations; Log-Likelihood = -7 903.203

The model is formally described by Formula 1.

$$\ln \frac{\pi}{1-\pi} = -1.847 - 0.015 * age + 0.437 * bred2 - 0.066 * bred3 + 0.688 * bred4 + 1.177 * bred5 + 0.885 * sex + 1.267 * type\_farm + 0.658 * region + 2.30 * sex * bred2 + 1.509 * sex * bred3 - 1.140 * sex * type\_farm - 0.636 * region * type\_farm \quad (1)$$

To better understand the results, we have presented the odds ratios for the significant groups.

Table 5 Estimated odds ratios and 95% Confidence interval for certain groups in the study CRB, n = 74,229

Variable	Subgroup	Odds ratio	95% CI	
age		0.638	0.618	0.651
breed4		1.989	1.667	2.370
breed5		3.243	2.825	3.730
male/female	breed2	24.067	16.960	34.630
male/female	breed3	10.957	5.330	22.53
male/female	dairy	2.423	1.970	2.980
male/female	beef	0.775	0.568	1.058
western/eastern	dairy	1.931	1.698	2.198
western/eastern	beef	3.624	2.650	4.970

Source: Authors' estimations

The estimated odds ratio for a positive difference of one month (30 days) is 0.638. This indicates that for each additional 30-day period, the odds of slaughter decrease by a factor of 0.638, representing a reduction in the odds of slaughter that could range from a 0.618-fold to a 0.651-fold decrease.

The analysis also reveals several key insights regarding the odds of slaughter across different *breeds* and *type of farms*. Firstly, the odds of slaughter for animals from the *beef breed* (*breed4*) are nearly twice as high – 1 989 times greater (95% CI: 1.667, 2.370) – compared to all *other breeds* in the study population. Similarly, the *other breed* (*breed5*), which consists of 517 less frequently occurring breeds, shows even higher odds of slaughter, with the odds being 3 243 times greater (95% CI: 2 825, 3 730) than those for the prevalent breed groups: *simmental, holstein, brown, and beef*.

When considering sex differences within specific breeds, male bovines from the *holstein breed* have 24 167 times higher odds of slaughter than females of the same breed (95% CI: 16 960, 34 630). A similar pattern is observed for the *brown breed*, where male bovines have 10 957 times greater odds of slaughter than their female counterparts (95% CI: 5 330, 22 530).

The *type of farm* also plays a critical role. On farms where more than 30% of cows are used for milking (classified as dairy farms), male bovines have 2 423 times greater odds of slaughter than females (95% CI: 1 970, 2 980). In contrast, on farms where fewer than 30% of cows are used for milking (classified as *beef farms*), male bovines face 0.775 times the odds of slaughter compared to females (95% CI: 0.568, 1.058).

Geography further influences these odds. Among *dairy farms* in Western Slovenia, the odds of slaughter are 1 931 times higher than those observed on *dairy farms* in Eastern Slovenia (95% CI: 1 698, 2 198). However, on *beef farms*, the odds of slaughter for animals in Western Slovenia are 3 624 times greater than those in Eastern Slovenia (95% CI: 2 650, 4 970). The confidence intervals for the interactions were estimated using the four-step method described by Hosmer et al. (2013).

These findings highlight significant differences in slaughter odds based on breed, sex, type of farm, and geographical location, with respect to the other variables in the model.

2.3 Assessment of model fit

To assess the accuracy and reliability of the model, we conducted both internal validation and external calibration. According to Hanley and McNeil (1982), there are two common methods for model validation: cross-validation and bootstrap validation, with the latter requiring fewer assumptions. For internal validation, we used bootstrap validation, in which samples were drawn with replacement from the original dataset, simulating the process of sampling from an underlying population. These bootstrap samples were the same size as the original dataset. As recommended by Steyerberg (2019), 100–200 bootstrap iterations are typically sufficient for stable estimates. In our analysis, we performed 200 bootstrap iterations to ensure robust performance estimates.

For each bootstrap sample, a prediction model was developed. The model was evaluated both within the bootstrap sample itself and on the original dataset. The difference in performance between the two reflects the "optimism" of the model – that is, how much better it performs on the bootstrap sample compared to the original data. This optimism is then subtracted from the apparent performance in the original sample to obtain a more accurate estimate of model fit (see Table 6).

Steyerberg (2019) claims that the advantages of bootstrap validation are numerous with the principal benefit being that the optimism-corrected performance estimate tends to be stable, as the same sample size is used for both model development and testing. This makes bootstrap validation a reliable method for assessing model performance in predictive modelling.

Model performance was assessed by calculating two of the many diagnostic measures suggested by Harrell (2015): Nagelkerke's coefficient of determination (R^2) and the Area Under the Receiver Operating Characteristic Curve (AUC). Both metrics range from 0 to 1, with higher values indicating better model fit and predictive accuracy. The model achieved a Nagelkerke's R^2 of 0.252. According to Field et al. (2012), Nagelkerke's R^2 values between 0.2 and 0.4 are considered acceptable.

Table 6 Bootstrap validation of model performance, as indicated by Nagelkerke's R2 in a subsample of the CRP dataset (n = 74,229)

	Original sample	Training sample	Test sample	Optimism	Corrected index	N
Dxy	0.680	0.680	0.679	0.001	0.678	200
R^2	0.253	0.254	0.253	0.001	0.252	200

Source: Authors' estimations

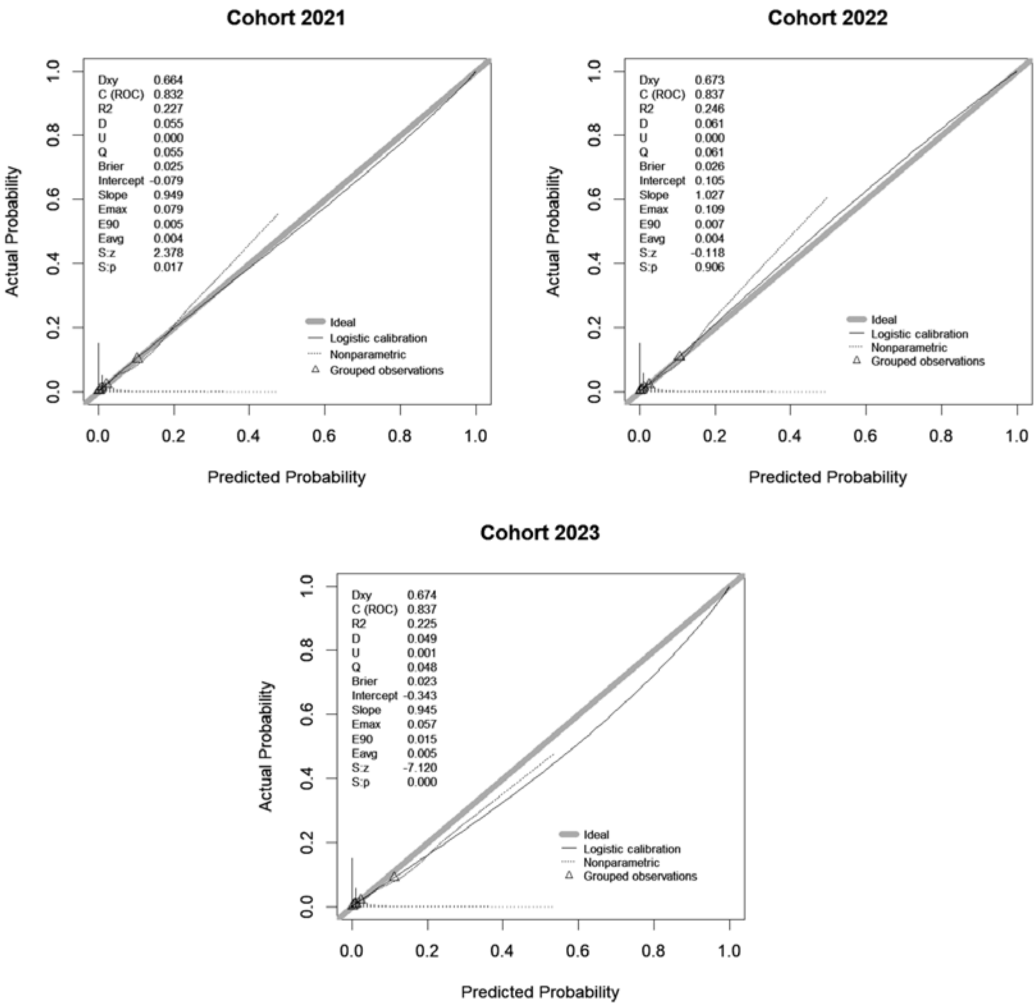
Regarding the second validation parameter (AUC; computed as $AUC = \frac{Dxy+1}{2}$), Hanley and McNeil (1982) provide basic guidelines for interpreting the AUC. Values between 0.8 and 0.9 suggest that the model has acceptable performance for many practical purposes. The model in our analysis achieved a validation AUC of 0.839.

Model calibration was evaluated using historical data, utilising the set from the immediately preceding year. Harrell (2015) stated that calibration refers to how well the predicted probabilities match the actual observed outcomes, reflecting the model's reliability in producing accurate predictions. To visually assess this calibration, we constructed three calibration plots for each dataset (see Figure 3).

In each plot, the predicted probabilities are plotted along the x-axis, while the observed outcomes are shown on the y-axis. The line of identity (the 45° diagonal line) serves as a reference point: perfectly calibrated predictions will fall along this line, indicating that the predicted probabilities align precisely with the actual outcomes.

The evaluation across three datasets – *Cohort2021*, *Cohort2022*, and *Cohort2023* –demonstrates consistent discriminatory ability in all models. Key metrics such as Dxy and C (ROC) provide clear evidence of the models' effectiveness in distinguishing between positive and negative outcomes. The Dxy values, ranging from 0.664 to 0.674, reflect a strong capacity that outcomes are correctly ranked, while a consistent range of C (ROC) values from 0.832 to 0.837 confirms the models' solid discriminatory power.

Figure 3 Calibration plot of actual outcome vs. predictions for three models, 2021 (n = 77,327), 2022 (n = 74,616), 2023 (n = 72,970)



Source: Authors' visualisation

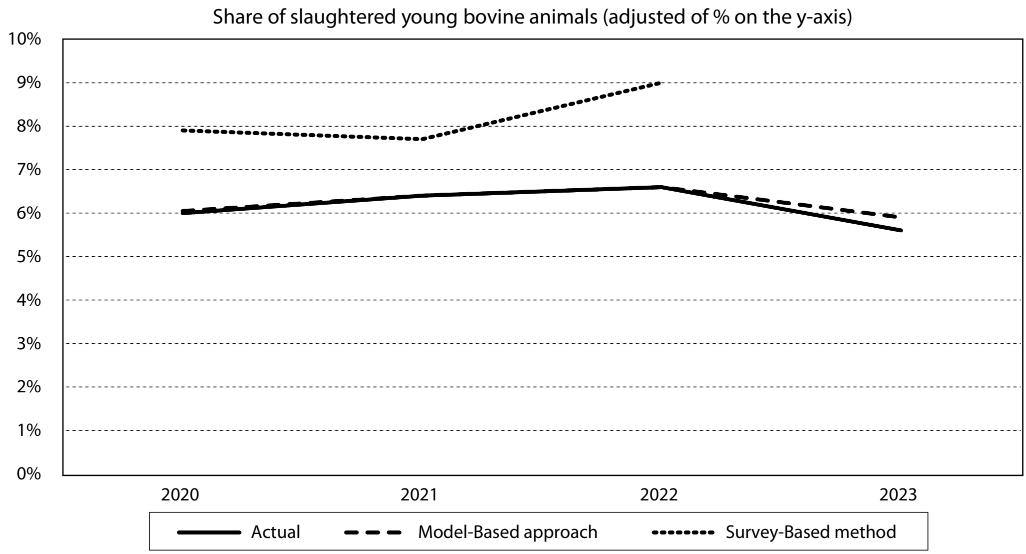
Regarding model calibration, the results demonstrate promising consistency in predictive performance. The models exhibit a reasonable fit, as indicated by moderate R^2 values ranging from 0.225 to 0.246. These values suggest that the models capture a meaningful portion of the data's variability, although they do not explain all of it. The relatively low residual deviance (D) values, between 0.049 and 0.061, suggest minimal unexplained error, supporting the conclusion that the models accurately reflect the underlying patterns in the data.

The Brier score values in this analysis range from 0.023 to 0.026, indicating strong predictive accuracy. The Brier score (Brier, 1950) ranges from 0 to 2, where 0 represents perfect predictions and 2 corresponds with the worst possible outcome. In practice, values typically fall between 0 and 1, with lower scores indicating better accuracy. These results suggest that the models are well-calibrated and perform effectively in probabilistic predictions. Furthermore, calibration slopes were consistently close to 1.0, which affirms that the predicted probabilities align closely with observed frequencies. This further reinforces the conclusion that the models are well-calibrated and their predictions reliably reflect real-world outcomes.

3 RESULTS

This chapter presents the practical perspective, comparing the estimated shares of slaughtered young bovines using the previous survey-based method with our new model-based approach.

Figure 4 Comparison of the share of slaughtered young bovines, computed from each year's cohort: survey-based method vs. model-based approach



Source: Authors' estimations

Figure 4 presents the actual and predicted shares of slaughtered young bovines by year, together with estimates from the previous survey-based method. According to data reported to Eurostat using the former method, the shares were 7.90% in 2020, 7.70% in 2021, and 9.00% in 2022. In contrast, the shares predicted by our model – 6.05% in 2020, 6.40% in 2021, and 6.60% in 2022 – are notably closer to the actual observed values: 6.00% in 2020, 6.40% in 2021, and 6.60% in 2022. This close alignment between predicted and actual shares highlights the model's strong predictive performance and its improvement over the previous methodology. Compared to the earlier approach, the difference between the survey results and the model's predictions is about 2%, or approximately 3 000 animals per year.

The graph also shows that in 2023, the predicted percentage was higher, likely because the model could not account for the unexpected decrease in the trend. Nevertheless, the difference between the observed and predicted percentages is minor. Since the survey was not conducted in 2023, there is no data point for that year on the graph.

4 DISCUSSION

The results indicate that the model demonstrates solid and consistent performance, exhibiting a strong ability to discriminate between outcomes, being suitably calibrated and showing low residual errors across all datasets.

From a practical perspective, the new model-based approach provides predicted slaughter shares for young bovines that closely align with the actual observed values. We have achieved our goal and have begun implementing the model in practice, as it outperforms the previous survey-based method reported to Eurostat and meets the expectations of our content specialists.

The model performs with high precision when external factors – such as weather conditions (droughts, floods, or extreme temperatures), crop yields, and disease or health factors – remain stable. However, when these external variables change, they can affect the percentage of slaughtered animals. For instance, it is often argued that if crop yields decrease, the number of slaughtered animals may increase, as farmers may have less feed available for their livestock. This concept was explored in a model developed by Breimyer (1952), where crop yields were treated as an external variable, making a significant contribution to the model's predictions.

The price of meat or livestock can significantly affect the number of animals slaughtered. Higher prices for meat may encourage producers to slaughter more animals to capitalise on profitable markets, while lower prices may reduce slaughter numbers. Many authors have analysed the factors that influence cattle prices (Gallardo et al., 2010; Goodwin, 1992; Kulshrestatha and Rosaasen, 1980; Weimar and Stillman, 1990; Tonsor and Schroeder, 2011; Zapata and Garcia, 1990) and have asserted in their research that potential additional elements to be incorporated in the model encompass supply and demand dynamics, production costs, weather and environmental conditions, market factors, economic conditions, government policies and trade, disease and health factors, technological improvements, and the influence of global markets and trade.

Although we are very pleased with the results, alternative approaches could be explored to simplify the methodology. Several researchers have forecasted meat production using time series approaches (Aujla and Sadiq, 2018; Duwalage et al., 2023; Gritsenko et al., 2023; Odru and Zengin, 2020), based on which a monthly time series of slaughtered animals could be created. This method would help mitigate seasonality effects (if present) and capture one-time events, such as disease outbreaks or droughts. While time series analysis is useful for generating estimates, it has limitations, as it does not account for the demographic characteristics of the animals or the influence of external factors like market fluctuations, weather patterns, or policy changes.

One promising approach is random forest prediction, a machine learning ensemble technique that does not rely on strict assumptions. For example, Ordu and Zengin (2020) proposed an interesting methodology for predicting animal production by comparing three forecasting approaches: Autoregressive integrated moving average (ARIMA), exponential smoothing (ES) and the function of the seasonal and trend decomposition using loess (STLF).

The percentage of slaughtered animals has been steadily increasing recently; however, in the coming years, content experts anticipate a decrease in the number of slaughtered animals. In this context, monitoring the model's performance will be necessary, as we will need to assess how accurately it predicts these shifting trends, and, if necessary, add more explanatory variables.

CONCLUSION

In this study, we demonstrated that the number of slaughtered animals can be predicted more accurately using administrative data. The findings indicate that employing detailed bovine data with statistical modelling can lead to more effective predictions and improved production outcomes. The methodology developed is directly applicable to the *Statistical Office of the Republic of Slovenia*. In conclusion, the model exhibited strong predictive performance, with a high AUC, a moderate Nagelkerke's R^2, and robust calibration diagnostics, reflecting both excellent predictive accuracy and reliable calibration.

References

AUJLA, K. M., SADIQ, N. (2018). Supply estimate and projection for poultry meat in Pakistan by the year 2030 AD [online]. *Pure and Applied Biology*, 7(2): 460–465. <http://dx.doi.org/10.19045/bspab.2018.70057>.

BREIMYER, H. F. (1952). Forecasting Annual Cattle Slaughter [online]. *Journal of Farm Economics*, 34(3): 392–398. <https://doi.org/10.2307/1233697>.

BRIER, G. W. (1950). Verification of forecasts expressed in terms of probabilities. *Monthly Weather Review, 78*(1): 1–3.

CRAWLEY, M. J. (2012). *The R Book*. 2nd Ed. Hoboken: John Wiley and Sons, Inc.

DUWALAGE, K. I. et al. (2023). Predicting *Carcass* Weight of Grass-Fed Beef Cattle before Slaughter Using Statistical Modelling [online]. *Animals*, 13(12): 1968. <https://doi.org/10.3390/ani13121968>.

FIELD, A., MILES, J., FIELD, Z. (2012). *Discovering Statistics Using R*. Los Angeles: Sage.

FRIEDMAN, J. H., HASTIE, T., TIBSHIRANI, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent [online]. Journal of Statistical Software, 33(1): 1–22. <https://doi.org/10.18637/jss.v033.i01>

FRIEDMAN, J. H., HASTIE, T., TIBSHIRANI, R. (2025). *glmnet: Lasso and elastic-net regularized generalized linear models* [software]. < https://CRAN.R-project.org/package=glmnet >.

GALLARDO, R. K., BRORSEN, B. W., LUSK, J. (2010). Prediction markets: an experimental approach to forecasting cattle on feed [online]. *Agricultural Finance Review*, 70(3): 414–426. <https://doi.org/10.1108/00021461011088521>.

GOODWIN, B. K. (1992). Forecasting Cattle Prices In The Presence Of Structural Change [online]. *Southern Journal of Agricultural Economics*, 24(2): 11–22. <https://doi.org/10.22004/ag.econ.29632>.

GRITSENKO, S. et al. (2023). On-Barn Forecasting Beef Cattle Production Based on Automated Non-Contact Body Measurement System [online]. *Animals*, 13(4): 611. <https://doi.org/10.3390/ani13040611>.

HANLEY, J. A., MCNEIL, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve [online]. *Radiology*, 143(1): 29–36. <https://doi.org/10.1148/radiology.143.1.7063747>.

HARRELL JR., F. E. (2015). *Regression Modeling Strategies: With Applications to Linear Models and Ordinal Regression, and Survival Analysis*. 2nd Ed. Cham: Springer.

HARRELL JR, F. E. (2025). *rms: Regression modeling strategies* [software]. <https://cran.r-project.org/package=rms>.

HOSMER JR., D. W., Lemeshow, S., Sturdivant, R. X. (2013). *Applied Logistic Regression.* 3rd Ed. Hoboken: John Wiley and Sons, Inc.

KUHN, M., JOHNSON, K. (2013). *Applied Predictive Modeling*. New York: Springer.

KULSHRESTATHA, S. N., ROSAASEN, K. A. (1980). A monthly price forecasting Model for cattle and calves [online]. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie*, 28(2): 41–62. <https://doi.org/10.1111/j.1744-7976.1980.tb01941.x>.

ORDU, M., ZENGIN, Y. (2020). A Comparative Forecasting Approach to Forecast Animal Production: A Case of Turkey [online]. *Livestock Studies*, 60(1): 24–31. <https://doi.org/10.46897/lahaed.719095>.

R CORE TEAM. (2025). *R: A language and environment for statistical computing* [software]. *Vienna: R Foundation for Statistical Computing.* <https://www.R-project.org>.

STEYERBERG, E. W. (2019). *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. 2nd Ed. Cham: Springer.

TONSOR, G. T., SCHROEDER, T. C. (2011). Multivariate forecasting of a commodity portfolio: application to cattle feeding margins and risk [online]. *Applied Economics*, 43(11): 1329–1339. <https://doi.org/10.1080/00036840802600517>.

WEIMAR, M. R., STILLMAN, R. P. (1990). A Long Term Forecasting Model of the Livestock and Poultry Sectors [online]. *Proceedings of the NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management*, *Chicago*, *IL,* 217–231. <https://doi.org/10.22004/ag.econ.285537>.

ZAPATA, H. O., GARCIA, P. (1990). Price Forecasting with Time-Series Methods and Nonstationary Data: An Application to Monthly U.S. Cattle Prices [online]. *Western Journal of Agricultural Economics*, 15(1): 123–132. <https://doi.org/10.22004/ag.econ.32505>.

What is the Relationship between University's Financial Resources and Student Perception of Institutional Attractiveness?

Hana Flusková[1] | *Prague University of Economics and Business, Prague, Czechia*
Karel Šafr[2] | *Prague University of Economics and Business, Prague, Czechia*

Received 26.8.2025, **Accepted** (reviewed) 18.11.2025, **Published** 12.6.2026

Abstract

We examine how pay and staffing conditions relate to the attractiveness of public-university faculties in Czechia. Using harmonised faculty-level administrative data for 2017–2024, we fit a covariance-based structural equation model (CB-SEM) with four latent constructs: structural conditions, research drive (including a time trend), wages and student interest. Student interest is positively associated with structural conditions and with research drive. The relationship between wages and student interest is not statistically significant once the other constructs are included. Research drive is also related to wages and structural conditions. Taken together, the results suggest that teaching capacity and research visibility are the main factors associated with faculties' attractiveness, while wage policy appears, if at all, related through those channels rather than directly. These insights can inform discussions of public funding formulas and institutional staffing strategies by emphasising teaching capacity and research visibility rather than undifferentiated wage increases.

Keywords	***DOI***	***JEL code***
CB-SEM, higher education institutions, institutional attractiveness, administrative data, academic staff workload, faculty-level funding	*https://doi.org/10.54694/stat.2025.44*	*I22, I23, J24, C39*

[1] Department of Economic Statistics, Faculty of Informatics and Statistics, Prague University of Economics and Business, W. Churchill Sq. 1938/4, 130 67 Prague 3, Czechia. Corresponding author: e-mail: hana.fluskova@vse.cz, phone: (+420)607964615. ORCID: <https://orcid.org/0009-0008-8675-6242>.

[2] Department of Statistics and Probability, Faculty of Informatics and Statistics, Prague University of Economics and Business, W. Churchill Sq. 1938/4, 130 67 Prague 3, Czechia. E-mail: karel.safr@vse.cz. The author is also working at the Czech Statistical Office, Na Padesátém 81, 100 82 Prague 10, Czech Republic.

INTRODUCTION

This article investigates how the economic and personnel conditions of academic staff relate to the attractiveness of faculties of public higher education institutions (HEI) in Czechia. While many international studies address the effects of institutional wealth, governance, or reputation on enrolment patterns, data and funding systems vary widely by country; even within the EU, which compiles statistics across member states, methods rarely transfer. For this reason, our analysis stays within the Czech higher education and relies on faculty-level administrative data for 2017–2024 that support comparable analyses across faculties. We also brought national evidence on university non-completion, showing that institutional support services, the study environment and time or workload pressures are linked to students' decisions to leave or stay. To study these links, we use the covariance-based structural equation modelling (CB-SEM), which evaluates both the indicators for each latent construct and their mutual interlink, including mediation.

1 LITERATURE SURVEY

The relationship between financial resources and institutional performance has been the subject of substantial scholarly attention. Avenali et al. (2024) examine 318 HEIs across five countries and identify total operating expenses and the net value of long-term physical capital as the strongest predictors of success in Shanghai ARWU rankings. Their machine learning approach (XGBoost) and complementary data envelopment analysis (DEA) demonstrate that financial investment efficiency contributes significantly to institutional reputation. Similarly, Mkhitaryan and Želvys (2025) show that top-ranked European universities are more likely to rely on negotiated funding models, in contrast to formula-based mechanisms, suggesting that flexible funding approaches align with high performance. Lassila (2011), using OLS regression on a large U.S. dataset, finds that higher revenue is modestly associated with a larger share of low-income students, but chiefly in private non-profits. This suggests that wealthier institutions have a competitive advantage in supporting financial aid, thereby influencing access. Bulman (2022) further shows that increases in endowment wealth at private U.S. colleges are associated with greater financial aid spending without lowering headline or net tuition. Overall prices stay flat, while selectivity rises and the share of Pell-eligible and underrepresented students drops slightly. Consistently, credit-rating evidence shows that student selectivity (higher SAT/ACT, lower admit rates) and resource strength map closely onto stronger HEI credit ratings, especially among private institutions, underscoring how reputation and resources reinforce each other (Gottesman and Ismailescu, 2020).

Beyond financial capacity, studies have also explored how universities convert resources into educational or reputational outcomes. Dorius et al. (2017) focus on return on investment (ROI) and find that graduation rates dominate; once controls enter, most spending categories – including instruction – yield weak or no effects, and prestige proxies are non-significant. Al-Hosaini et al. (2023) apply a Balanced Scorecard framework and PLS-SEM to show that non-financial institutional strengths, such as internal processes and customer perspective, directly contribute to financial performance in private Yemeni universities. Won and Chelladurai (2016) use SEM to show that intangible resources (reputation, culture) indirectly improve performance in U.S. athletic departments by enabling the accumulation of tangible resources. Similarly, Nafisi et al. (2017), using survey data and SEM, find that quality in educational services, social interaction, and the academic environment all positively influence goal attainment, and also emphasise the role of financing in support of these effects (modelled as direct relations rather than a formal moderation test).

The second body of literature addresses factors that influence the level of how attractive students perceive higher education institutions to be. Yadav et al. (2022), in a PLS-SEM study of Indian private HEIs, find that attractiveness is caused by teaching quality, branding, and research activity. This attractiveness, in turn, enhances institutional sustainability, mediated partly by student belongingness. Nguyen (2023), based on Vietnamese data, similarly highlights curricula and facilities as major determinants of university choice, with PLS-SEM results confirming their predictive value. Cingillioglu et al. (2022) focus on digital engagement, finding that Facebook shares and comments predict university preference among Australian students, and that this engagement interacts with institutional status indicators like top eight Australian universities membership (Go8) and global ranking. Foster (2024) investigates private HEIs in Jakarta and reveals that segmentation and market positioning strategies are key to attracting students, especially in urban areas.

Attractiveness also intersects with broader structural factors shaping student access. Hervás et al. (2020) show that social perceptions, including perceived employability, and individual/institutional factors, such as geographic accessibility, drive student demand for public universities in Spain. Giambona et al. (2017) employ a spatial modelling approach to show that students prefer territories with better living standards, academic services, and labour market prospects. Wörner (2011) uses Gini-type concentration measures to show that a small set of prestigious Chilean universities captures a disproportionate share of high--performing students, evidencing reputational sorting in student inflows. These findings resonate with broader concerns about stratification. For instance, Lassila (2011) and Bulman (2022) provide evidence that financial aid can mitigate, but not eliminate, enrolment disparities for low-income students. Bulman especially highlights how increases in financial resources are more likely to benefit higher-income applicants, reinforcing exclusivity. Giambona et al. (2017) confirm that socio-economic conditions of the region significantly influence student migration in Italy, with more affluent regions attracting more students.

These dynamics often reproduce social inequalities unless explicitly corrected. Dorius et al. (2017) show that institutional ROI is positively associated with the share of STEM degrees and negatively associated with private status and female student share, indicating deep demographic inequalities in outcomes. These studies demonstrate that financial and institutional characteristics often compound rather than reduce inequalities in access and returns. The growth of the use of structural equation modelling (SEM) and partial least squares SEM (PLS-SEM) across contexts underscores its utility in modelling these complex relationships. Yadav et al. (2022), Nguyen (2023) and Foster (2024) apply PLS-SEM to student perception and university attractiveness, highlighting indirect and mediating effects. Won and Chelladurai (2016), Nafisi et al. (2017), and Angulo-Ruiz and Pergelova (2013) use SEM to explore institutional commitment, service quality, and competitive advantage.

Huang (2021) applies PLS-SEM in a blended learning context and finds that perceived usefulness and motivation significantly predict learning satisfaction. Hervás et al. (2020) and Cingillioglu et al. (2022) use SEM for modelling student decision-making and digital engagement. This literature collectively demonstrates SEM's capacity to capture complex interdependencies relevant to university growth, offering a robust analytical framework for future research on how financial resources, attractiveness, and enrolment interact.

A complementary stream of Czech research has focused on the reasons for non-completion of university studies. A study commissioned by the Ministry of Education (Vlk et al., 2017) focuses on synthesising evidence on three salient groups of causes (academic preparedness, socio-economic disadvantage, and motivational/personal factors) while also highlighting institutional shortcomings (e.g., academic support, feedback, staff–student connection). Among the most frequently cited causes are weak academic performance, mismatched expectations about the study field, and a lack of time due to working duties. These issues are especially pronounced in certain fields, most notably technical disciplines, and are often associated with workload pressures (e.g., non-university work duties), as well as institutional support gaps. Institutional shortcomings, such as poor academic support, inadequate feedback, and a lack of connection with teaching staff, also play a significant role. This indicates that study failure in Czechia is not solely an individual issue, but reflects structural conditions that may undermine long-term institutional attractiveness.

In summary, prior studies indicate that financial resources, institutional characteristics and territorial context jointly shape university performance and student demand. Studies on endowments, revenues and funding models document how wealth and flexible funding reinforce institutional reputation and selectivity, often in ways that reproduce social inequalities rather than mitigating them. A second strand highlights how perceived quality, branding, facilities and digital engagement contribute to institutional attractiveness, again interacting with broader socio-economic and spatial factors. At the same time, the growing use of SEM and PLS-SEM demonstrates the value of modelling indirect, mediating and contextual effects in this domain. Czech research on non-completion further indicates that structural and institutional conditions are central to student success, but it does not explicitly connect detailed staffing, financial and research conditions to student interest within a single causal framework. Existing evidence suggests that comparable studies combining SEM with faculty-level data from Czech public universities have not yet been established in the literature.

2 METHODS AND DATA

2.1 Data Sources

The analysis is based on harmonised administrative data from multiple official sources provided by the Ministry of Education, Youth and Sports (MEYS) and the Government Office of the Czech Republic (GOV Office). All variables relate to public higher education institutions (HEIs) at the faculty level in Czechia between 2017 and 2024. State HEIs, such as the University of Defence, and private HEIs were excluded due to different rules, methods and funding sources, and limited data availability. We also excluded university-wide units that offer study programmes independently of faculty structures, except non-university colleges that are not divided into faculties. These units either provide programmes outside specific faculties (e.g., they are part of non-faculty institutes) or belong to institutional types that do not implement a faculty structure. Data from university-wide units were excluded due to administrative and data collection inconsistencies across the public HEIs in Czechia. After data cleaning and preparation, we worked with a data sample of 1 233 observations out of 157 faculties of public HEIs, with no missing values. All variables were normalised before calculations.

The first dataset, obtained from MEYS's Integrated Student Information System (SIMS), includes student enrolment data such as recalculated enrolment data, numbers of newly admitted students, total student counts, numbers of graduates, and study-programme information at the faculty level. Each study programme is assigned a coefficient of economic demands (CED). The average faculty-level CED is the ratio of the normative number of student enrolments to the recalculated number of student enrolments at a given faculty. The normative number is calculated as a weighted average of the recalculated number of student enrolments across the faculty's study programmes, with weights given by the coefficients of economic demands (CEDs) assigned to each study programme.

$$normative\,number\,of\;student\,enrolments = recalculated\;number\;of\;student\,enrolments * CED \qquad (1)$$

The second dataset, also from MEYS, relates to higher education funding and includes data about academic, research and other staff and their wage resources. Data on research staff were excluded, as they lie outside the scope of this study. The category of other employees was also excluded from the analysis due to inconsistencies in their classification across faculties or university-wide units within the structure of individual public universities. The dataset includes information on average monthly salaries, total wage expenditures and numbers of academic staff in full-time equivalents disaggregated by rank: research, development and innovation (RDI) teaching staff, professors, associate professors, assistant professors, assistants and lecturers. The reported values are aggregated regardless of the source of funding (MEYS including supplementary activities, EU structural funds and other sources). In practice, an employee's position may be financed through various sources. For example, a single full-time equivalent (FTE) position may consist of 0.1 full-time position at the department, 0.4 full-time position on a National Recovery Plan (NRP) project and 0.5 full-time position on a Grant Agency of the Czech Science Foundation (GACR) grant. Compared to the enrolment dataset, this personnel dataset covers significantly more units within public HEIs, such as research institutes, university-wide units, centres and publishing houses, because all of these parts of public HEIs employ staff, even though not all of them enrol students in study programmes outside the faculty structure.

Institutional location was incorporated via the binary variable "metropolis", which categorises HEIs as metropolitan (Prague or Brno) or regional. Data were obtained from MEYS's Register of Higher Education Institutions and Study Programs Offered (REGVSSP).

To capture academic hierarchy and roles within faculties, academic staff were grouped into two composite categories. The first group includes professors and associate professors, who are typically involved in administrative responsibilities related to teaching such as guaranteeing courses or study programmes, or chairing final examination committees. The second group includes all other academic ranks (RDI staff, assistant professors, assistants, lecturers), who might have fewer administrative roles and can focus more on teaching.

We also incorporated data on completed grants from the Research, Development and Innovation Information System (IS VaVaI) provided by the GOV Office. We treat the number of completed grants as a proxy for faculty-level research quality at the Czech public HEIs. For this research, we considered two types of project-funding providers: the Czech Science Foundation (GACR), the Technology Agency of the Czech Republic (TACR). The data was filtered as "completed grants" and consists of successfully completed and discontinued (prematurely terminated) multi-year projects. Discontinued projects represent only a negligible proportion (under 1%) of the considered grants during the period under review. Both variables are strong indicators of research prestige as they are highly competitive and strictly peer-reviewed funding schemes, the selection process is very strict and failure to complete the grant could put the research institute at high reputational risk.

Table 1 Descriptive statistics and key information about the variables

Variable	Abbrev.	Unit	Mean	Median	SD
Average monthly wages of academic staff	avg_m_w_AcadS	CZK	58 207.42	53 357.62	29 603.93
Average monthly wages of professors and associate professors	avg_m_w_P_AP	CZK	82 872.19	68 961.92	152 992.95
Average monthly wages of other academic staff	avg_m_w_other	CZK	48 733.06	46 202.03	20 290.52
Number of projects provided by the Czech Science Foundation	GACR	–	1.37	0.00	3.97
Number of projects provided by the Technology Agency of the Czech Republic	TACR	–	0.75	0.00	2.01
Average number of lecturers	avg_amount_lecturers	FTE	5.17	0.87	10.29
Normative number of student enrolments	normative	–	2 828.12	2 437.57	2 057.44
Number of graduates	grad	–	370.14	279.00	314.12
Total student counts	total_stud	–	1 808.83	1 452.00	1 455.65
Number of newly admitted students	new_stud	–	323.27	271.00	241.97
Year	year	–	–	–	–
Metropolis	metropolis	0/1 type	–	–	–

Source: Own construction, MEYS, GOV Office

2.2 Methodology

2.2.1 Model construction

We specified the model deductively, starting from a theory-led view of how Czech faculties convert resources into reputation, teaching capacity, and student demand. Four latent constructs were defined and measured from harmonised administrative indicators for 2017–2024: structural conditions of teaching, research drive (including a time trend), wages (economic conditions of academic staff), and student interest (attractiveness/retention). This mapping reflects the idea that students respond to a bundle of capacity, location and perceived quality, while institutions' financial and reputational positions feed back into staffing and wages. Evidence that financial and reputational strength shape student markets and selection comes from multiple contexts: endowment growth and prestige dynamics in the U.S., where wealthier institutions spend more and become more selective, shifting composition rather than scale (Bulman, 2022; Gottesman and Ismailescu, 2020), territorial context and services shaping attractiveness in Italy (Giambona et al., 2017), and prestige-driven sorting of high-performing students in Chile (Wörner 2011).

Because staff pay can affect recruitment/retention and role mix yet trade off with headcount, we model wages and structural conditions separately. Finally, given wide structural changes over 2017–2024, the research-drive construct includes the calendar year to absorb shared period dynamics while GACR/TACR grants proxy faculty-level research intensity.

The CB-SEM measurement model is assessed using standardized factor loadings and SEM-consistent diagnostics. Convergent validity is evaluated based on the magnitude and statistical significance of the loadings alongside the Average Variance Extracted (AVE). Internal consistency is estimated via composite reliability indices; Cronbach's alpha is excluded as its assumptions are incompatible with the heterogeneous administrative indicators employed. Discriminant validity is examined using the Fornell-Larcker criterion, complemented by the Heterotrait-Monotrait ratio (HTMT) for sensitivity analysis. Finally, reliability metrics for the "research drive" construct are interpreted strictly as diagnostic evidence rather than as absolute pass-fail thresholds since we work with time series.

2.2.2 Variables

The empirical model incorporates four latent constructs. All latent constructs were selected to investigate the relationship between research, financial and structural conditions of academic staff and students' willingness to enrol, continue and graduate from the faculty of public HEI in Czechia.

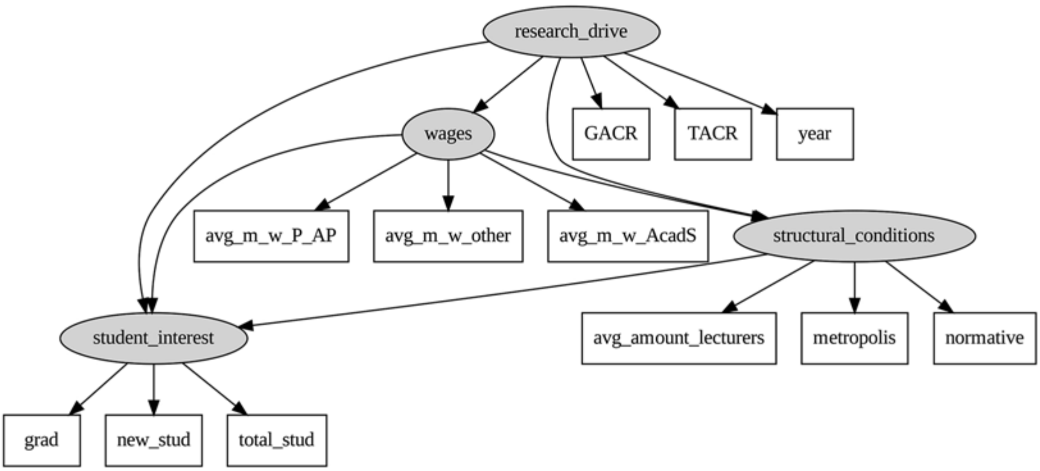
The first latent variable, "structural conditions", might be considered as a proxy of the faculty size and location. The variable shows the size of the faculty through the number of lecturers in full-time equivalents as a primarily teaching academic employee (avg_amount_lecturers) and the normative number of student enrolments as the amount of public funding a faculty receives per student, adjusted for programme cost and type (normative). These two variables are supposed to show the number of teaching personnel and the money the faculty has to provide education. We also include a binary location indicator (metropolis) to distinguish faculties of HEIs in two largest cities in Czechia (Prague and Brno) from regional faculties as a proxy of students' willingness to move to bigger cities, for example, to increase future job opportunities.

The second latent variable, "research drive", should represent the prestige of the university in terms of research quality in Czechia and the trend in data. The GACR and TACR variables represent the number of completed grants per faculty of HEIs as an indicator of high-quality and prestigious research. The variable year helps us show the trend in our data.

"Wages" capture the economic conditions facing academic personnel and the ability of faculties to attract and retain staff. "Wages" consists of various variables referring to weighted average monthly salaries: academic staff in total (avg_m_w_AcadS), professors and associate professors (avg_m_w_P_AP) and all other academic ranks, such as RDI staff, assistant professors, assistants, lecturers (avg_m_w_other).

The last latent variable "student interest" is captured through newly enrolled students (new_stud), the total number of students (total_stud) and the number of graduates (grad). "Student interest" is a proxy of the faculty attractiveness and should provide answers to questions such as the ability of the faculty to attract and keep students and give them the appropriate conditions to successfully graduate from the faculty.

Figure 1 Proposed model structure



Source: Own construction

Based on the literature, Czech public data structures, and our construct definitions, we posit the following directional expectations. Student interest should be positively associated with structural conditions and with research drive, which includes the time trend. The direct effect of wages on student interest is expected to be small and positive, and may be non-significant. Research drive is expected to be positively associated with wages. By contrast, the effects of research drive on structural conditions and of wages on structural conditions are theoretically ambiguous: research-intensive units may either expand staffing and infrastructure or maintain lean teaching footprints, and higher salaries can facilitate recruitment yet crowd out headcount. Because the research drive combines competitive grant activity (GACR/TACR) with the time trend (year), these relations should be interpreted as the joint influence of research intensity and system dynamics.

2.2.3 Used statistical methods

Covariance-based structural equation modelling (CB-SEM) is the classic SEM framework developed the 1960s and formalised in the social sciences in 1975 by Duncan (1975). This method is mainly connected with software solutions due to the complexity of this model. One of the first popular solutions was by Karl Jöreskog (LISREL software – Jöreskog and Van Thillo, 1972). In our paper, we are using the widely applied SEMOPY (Igolkina and Meshcheryakov, 2020) package in Python, which outperforms other available software in terms of estimation quality.

CB-SEM estimates model parameters so that the model-implied covariance matrix reproduces the observed covariance matrix as closely as possible, enabling rigorous tests of both the measurement and structural parts of a theory. Model adequacy is judged with well-established global fit indices – e.g., the Comparative Fit Index (CFI), Tucker–Lewis Index (TLI), and the RMSEA – which collectively evaluate how well the overall theoretical structure captures the data's dependence patterns. This tradition underpins mainstream SEM texts and software and remains the dominant approach for theory testing in the social sciences. (Bentler, 1990; Igolkina and Meshcheryakov, 2020).

Why CB-SEM is a strong choice for our study (vs. PLS-SEM), when the goal is to evaluate a new, theory-driven model – including its latent constructs, cross-construct paths, and the global plausibility of the entire system – CB-SEM is preferable because it provides unbiased, consistent parameter estimates under correct specification, formal global fit testing and nested-model comparisons, and rich diagnostics for the measurement model.

In contrast, variance-based PLS-SEM is optimised for prediction, handles small samples and complex composites well, but historically lacked consistent estimation for reflective constructs (addressed by PLSc) and does not emphasise global theory fit in the same way; leading PLS guidance itself recommends CB-SEM when the research objective is theory confirmation rather than prediction. Given our reflective constructs and confirmatory aims, CB-SEM is the appropriate, widely recommended choice for creating new theoretical models and testing the theories behind them (Hair et al., 2022).

3. RESULTS

Based on the CB-SEM output, student interest is significantly and positively related to structural conditions ($\beta = 1.44$, SE = 0.16, z = 9.13, $p < .001$) and to research drive ($\beta = 2.27$, SE = 0.63, z = 3.58, $p < .001$). In contrast, the direct association between wages and student interest is null ($\beta = -0.01$, SE = 0.03, z = –0.31, p = .76). Upstream, research drive is significantly (negatively) associated with structural conditions ($\beta = -2.58$, SE = 0.41, z = –6.29, $p < .001$) and with wages ($\beta = -0.69$, SE = 0.18, z = –3.76, $p < .001$). In contrast, wages do not predict structural conditions ($\beta = -0.01$, SE = 0.03, z = –0.24, p = .81). Taken together, the only statistically meaningful correlates of student interest in this specification are structural capacity and research drive, with wages operating – if at all – indirectly and without a detectable direct path to demand.

Table 2 Statistics summary of the structural model

Right latent variable		Left latent variable	Estimate	Std. error	z-value	P-value
Structural conditions	->	Student interest	1.44	0.16	9.13	0.00
Research drive	->	Student interest	2.27	0.63	3.58	0.00
Wages	->	Student interest	–0.01	0.03	–0.31	0.76
Research drive	->	Structural conditions	–2.58	0.41	–6.29	0.00
Wages	->	Structural conditions	–0.01	0.03	–0.24	0.81
Research drive	->	Wages	–0.69	0.18	–3.76	0.00

Source: Own construction

Also, our overall statistics suggest that our model is statistically acceptable and the results are aligned with our theory. Assessing global model fit, six indices point to acceptable overall performance, we also want to emphasise approximate and incremental indices given the large N.

The χ^2 test is significant – $\chi^2(47) = 613.66$, $p < .001$ – with $\chi^2/df = 13.06$; because χ^2 is highly sensitive to large samples, we rely more on incremental indices (which is expected with large samples; we therefore do not rely on the normed χ^2/df ratio; Kline, 2023). The CFI = 0.945 (or 94.5% better than baseline model) and TLI = 0.923 exceed the conventional .90 threshold (with .95 often cited as "very good"), indicating the model reproduces the covariance structure reasonably well. NFI = 0.941 is consistent with this verdict. The RMSEA = 0.097 sits at the boundary of what is typically regarded as mediocre-to-acceptable (extreme close fit ≤ .05; reasonable .05–.08; mediocre .08–.10; poor-fit hypothesis> .10; Kline, 2023), suggesting limited data (variables) rather than failure of the model. RMSEA sits near the boundary (0.10) where many guidelines move from "reasonable" to "problematic," but Kline cautions that such thresholds are only heuristics. Finally, AIC = 61.00 and BIC = 219.64 help compare rival specifications (lower is better) but have no absolute cut-off.

Overall, this pattern – incremental fit in the mid 90s paired with acceptable RMSEA in a large-N panel – supports the adequacy of the proposed CB-SEM for theory testing while leaving scope for minor, theory-guided refinements. The model captures the hypothesised structure (strong positive effects of structural teaching conditions and research drive on student interest, and a null direct wage effect), aligning with your theoretical expectations.

Table 3 Model variables measurements

Measured variable	Latent variable	Estimate	Std. error	z-value	P-value
Normative	L1	1.00	–	–	–
avg_amount_lecturers	L1	0.62	0.03	20.27	0.00
Metropolis	L1	0.19	0.02	11.36	0.00
Year	L2	1.00	–	–	–
GACR	L2	–3.08	0.51	–6.05	0.00
TACR	L2	–0.47	0.16	–3.05	0.00
avg_m_w_other	L3	1.00	–	–	–
avg_m_w_P_AP	L3	0.31	0.03	10.52	0.00
avg_m_w_AcadS	L3	1.06	0.08	13.82	0.00
Grad	O1	1.00	–	–	–
new_stud	O1	0.97	0.02	63.19	0.00
total_stud	O1	1.06	0.01	94.54	0.00

Source: Own construction

All estimated measurement loadings are statistically significant ($p < .001$ or $p < .01$), supporting the convergent validity of all reflective constructs. For structural conditions (L1) we observe positive loadings for avg_amount_lecturers ($\beta = 0.62$, $z = 20.27$) and metropolis ($\beta = 0.19$, $z = 11.36$); for wages (L3) both avg_m_w_P_AP ($\beta = 0.31$, $z = 10.52$) and avg_m_w_AcadS ($\beta = 1.06$, $z = 13.82$) are significant; and for student interest (O1), new_stud ($\beta = 0.97$, $z = 63.19$) and total_stud ($\beta = 1.06$, $z = 94.54$) are very strong indicators. Within research drive (L2), GACR ($\beta = -3.08$, $z = -6.05$) and TACR ($\beta = -0.47$, $z = -3.05$) are also significant. Their negative signs reflect factor orientation relative to the year anchor fixed at 1.00, while the reference indicators (normative, year, avg_m_w_other, grad) were fixed to 1.00 for identification and thus not tested.

We assessed the CB-SEM measurement model using standardized factor loadings and SEM--consistent diagnostics. Convergent validity was evaluated using the magnitude and statistical significance of standardized loadings and the Average Variance Extracted (AVE). Internal consistency was summarized using composite reliability indices derived from standardized loadings and indicator error variances. We do not report Cronbach's alpha because our constructs are measured by a small number of heterogeneous administrative indicators, for which alpha is strongly driven by the number of indicators and relies on assumptions (e.g., tau-equivalence and item homogeneity) that are not well matched to this measurement setting. We therefore prioritize SEM-aligned indices and interpret them as diagnostic evidence rather than strict pass-fail thresholds.

Since our research intentionally combines competitive grant activity with a time trend, high internal consistency is not a conceptual requirement; accordingly, reliability indices are interpreted as diagnostic only.

4 DISCUSSION

Consistent with the a priori expectation, student interest is positively and statistically significantly associated with structural conditions. Faculties with greater teaching capacity, represented by the number of lecturers, higher normative resources per student and metropolitan or regional location, exhibit higher inflow and persistence of students, conditional on the other constructs.

Student interest is statistically significantly associated with research drive, which is formed by the number of completed grants from two Czech grant agencies together with a time trend (2017–2024). Our results indicate that higher research intensity and system improvements over the period move alongside higher student demand and progression. Because the research drive combines grant activity with the time trend, this effect should be read as their joint influence.

In line with the expectation of a small direct effect, wages show no statistically significant direct association with student interest once structural conditions and research drive are accounted for. This is consistent with wages not operating as a demand signal by itself; any indirect role via capacity and reputation remains tentative.

There is a statistically significant association between research drive and wages, indicating that compensation systematically varies with the research/time profile of faculties. Given the composite nature of the research drive, the sign should be interpreted with caution. The result nevertheless indicates a robust linkage between research profile, period dynamics and pay setting.

The association between research drive and structural conditions is statistically significant, indicating that research/time dynamics are systematically related to teaching capacity and metropolitan versus regional positioning of the faculty. The direction observed in our data suggests that research orientation is linked to a distinct organisational footprint, consistent with strategic differentiation over the study period.

Wages show no statistically significant direct association with structural conditions. This aligns with the theoretical ambiguity: higher salaries may facilitate recruitment yet be offset by headcount constraints, yielding no net direct effect on teaching capacity once the research drive is considered.

Taken together, structural conditions and research drive emerge as the principal correlates of student interest, while wages appear consequential mainly through their associations with these constructs. The significant links from research drive to both wages and structural conditions indicate that research and period dynamics shape how faculties remunerate and organise teaching, whereas wages alone do not translate into direct changes in capacity or demand.

Our model adds theoretical clarity by aligning four institution-level constructs (structural conditions, research drive, wages, and student interest) with how Czech faculties convert resources into attractiveness and progression. Treating student interest as a composite of inflow, stock, and graduation is consistent with evidence that graduation rates are a primary driver of institutional returns, whereas prestige proxies are not robust once controls are included (Dorius et al., 2017). Separating structural conditions from wages avoids conflating scale with compensation and allows mediation: pay can shape who is hired, while capacity and place transmit what students experience. In this sense, the chain we specify is consonant with work showing that non-financial organisational strengths (customer perspective, internal processes, learning and growth) can underpin financial performance (Al-Hosaini et al., 2023), and with evidence that intangible resources such as reputation tend to operate indirectly by enabling the accumulation of tangible resources before influencing outcomes (Won and Chelladurai, 2016).

The pattern we observe, strong associations of structural conditions and research drive with student interest, coupled with a weak direct role for wages, is consistent with international findings. In the U.S., endowment shocks are associated with higher spending and increased selectivity, without expanding scale and with slight declines in Pell/URM shares (Bulman, 2022). Financial-market evidence shows that credit ratings co-move with student demand and quality, especially among private universities (Gottesman and Ismailescu, 2020). Territorial amenities, public services and labour-market conditions are associated with where students enrol (Giambona et al., 2017) and prestige mechanisms can concentrate top-performing students within a small set of universities (Wörner, 2011). Read against this backdrop, our results suggest that attractiveness in Czech faculties is primarily associated with teaching capacity and research visibility, while wage policy appears, at most, indirectly related through those channels; the research drive construct should be read as the joint influence of competitive grant activity and system-wide time dynamics between 2017 and 2024.

CONCLUSION

This study advances how Czech faculties of public higher education institutions convert resources into institutional attractiveness by aligning four constructs considering financial, personnel, spatial and research conditions of academic staff within a CB-SEM framework. The pattern of associations indicates that student interest relates most strongly to structural conditions and research drive, while wages matter chiefly indirectly through these channels. Conceptually, the separation of structural conditions from wages clarifies how capacity and place transmit what students experience, whereas compensation primarily conditions who can be hired. Taken together, the results provide a parsimonious structure for interpreting institutional attractiveness as the interplay of teaching capacity, research visibility and staffing economics rather than as a function of any single input.

Our findings are consistent with the existing literature on the roles of capacity, reputation and territorial context in shaping student demand. Two qualifications are important for interpretation: research drive embeds both competitive grant activity and an institutional time trend, so its effects capture joint institutional and period dynamics; and all relationships are read as associations contingent on measurement and context. These features point to natural extensions for future work (e.g., alternative operationalisations of research reputation, explicit policy shocks), but they do not alter the core theoretical implication: in this setting, strengthening teaching capacity and research visibility appears more consequential for institutional attractiveness than direct wage levels alone.

The findings suggest that, for government, ministry-level and institutional stakeholders, institutional attractiveness is most strongly associated with stable teaching capacity, research visibility and student-facing support. For universities and faculties, this highlights the role of strategic staffing and research investments, while wage policy – shaped by institutional autonomy and discipline-specific differences in research funding – appears to influence student interest mainly indirectly through its effects on capacity and reputation.

References

AL-HOSAINI, F., ALI, B. J. A., BAADHEM, A. M, JAWABREH, O., ATTA, A. A. B., ALI, A. (2023). The Impact of the Balanced Scorecard (BSC) Non-Financial Perspectives on the Financial Performance of Private Universities [online]. *Information Sciences Letters*, 12(9). ISSN 2090-9551 (Print), ISSN 2090-956X (Online). <https://www.naturalspublishing.com/Article.asp?ArtcID=27566>.

ANGULO-RUIZ, L. F., PERGELOVA, A. (2013). The Student Retention Puzzle Revisited: the Role of Institutional Image [online]. *Journal of Nonprofit & Public Sector Marketing*, 25(4): 334–353. ISSN 1049-5142. <https://doi.org/10.1080/10495142.2013.830545>.

AVENALI, A., DARAIO, C., DI LEO, S., WOLSZCZAK-DERLACZ, J. (2024). Heterogeneity of national accounting systems, world-class universities and financial resources: What are the links? [online]. *Journal of Informetrics*, 18(2): 101502. ISSN 1751-1577. <https://doi.org/10.1016/j.joi.2024.101502>.

BENTLER, P. M. (1990). Comparative fit indexes in structural models [online]. *Psychological Bulletin*, 107(2): 238–246. ISSN 1939-1455. <https://doi.org/10.1037/0033-2909.107.2.238>.

BULMAN, G. (2022). *The Effect of College and University Endowments on Financial Aid, Admissions, and Student Composition* [online]. Working paper, Cambridge (Mass.). <http://www.nber.org/papers/w30404>.

CINGILLIOGLU, I., GAL, U., PROKHOROV, A. (2022). Facebook Engagement and Student Preferences for Universities [online]. In: *2022 20th International Conference on Information Technology Based Higher Education and Training (ITHET)*, 1–10. [cit. 5.6.2025]. ISSN 2380-1603. <https://doi.org/10.1109/ITHET56107.2022.10032046>.

DORIUS, S., TANDBERG, D., CRAM, B. (2017). Accounting for institutional variation in expected returns to higher education [online]. *Education Policy Analysis Archives*, 25: 110–110. ISSN 1068-2341. <https://doi.org/10.14507/epaa.25.3238>.

DUNCAN, O. D. (1975). *Introduction to Structural Equation Models*. New York: Academic Press. ISBN 978-0-12-224150-5.

FOSTER, B. (2024). Model for increasing interest in enrolling in private universities based on a segmentation, targeting, and market positioning strategy in Indonesia [online]. *International Journal of Advanced and Applied Sciences*, 11(9): 164–172. ISSN 2313-626X. <https://doi.org/10.21833/ijaas.2024.09.018>.

GIAMBONA, F., PORCU, M., SULIS, I. (2017). Students Mobility: Assessing the Determinants of Attractiveness Across Competing Territorial Areas [online]. *Social Indicators Research*, 133(3): 1105–1132. ISSN 1573-0921. <https://doi.org/10.1007/s11205-016-1407-1>.

GOTTESMAN, A., ISMAILESCU, I. (2020). *Student Selectivity and Higher Education Institutions Credit Ratings* [online]. SSRN Scholarly Paper, Feb 17th 2020, Rochester, NY: Social Science Research Network. [cit. 18.8.2025]. <https://doi.org/10.2139/ssrn.3458916>.

HAIR, J., HULT, G. T. M., RINGLE, C., SARSTEDT, M. (2022). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. 3rd Ed. USA: Sage Publishing. ISBN 978-1-5443-9640-8.

HERVÁS, A., SORIANO, P. P., OLMOS, J. G., PERÓ, M., CAPILLA, R., MONTAÑANA, J. M. (2020). Modelling the Process to Access the Spanish Public University System Based on Structural Equation Models [online]. *Mathematical and Computational Applications*, 25(2): 31. ISSN 2297-8747. <https://doi.org/10.3390/mca25020031>.

HUANG, C.-H. (2021). Using PLS-SEM Model to Explore the Influencing Factors of Learning Satisfaction in Blended Learning [online]. *Education Sciences*, 11(5): 249. ISSN 2227-7102. <https://doi.org/10.3390/educsci11050249>.

IGOLKINA, A. A., MESHCHERYAKOV, G. (2020). Semopy: a Python Package for Structural Equation Modeling [online]. *Structural Equation Modeling: a Multidisciplinary Journal*, 27(6): 952–963. ISSN 1070-5511. <https://doi.org/10.1080/10705511.2019.1704289>.

JÖRESKOG, K. G., VAN THILLO, M. (1972). *LISREL: a General Computer Program for Estimating a Linear Structural Equation System Involving Multiple Indicators of Unmeasured Variables* [online]. B.M.: Educational Testing Service, Princeton, N.J. [cit. 20.8.2025]. <https://files.eric.ed.gov/fulltext/ED073122.pdf>.

KLINE, R. B. (2023). *Principles and Practice of Structural Equation Modeling.* 5th Ed. New York: Guilford Press. ISBN 978-1-4625-5191-0.

LASSILA, N. (2011). Effects of Tuition Price, Grant Aid, and Institutional Revenue on Low-Income Student Enrollment [online]. *Journal of Student Financial Aid*, 41(3). ISSN 0884-9153. <https://doi.org/10.55504/0884-9153.1013>.

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY (MŠMT). (2025). *Přepočtené počty studií po fakultách a SP vč. KEN k 31.10.* [online]. [cit. 26.8.2025]. <https://msmt.gov.cz/vzdelavani/vysoke-skolstvi/sdruzene-informace-matrik-studentu-sims>.

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY (MŠMT). (2025). *Registr vysokých škol a uskutečňovaných studijních programů* [online]. [cit. 26.8.2025]. <https://regvssp.msmt.cz/registrvssp/cvslist.aspx>.

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY (MŠMT). (2025). *Statistická ročenka školství – Zaměstnanci a mzdové prostředky* [online]. [cit. 26.8.2025]. <https://msmt.gov.cz/vzdelavani/skolstvi-v-cr/statistika-skolstvi/statistickerocenky-skolstvi-zamestnanci-a-mzdove-prostredky>.

MKHITARYAN, A., ŽELVYS, R. (2025). Exploring the Relationship between Higher Education Financing Mechanisms and University Ranking [online]. *Economics, Finance and Accounting*, 1(15): 5. ISSN 2953-8068. <https://doi.org/10.59503/29538009-2025.1.15-5>.

NAFISI, G., TAEE, H., ARASTEH, H. R. (2017). Investigating the Relationship between the Role of the Quality of the Educational Environment, the Quality of Educational Services and Social Interactions and Achieving Educational Goals by Considering the Role of Financing. *Iranian journal of educational sociology*, 1(5): 78–91.

NGUYEN, T. H. (2023). Factors affecting students' decision to choose a university: a case study of Vietnam National University, Hanoi [online]. *Journal of Economic and Banking Studies*, 6: 52–66. ISSN 2734-9853. <https://doi.org/10.59276/JEBS.2023.12.2546>.

ÚŘAD VLÁDY ČESKÉ REPUBLIKY, RADA PRO VÝZKUM, VÝVOJ A INOVACE. (2025). *Projekty VaVaI – Centrální evidence projektů* [online]. [cit. 26.8.2025]. <https://www.isvavai.cz/cep>.

VLK, A., DRBOHLAV, J., FLIEGL, T., HULÍK, V., STIBUREK, Š., ŠVEC, V. (2017). S*tudijní neúspěšnost na vysokých školách: Teoretická východiska, empirické poznatky a doporučení.* Prague: SLON. ISBN 978-80-7419-248-7.

WON, D., CHELLADURAI, P. (2016). Competitive Advantage in Intercollegiate Athletics: Role of Intangible Resources [online]. *PLOS ONE*, 11(1): e0145782. ISSN 1932-6203. <https://doi.org/10.1371/journal.pone.0145782>.

WÖRNER, C. H. (2011). The pursuit of prestige: the distribution of talented students in Chile's universities [online]. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, 16: 463–476. ISSN 1414-4077, 1982-5765. <https://doi.org/10.1590/S1414-40772011000200012>.

YADAV, R., SHIVA, A., NARULA, S. (2022). Exploring private university attractiveness from students' perspective to ensure sustainable institutes: an empirical investigation from Indian perspective [online]. *Asia-Pacific Journal of Business Administration*, 16(1): 170–203. ISSN 1757-4323. <https://doi.org/10.1108/APJBA-04-2021-0165>.

International Equilibrium: the Stories We Tell

Václav Rybáček[1] | *University of Economics and Management, Prague, Czechia*
Dan Šťastný | *University of Economics and Management, Prague, Czechia*

Received 15.7.2025 (revision received 30.11.2025), **Accepted** (reviewed) 10.1.2026, **Published** 12.6.2026

Abstract

Macroeconomic theory and macroeconomic aggregates are a sine qua non for the economic policy decision-making. It is of vital importance not only that macroeconomics aggregates are measured as correctly as possible but also that their content constitutes an appropriate input into the macroeconomic models, in line with the expectations and intentions of modellers. In this paper, we focus on the current methodological treatment of respective statistical indicators serving as an illustration of conditions under which external balance in the IS-LM-BP model is achieved and where pressures on foreign exchange rate ensue. We will investigate the consequences of the current content of external statistics indicators for their interpretation in the stories economists tell about the international equilibrium and its adjustments, whether in class or in policy-making.

Keywords	***DOI***	***JEL code***
International equilibrium, BP curve, exchange rate	*https://doi.org/10.54694/stat.2025.35*	*F31, F32, E40*

INTRODUCTION

Macroeconomic theory and macroeconomic data are a sine qua non for the economic policy decision-making. The former provides a general framework for how large units, such as the entire economy or its constituents (sectors or industries), operate and how they are affected by changes in economic conditions or the policy setting. The latter supplies ways of quantifying the theoretical concepts used by the former, making testing and illustrating their validity, and hence the whole policy decision-making based on them, possible.

Now the implicit assumption here is that the metric, or the indicator, is designed and its data collected to capture the phenomenon, whose quantification it is assumed to provide, reasonably well. That is to say, that e.g. a price indicator embraces all prices policymakers are interested in, or that all value-creating activities are appropriately counted in GDP. In any case, the methodology of indicators dictates how a respective indicator can be read and what it actually constitutes. The way of quantification also affects the choice of a respective indicator to illustrate or to examine the validity of a theory.

It is of note that, especially after the World War II, the evolution of macroeconomic statistics gradually started to follow its own course without being fully constrained by the objective originally pursued

[1] University of Economics and Management, Nárožní 2600/9A, 15800 Prague 5, Czechia. Corresponding author: e-mail: vaclav.rybacek@vsem.cz.

by the economic theory.[2] Such deviation might have led to a situation where the content of macroeconomic aggregates provides the policymakers with information that is interpreted in a way divorced from its actual content.[3]

One of the fields of macroeconomic statistics that has experienced very substantial modifications over time is the methodology and the framework of external statistics. This paper aims to discuss one particular set of external statistics whose use spans both policy-making and education, namely the balance of payments and its role in, and relation to, the concept of international equilibrium.

In formulating the international or external equilibrium, we rely, in the classroom as in policy-making, on an intuitive story we like to tell. Analogous to the original price-specie-flow mechanism described early, e.g., by Hume in his treatise On the Balance of Trade published in 1752, adapted for a world of fiat currencies and based on the monetary approach (e.g., Mundell, 1968), is predicated on exchange rate adjustments as the equilibrating force. The macroeconomic statistics reflecting here the notion of international equilibrium is the balance of payments, whose deficits and surpluses result from imbalances in the supply of and demand for money (Bijan and Moshin, 1977). However, Obstfeld and Taylor (2002) pointed out that with the growing international capital market, relying on the balance of payments statistics, as a reflection of external balance was increasingly questionable.

Indeed, the methodology of external statistics does not follow the monetary approach, reflecting therefore an international flow of money in scope unknown to users. Instead, external statistics, as currently defined, is rather focused on the international flows of economic values. Against this backdrop in the methodological changes that have been gradually taking place over decades, we will investigate in this text the consequences of the current content of external statistics indicators for their interpretation in the stories we tell about the international equilibrium and its adjustments. On the theoretical as well as practical grounds, we will discuss and investigate the contention put already by Machlup (1950) that there is not necessarily a relationship between accounting balance of payments and market balance of payments.[4]

To that end, the paper is organized as follows. First, the current statistical method of external statistics will be examined in detail necessary for further discussion of the object of this paper. Then, we will proceed to the definition of external balance in the IS-LM-BP model, concretely the BP curve and statistical indicators serving as an illustration of conditions under which external balance is achieved or where pressures on foreign exchange rate ensue. In the last chapter, a presumed misalignment between theoretical and statistical approach is outlined.

However, before proceeding further, the content of the term "external balance" as used throughout the text is to be clarified. In its broader sense, the term "external balance" refers to the total balance of current and capital accounts of the balance of payments. In methodological terms, this corresponds to the item "net lending/borrowing" as defined in the statistical methodology. Besides, when examining the main topic of the paper and in line with the logic of the Mundell-Fleming model, the term "external balance" will also be identified with foreign trade balance only, i.e. a narrower sense of "external balance" routinely employed in macroeconomic theories and textbooks, is also utilized.

[2] This is not least the case of GDP, which was originally designed to define taxable capacity of economy (Coyle, 2014).

[3] Consider the most prominent indicator, the GDP, which is routinely employed as an indicator of the level of economic welfare or the value of all goods and services created in an economy and its changes in time (economic growth). In fact, the GDP might tell us little about the flow of cash in the economy due to its accrual nature. It therefore might be the case that while the GDP figure is growing, insolvency is spreading across the economy as suppliers don´t get paid as contractually agreed, which was indeed the case in the Czech economy in 2017.

[4] Accounting balance payments is defined as difference between credit and debit transactions as recorded in the balance of payments statistics, while market balance of payments reflects the amounts supplied and demanded on the currency markets (Machlup, 1950).

1 BALANCE OF PAYMENTS STATISTICS AND THE EXCHANGE RATE

Balance of payments (henceforth as "BoP") statistics is an integral part of the general macroeconomic framework and an input into economic policy decision-making. The standard macroeconomic narrative told chiefly in economic textbooks, but assumed in academic papers as well, is that the imbalances of the respective BoP macroeconomic aggregates, such as the current account balance, primarily create pressure on foreign exchange rate (henceforth as "ER") changes through the interactions between residents and non-residents[5]. For instance, a large foreign trade surplus is accompanied by an appreciating exchange rate, high volumes of dividends or interest paid abroad result in exchange rate depreciation, etc.

Empirically, however, such mutual interdependence does not appear to hold. A rather weak relation between prominent BoP aggregates and exchange rate[6] movements is increasingly getting the attention of economists.[7] Camacho and Lindström (2021), when analysing the Swedish economy, point to the fact that although the Swedish current account has experienced a surplus for the past 27 years, the krona´s exchange rate was on a weakening path. As the authors concluded, this trend can be explained, *inter alia*, by a higher level of domestic savings and lower domestic interest created an incentive to invest free capital abroad rather than in the domestic economy.

Müller-Plantenberg (2010) points out that the ER movements depend on whether financial flows recorded in the BoP only accommodate transactions in the current and capital accounts or whether financial flows move in a rather autonomous fashion, depending on the extent of capital flows restrictions. In formulating a simple model of economic adjustments, the author therefore employs the BoP flows recorded in both non-financial and financial accounts that might be ER-relevant. Apart from that, a time lag between a flow of value (accrual point) and a related movement of cash may defer the impact of given transaction in the external balance on the ER.

As pointed out by Drahozalova and Rybacek (2025), a deviating pattern between external balance, in its broader sense (current and capital accounts), and the ER adjustments has been clearly observable in the Czech economy as well. Indeed, as shown in the following chart, the Koruna´s (CZK) ER adjustments towards global currencies mostly did not follow the trend in the external balance. There are only few exceptions such as the year 2017 where 2.4% surplus in external balance was accompanied by the appreciation of the CZK towards both EUR and USD. In the most recent years, a deficit in 2022 was followed by a surplus in 2023 while the ER against EUR, which takes the major fraction in the external flow in and out of the Czech economy, developed quite contrarily, i.e. the CZK appreciated in 2022 and then depreciated in 2023.

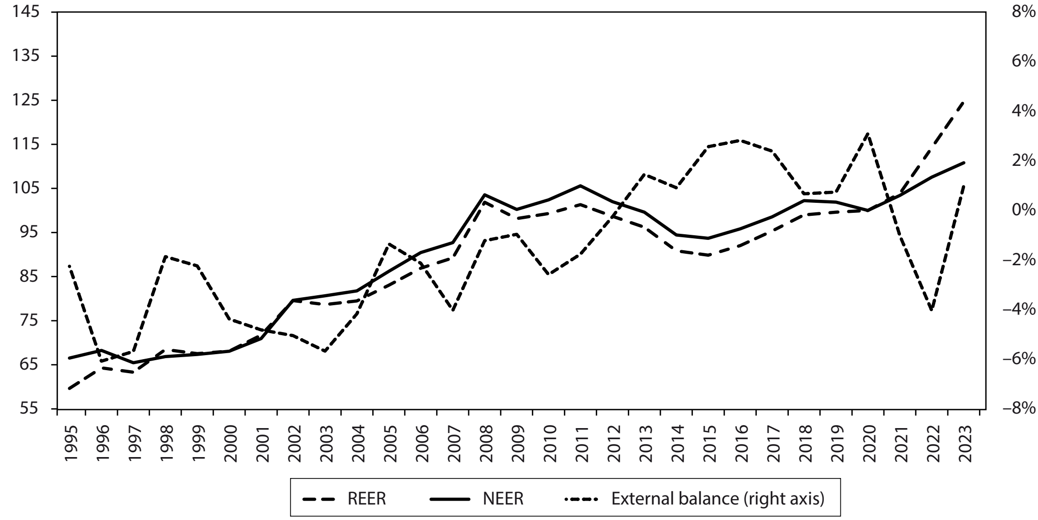
There is a number of explanations for this seemingly counterintuitive development. As pointed out in Drahozalova and Rybacek (2025), the transactions recorded in the BoP are not necessarily coupled with a transfer of cash and thus a conversion from one currency to another, as the BoP statistics applies the accrual principle. Even more importantly, the ER changes have been dominated by capital flows such as portfolio investments in the first place, which are, however, recorded outside of the external balance. Furthermore, Blanchard, Chamon, Gosh and Ostry (2015) and Gardberg (2018) further argued that also the form of capital (in terms of instruments) entering and leaving the economy matters, e.g. financing through equity exhibits lower sensitivity of the ER to international flows than debt financing.

[5] See e.g. Makin (2005), Dornbusch and Fischer (1994), Mach (2001), Johnson, Zuber and Gandar (2005), or Leightner (2024).

[6] The US dollar exchange rate can serve a textbook example when the US external balance only weakly affects the USD exchange rate, as the US dollar serves as global currency and is therefore routinely employed in the external transactions between non-US individuals and companies.

[7] See e.g. Cheung, Chinn and Fujii (2010), or Kincaid et al. (2001).

Figure 1 The Czech external balance (current and capital accounts; % of GDP), changes in the exchange rates CZK/EUR and CZK/USD (index, 1999–2023)



Source: <*www.cnb.cz*>

Data compilers and users have also been increasingly aware of the weakening relation between the external balance and the ER, which also followed substantial methodology changes the BoP statistics has gone through over the last decades. To develop new statistical tools, the ECB staff proposed the "monetary presentation of the balance payments", first suggested in 2008 (Buc, Mayerlen and Sola, 2008) and further updated in 2020 (Aguilar, Soares and Alidad, 2020). In principle, the monetary presentation separates the external transactions within the BoP statistics carried out through the banking industry. While still suffering from many weaknesses, such as missing information on the conversions between banks themselves, the creation of alternative tools clearly represents an aim to overcome the limited power of the BoP aggregates in explaining (and predicting) the ER adjustments.

Hence, there is a growing body of evidence that the relation between the external balance aggregates and the ER changes is rather weak. At a practical level, the compilers aim to create further statistical tools, and the users have discovered that the financial account, as an integral part of the BoP statistics, provides better insights into the development of ER. From another perspective, the issue we are dealing with in this paper is often referred to as "adequacy problem" (Kostakova, 2019). It arises when an aggregate employed in the analysis does not capture the respective phenomena adequately so that there is a gap between the development of the aggregate believed to be related to the phenomenon and the development in the phenomenon itself. Adequacy problem constitutes one of fundamental reasons why postulates of economic theory are not necessarily demonstrable by statistical data. The issue analysed in this paper can be thus also observed as a dimension of the adequacy problem, here as an ability of the core BoP aggregates to contribute to explanation of the exchange rate changes.[8]

[8] Yet, there is another important implication of this weak relationship that transcends mainstream policy-making, and that is the pedagogical use of these aggregates we make in classrooms. From simple stories about the international equilibrium to more formal macroeconomic models (such as the IS-LM-BP), we hate when our story is contrasted with actual data by an avid student, forcing us to engage in convoluted attempts to square the just presented theory with empirical data, at which point we typically lose the audience altogether.

2 BOP THEORY: MUNDELL-FLEMMING MODEL AND THE BP CURVE

BP curve is an indispensable part of the Mundell-Flemming model (or IS-LM-BP model), widely employed in macroeconomic modelling,[9] which portrays the relationship between the level of output, and the interest rate and the nominal ER for a small open economy in the short run. The BP curve constitutes an extension of the IS-LM model by displaying the points of overall balance of flows between domestic economy and the rest of the world. In the framework of the model, the BP curve describes the functional relationship between the interest rate and output which maintains the flows between residents and non-residents balanced.

Let´s briefly describe the logic behind the BP curve. The BP curve is a set of points showing the combinations of the interest rate and the level of output that are compatible with international equilibrium. That implies that the balance of payments is in fact in perfect balance, meaning that the balance of capital accounts precisely compensates for the balance of the current account *ceteris paribus*.[10] In the current terminology, with the introduction of the manual BPM6, this corresponds to a situation where the sum of the balances recorded in the current and the capital account equals the balance in the financial accounts.

Further in the text, we will refer to the sum of current and capital accounts as "non-financial accounts balance (NFAB)", while to the balance of financial flows as "financial account balance (FAB)". Formally, following the methodology of BPM6 being currently in place, the BP curve can be written as follows:

$$\text{NFAB}\ (i, Y) = -\ \text{FAB}\ (i, Y), \qquad (1)$$

where NFAB and FAB are both functions of interest rate (i) and domestic output (Y). The macroeconomic textbooks routinely employ a simplified view of the NFAB as foreign trade (NX) and FAB as portfolio investment. This then allows making a simplified presentation of the Formula (1) as the left-hand side being a function of output and the right-hand side being a function of the interest rate differential between the domestic economy and the rest of the world (Mach, 2001). This simplification, however, results in understanding of the adjustment process by means of ER changes that is generally not observable in the official macroeconomic statistics, as demonstrated below.

Let´s look at the combinations of the interest rate and output level that lie off the BP curve. All of these combinations indicate an international disequilibrium: if the economy operates at a combination lying to the left of the curve, then its BoP must be in surplus. In that case, the level of output generates relatively low demand for imports, affecting the trade balance in a positive direction. On the other hand, comparatively higher interest rate gives rise to a net inflow of capital seeking higher-yield opportunities. The combined effect on both NFAB and FAB, resulting in a BoP surplus will then create pressure, as the story goes, on the ER to appreciate.

If the economy operates in the flexible ER regime, then the ER appreciation pressure will materialize, initiating the self-correcting mechanism to reduce the original BoP surplus.[11] For completeness,

---

[9] See e.g. Céspedes, Chang and Velasco (2008), Dvoskin, Feldman and Landau (2024), Huh (1999), Azar, Bolbol and Mouradian (2020).

[10] The curve is constructed for a given level of domestic prices, exchange rate and net foreign debt (Melvin and Norrbin, 2023). Any change in these variables results in a shift of the curve right- or leftwards. Our discussion is, however, focused on the reflection of the BP curve in the macroeconomic statistics, so the discussion on the determinants of the slope and the shape is beyond our topic.

[11] For the sake of completeness, if the central bank commits to the fixed exchange rate, the reserve assets held by the central bank are bound to increase as the central bank, through its direct purchases of foreign currencies on the FX market, attempts to prevent the appreciation from taking place. Whether increasing international reserves in the balance sheet of the central bank will be translated into growing money supply in the domestic economy is dependent on whether the central bank will resort to sterilization of these operations or not.

if the economy operates at a combination of out and interest rate differential lying to the right of the BP curve, then the BoP is in deficit, and an inverted set of consequences triggered by the pressure for the ER to depreciate ensues. Therefore, the model story heavily relies on the BoP aggregates to illustrate how imbalances of the domestic economy towards the rest of world are to be remedied by the ER adjustments (when left to float freely). Exchange rate adjustment are thus seen as balancing mechanism in the model.

In the following section, we will confront the logic of the model on the empirical, as well as theoretical grounds. We will make use of the actual data to find out whether the model conclusions about the current account balance and net portfolio investment flows can be upheld, considering the content of the concerned aggregates as currently defined.

3 BOP EMPIRICALLY: THE CZECH EXPERIENCE

The definitional inference from the BP curve is that if international payments are not in balance, i.e. the economy operates "off the curve", there will be pressure on the ER to either appreciate or depreciate, thus tending to eliminate the initial imbalance. To demonstrate this development with respective statistics thus requires finding an aggregate, or a combination of aggregates, that would adequately reflect the pressure stemming from the conversion of domestic currency into foreign ones, or vice versa.

Unsurprisingly, there is vast historical evidence of situations in which non-financial accounts balance in surplus was accompanied by an ER depreciation and *vice versa*, which contradicts the theoretical logic. To provide an empirical argument, in the following paragraphs we will confront actual data with the construction of the BP curve. Besides, in seeking an aggregate presumably better fitting the logic behind the curve, we will also analyse an alternative indicator sourced from a newly developed statistical tool commonly referred to as "monetary presentation of balance of payments" (Aquilar, Soares and Alidad, 2020).

Let us look at recent empirical evidence of the Czech economy.[12] Lest this analysis be distorted by the FX intervention regime launched by the CNB in the fall of 2013, the examined time series starts in mid-2017 when the FX interventions with lower-bound were discontinued. The time series then extends until November 2024, which leaves us with 90 monthly observations. We will primarily focus on situations of aggregate imbalances in which the ER is supposed to be under pressure to adjust. To check the conformity of macroeconomic data with the model, several modalities are considered. First, the model is standardly presented in its simplified version so that the current account balance is associated with foreign trade and financial accounts balance with portfolio investment. Alternatively, however, we will also scrutinize the conformity by considering the actual scope of the respective aggregates as these are defined in the relevant manuals and quantified by statistics around the world.

Let´s start with the simplified version of current and financial accounts. The following table summarizes the number of cases where the ER adjustments were in line with the conclusion of the model, broken down by specific types of adjustment: that is a) when a surplus led to domestic currency appreciation, or b) when a deficit led to domestic currency depreciation.[13]

Table 1 Simplified version of current and financial accounts, monthly data (7/2017–11/2024, Czechia)

Conformity with theory	Surplus	Deficit
Yes	44	5
No	46	85

Source: Own calculation, <*www.cnb.cz*>

[12] Using the Czech data retrieved on December 15, 2024.
[13] For the sake analysis, the indicate nominal exchange rate (NEER) is employed.

For the situation of a surplus, the model prediction conforms to the empirical data in less than 50 percent of cases, while for deficits, it is less than 5 percent.[14] While a certain leeway in the interpretation of data for educational purposes might be understandable, such results hardly exhibit convincing support for the theory. These aggregates then do not appear to be a good fit for measuring the external equilibrium.

Let us then make the aggregate larger and proxy the international equilibrium by the BoP in its entirety as published by national statistical authorities. The results, as shown in the table below, are on average even more disappointing: while deficit situations conformed to the theory in 20 % of cases, less than 10 percent of the surplus situations did so.

Table 2 Full version of current and financial accounts, monthly data (7/2017–11/2024, Czechia)

Conformity with theory	Surplus	Deficit
Yes	7	18
No	83	72

Source: Own calculation, <*www.cnb.cz*>

For the sake of completeness, and perhaps more out of curiosity, the following table shows the results of the same analysis for a host of different components of the BoP, as if each of them represented a different metric of international equilibrium. Interestingly enough, the relations between values of indicators and simultaneous changes in the ER show results that are more conforming to the theory than when aggregated into larger wholes. All of them were correct in at least 50 percent of cases:

Table 3 Balances of individual section of BoP vs ER changes, monthly data (7/2017–11/2024, Czechia)

Conformity with theory	Trade balance	Current account	Financial account	Portfolio investment
Yes	50	45	52	49
No	40	45	38	41

Source: Own calculation, <*www.cnb.cz*>

The analysis suggests that the relation between BoP indicators and ER changes is more complex and cannot be simplified in any meaningful way. The range of transactions affecting ER and the complexity of the financial system is much broader than foreseen in the model, and there seems to be no single standard BoP indicator, or even a combination of indicators, that can be predictably associated with changes in the ER.

This weaker link between the external balance and the concomitant ER changes may have a number of explanations. The Mundell-Flemming framework assumes producers´ mark-ups and marginal costs being responsive to ER movements, unlike the producers' prices themselves, which are sticky (Boz, Gopinath and Plagborg-Moller, 2018). On the other hand, the prices an importing country faces, when expressed in its domestic currency, do fluctuate in response to the bilateral exchange rate. In this respect, the model is consistent with the producer currency pricing where the pass-through to prices in the importer´s

[14] Which broadly suggests an independency between the two variables.

currency is one, while the pass-through to the exporter´s currency is zero. However, this assumption might be challenged in many ways, briefly elaborated below. Importantly, relaxing the assumption on the application of producer currency pricing weakens the exchange rate adjustment mechanism formulated in the model, as the terms of trade becomes less sensitive to the bilateral exchange rate movements.

As the IMF paper discusses (2019), there are several alternatives to producer currency pricing assumption commonly observable in economic reality. First alternative to be mentioned is the so-called local currency pricing (Deveraux and Engels, 2000) where prices are rigid in the currency of the importer. In this case, no pass-through in the importers currency stemming from the bilateral exchange rate movements is observable, while a complete pass-through in the exporter´s currency ensue. As a result, a nominal depreciation in the bilateral exchange rate leads to a deterioration in the price competitiveness of the exporters which is in a sharp contrast to the Mundell-Flemming paradigm.

Another alternative is pricing at dominant currency, typically USD or EUR, where prices are set in a third currency and the mutual trade is therefore affected by exchange rates of both trading economies vis-à-vis the dominant currency. As the study of Boz (2018) finds, pass-through effect is from the dominant currency into export and import is high, while comparatively lower in the cases of producer and local currency pricing. These findings further weaken the external adjustment of the balance of payments by way of exchange rate movements as the flows with the trading partners by less affected by the bilateral exchange rate, or at all.

Another aspect making the foreign trade balance to a certain extent independent of exchange rate movements is the changing nature of the global trade and production chains getting increasingly spreading across several currency areas. Against this background, the country´s marginal costs in producing semi-products might be affected not only by bilateral exchange rate changes with their trading partners but also by bilateral exchange rate movement between countries trading at earlier production stage (backward integration). Furthermore, if produced semi-products are consequently exported further down the production chain, these trade flows will also be affected by movements in the bilateral exchange rate between economies at later production stage (forward integration), representing a demand shock[15] (IMF, 2019).

Last but not least, the ER may be affected asymmetrically depending on the types of financial flows. The sensitivity of the ER in countries with large foreign debt financing tend to vary more if the debt consists of bonds and bank loans than in countries where the debts are financed via equity (Gardberg, 2018; Blanchard, Chamon, Gosh and Ostry, 2015). Gardberg (2018) – FDI flows being less influenced by the global financial cycle.

Overall, as it appears from the previous paragraphs, there has been a growing body of literature suggesting that the model might have been getting outdated and the operation of the ER mechanism might not be decisive in the readjustment process.

Considering all objections mentioned above, let us consider an alternative metric of international equilibrium based on the "net financial assets" (henceforth NFA) of banking institutions, which shows a change in the position of the domestic banking system toward the non-resident banking system. The following table shows the number of cases where a change in NFA[16] was associated with a change in ER in a way that conforms to the theory, i.e. an increase in NFA was accompanied by an appreciating ER, and *vice versa*.

[15] For more details, please see IMF (2019).

[16] The respective indicator was retrieved from so-called monetary presentation of balance of payments, which is meant to separate from the BoP indicators actual flow of money in and outside of the economy.

Table 4 NFA vs ER changes, monthly data (7/2017–11/2024, Czechia)

Conformity with theory	Change in NFA
Yes	35
No	55

Source: Own calculation, <*www.cnb.cz*>

The results do not show much of an improvement over the orthodox BoP statistics, with less than 40 percent of cases conforming to the theory. There may be multiple reasons for this finding. First, not all ER-relevant transactions are necessarily channelled through the banking system. If they are not, then they are not tracked in the banking statistics and in NFA. Secondly, transactions captured in NFA can also be transactions in domestic currency. The reason is that this statistics is based on the sectoral and not on the currency approach. Thirdly, and apparently most importantly, the NFA indicator does not cover ER-relevant transactions carried out within the domestic banking system, as the indicator is quantified towards non-residents only.

To wrap up, none of the established indicators of international equilibrium seem to provide reasonably clear empirical evidence of the story we like to tell about that equilibrium. In other words, we not only have no statistical tool to illustrate the mechanism behind the international equilibrium, but we, therefore, also lack any statistical tool to incorporate the notion of international equilibrium (such as the BP curve in IS-LM-BP) into our analytical framework and econometric models.

CONCLUSION

We can summarize that there is neither a single item nor a combination of items that can be sourced from the BoP system which would be empirically indicative or predictive of the pressures on the ER adjustments. The relation between ER adjustments and current account balance and financial account balance was found to be rather weak. The same holds true for the indicators incorporated in the IS-LM-BP model, i.e. foreign trade balance and net portfolio investment flows. The limiting factor as for the strength of this conclusion is the fact that the analysis was carried out on data for the Czech economy only, so the empirical evidence presented in the paper does not necessarily leads to a general conclusion valid for all economies.

Still, as discussed above, the clarity of inferences from the BoP statistics concerning ER changes are widely complicated by methodological aspects, to name the ones we find the most crucial:

– the BoP statistics is not based on currency but on sectoral approach, therefore the BoP transactions do not necessarily involve conversion between currencies as they can be settled in domestic currency,
– the accrual principle is applied in the compilation process, so similarly to the previous case, transactions in the system do not necessarily involve conversion between currencies as there is no flow of cash,
– ER-relevant transactions in the BoP statistics are to a large extent recorded off the current or capital account,
– ER-relevant transactions between residents having an effect on exchange rates are not addressed in the external statistics altogether.

To interpret the point off the BP curve by means of selected BoP aggregates is therefore misleading due to adequacy problem. While to do so may serve as an intellectual exercise to serve educational purposes for students to understand the operation of the economy and ER adjustment pressures ensuing from certain economic conditions, the observable changes in ER cannot be explained by the use of the BoP aggregates in question. Importantly, the fact that no single aggregate (or a combination of aggregates) sufficiently reflects or covers ER adjustments makes any model involving international equilibrium empirically untestable, implying also its limited use in practice for analysing and determining the position of the economy in terms of external balance. To illustrate the point raised in the model, we shall rather turn to the currency markets itself where, at the points off the BP curve, the demand for foreign currencies is outstripping the supply, or *vice versa*.[17]

However, we in no way dispute the conclusions made by the authors of the Mundell-Flemming model such as impossible trinity.[18] We only intend to flag the fact that economic policy faces serious data constraints, consciously or not, when employing the model in the decision-making. And, not less serious inference we can draw, that the BP curve is practically impossible to test empirically, implying that there is no way of realizing whether the economy reached the general economic equilibrium in the IS-LM-BP model or not.

References

AGUILAR, C. P., SOARES, R. O., ADILAD, R. (2020). *Revisiting the monetary presentation of the euro area balance of payments* [online]. ECB Occasional Paper No. 238, <https://doi.org/10.2139/ssrn.3536330>.

AZAR, S. A., BOLBOL, A., MOURADIAN, A. (2020). IS-LM-BP Model for Lebanon: a Simple Empirical Analysis. *International Research Journal of Finance and Economics*, 177: 97–110.

BE DUR, L., MAYERLAN, F., SOLA, P. (2008). *The Monetary Presentation of the Euro Area Balance of Payments* [online]. Occasional Paper Series, No. 96, ECB, <https://doi.org/10.2139/ssrn.1144467>.

BIJAN, A., MOSHIN, K. (1977). The Monetary Approach to Balance of Payments Determination: an Empirical Test [online]. In: *A Collection of Research Papers by Members of the Staff of the International Monetary Fund*, IMF. ISBN 9781557752772. <https://doi.org/10.5089/9781557752772.071>.

BLANCHARD, O., CHAMON, M., GOSH A. R., OSTRY, J. D. (2015). *Are capital inflows expansionary or contractionary? Theory, policy implications and some evidence* [online]. IMF Working Paper, No. 15/226. <https://doi.org/10.3386/w21619>.

BOZ, E., GOPINATH, G., PLAGBORG-MøLLER, M. (2018). *Global Trade and the Dollar* [online]. IMF Working Paper No. 2017/239. <https://doi.org/10.5089/9781484327975.001>.

CAMACHO, J., LINDSTRÖM, A. (2021). *Exchanged rate and balance of payments – a correlation that got lost?* Monetary Policy Department, Sverige Riksbank.

CESPEDES, L. F., CHANG, R., VELASCO, A. (2008). *IS-LM-BP in the Pampas* [online]. IMF Staff Papers, Vol. 50, Special Issue. <https://doi.org/10.3386/w9337>.

COYLE, D. (2014). *GDP: A Brief but Affectionate History* [online]. Princeton University Press. ISBN 9780691156798. <https://doi.org/10.1515/9781400873630>.

DEVERAUX, M., ENGEL, Ch. (2000). Monetary policy in the Open Economy Revisited: Price Setting and Exchange Rate Flexibility [online]. *Review of Economic Studies*, 70: 765–84. <https://doi.org/10.3386/w7665>.

DORNBUSCH, R., FISCHER, S. (1994). *Macroeconomics*. Mc-Graw-Hill. ISBN 0071125906

DRAHOZALOVA, A., RYBÁČEK, V. (2025). Making Sense of Exchange Rate Pressures in Macroeconomic Statistics [online]. *Statistika: Statistics and Economy Journal*, 105(1): 33–48. <https://doi.org/10.54694/stat.2024.33>.

DVOSKIN, A., FELDMAN, G. D., LANDAU, M. T. (2024). The Argentine economy through the lens of an adapted Mundell–Fleming model for small open peripheral economies [online]. *Review of Keynesian Economics*, 12(2): 239–264. <https://doi.org/10.4337/roke.2024.02.07>.

FLOOD, R. (2003). *IS-LM-BP in Pampas* [online]. IMF. ISBN 9781589062047. <https://doi.org/10.5089/9781589062047.024>.

FROHM, E. (2018). *How do global value chains influence the effects of the krona exchange rate on export?* Economic Commentaries No. 9, Sveriges Riksbank.

GARDBERG, M. (2018). *Linking net foreign portfolio debt and equity to Exchange rate movements*. IFN Working Paper 1246.

HUH, H. S. (1999). How well does the Mundell-Fleming model fit Australian data since the collapse of Bretton Woods? [online]. *Applied Economics*, 31(3): 397–407. <https://doi.org/10.1080/000368499324372>.

[17] In this context, Machlup (1950) distinguishes between market and accounting balance of payments, pointing to the fact that both are not necessarily related. Country may very well experience accounting deficit, while its currency might be appreciating on the foreign exchange market if the receipt of the foreign funds exceeds the payments abroad.

[18] The term refers to the impossibility of maintaining fixed exchange rate, free capital movement and an independent monetary policy.

CHEUNG, Y-W., CHINN, M. D., FUJII, E. (2010). China´s Current Account and Exchange Rate. In: FEENSTRA, R. C., WIE, S. J. *China´s Growing Role in the Worlds,* University of Chicago Press. ISBN 9780226239743

IMF. (2019). *Exchange Rates and the External Adjustments* [online]. Chapter 2 in External sector report. <https://doi.org/10.5089/9781498318976.085>.

JOHNSON, R., ZUBER, R., GANDAR, J. (2005). A Re-examination of the Mundell-Flemming Flexible Exchange-Rate Model. *Journal of Economics and Finance Education*, 4(2): 22–38.

KINCAID, M., FETHERSTON, M., ISARD, P., FARUQEE, H. (2001). *Methodology for Current Account and Exchnage Rate Assessments* [online]. IMF. ISBD 9781589060914. <https://doi.org/10.5089/9781589060814.084>.

KOŠŤÁKOVÁ, T. (2019). *O složitém jednoduše*. Prague: Czech Statistical Office. ISBN 978-80-250-2908-4

LEIGHTNER, J. (2024). How US Fiscal and Monetary Policy affect the GDP of Countries with Fixed and Flexible Exchange Rates: Estimates using Korean Data from 1963 to 2022 [online]. *Journal of Economic Integration*, 39(1): 86–106. <https://doi.org/10.11130/jei.2024002>.

MACH, M. (1998). *Makroekonomie 2*. Melandrium. ISBN 9788086175041

MACHLUP, F. (1950). Three Concepts of the Balance of Payments and the So-called Dollar Shortage [online]. *The Economic Journal*, 60(237): 46–68. <https://doi.org/10.2307/2227208>.

MAKIN, A. J. (2005). A Monetary Model of Exchange Rate and Balance of Payments. *Economic Issues*, 10(1): 25–36.

MÜLLER-PLATENBERG, N. A. (2010). Balance of payments accounting and exchange rate dynamics [online]. *International Review of Economics and Finance*, 19(1): 46–63. <https://doi.org/10.1016/j.iref.2009.02.010>.

MUNDELL, R. (1968). *International Economics*. New York: Macmillan.

OBSTFELD, M., TAYLOR, A. (2002). *Globalization and Capital Market* [online]. NBER Working Paper Series No. 8846. <https://doi.org/10.3386/w8846>.

Upcoming Events

Conferences

The ***28th Applications of Mathematics and Statistics in Economics Conference (AMSE 2026)*** will take place **from 25th to 29th August 2026 in Martin, Slovakia.** The conference seeks to acquaint its participants with the latest mathematical and statistical methods useful in addressing theoretical and practical problems or challenges of economics. The conference gives a unique opportunity to present research results achieved in the field where mathematics and statistics meet economics. More at: <*https://www.amse-conference.eu*>.

The ***34th Interdisciplinary Information Management Talks (IDIMT 2026)*** will be held **during 2–4 September 2026 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 ***20th International Days of Statistics and Economics (MSED 2026)*** will take place **during 3–5 September 2026 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*>.

The ***44th International Conference on Mathematical Methods in Economics (MME 2026)*** will be held **from 9th to 11th September 2026 in Ostrava, 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.ekf.vsb.cz/mme/en*>.

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 22 pages (Times, 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 10 pages (Times, 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 22 pages (Times, 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 10 pages (Times, 1.5-spaced pages).

Book Review evaluates selected titles of recent books from the official statistics field. Reviews shall have 2 pages (Times, 1.5-spaced pages).

Information section contains informative (descriptive) texts, latest publications, or recent and upcoming scientific conferences. Recommended range of information is 10 pages (Times, 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 <u>underline</u> 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…".

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). **All titles from the list of references must be cited in the text.**

Tables and numbers
Provide each table on a separate page. Indicate position of the table by placing in the text "<u>insert Table 1 about here</u>". Number tables in the order of 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). 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 "<u>insert Figure 1 about here</u>". 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.
Authors are also required to read our **Principles of using generative artificial intelligence (AI) and artificial intelligence tools** (published at: http://www.csu.gov.cz/statistika_journal) and respect them.

Managing Editor: Vladimír Nedvídek
Phone: (+420) 274 054 299
E-mail: statistika.journal@csu.gov.cz | **web:** 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
E-mail: objednavky@csu.gov.cz

Design: Toman Design
Layout: Ondřej Pazdera
Typesetting: Václav Adam
Print: Czech Statistical Office

106th 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