STATISTIKA

STATISTICS
AND ECONOMY
JOURNAL

VOL. **106** (3) 2026

CZECH
STATISTICAL
OFFICE

CONTENTS

ANALYSES

INFORMATION

BOOK REVIEW

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

Czech Regional GDP: Challenges and Comprehensive Estimates[1)]

Petr Musil[2] | *Prague University of Economics and Business, Prague, Czechia*
Jana Fischerová[3] | *Prague University of Economics and Business, Prague, Czechia*
Jaroslav Kahoun[4] | *Czech Statistical Office, Prague, Czechia*

Received 18.12.2025 (revision received 24.4.2026), **Accepted** (reviewed) 19.5.2026, **Published** 11.9.2026

Abstract

Statistical descriptions of regions or even cities are usually more popular with users than those at the national level. Users can more easily identify themselves with regional indicators such as the average wage. However, the production of regional indicators is more demanding in terms of data sources. National statistical institutes mostly do not officially publish all components of regional GDP, but rather select a few, such as gross value added or gross fixed capital formation, due to a lack of data. This paper aims at regional GDP in Czechia in particular to estimate missing components of regional GDP as well as analyse the results. It uncovers difficulties connected with the estimates. In addition, the effect of different regional price levels is assessed, regional indicators are adjusted accordingly. The findings indicate substantial differences in the regional GDP structure as well as notable variations of regional price levels. The analysis is conducted for the year 2020.

Keywords	***DOI***	***JEL code***
Regional expenditures, GDP, national accounts, regional accounts, experimental estimates, NUTS 3 regions	*https://doi.org/10.54694/stat.2025.61*	*O11, R10*

INTRODUCTION

Attempts to estimate regional aggregates are pretty common and are spread throughout the world – e.g. Billé et al. (2023), Secondi (2021) in Italy, Gough (2017) in the UK, Rokicki and Hewings (2016)

[1] This paper is prepared under the support of the Technology Agency of the Czech Republic within the Project No. SS04030013 (Center for Socio-Economic Research on Environmental Policy Impact Assessment). Part of the paper was presented at AMSE 2025 conference in Hradec Králové.

[2] 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: petr.musil@vse.cz, phone: (+420)224095435. The author is also working at the Czech Statistical Office, Na Padesátém 81, 100 82 Prague 10, Czechia. ORCID: 0000-0002-8457-4559

[3] 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. E-mail: jana.fischerova@vse.cz. The author is also working at the Czech Statistical Office, Na Padesátém 81, 100 82 Prague 10, Czechia.

[4] Czech Statistical Office, Na Padesátém 81, 100 82 Prague 10, Czechia. E-mail: jaroslav.kahoun@csu.gov.cz.

in Poland, or Aten and D'Souza (2008) in the USA. However, the difficulties are connected mainly with the limited data at the regional level (missing detailed structure as well as missing continual availability in time). It means that when researchers decide to do the estimation, it is usually just once or once in a longer time. In the meantime, there are often some methodological changes, which may be feasible to implement rather at the national level, but may bring more troubles at the regional level (e.g. scanner data or a change in household budget survey (hereinafter HBS) – from a quote to a random sample). This results in increased volatility of estimates at the regional level.

Our motivation is based on the fact that the topic of regional expenditure and regional price levels remains a subject of public interest, and there is a demand for updated results. The start of our research in this field dates back to 2009, when we began with the first estimates for 2007, using the expenditure approach to gross domestic product (hereinafter GDP) for 14 Czech NUTS[5] 3 regions. The regional price levels were estimated for the years 2012 and 2017, but regional expenditures are being updated now for the year 2020 within the current project.

The updates are not always straightforward, because many things are changing and improving over time, so it was necessary to make changes to many parts of the computation differently. The system of national accounts is dynamic and constantly evolving. It means that every further effort on our part involves particular changes in the methodology and calculations, which, on the one hand, brings a better approach to reality, but on the other hand, may result in less comparability of results. Each trade-off means some limitation elsewhere. In this research, we faced changes in data sources, classifications, methodologies and also in the entire economy.

At the national level, there is a detailed manual called the European System of Accounts or ESA 2010 (Eurostat, 2013a), a comprehensive methodological book exceeding 750 pages. This manual is regularly updated, and the adoption of new ESA 2028[6] is currently underway (see, e.g., Hronová, 2024), which means that our future estimates will again be connected with some methodological novelties. In (effective) ESA 2010, there is a methodological guide inside not only for estimates of aggregates on the national level, but for the whole system of national accounts. It should be noted that ESA 2010 is the Regulation of the European Parliament and of the Council, which is binding for all member states. Users benefit from a high level of comparability and harmonisation.

There are three different approaches to estimating GDP at the national level – the production method, the expenditure method, and the income method - that are briefly described e.g. in the methodology under the Czech Statistical Office (2025) or in Gough (2017, p. 2). The aim of this paper is to move to the regional level, i.e. to the 14 Czech NUTS 3 regions, and estimate regional GDP (RGDP) using all three approaches. We also discuss the main difficulties we encountered, the availability of data, and the need for some assumptions. At the regional level, there is officially published just part of the GDP components (e.g., gross value added – hereinafter GVA); as will be shown further in the paper, the rest are our experimental estimates. The availability of GVA at the regional level is quite common, e.g. Billé et al. (2023) work with data for 103 Italian provinces (NUTS 3 level) obtained from the National and Regional Accounts of the Italian National Institute of Statistics (ISTAT).

We decided to choose for this analysis the year 2020. There were more reasons for this. Firstly, when we started our effort, there was a new, semi-definitive version at the national level for the year 2022.

[5] Regional classification (Nomenclature of Territorial Units for Statistics) CZ-NUTS, for detailed structure see https://apl2.czso.cz/iSMS/en/klasstru.jsp?kodcis=80113

[6] In the meantime, it was decided that the manual would be entitled ESA 2030. "A draft legislative proposal for the ESA 2030 is being prepared, which includes the methodological framework as well as the revised data transmission programme. It is currently undergoing a thorough consultation process with EU countries and other stakeholders. The European Commission is expected to adopt the proposal once this process has been completed." (Eurostat, 2026).

However, the years 2021 and 2022 were extensively affected by COVID-19 disease (especially household and government expenditure, but to some extent the whole economy). In 2020, the first quarter was still under usual conditions, however it is clear that there was a significant influence of the COVID-19 pandemic in the rest of the year. What is important for this paper, the impact of the pandemic on all Czech regions was likely similar. This allows us to compare regions with each other as we intended. Of course, comparison with the previous analysis may be more difficult, as on one hand household expenditures were probably lower at the national level (especially on vacation or restaurant services) and on the other hand government expenditures were higher (especially in healthcare). The entire calculation is complicated and time-consuming, especially in terms of data sources therefore time – consuming. However, we plan to update the results to a more suitable year, 2023, in the future.

The detailed estimates of expenditure method in the Czech regions were published in 2013 (Kramulová and Musil, 2013); therefore, we aimed to carry out new estimates for the most recent possible year. National accounts are being regularly revised – the latest harmonised benchmark revision of the entire available time series of national accounts (1993–2024), was published in June 2024. Therefore, the year 2020 (after the revision) was chosen and it was necessary to consider new national numbers. The revision was aimed e.g. at imputed rent, paid rent, the incorporation of the new population census from 2021 and other data sources, the service life of buildings, or individual residential construction. Above that, the 2020 harmonised benchmark revision and the 2014 methodological revision have been published since our previous publication was issued. The latter changed concepts and definitions of national accounts indicators.

The paper is divided into several sections. The first one is aimed at concerning what is available in official statistics at the regional level and how the officially published regional GDP is compiled together with challenges that have to be faced. The second section is dedicated to the expenditure approach of GDP estimate at the regional level. The third section describes the income approach to regional GDP. Section 4 contains the discussion and the final one conclusion.

1 OFFICIAL REGIONAL ACCOUNTS STATISTICS IN CZECHIA

1.1 Which regional indicators are available in official statistics?

Official regional accounts in Czechia are limited to compiling only selected macroeconomic indicators: GVA, GDP in current and constant prices, gross fixed capital formation (GFCF), employment (employees, employed persons, hours worked, compensations of employees) and household sector accounts necessary for the allocation of disposable income (distribution of primary income account, secondary distribution of income account). The requirements for regional accounts statistics at the European level determine the scope of compiled regional indicators. Regional accounts in Czechia do not offer any additional indicators beyond those required by EU regulations.

Regional accounts are based on transactions of units resident in the territory of a region. If information for local units resident in the territory is not available, a distribution key close to the measured indicator is used to allocate data for a higher level (enterprise, total economy). Such approach is called either "top-down" (distribution of data from the level of total economy to regions) or "pseudo-bottom-up" (distribution of data from the enterprise or institutional unit level to regions), see Eurostat (2013b).

1.2 Compilation of the regional GVA and GDP in Czechia

Compilation of the regional GVA in Czechia is carried out on the basis of the production approach for sectors of non-financial corporations and enterprises in the sector of households (see Kahoun and Sixta, 2013 and Kahoun, 2009). Regional indicators of output and intermediate consumption are obtained from the same data sources used to calculate the national accounts. It means for the case of production, collected data about sales of products and services, sales of goods for resale minus costs of goods sold, changes in inventories of own production and production for own final use. For intermediate consumption, it means collected data on material consumption, energy consumption and expenses for services incurred in making production. Both information on production and intermediate consumption is only available at the enterprise level in Czechia. The general problem in most countries is the valuation of output for the local units in case of multi-regional organisations (e.g. corporate headquarters, local units performing only administrative activity etc.).

The calculation is carried out at the level of divisions of NACE classification[7] and published at the level of sections. A special regionalisation approach is used for selected methodological adjustments in the national accounts, such as imputed rent, individual household construction, consumption of fixed capital in the general government sector (depreciation of roads, motorways, municipal roads and railways), capitalisation of research and development, geological exploration and part of illegal activities. In addition, FISIM[8], financial leasing, subsidies for mining reduction, and activities of households as employers of domestic personnel are also regionalised separately.

The predominant approach for the production method of regional GDP in Czechia is the "pseudo-bottom-up" method. This procedure is based on estimates of data on the activity of local units by disaggregating data from individual enterprises or companies. It is similar to the "top-down" method, but not at the national level, rather on aggregated data for a single multi-regional institutional unit. The obtained data are then aggregated into a total regional data in the same way as in the classic "bottom-up" method, which obtains the necessary information directly from the local units.

The method of "pseudo-bottom-up" is used for GVA of the most significant institutional sectors – non-financial corporations (S.11) and household entrepreneurs (S.14). The allocation of regional GVA for remaining institutional sectors (financial corporations, general government and non-profit institutions serving households (hereinafter NPISH) – S.12, S.13 and S.15) is carried out on the basis of wages and salaries broken down by types of statistical surveys and required regional and NACE breakdown. The regional structure of wages and salaries is used as the key for the "top-down" distribution of the remaining value of recorded GVA in the national accounts, i.e. for S.12, S.13 and S.15 (although this is still before methodological and conceptual adjustments that are made in the compilation of the national accounts). A pure "top-down" allocation based solely on wages and salaries was applied to 13.7% of the total national gross value added in 2020.

After adding up all the regional allocated values described above, there is still approximately 12% of GVA from national accounts that remains unallocated. These are mainly parts of the methodological adjustments of national accounts (except those mentioned above) and adjustments for the exhaustiveness of the economy, which includes intentional misrepresentation of most of the illegal activities, units that are not subject to identification, etc. These data are regionally allocated according to the structure of the regional GVA found in each NACE division separately. Estimates of these adjustments are usually made using models without the possibility of using currently available and sufficient quality regional data.

There are notable differences in the approaches to regional allocation of GVA among individual NACE industries, which are due to the different shares of institutional sectors in each NACE division and the different allocation approaches to institutional sectors. The 'top-down method' is used for the general government sector, financial corporations and NPISH sector, while the 'bottom-up method'

---

7 Classification of Economic Activities CZ NACE or CZ NACE 2025, for detailed structure see https://apl2.czso.cz/iSMS/en/klaslist.jsp?fnazev=NACE

8 Financial intermediation services indirectly measured.

is used for the non-financial corporations and entrepreneurs in households' sectors, as described above. So, while the 'bottom-up method' is used almost exclusively in manufacturing, mining and energy, as well as in trade and construction, the 'top-down method' is used almost exclusively in public administration and financial intermediation industry.

Location of corporate headquarters and central administrative activities is a key factor that affects data interpretation for capital cities or metropolitan areas in general. Regional accounts in Czechia assume that the value added created by corporate headquarters and individual workplaces can be allocated in the structure of wages and salaries (in accordance with available data sources). It is expected that the contributions of individual corporate workplaces or local units to the overall economic performance of the company are adequately reflected in the wages of their employees. Higher GDP in such regions is in this way neither overrated nor underrated by the presence of corporate headquarters, since wages correctly reflect the contribution of local units to overall economic performance at the company level.

Morillas-Jurado et al. (2026) propose to use more complex models for the analysis of the influence of the headquarters on GVA and GDP, for example input-output analysis, which would go beyond the scope of this article. However, the topic of regional input-output tables in Czechia has been addressed, for example, by Sixta and Vltavská (2016).

1.3. Challenges of other regional accounts indicators beyond GDP

In addition to regional GDP, which is the subject of this paper, regional accounts also face challenges with other officially published indicators, especially disposable income of households and GFCF. When regionally allocating disposable income of households, one needs to allocate all household income according to their place of residence (more precisely, mid-period population published by the Czech Statistical Office by regions), while, for example, wages are monitored in statistical surveys according to the place of work, possibly also the headquarters of the institutional unit. The distribution of households' income from the place of work to the place of residence thus requires the application of estimates and the linking of multiple data sources, which are usually not combined in other statistics.

In the case of regional GFCF, one is mainly confronted with the problem of correct territorial recording. First, there is the problem of central purchases of multi-regional units in the headquarters residence, when later the distribution of individual fixed assets to other regions may occur, which accounting units will not properly record in statistical surveys. Then there is also a problem of correct allocation in the case of large infrastructure projects that are carried out in several regions from one workplace (local unit) located in a single region. In this case, instead of statistical surveys for local units, information is provided directly by the transport infrastructure companies (Railway Administration or Road and Motorway Directorate) about actually implemented construction projects in the territory of individual regions. At this point, it is also worth noting that the GFCF does not enter into the calculation of officially published regional GDP (since it is not calculated by the expenditure approach).

Eurostat publishes also primary income of households and household disposable income by NUTS 2 regions for EU countries. Regional population is taken over from demographic statistics that is on one hand based on usual residency concept (census), on the other hand migration, births and deaths follow a permanent residency concept. The difference between the concepts is not negligible in several regions (Czech Statistical Office, 2021). With respect to the new EU Regulation on European statistics on population and housing (for details see EUR-Lex Document 52023PC0031), the upcoming census will be likely based on the permanent residency concept which may cause some differences between the count of people registered in and actually living in the region. Regional incomes by place of residence rely mainly on Statistics on Income and Living Conditions (SILC) that is a sample survey carried out by the Czech Statistical Office. Single Monthly Employer Report, which is being implemented nowadays, represents a new administrative data source that may substantially improve various statistical domains including regional accounts. Data on each single employee will be collected that allow detailed analysis of the labour market. In particular, information about employees' places of work, places of residence and paid salaries and wages may improve quality of regional GVA as well as regional household incomes. Above all, the new data source may represent a significant improvement in the regional allocation of wages and salaries as the most important component of disposable income of households (the limitations of which were mentioned above).

2 EXPENDITURE APPROACH TO REGIONAL GDP

2.1 General introduction to chapter

Expenditure approach to GDP is very well known in Economics as a sum of consumption (C), investment (I), government spending (G) and net exports (NX):

$$GDP = C + I + G + NX \tag{1}$$

In statistical language, the GDP by expenditure approach is defined as the sum of final consumption expenditure made by households, the government, and NPISH, plus gross capital formation and net exports (total exports minus total imports), see e.g. Gough (2017). Nevertheless, the above-mentioned approach is not the primary one, as most National Statistical Offices (hereinafter NSOs) rely primarily on the production approach to GDP at the national level, including the Czech Statistical Office[9]. Pursuant to ESA 2010 Regulation (Eurostat, 2013a), the NSO is obliged to publish GDP by production, expenditure, and income approaches for a respective member state, while regional indicators are limited to the production approach, employment, household income, and gross fixed capital formation. The lack of relevant data sources, particularly for foreign trade and capacities, prevents statistical authorities from compiling regional GDP using all approaches (Gough, 2017). It should be noted that international and interregional exports and imports would have to be distinguished, including travel (purchases in other regions).

Nevertheless, various users need that regional data for analysis. A team of researchers of the University of Economics and Business estimated the first results a decade ago (Kramulová and Musil, 2013). Several methods could be reused, while others have been replaced due to changes in the economy, data collection, and other improvements.

The expenditure method of GDP includes, from the point of view of national accounts, three main parts:

- Final consumption expenditure (Households, General Government, NPISH),
- Gross capital formation (Gross fixed capital formation, Changes in inventories, Acquisition of valuables less disposals) and
- External balance of imports and exports of goods and services.

Regionalisation of final consumption expenditure is based on our research. All expenditure groups of households need to have their regionalisation keys.

2.2 Final household consumption expenditures

Final household consumption expenditures form approximately 2/3 of total final consumption expenditures at the national level. They are divided into 13 divisions according to the COICOP[10] classification

[9] Detailed description of national accounts compilation in Czechia can be found on dedicated webpage: https://csu.gov.cz/gdp-national-accounts-reports-on-quality

[10] CZ-COICOP or CZ-COICOP 18 – Classification of Individual Consumption According to Purpose, for detailed structure see https://apl2.czso.cz/iSMS/en/klaslist.jsp?fakronym=CZ-COICOP

and further categorised into 52 groups. We use the version COICOP 18, which is valid as of January 1, 2024. It is slightly different from the previous version. National accounts are already compiled in this classification, whereas HBS – one of our main data sources – is published in the previous ECOICOP classification. Therefore, we had to transform the data into the new classification using a correspondence table between ECOICOP and COICOP 18. In the previous classification, there were 12 divisions; now there are 13 divisions for households' expenditures and newly also one for NPISH and one for general government. These divisions are hierarchically split into groups, which is the level we are using for our computations. Instead of the previous 47 groups, there are now 52 for households. Some have just moved, some are new, one has been removed, and some are nested under a different division. Households' expenditures can be recorded either in the national or domestic concept. The former refers to all expenditures spent by a resident household whereas the latter refers to all spending in a given territory by resident and non-resident households. The national concept is used in sector accounts allowing a comparison of income expenditures and a calculation of savings. As regards regional accounts, the 'national' concept consists of spending in a particular region, spending in other regions of Czechia and abroad. Following the choice of the concept of households' expenditures the coherent concept of foreign trade, in particular tourism, is applied. For instance, expenditures of citizens of Středočeský kraj in Hlavní město Praha are booked as imports and households' expenditures in Středočeský kraj while as output and exports in Hlavní město Praha.

Concerning the regionalisation keys, some are derived from regional national accounts, but most of them are from HBS. From our last estimates, there is a substantial change that HBS switched from a quote sample to a random sample in the year 2017 (Czech Statistical Office, 2017). The majority of countries use for HBS random sample as well, from the EUROSTAT quality report in 2015 it is evident that in European Union "*[e]xcept for CZ and DE (which resort to quota sampling) all the HBS samples were selected according to a probability sampling scheme.* " (Eurostat, 2015, p. 9). Now, the only country with a quota sample in the EU remains Germany. Ayyash and Sek (2020) investigating Malaysian household consumption expenditure also used data from Household Expenditure Survey based on a probability sampling design. The switch to a random sample presents some challenges for us, as it introduces new attributes and pros and cons. The regional data (for NUTS 2 regions, which we must transform to the NUTS 3 regional level) we used are also no longer publicly available; earlier, it was available for download. Together with the change to a random sample, the regional structure may be less reliable. This is quite common, e.g. in Italy there are regional HBS data available for NUTS 1 and NUTS 2 levels, however no more for NUTS 3 or LAU 1 administrative level due to low reliability (Marchetti and Secondi, 2017). Other data sources such as health statistics or ministry data remain more or less the same.

After applying the keys to each of 52 groups of household expenditure, we obtained the weights, that were used to regionalise the national values. As we already mentioned above, some groups are completely new – e.g. groups 01.3 Services for processing primary goods for food and non-alcoholic beverages, 02.2 Alcohol production services, 06.4 Other health services, 07.3 Passenger transport services, 09.5 Cultural goods or 09.6 Cultural services requiring new keys to be assigned. For some groups, we tried to adjust previously assigned keys to better reflect economic reality. One example – for the groups 07.1 Purchase of vehicles and 7.2 Operation of personal transport equipment, data from the Ministry of Transport's yearbook on registered cars and motorcycles was previously used, but because there were doubts about the consistency of data on the region of registration, region of purchase and region of subsequent operation, this time we used data from HBS, which we consider more relevant. Some slight additional corrections were used. The results are in Table A1 in the Annex.

2.3 Government consumption expenditure

Government consumption expenditure can be divided into collective and individual parts. The former is sourced from government non-market output, while the latter is sourced from either government non-market output or purchased market production. Consequently, government consumption expenditure is regionalised consistently with the production approach, i.e., compensations of employees, assuming that other non-market components (intermediate consumption, consumption of fixed capital, and net taxes on production) are distributed alike (Kahoun and Sixta, 2013). Regarding purchased market production, the most important element is health, including health services and pharmaceutical products. The latter is allocated to regions using value of prescriptions and vouchers that is an improvement compared to the previous estimate where hospital consumption of pharmaceutical products was used. The former is allocated according to hospital revenues. As Lequiller and Blades (2007, p. 248) say, social benefits in kind (D.631) "*include repayments from the health sector of social security payment of household and medical assistance in the home and housing allowances*". The remaining part (less than 20%) is distributed pro rata, assuming the same consumption per capita.

2.4 Consumption expenditure of NPISH

Consumption expenditure of NPISH represents a negligible component of GDP. Similarly to government consumption expenditure it consists of non-market output produced by NPISH and purchased market output. Regionalisation is carried out analogously.

2.5 Foreign trade

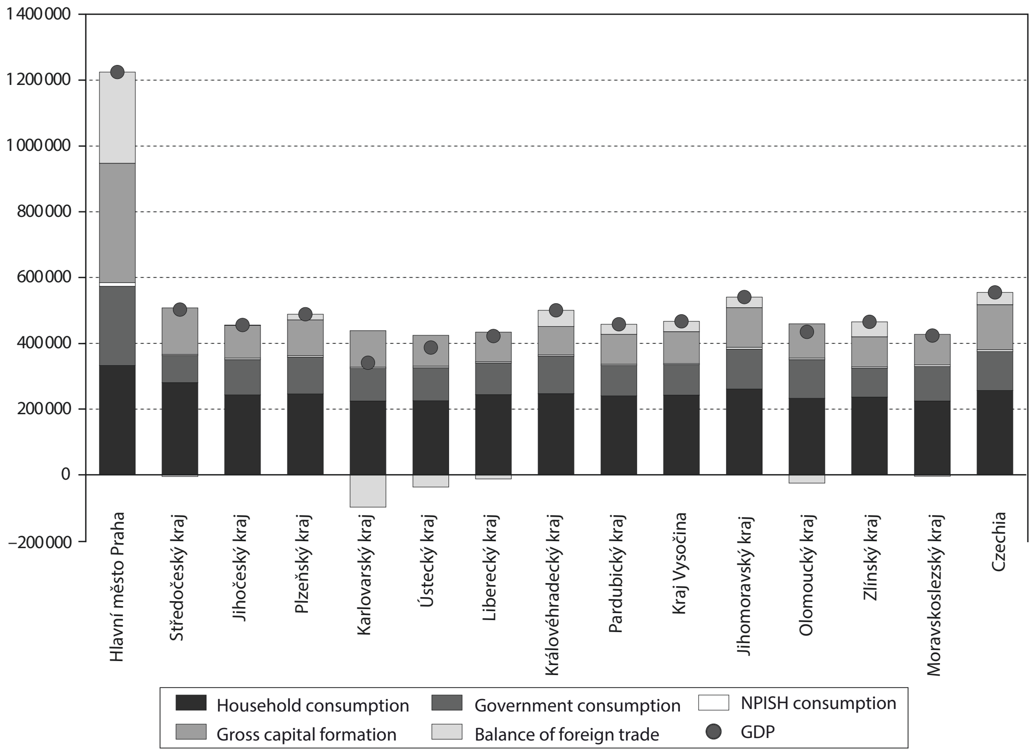
Exports of a region consists of international exports of goods and services and intraregional exports of trade and services. Those components can be hardly estimated as international trade is realised also by foreign companies and traders located in other countries or regions. Additionally, trade in services includes also international and intraregional tourism. Total imports can be divided similarly. Consequently, the balance of foreign trade is calculated as the difference between the official regional GDP published by the Czech Statistical Office and all other expenditure components. Net exports are therefore derived residually rather than estimated directly. Potential further disaggregation is discussed by Holý and Šafr (2022).

2.6 Results

The highest GDP per capita is registered in the capital city, where economic performance per capita is more than two times higher than the national average, while differences among other regions are substantially smaller. This is a typical picture of many countries where GDP per capita is the highest in the region including the capital city. Alves et al. (2021) suggest that cities generate wealth with the help of their own population and the population of workers attracted from neighbouring cities. GDP per capita of regions may suffer from inconsistency of nominator and denominator of the indicator (see also Dunnell, 2009). GDP in a given region may be produced using 'domestic' resources, in particular people living in a region in question, but also commuters from other regions. Likewise, residents encompassed in denominator may contribute to other regions' GDP. The balance of commuting to work of Hlavní město Praha is about 180 thousand people mainly with Středočeský kraj (surroundings of the capital city, Czech Statistical Office, 2023). Assuming the same productivity of commuters and residents, the contribution of employed persons commuting to Hlavní město Praha is estimated as GDP per employed person in the place of work multiplied by the number of commuting persons. The estimate should be interpreted as illustrative, as it relies on the simplifying assumption that commuters and residents exhibit the same average productivity. Differences in occupational structure, industry composition, or skill levels between the two groups may affect the resulting values. The result suggests that about ¼ GDP is created by commuters while negative contribution of residents working in other regions is about 3%. 'Modified'

GDP generated with 'domestic' resources exhibits lower variability among regions. The highest value is still observed in Hlavní město Praha, where GDP per capita is 73% above the national average. Besides the level of GDP, the structure of GDP varies across Czech regions. Household consumption expenditure represents only 27% GDP in Hlavní město Praha while it constitutes 2/3 of GDP in Karlovarský kraj. In other regions, the share of household consumption fluctuates around half of GDP. Government consumption expenditure constitutes about one fifth of GDP in Czech regions, nevertheless the amount in nominal term is two times higher in Hlavní město Praha than national average. Net exports are positive in most regions, but several regions registered deficits, with the highest deficit in Karlovarský kraj. The leading net exporter is Hlavní město Praha. It may be assumed that expenditures by foreign tourists as well as Czech tourists and commuters notably contribute to net exports in the capital city.

Figure 1 Regional GDP by expenditure approach per capita (in CZK)



Source: Czech Statistical Office, own calculations.

3 INCOME APPROACH TO REGIONAL GDP

The income method represents the aggregation of variables of the income generation account: compensations of employees, gross operating surplus and net taxes on production and imports (excluding subsidies) at the regional level. Information on compensations of employees by industry is obtained from official regional accounts statistics. Compensations of employees are concurrently used in combination with the production method indicators to estimate regional gross value added. However, the availability of information on gross operating surplus and mixed income is the biggest weakness of the regional GDP calculation by income method. Operating surplus is not directly calculated at the national level let alone broken down by region, thus complicating the use of the full income method for estimating regional

GDP. There is not only the issue of data availability, but also the issue of where actual operating surplus or profit is created (if it is only in the headquarters of companies or individual local units in regions, and what is the share of local units on company operating surplus).

Net taxes on production and imports consist of *net taxes on products* (such as VAT) and *net other taxes on production* (such as income from sales of emission trading permits). An allocation of *net taxes on products* is made in official regional accounts for GDP compilation according to the regional structure of total regional GVA. For *net other taxes on production,* the allocation into the regions can be done in the same way, i.e. in the proportion of total regional GVA.

3.1 Regional allocation of individual GDP components of the income approach

The calculation of regional GDP for Czechia is mostly carried out using the production method at the Czech Statistical Office. However, the regional structure of GVA in the case of multi-regional organisations (bodies with operations in several regions) is based on the regional structure of wages and salaries paid by these organisations. The production method of calculation is thus combined with the compensations of employees as a key indicator of the income method. The reason is that compensations of employees is easily ascertainable by region and, while indicators characterising the production method, i.e. production and intermediate consumption, are easily available by enterprises from business survey. Therefore, the regional structure of wages serves as a key for the regionalisation of GVA (output minus intermediate consumption) for multi-regional organisations.

For the calculation of regional GVA directly by the income approach, it is necessary to work beside compensations of employees as well as with reliable data on regional consumption of fixed capital (CFC) and data on regional net operating surplus (OS), which is another essential part (the most difficult to detect) of the income components of GVA.

The experimental calculation of the income components of regional GVA is made below on the example of data from the regional accounts of Czechia for the year 2020. It follows the same experimental calculation for 2011 carried out last time in 2013 (see Kahoun and Sixta, 2013).

Assuming that the results of production and income approaches must be balanced, the operating surplus and mixed income is equal to the difference between the total GVA and the remaining components of the income approach (compensations of employees, consumption of fixed capital and net other taxes on production and imports). The formula for the regional calculation of operating surplus and mixed income is so determined:

$$OS = GVA - W - CFC - (T - S) \tag{2}$$

where:

OS – *regional net operating surplus and mixed income*
GVA – *regional gross value added (available in regional accounts)*
W – *compensations of employees paid by local units (available in regional accounts)*
CFC – *regional consumption of fixed capital (to be estimated)*
T – S – *regional other taxes on production and imports less subsidies (to be estimated)*

Regional GVA is known from regional accounts, as well as compensations of employees, but there is also necessary to make estimates for regional consumption of fixed capital and regional net taxes on production and imports.

Regional CFC (K.1 according to ESA 2010) based on local units is not directly calculated in Czechia and its regional values are not available anywhere. Therefore, it is necessary to find a suitable key for a regional distribution of national values using the top-down method. The available statistical indicator of regional accounts regularly published is gross fixed capital formation (GFCF). Between 2002 and 2004, the Structural Business Survey in Czechia included a specific section on stocks of fixed assets by region. Data on fixed assets since 2004 and their regular update on the new GFCF and estimated depreciation of assets until 2020 can therefore serve as a basis for an experimental calculation of the CFC. Only total assets acquired can be depreciated and the regional structure of total fixed assets should therefore be close to the regional structure of the estimated indicator. The estimated structures of regional fixed assets in 2020 are used as weights for the regional top-down distribution of CFC from national accounts for 2020. It should be noted that the estimates rely on historical information on regional fixed assets updated by subsequent investment flows, which may not fully capture all structural changes in the regional distribution of assets over time.

A relatively small remaining item is the indicator of net other taxes on production and imports (D.29 minus D.39 according to ESA 2010). Similarly to the case of net taxes on products (D.21 minus D.31), regional allocation can be made according to the regional structure of total gross value added. However, the regional allocation of net other taxes on production is made for each two-digit NACE separately, which allows considering the different meanings of subsidies in each NACE – e.g. a significant part of subsidies is concentrated in energy or agriculture. The remaining value to total gross value added represents net operating surplus (including mixed income).

While the methodology of regional allocation and the data sources used in the current calculation were the same, the interpretation of items slightly changed due to the inclusion of indicators, which were revised with the revisions of national accounts since 2013. Compensations of employees and net taxes on production and imports changed only very little in official statistics after the national accounts' revisions. More significant methodological changes occurred in the consumption of fixed capital, which mainly reflected the new valuation of fixed assets in the national accounts. The revision in 2014 represented the transition to the updated standard ESA 2010. In fixed assets there were reflected mainly the changes in capitalisation of research and development or capitalisation of small tools previously not included. The revisions of the national accounts in 2020 and 2024 had an impact on the valuation of dwellings and other buildings and structures.

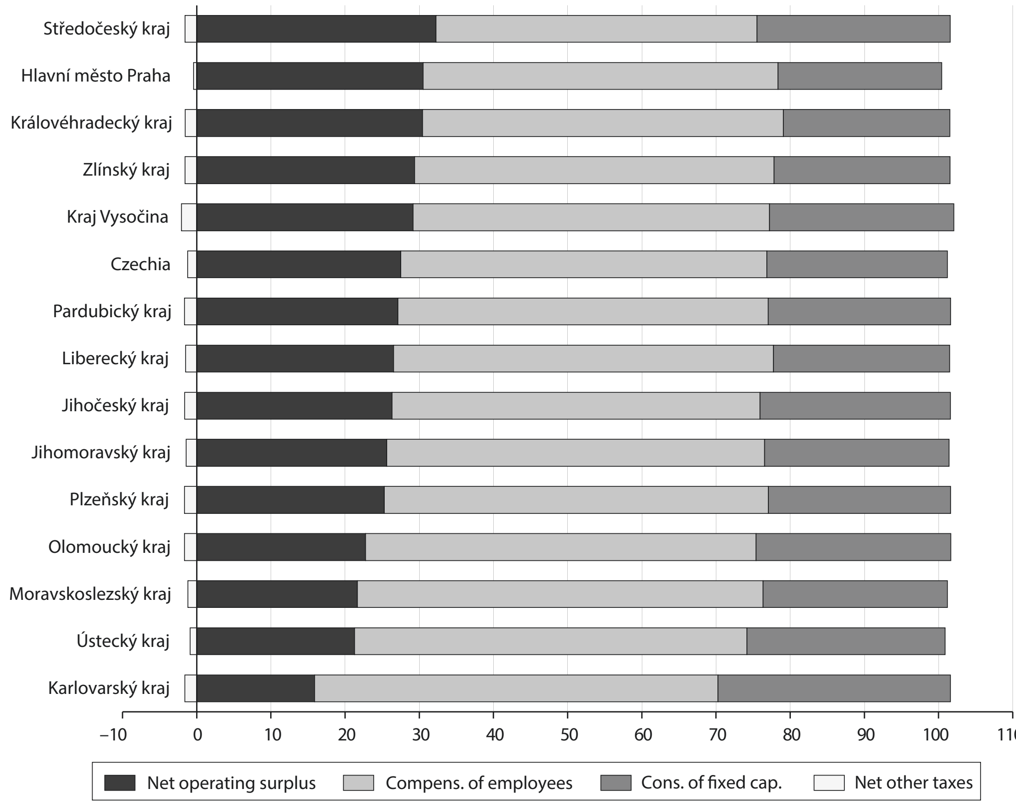
In regional weights used for this experimental calculation, the revisions of national accounts are reflected in the form of data on the regional GFCF applied into the distribution of regional total fixed assets, and therefore indirectly also in the allocation of CFC.

3.2 The resulting regional allocation of the operating surplus

Figure 2 illustrates the structure of GVA, broken down by individual components, ranked by the share of operating surplus in gross value added in 2020. The results logically indicate a relatively higher share of net operating surplus in regions with a lower share of compensations of employees or a lower share of CFC on GVA (Středočeský kraj, Hlavní město Praha and Královéhradecký kraj). In particular, a historically lower level of wages and salaries may strengthen the share of operating surplus, if the region is currently developing economically. On the other hand, a historically higher level of wages and salaries may currently be limiting for the operating surplus (Moravskoslezský, Ústecký), especially if the region is in a phase of economic slowdown.

Compared to the calculations made in 2013 for 2011, the share of the regions on national net operating surplus decreased significantly for Karlovarský kraj (–0.8 p.p.), Ústecký kraj (–2.3 p.p.) and Moravskoslezský kraj (–3.8 p.p.). On the other hand, the share increased in Hlavní město Praha (+ 5.2 p.p.) and Středočeský (+ 1.9 p.p.) kraj.

Figure 2 Structure of GVA by NUTS 3 regions, items of the income approach in %, year 2020



Note: Ranking by the share of the estimated net operating surplus on GVA.
Source: Czech Statistical Office, own calculations.

In Table 1, the last column shows the regional structure of net operating surplus, which is based on the assumption of equilibrium of total GVA known from regional accounts and sum of individual components of the income method. However, it should be mentioned that the calculation of regional GVA for multiregional organisations in regional accounts is based on the regional structures of wages and salaries of individual companies. This means that the remaining components of the income calculation of GDP are also distributed for multiregional organisations in the structure of wages and salaries (including their operating surpluses). Therefore, the experimental calculation is relatively simplified and carried out only with regard to the absence of the necessary regionally broken-down data. At the same time, it can be assumed that operating surpluses are more important in the cases of multiregional organisations (energy companies, financial intermediaries, telecommunications companies etc.).

The main challenge for the regional allocation of income components of GVA remains in finding a suitable key for the regional distribution of operating surplus. This is what mainly complicated further progress towards the full use of the income method, and therefore, the calculation of regional GDP in this way is not implemented in most countries, or only for a selected part of GVA.

Table 1 Regional estimated structures of the income components of GDP in %, 2020

Regions of Czechia, 2020	Compensations of employees	Consumption of fixed capital	Net other taxes	GVA	Net operating surplus
ESA codes	**D.1**	**K.1**	**D.29-D.39**	**B.1g**	**B.2n**
Hlavní město Praha	25.8	24.1	9.4	26.6	29.6
Středočeský kraj	10.4	12.7	15.5	11.9	14.0
Jihočeský kraj	5.0	5.2	6.7	5.0	4.8
Plzeňský kraj	5.1	4.9	6.6	4.8	4.4
Karlovarský kraj	1.8	2.1	2.2	1.7	1.0
Ústecký kraj	5.7	5.8	3.9	5.3	4.1
Liberecký kraj	3.3	3.1	3.9	3.2	3.1
Královéhradecký kraj	4.6	4.3	5.9	4.7	5.2
Pardubický kraj	4.1	4.1	5.5	4.0	4.0
Kraj Vysočina	3.9	4.1	6.9	4.0	4.3
Jihomoravský kraj	11.3	11.2	12.9	11.0	10.2
Olomoucký kraj	5.0	5.0	6.4	4.7	3.9
Zlínský kraj	4.5	4.5	5.9	4.6	4.9
Moravskoslezský kraj	9.5	8.8	8.4	8.6	6.8
Czechia	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**

Source: Czech Statistical Office, own calculations.

4 DISCUSSION – WHY IS THIS TOPIC IMPORTANT?

GDP is one of the most important indicators of national accounts; however, it is sometimes misused. Users should keep in mind that GDP is an indicator of economic performance, but not of the wealth or well-being of households, which is rather measured by household consumption expenditure (see Secondi, 2021, p. 652). Stiglitz et al. (2009) pointed out that other indicators focused on households should be promoted or developed. The highest GDP per capita in purchase power standard (PPS) in the EU was recorded in Luxembourg (242% EU average in 2023) and Ireland (213%). The latter is mainly caused by operations of globally active firms in Ireland (Lane, 2017). The former case is very well known, as the headquarters of many financial corporations and EU institutions are seated in Luxembourg, a very small country. In addition, commuters from neighbouring countries contribute to Luxembourg's GDP. Such cities as e.g. Luxembourg, Basel, Geneva, Vienna are attracting qualified foreign labour (for details see Decoville et al., 2013). In these cities, the issue of the different structure of cross-border consumption expenditure also arises (Math*ä* et al., 2014). However, this is probably not the case of Hlavní město Praha (and Czechia) in such extent, also due to the language or common consumption expenditure habits of commuters from neighbouring countries (e.g. Slovakia).

Therefore, a headline indicator of international comparison published by Eurostat is actual individual consumption expressing amount of goods and services actually consumed by households. The highest value is likewise the highest in Luxembourg (140% of the EU average in 2023), which is substantially lower than the GDP. Similarly, the highest regional GDP per capita in PPS is recorded in regions, where large companies are located: South-West (Ireland), Dublin or Wolfsburg. GDP per capita in PPS is more than three times higher than the EU average in these areas. Obviously, GDP in small areas can be affected by local circumstances that are not necessarily reflected in household consumption.

Above mentioned cases pose challenges to interpretation of GDP in Ireland or Luxembourg to users that might not be familiar with those phenomena. Regional GDP may be affected to a large extent. Various public policies have been introduced to support so called structurally affected regions or in general to diminish regional differences. Progress could be assessed using a wide range of indicators, nevertheless regional national accounts fail in providing a complete macroeconomic picture of regions. The results presented in the paper may help policy makers to target public support to appropriate regions and area of funding. In particular, programmes focused on households may benefit from a consistent picture of regional incomes, expenditures and savings depicting living standards.

It should be noted that indicators have conceptual limits that are negligible at the national level but relevant at the regional level. Firstly, ESA 2010 stipulates that non-market output is consumed by government and NPISH and cannot be exported. Ministries and other government agencies are mainly located in Hlavní město Praha where non-market output as well as government consumption expenditure are booked even though citizens in other regions also benefit from their activities. Notwithstanding, the next ESA may allow to record international trade in government services.

Regional indicators published by NSOs and Eurostat are valued either in national currency/Euro or in PPS, i.e. no adjustment for a difference in regional prices levels is made. Gough (2017) pointed out that an assumption of the same price level in a given country is not held in many countries. Čadil et al. (2014) proved that this is also the case in Czechia. The updated estimates are presented below in Table 2. Regional price levels are based on final household consumption expenditure; considerable differences are recorded in housing, personal and restaurant services. Unsurprisingly, the highest prices were recorded in Hlavní město Praha, Středočeský kraj (a region surrounding Hlavní město Praha) and Jihomoravský kraj, including the second largest city Brno. Nominal household expenditure is revalued to the regional purchase power standard (RPPS), eliminating variation in prices among Czech regions. Differences in consumption in RPPS are substantially smaller than for nominal indicators.

Table 2 Impact of regional price levels on household consumption expenditures, 2020

Region	Household consumption expenditures per capita, %	RPPP	Household consumption expenditures per capita in RPPS, %
Hlavní město Praha	129.9	124.6	104.3
Středočeský kraj	109.6	108.2	101.3
Jihočeský kraj	94.9	98.1	96.8
Plzeňský kraj	96.2	98.3	97.8
Karlovarský kraj	87.7	97.6	89.8
Ústecký kraj	87.9	95.8	91.8
Liberecký kraj	95.1	100.5	94.7
Královéhradecký kraj	96.3	98.1	98.2
Pardubický kraj	93.6	97.0	96.5
Kraj Vysočina	94.5	95.4	99.0
Jihomoravský kraj	102.0	103.1	98.9
Olomoucký kraj	90.8	95.5	95.1
Zlínský kraj	92.3	96.9	95.3
Moravskoslezský kraj	87.7	94.5	92.8
Czechia	**100.0**	**100.0**	**100.0**

Source: Own calculations.

CONCLUSION

The Czech Statistical Office publishes just the required set of regional accounts' indicators specified in the ESA 2010 Transmission Program. The aim of this paper was to extend and complement regional indicators, in particular RGDP using all three methods – production, expenditure and income. The results show that despite Czechia being a relatively small country, the differences among regions are quite substantial. It is evident that regions surrounding or containing the capital city perform differently from remote regions. The distinction occurs in the expenditure approach (e.g., a share of housing costs or food and non-alcoholic beverages), and disparities also exist in the structure of income components of GVA. The officially published data on GDP suggest that economic performance in Hlavní město Praha is more than twice as high as in other relatively homogeneous regions. The presented results shed light on economic reasoning, especially the demand side of economy.

All the calculations were done for the year 2020 that was partly affected by COVID-19 pandemic; however, it is intended to update the results for the more suitable year 2023. Especially the expenditures were probably highly influenced (e.g. lockdowns and related government containment measures). In the case of the income method, the biggest challenge remains in the choice of an appropriate key for regional allocation of operating surplus for multiregional companies.

Finally, it is necessary to emphasize and remind that most of the values of RGDP are not recalculated to respective regional price levels, which worsens the right interpretation. Table 2 shows, that after taking regional price levels into account, the results may significantly differ. Household consumption expenditures variance among Czech NUTS 3 regions decreased considerably and the same could be naturally in the case of RGDP. Use of the real RGDP can help to consecutively assess the tendency of convergence or divergence of regions and to take political decisions in order to improve the economic performance of the territories. However, the vast topic of regional price levels and the broad consequences of this recalculation exceeds this paper and will be part of our further research and discussed and published thereafter.

ACKNOWLEDGMENT

This paper is prepared under the support of the Technology Agency of the Czech Republic within the Project No. SS04030013 (Center for Socio-Economic Research on Environmental Policy Impact Assessment).

References

ATEN, B., D'SOUZA, R. J. (2008). Regional Price Parities: Comparing Price Level Differences Across Geographic Areas. *Survey of Current Business*, 88(11): 64-74.

ALVES, L. G. A., RYBSKI, D., RIBEIRO, H. V. (2021). Commuting network effect on urban wealth scaling [online]. *Scientific Reports*, 11, 22918. <https://doi.org/10.1038/s41598-021-02327-7>.

AYYASH, M., SEK, S. K. (2020). Decomposing Inequality in Household Consumption Expenditure in Malaysia [online]. *Economies*, 8(4): 1-15. <https://doi.org/10.3390/economies8040083>.

BILLÉ, A. G., TOMELLERI, A., RAVAZZOLO, F. (2023). Forecasting regional GDPs: a comparison with spatial dynamic panel data models [online]. *Spatial Economic Analysis*, 18(4): 530–551. <https://doi.org/10.1080/17421772.2023.2199034>.

CZECH STATISTICAL OFFICE. (2017). *Household Budget Survey (HBS) – Methodology since 2017* [online]. Prague: Czech Statistical Office. [cit. 19.11.2025] <https://csu.gov.cz/household-budget-survey-hbs-methodology>.

CZECH STATISTICAL OFFICE. (2021). *Obvyklý a registrovaný pobyt ve sčítání lidu* [online]. Prague: Czech Statistical Office. [cit. 6.2.2026]. <https://csu.gov.cz/docs/107508/3141b8b4-14de-61f7-20a8-17e7bd78cce7/17101025.pdf?version=1.0>.

CZECH STATISTICAL OFFICE. (2023). *Dojížďka mezi obcemi včetně denní dojížďky (data a schema)* [online]. Prague: Czech Statistical Office. [cit. 9.4.2026]. https://csu.gov.cz/produkty/vysledky-scitani-2021-otevrena-data

CZECH STATISTICAL OFFICE. (2025). *Gross domestic product (GDP) – Methodology* [online]. Prague: Czech Statistical Office. [cit. 20.8.2025]. <https://csu.gov.cz/gross_domestic_product_-gdp-methodology>.

ČADIL, J., MAZOUCH, P., MUSIL, P., KRAMULOVÁ, J. (2014). True regional purchasing power: evidence from the Czech Republic. *Post-Communist Economies*, 26(2): 241–256. < https://doi.org/10.1080/14631377.2014.904109>.

DECOVILLE, A., DURAND, F., SOHN, C., WALTHER, O. (2013). Comparing Cross-border Metropolitan Integration in Europe: Towards a Functional Typology [online]. *Journal of Borderlands Studies*, 28(2), 221-237. <https://doi.org/10.1080/08865655.2013.854654>.

DUNNELL, K. (2009). National Statistician's article: measuring regional economic performance [online]. *Economic & Labour Market Review*, 3(1), 18–30. <https://doi.org/10.1057/elmr.2009.6>.

EUROSTAT. (2013a). *ESA 2010: European System of Accounts.* Luxembourg: Publications Office of the European Union. <https://doi.org/10.2785/16644>.

EUROSTAT. (2013b). *Manual on regional accounts methods. Luxembourg: Publications Office of the European Union*. <https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-gq-13-001>.

EUROSTAT. (2015). *Household Budget Survey – 2015 Wave – EU Quality Report (Version 1)*. [online] Luxembourg. [cit. 19. 11. 2025]. <https://ec.europa.eu/eurostat/documents/54431/1966394/HBS_EU_QualityReport_2015.pdf>.

EUROSTAT. (2026). *Towards ESA 2030.* [online] Luxembourg. [cit. 30. 06. 2026]. <https://ec.europa.eu/eurostat/web/esa-2010/towards-esa-2030>.

GOUGH, J. (2017). New Estimates of Regional GDP in the UK. *World Economics*, 18(2): 1-20.

HOLÝ, V., ŠAFR, K. (2022). Disaggregating input–output tables by the multidimensional RAS method: a case study of the Czech Republic [online]. *Economic Systems Research*, 35(1): 95–117. <https://doi.org/10.1080/09535314.2022.2091978>.

HRONOVÁ, S. (2024). Current Issues in the Development of the National Accounts System [online]. *Statistika: Statistics & Economy Journal*, 104(3): 371–375. <https://csu.gov.cz/docs/107508/21655648-5772-9bd1-13fa-574902578e74/32019724q3_hronova_information.pdf?version=1.0>.

KAHOUN, J. (2009). Metoda výpočtu regionálního HDP v České republice. *Statistika*, 46(6): 518–530.

KAHOUN, J., SIXTA, J. (2013). Regional GDP Compilation: Production, Income and Expenditure Approach [online]. *Statistika: Statistics & Economy Journal*, 93(4): 24–36. <https://csu.gov.cz/docs/107516/a86cf82c-d088-deb5-cd47-81604a6babbd/180213q4024.pdf?version=1.0>.

KRAMULOVÁ, J., MUSIL, P. (2013). Experimentální odhad složek výdajové metody regionálního HDP v ČR [online]. *Politická ekonomie*, 61(6): 814–833. <https://polek.vse.cz/pdfs/pol/2013/06/06.pdf>.

LANE, P. (2017). The Treatment of Global Firms in National Accounts. *Economic Letter Series*, Central Bank of Ireland, 2017 (1): 1–6.

LEQUILLER, F., BLADES, D. (2007). *Understanding National Accounts*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264027657-en>.

MARCHETTI, S., SECONDI, L. (2017). Estimates of Household Consumption Expenditure at Provincial Level in Italy by Using Small Area Estimation Methods [online]. *Social Indicators Research* 131(1): 215–234. <http://doi.org/10.1007/s11205-016-1230-8>.

MATHÄ, T. Y., PORPIGLIA, A., ZIEGELMEYER, M. (2014). Cross-border commuting and consuming: An empirical investigation. *ECB Working Paper Series.* No. 1661.

MORILLAS-JURADO, F., RICO-BELDA, P., CABRER-BORRAS, B. (2026). Does the location of company headquarters affect regional GDP? [online]. *Investigaciones Regionales - Journal of Regional Research*, 64(1): 67-89. <https://doi.org/10.38191/iirr-jorr.26.001>.

ROKICKI, B., HEWINGS, G.J.D. (2016). Regional convergence within particular country — An approach based on the regional price deflators [online]. *Economic Modelling*, 57(C): 171-179. <https://doi.org/10.1016/j.econmod.2016.04.019>.

SECONDI, L. (2021). Estimating Household Consumption Expenditure at Local Level In Italy: The Potential of the Cokriging Spatial Predictor. [online]. *Social Indicators Research*, 153(2): 651–674. <https://doi.org/10.1007/s11205-020-02510-9>.

SIXTA, J., VLTAVSKÁ, K. (2016). Regionální input-output tabulky: praktické aspekty jejich sestavování pro kraje České republiky. *Ekonomický časopis*. 64 (1): 56-69.

STIGLITZ, J.E., SEN, A. and FITOUSSI, J.P. (2009). *Report by the commission on the measurement of economic performance and social progress* [online]. Paris: Commission on the measurement of economic performance and social progress. [cit. 12.11.2025]. <https://ec.europa.eu/eurostat/documents/8131721/8131772/Stiglitz-Sen-Fitoussi-Commission-report.pdf>.

ANNEX / APPENDIX

Table A1 Final household consumption expenditures in % according to NUTS 3 region and COICOP18 classification

ANNEX / APPENDIX	Hlavní město Praha	Středo-český kraj	Jiho-český kraj	Plzeň-ský kraj	Karlo-varský kraj	Ústecký kraj	Libe-recký kraj	Králové-hradecký kraj	Pardu-bický kraj	Kraj Vysočina	Jiho-moravský kraj	Olomoucký kraj	Zlínský kraj	Moravsko-slezský kraj	Czechia
01 – Food and non-alcoholic beverages	12.88	15.97	17.50	17.27	17.41	17.37	16.48	16.28	16.74	17.40	16.65	17.74	17.46	17.55	**16.31**
02 – Alcoholic beverages, tobacco and narcotics	7.06	7.63	8.24	8.21	8.54	8.72	8.45	8.27	8.43	7.76	7.51	8.76	8.62	9.88	**8.13**
03 – Clothing and footwear	3.33	3.19	3.70	3.65	3.60	3.60	3.26	3.22	3.31	3.59	3.52	3.50	3.45	3.48	**3.43**
04 – Housing, water, electricity,gas and other fuels	38.03	38.34	32.66	32.50	31.25	31.63	32.78	32.84	31.75	31.58	33.93	30.92	31.62	29.90	**33.86**
05 – Furnishings, household equipment and routine household maintenance	4.44	4.23	5.26	5.21	4.86	4.85	5.34	5.35	5.48	5.76	5.49	5.72	5.63	4.97	**5.02**
06 – Health	1.97	1.61	2.13	2.19	2.61	2.25	2.12	2.33	2.13	2.19	2.20	2.55	2.54	2.49	**2.15**
07 – Transport	8.21	7.64	8.83	8.90	8.80	8.61	8.50	8.62	8.71	8.86	8.09	8.03	7.82	7.98	**8.26**
08 – Information and communication	4.11	4.17	4.86	4.95	5.60	5.53	4.98	4.82	4.80	4.56	4.52	5.09	4.78	5.38	**4.72**
09 – Recreation, sport and culture	5.76	5.95	5.07	5.00	5.54	5.68	5.89	5.86	6.02	5.82	5.65	5.60	5.51	5.67	**5.68**
10 – Education services	0.81	0.51	0.41	0.36	0.36	0.37	0.37	0.39	0.40	0.45	0.40	0.42	0.42	0.44	**0.48**
11 – Restaurants and accommodation services	5.77	4.15	3.94	3.88	3.74	3.73	4.82	4.79	4.92	4.92	4.71	4.34	4.27	5.02	**4.64**
12 – Insurance and financial services	4.20	2.66	3.82	4.32	3.78	3.48	3.73	3.86	3.81	3.41	3.76	3.34	3.98	3.38	**3.63**
13 – Personal care, social protection and miscellaneous goods and services	3.43	3.95	3.59	3.54	3.92	4.18	3.29	3.36	3.50	3.71	3.56	3.99	3.91	3.86	**3.70**
Total	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**

Source: Czech Statistical Office, own calculations.

Evolution of the GEKS index

Jacek Białek[1] | *University of Lodz, Łódź, Poland*

Received 27.10.2025 (revision received 2.3.2026), **Accepted** (reviewed) 10.3.2026, **Published** 11.9.2026

Abstract

The GEKS index is a well-known multilateral index used by many statisticians to measure inflation based on scanner data. As a rule, it is assumed that the underlying index within the GEKS formula satisfies the time reversal test. Most often, this underlying index is assumed to be superlative, and therefore, the GEKS index based on the Fisher, Törnqvist (GEKS-T or CCDI) or Walsh (GEKS-W) formulas are most often considered. However, the 'classic' GEKS index does not meet the stringent identity test. Unfortunately, most of the known multilateral indices do not meet the time reversal test, which many researchers consider a weakness. This paper reviews both well-known and lesser-known modifications and generalizations of the GEKS index. We discuss three new and general classes of indices based on the GEKS method, as well as some special cases of these classes, which include the GEKS, GEKS-T, and GEKS-W indices and, additionally, the GEKS-L and GEKS-GL indices. One class uses elasticity of substitution, so it is close to the economic approach, while another class - like the multilateral Geary-Khamis index- uses quality-adjusted prices and quantities. Not all the GEKS modifications discussed require the underlying index to meet the time reversal test. It should be noted that in cases where this assumption has been dropped, an identity test has been gained. We present the basic axiomatic properties of the proposed indices and compare them empirically based on real and available scanner datasets. The simulation study verifies the impact of price and quantity volatilities on the differences between the price indices discussed.

Keywords	***DOI***	***JEL code***
inflation measurement, Consumer Price Index (CPI), scanner data, multilateral indices, identity test, GEKS method	*https://doi.org/10.54694/stat.2025.58*	*C43, E52*

INTRODUCTION

Among the multitude of challenges facing statistical offices that use scanner data to measure inflation, the choice of an appropriate price index formula is a key issue (see ABS (2016), CPI Manual (2004). In general, bilateral unweighted and weighted price indices, chain indices, and multilateral indices can be considered for this purpose. However, the choice of price index should lead to a reduction in both substitution bias and chain drift bias (Chessa, Verburg, and Willenborg, 2017). It is important to note that in literature, there is a set of criteria that can be taken into account when choosing a price index formula. These include the *axiomatic*, *economic* and *stochastic approaches* (Eurostat, 2022), as well as the complementary *time-consuming approach* and the *new stochastic approach* (Białek and Beręsewicz, 2021). The consensus among statisticians is that for scanner data, the most rational choice of price index

[1] Department of Statistical Methods, University of Lodz, Łódź POW 3/5 90-255 Łódź, Poland. E-mail: jacek.bialek@uni.lodz.pl, Statistics Poland, Department of Price and Services. E-mail: J.Bialek@stat.gov.pl.
ORCID: <https://orcid.org/0000-0002-0952-5327>.

is a multilateral index, which operates on a time window and thus takes into account the high turnover of scanned products (see Eurostat (2022) for details).

It is well known that GEKS-type indices (e.g., the GEKS, CCDI, and GEKS-W indices) have good axiomatic properties and are close to the *economic approach* since they are based on superlative indices. The GEKS index has been shown (Gini, 1931; Eltetö and Köves, 1964) to be exact for a "flexible" functional form, i.e., it expresses the price differences experienced by optimizing consumers without imposing restrictive assumptions about how they can substitute between products. Moreover, GEKS-type indices are relatively fast in computations, which can be of practical importance when dealing with large scanner datasets (Białek, 2022b). This paper reviews more or less well-known modifications and generalizations of the GEKS index. We discuss several general classes of indices based on the idea of the GEKS method and some special cases of these classes, which include GEKS, GEKS-T and GEKS-W indices and additionally GEKS-L and GEKS-GL indices. One class uses elasticity of substitution, so it is close to the economic approach, while another class - like the multilateral Geary-Khamis index - uses quality-adjusted prices and quantities. Not all the GEKS modifications discussed require the underlying index to meet the time reversal test. Note that where this assumption has been dropped, an identity test has been gained (see Theorems 1, 2, and 5). We present the basic axiomatic properties of the proposed indices, and in an empirical study, we compare them based on real scanner datasets. The simulation study verifies the impact of price and quantity volatilities on the differences between the price indices discussed.

It should be noted that GEKS-type indices are not the only multilateral indices designed for scanner data. For example, the Time-Product Dummy (TPD) index is based on the regression framework discussed in De Haan and Krsinich (2014), who analyzed scanner data and addressed the treatment of quality change in non-revisable price indexes using time and product dummy variables. Nevertheless, a discussion of other classes of multilateral price indices is beyond the scope of this study.

1 THE GEKS-TYPE INDEX IDEA

Let us denote sets of homogeneous products belonging to the same product group in the months 0 and t by G_0 and G_t, respectively, and let $G_{0,t}$ denote a set of matched products in both moments 0 and t. Let p_i^τ and q_i^τ denote the price and quantity of the i-th product at the time τ and let $N_{0,t}$ be the number of elements of set $G_{0,t}$.

The GEKS price index between the months 0 (the base period) and t (the current period) is an un-weighted geometric mean of the $T + 1$ ratios of bilateral price indices $P^{\tau,t}$ and $P^{\tau,0}$, which are based on the same price index formula. Typically, the GEKS method uses the superlative Fisher price index (Fisher 1922), resulting in the following formula:

$$P_{GEKS}^{0,t} = \prod_{\tau=0}^{T} \left(\frac{P_F^{\tau,t}}{P_F^{\tau,0}} \right)^{\frac{1}{T+1}}, \tag{1}$$

where $P_F^{\tau,t}$ denotes the Fisher price index calculated for products from the set $G_{\tau,t}$.

The GEKS method for making international index number comparisons was proposed by (Gini, 1931), although it was derived in a different manner by Eltetö and Köves (1964) and Szulc (1964). The CCDI price index, based on the Törnqvist (1936) index, can be expressed as follows:

$$P_{CCDI}^{0,t} = \prod_{\tau=0}^{T} \left(\frac{P_T^{\tau,t}}{P_T^{\tau,0}} \right)^{\frac{1}{T+1}}, \tag{2}$$

where $P_T^{\tau,t}$ denotes the Törnqvist price index calculated for products from the set $G_{\tau,t}$.

In Chessa, Verburg, and Willenborg (2017), a hint is provided for selecting a base index formula for the GEKS method: "the bilateral indices should satisfy the time reversal test." However, it is most often assumed that the price index formula found in the body of the GEKS index is a superlative formula (van Loon and Roels, 2018; Diewert and Fox, 2018). For this reason, a GEKS index based on the superlative Walsh (1901) index, which we will refer to as GEKS-W, is also often considered. It can be written as follows:

$$P_{GEKS-W}^{0,t} = \prod_{\tau=0}^{T} \left(\frac{P_W^{\tau,t}}{P_W^{\tau,0}} \right)^{\frac{1}{T+1}}, \tag{3}$$

where $P_W^{\tau,t}$ denotes the Walsh price index calculated for products from the set $G_{\tau,t}$.

2 SEMI GEKS INDICES

New GEKS-type price indices have recently appeared in the literature that challenge traditional assumptions associated with the underlying index found in the body of the GEKS index. These novel GEKS-type indices are based neither on a superlative price index nor on an index that meets the *time reversal test*. In Białek (2022b) and Białek (2022a), a general class of such indices (denoted by GS-GEKS) is proposed, and its two special cases, i.e., the GEKS-L and GEKS-GL index, are discussed (see Sections 3.1 and 3.2). Without elaborating further here, the papers cited demonstrate that under appropriate assumptions for the base formula, indices from the general class of semi-GEKS methods (GS-GEKS) satisfy many valuable tests (including the *identity test*).

2.1. The GEKS-L index

Let $P_L^{\tau,s}$ denote the Laspeyres (1871) price index, which compares period s to period τ using data from $G_{\tau,s}$. This index is defined as:

$$P_L^{\tau,s} = \frac{\sum_{i \in G_{\tau,s}} q_i^\tau p_i^s}{\sum_{i \in G_{\tau,s}} q_i^\tau p_i^\tau}. \tag{4}$$

Under this notation, the GEKS-L index can be defined as follows (Białek, 2022a):

$$P_{GEKS-L}^{0,t} = \prod_{\tau=0}^{T} \left(\frac{P_L^{\tau,t}}{P_L^{\tau,0}} \right)^{\frac{1}{T+1}} = \prod_{\tau=0}^{T} \left(\frac{\dfrac{\sum_{i \in G_{\tau,t}} q_i^\tau p_i^t}{\sum_{i \in G_{\tau,t}} q_i^\tau p_i^\tau}}{\dfrac{\sum_{i \in G_{\tau,0}} q_i^\tau p_i^0}{\sum_{i \in G_{\tau,0}} q_i^\tau p_i^\tau}} \right)^{\frac{1}{T+1}}. \tag{5}$$

The GEKS-L index can be treated as a generalization of the Fisher price index formula ($P_F^{0,t}$) to a multi-period case. Specifically, in a static item universe G observed over the two-period time interval [0,1], we obtain:

$$P_{GEKS-L}^{0,1} = \prod_{\tau=0}^{1} \left(\frac{\sum_{i \in G} q_i^\tau p_i^1}{\sum_{i \in G} q_i^\tau p_i^0} \right)^{\frac{1}{1+1}} = \left(\frac{\sum_{i \in G} q_i^0 p_i^1}{\sum_{i \in G} q_i^0 p_i^0} \times \frac{\sum_{i \in G} q_i^1 p_i^1}{\sum_{i \in G} q_i^1 p_i^0} \right)^{\frac{1}{2}} = P_F^{0,1}, \tag{6}$$

since $G_0 = G_1 = G_{0,1} = G$.

2.2. The GEKS-GL index

Let $P_{GL}^{\tau,s}$ denote the geometric Laspeyres price index formula, which compares period s to period τ using data from $G_{\tau,s}$. This index is defined as:

$$P_{GL}^{\tau,s} = \prod_{i \in G_{\tau,s}} \left(\frac{p_i^s}{p_i^\tau} \right)^{w_i^{\tau,s}(\tau)}, \tag{7}$$

where

$$w_i^{\tau,s}(\tau) = \frac{q_i^\tau p_i^\tau}{\sum_{k \in G_{\tau,s}} q_k^\tau p_k^\tau}. \tag{8}$$

Using this notation, the GEKS-GL index can be defined as follows (Białek, 2022a):

$$P_{GEKS-GL}^{0,t} = \prod_{\tau=0}^{T} \left(\frac{P_{GL}^{\tau,t}}{P_{GL}^{\tau,0}} \right)^{\frac{1}{T+1}} = \prod_{\tau=0}^{T} \left(\frac{\prod_{i \in G_{\tau,t}} \left(\frac{p_i^t}{p_i^\tau} \right)^{w_i^{\tau,t}(\tau)}}{\prod_{i \in G_{\tau,0}} \left(\frac{p_i^0}{p_i^\tau} \right)^{w_i^{\tau,0}(\tau)}} \right)^{\frac{1}{T+1}}. \tag{9}$$

It is noteworthy that the GEKS-GL index can be treated as a generalization of the Törnqvist price index ($P_T^{0,t}$) to a multi-period case. Specifically, in a static item universe observed over the two-period time interval [0,1], we obtain:

$$P_{GEKS-GL}^{0,1} = P_T^{0,1}. \tag{10}$$

since $G_0 = G_1 = G_{0,1} = G$, and consequently $w_i^{\tau,0}(\tau) = w_i^{\tau,1}(\tau) = w_i(\tau)$ for any τ.

Białek (2022a) proves the following theorem:

Theorem 1. *The GEKS-L and GEKS-GL indices satisfy the following tests: transitivity, identity, multi-period identity, responsiveness, continuity, positivity and normalization, commensurability, price proportionality, homogeneity in prices, and homogeneity in quantities.*

3 QUALITY-ADJUSTED GEKS-TYPE INDICES

Two new multilateral indices, which superficially resemble the GEKS index, were proposed in Białek (2023) (building upon discussions in Białek (2022b) and Białek and Pawelec (2025)). However, their base index structure differs completely from the established convention of using superlative indices. Moreover, calculating the base index requires *quality adjusting*, which, in turn, is more like the Geary-Khamis index idea. These proposed indices represent a hybrid approach, bridging the gap between quality-adjusted unit value methods and the GEKS method.

3.1. The GEKS-AQU index

In the unit value concept, prices of homogeneous products are equal to the ratio of expenditure to quantity sold (CPI Manual, 2004; Chessa, Verburg, and Willenborg, 2017). However, unlike homogeneous products, quantities of different products cannot be added together. That is why the idea of quality-adjusted unit value assumes that prices p_i^s of different products $i \in G_s$ in month s are transformed into "quality-adjusted prices" $\frac{p_i^s}{v_i}$, and quantities q_i^s are converted into "common units" $v_i q_i^s$ by using a set of factors $v = \{v_i : i \in G_s\}$ (Chessa, Verburg, and Willenborg, 2017). Thus, the "classical" quality-adjusted unit value $QUV_{G_s}^s$ of a set of products G_s in month s can be expressed as follows:

$$QUV_{G_s}^s = \frac{\sum_{i \in G_s} q_i^s p_i^s}{\sum_{i \in G_s} v_i q_i^s}. \tag{11}$$

The term "Quality-adjusted unit value method" (QU method for short) was introduced by Chessa (2015). The QU method is a family of unit value-based index methods, and its general form can be expressed by the following ratio:

$$P_{QU}^{0,t} = \frac{QUV_{G_t}^t}{QUV_{G_0}^0}. \tag{12}$$

In practice, consumer responses to price changes can be delayed or even accelerated, as consumers not only react to current price changes but also use their own "forecasts" or concerns about future price increases. For example, the consumption of thermophilic (seasonal) fruit is likely to be higher in summer because they are cheaper than in winter, when the season is almost over. For instance, some interesting studies on "unconventional" consumer behavior, such as stocking and delayed quantity responses to price changes, and their impact on chain drift bias, can be found in (von Auer, 2019). Given that, in practice, prices and quantities are often not perfectly synchronized in time, the following form of the "asynchronous quality-adjusted unit value" is proposed:

$$AQUV_{G_{\tau,s}}^{\tau,s} = \frac{\sum_{i \in G_{\tau,s}} q_i^\tau p_i^s}{\sum_{i \in G_{\tau,s}} v_i q_i^\tau}, \tag{13}$$

where τ is any period from the considered time interval $[0,T]$. It is thus evident that $AQUV_{G_{s,s}}^{s,s} = QUV_{G_s}^s$. Let us now define the function $P^{\tau,s}\left(v, q^\tau, p^\tau, p^s\right)$ as follows:

$$P^{\tau,s}\left(v, q^\tau, p^\tau, p^s\right) = \frac{AQUV_{G_{\tau,s}}^{\tau,s}}{AQUV_{G_{\tau,\tau}}^{\tau,\tau}}. \tag{14}$$

Substituting (14) into the GEKS formula, we obtain:

$$P_{GEKS-AQU}^{0,t} = \prod_{\tau=0}^{T} \left(\frac{\dfrac{\sum_{i \in G_{\tau,t}} q_i^\tau p_i^t}{\sum_{i \in G_{\tau,t}} v_i q_i^\tau}}{\dfrac{\sum_{i \in G_{\tau,0}} q_i^\tau p_i^0}{\sum_{i \in G_{\tau,0}} v_i q_i^\tau}} \right)^{\frac{1}{T+1}}. \tag{15}$$

It is important to note that the proposed index behaves like a GEKS index based on the Laspeyres index when the item universe G is static. Specifically, if the item universe is static, we obtain

$$P_{GEKS-AQU}^{0,t} = \prod_{\tau=0}^{T} \left(\frac{\frac{\sum_{i \in G} q_i^\tau p_i^t}{\sum_{i \in G} v_i q_i^\tau}}{\frac{\sum_{i \in G} q_i^\tau p_i^0}{\sum_{i \in G} v_i q_i^\tau}} \right)^{\frac{1}{T+1}} = \prod_{\tau=0}^{T} \left(\frac{\sum_{i \in G} q_i^\tau p_i^t}{\sum_{i \in G} q_i^\tau p_i^0} \right)^{\frac{1}{T+1}} = \prod_{\tau=0}^{T} \left(\frac{\frac{\sum_{i \in G} q_i^\tau p_i^t}{\sum_{i \in G} q_i^\tau p_i^\tau}}{\frac{\sum_{i \in G} q_i^\tau p_i^0}{\sum_{i \in G} q_i^\tau p_i^\tau}} \right)^{\frac{1}{T+1}} = P_{GEKS-L}^{0,t}. \tag{16}$$

Finally, it should also be noted that the class of indices is theoretically infinite since different choices of v_i factors lead to different index values. In this paper, we adopt the system of weights v_i that correspond to the augmented Lehr index (Lamboray, 2017; van Loon and Roels, 2018), where

$$v_i = \frac{\sum_{t=0}^{T} p_i^t q_i^t}{\sum_{t=0}^{T} q_i^t}. \tag{17}$$

3.2. The GEKS-AQI index

Formula (13) can be expressed by using quality-adjusted prices and quantities:

$$AQUV_{G_{\tau,s}}^{\tau,s} = \frac{\sum_{i \in G_{\tau,s}} v_i q_i^\tau \frac{p_i^s}{v_i}}{\sum_{i \in G_{\tau,s}} v_i q_i^\tau}. \tag{18}$$

If we replace all the adjusted prices ($\frac{p_i^s}{v_i}$) with the relative prices ($\frac{p_i^s}{p_i^\tau}$), then we obtain an "asynchronous quality-adjusted price index" (AQI), i.e.,

$$AQI_{G_{\tau,s}}^{\tau,s} = \frac{\sum_{i \in G_{\tau,s}} v_i q_i^\tau \frac{p_i^s}{p_i^\tau}}{\sum_{i \in G_{\tau,s}} v_i q_i^\tau}. \tag{19}$$

This means that the AQI formula can be treated as a weighted arithmetic mean of partial indices $\frac{p_i^s}{p_i^\tau}$, where the weights are proportional to the relative share of the product's adjusted quantities (from the base period τ) in the sum of all adjusted quantities.

After inserting (19) into the GEKS formula, we obtain the following:

$$P_{GEKS-AQI}^{0,t} = \prod_{\tau=0}^{T} \left(\frac{\frac{\sum_{i \in G_{\tau,t}} v_i q_i^\tau \frac{p_i^t}{p_i^\tau}}{\sum_{i \in G_{\tau,t}} v_i q_i^\tau}}{\frac{\sum_{i \in G_{\tau,0}} v_i q_i^\tau \frac{p_i^0}{p_i^\tau}}{\sum_{i \in G_{\tau,0}} v_i q_i^\tau}} \right)^{\frac{1}{T+1}}. \tag{20}$$

Note that while the GEKS-AQI index takes into account prices and quantities directly from all-time window periods, the GEKS-AQU index takes into account all quantities but only prices from the reference and base periods. However, both formulas indirectly need information about the prices (and quantities) of products from each period in the time window to determine the factors v_i defined by formula (17). In this way, each new product in the analyzed time window has an impact on the final value of the proposed indices.

3.3. The GEKS-GAQI index

Please note that the AQI defined in (19) can be viewed as a weighted version of the Carli (1804) price index. Currently, there is a trend in price statistics toward shifting from the Carli index to the Jevons (1865) index, which replaces the arithmetic mean of the sub-indices with the geometric mean. This shift serves as the foundation for a definition of *the geometric version of the asynchronous quality-adjusted price index* (GAQI), which was proposed by Białek and Pawelec (2025). The GAQI price index can be expressed as follows:

$$GAQI_{G_{\tau,s}}^{\tau,s} = \prod_{i \in G_{\tau,s}} \left(\frac{p_i^s}{p_i^\tau} \right)^{\frac{v_i q_i^\tau}{\sum_{i \in G_{\tau,s}} v_i q_i^\tau}} . \tag{21}$$

On the basis of (21), the following form of the GEKS-GAQI multilateral price index was proposed (Białek and Pawelec, 2025):

$$P_{GEKS-GAQI}^{0,t} = \prod_{\tau=0}^{T} \left(\frac{\prod_{i \in G_{\tau,s}} \left(\frac{p_i^t}{p_i^\tau} \right)^{\frac{v_i q_i^\tau}{\sum_{i \in G_{\tau,t}} v_i q_i^\tau}}}{\prod_{i \in G_{\tau,s}} \left(\frac{p_i^0}{p_i^\tau} \right)^{\frac{v_i q_i^\tau}{\sum_{i \in G_{\tau,0}} v_i q_i^\tau}}} \right)^{\frac{1}{T+1}} . \tag{22}$$

The following theorem can be proved (see, for instance, Białek (2023) or Białek and Pawelec (2025)):

Theorem 2. *The GEKS-AQU, GEKS-AQI, and GEKS-GAQI indices satisfy the following tests: transitivity, identity, multi-period identity, responsiveness, continuity, positivity and normalization, price proportionality, and weak commensurability. If the item universe is the same in the compared periods and , then the GEKS-AQU, GEKS-AQI, and GEKS-GAQI indices also satisfy the tests of homogeneity in prices and homogeneity in quantities.*

4 GENERALIZATIONS OF THE GEKS METHOD

Sections 5.1, 5.2, and 5.3 present proposals for certain generalizations of the GEKS index whose constructs originate from the economic approach. Therefore, these proposals will be preceded by a brief introduction that contains basic concepts and definitions related to the Cost of Living Index (COLI). Considerations regarding the following general classes of indices can be found in Białek (2025).

4.1. The Cost of Living Index (COLI)

The theory of the Cost of Living Index (COLI) for a single consumer (or household) was first developed by Konüs (1924). In this economic approach, the period τ quantity vector $q^\tau = \left[q_1^\tau, q_2^\tau, ..., q_n^\tau \right]$ is determined by the consumer's preference function *f* and the period τ price vector $p^\tau = \left[p_1^\tau, p_2^\tau, ..., p_n^\tau \right]$ that the consumer faces when observing commodities or items. The consumer's preferences over the given consumption vector q are assumed to be represented by a continuous, non-decreasing, and concave utility function *f*

(CPI Manual, 2004). The main assumption here is that the consumer minimizes the cost of achieving the period τ utility level $u^{\tau} \equiv f\left(q^{\tau}\right)$ for the base period 0 and the current period t. Consequently, the consumer's cost function is defined as follows:

$$C\left(u^{\tau}, p^{\tau}\right) = min_{q}\left\{\sum_{i=1}^{n} p_i^{\tau} q_i : f\left(q\right) = u^{\tau}\right\}, \tau = 0, t. \tag{23}$$

The family of true cost-of-living indices (Konüs, 1924), which compares the current period with the base one, is defined as the ratio of minimum costs of achieving the same utility level $u \equiv f\left(q\right)$:

$$P_K\left(p^0, p^t, q\right) = \frac{C\left(f\left(q\right), p^t\right)}{C\left(f\left(q\right), p^0\right)}, \tag{24}$$

where $q = (q_1, q_2, \ldots, q_n)$ is a positive reference quantity vector. The homothetic preferences assumption simplifies the family of true cost-of-living indices as follows (CPI Manual, 2004):

$$P_K\left(p^0, p^t, q\right) = \frac{c\left(p^t\right)}{c\left(p^0\right)}, \tag{25}$$

where $c\left(p\right) \equiv C\left(1, p\right)$ denotes the unit cost function.

If the $c(p)$ function has a flexible functional form, i.e., it can approximate an arbitrary, twice-differentiable linearly homogeneous cost function to the second order, then the corresponding index is termed superlative (Diewert, 1976). It can be shown that the following quadratic mean of order r price index:

$$P^r\left(p^0, p^t, q^0, q^t\right) = \frac{\sqrt[r]{\sum_{i=1}^{n} s_i^0 \left(\frac{p_i^t}{p_i^0}\right)^{r/2}}}{\sqrt[r]{\sum_{i=1}^{n} s_i^t \left(\frac{p_i^t}{p_i^0}\right)^{-r/2}}}. \tag{26}$$

is a superlative price index formula for any $r \neq 0$ (Diewert, 1976; CPI Manual, 2004).

The *implicit* quadratic mean of order price index:

$$P_{im}^r\left(p^0, p^t, q^0, q^t\right) = \frac{\sum_{i=1}^{n} p_i^t q_i^t}{Q^r\left(p^0, p^t, q^0, q^t\right)\sum_{i=1}^{n} p_i^0 q_i^0}, \tag{27}$$

where the quadratic mean of order r quantity index Q^r can be written as follows:

$$Q^r\left(p^0, p^t, q^0, q^t\right) = \frac{\sqrt[r]{\sum_{i=1}^{n} s_i^0 \left(\frac{q_i^t}{q_i^0}\right)^{r/2}}}{\sqrt[r]{\sum_{i=1}^{n} s_i^t \left(\frac{q_i^t}{q_i^0}\right)^{-r/2}}}, \tag{28}$$

is also a superlative price index formula (Diewert, 1976; CPI Manual, 2004). We have: $P^1_{im} = P_W$, $P^2_{im} = P_F$, $P^2 = P_F$ and P^r (r $\rightarrow$ 0) = P_T, where P_W, P_F and P_T denote the superlative Walsh, Fisher, and Törnqvist price indices, respectively.

4.2. The GEKS-IQM index class

Let us define a general GEKS-type index family as follows:

$$P^{0,t}_{GEKS-IQM}\left(r\right)=\prod_{\tau=0}^{T}\left(\frac{P^{\tau,t}_{IQM}\left(r\right)}{P^{\tau,0}_{IQM}\left(r\right)}\right)^{\frac{1}{T+1}}, \tag{29}$$

where $P^{\tau,t}_{IQM}\left(r\right)\equiv P^{r}_{im}\left(p^{\tau},p^{t},q^{\tau},q^{t}\right)$.

The indices satisfy the *time reversal test*:

$$P^{0,t}_{IQM}\left(r\right)P^{t,0}_{IQM}\left(r\right)=\frac{\sum_{i\in G_{0,t}}p_i^t q_i^t}{Q^r\left(p^0,p^t,q^0,q^t\right)\sum_{i\in G_{0,t}}p_i^0 q_i^0}\frac{\sum_{i\in G_{0,t}}p_i^0 q_i^0}{Q^r\left(p^t,p^0,q^t,q^0\right)\sum_{i\in G_{0,t}}p_i^t q_i^t}=$$

$$=\frac{1}{Q^r\left(p^0,p^t,q^0,q^t\right)Q^r\left(p^t,p^0,q^t,q^0\right)}=\frac{\sqrt[r]{\sum_{i\in G_{0,t}}s_i^t\left(\frac{q_i^t}{q_i^0}\right)^{-r/2}}}{\sqrt[r]{\sum_{i\in G_{0,t}}s_i^0\left(\frac{q_i^t}{q_i^0}\right)^{r/2}}}\frac{\sqrt[r]{\sum_{i\in G_{0,t}}s_i^0\left(\frac{q_i^0}{q_i^t}\right)^{-r/2}}}{\sqrt[r]{\sum_{i\in G_{0,t}}s_i^t\left(\frac{q_i^0}{q_i^t}\right)^{r/2}}}=$$

$$=\frac{\sqrt[r]{\sum_{i\in G_{0,t}}s_i^t\left(\frac{q_i^0}{q_i^t}\right)^{r/2}}}{\sqrt[r]{\sum_{i\in G_{0,t}}s_i^0\left(\frac{q_i^t}{q_i^0}\right)^{r/2}}}\frac{\sqrt[r]{\sum_{i\in G_{0,t}}s_i^0\left(\frac{q_i^t}{q_i^0}\right)^{r/2}}}{\sqrt[r]{\sum_{i\in G_{0,t}}s_i^t\left(\frac{q_i^0}{q_i^t}\right)^{r/2}}}=1.$$

It can be proved that the following theorem holds (Białek, 2025):

Theorem 3. *For each r ≠ 0, the multilateral price index $P^{0,t}_{GEKS-IQM}\left(r\right)$ satisfies the following tests: transitivity, multi-period identity, responsiveness, continuity, positivity and normalization, homogeneity in prices, and homogeneity in quantities, as well as commensurability.*

We can obtain some known particular cases of this index class:

$$P^{0,t}_{GEKS-IQM}\left(1\right)=\prod_{\tau=0}^{T}\left(\frac{P^{\tau,t}_{W}}{P^{\tau,0}_{W}}\right)^{\frac{1}{T+1}}=P^{0,t}_{GEKS-W}, \tag{30}$$

$$P^{0,t}_{GEKS-IQM}\left(2\right)=\prod_{\tau=0}^{T}\left(\frac{P^{\tau,t}_{F}}{P^{\tau,0}_{F}}\right)^{\frac{1}{T+1}}=P^{0,t}_{GEKS}, \tag{31}$$

$$P^{0,t}_{GEKS-IQM}\left(r\rightarrow 0\right)\approx\prod_{\tau=0}^{T}\left(\frac{P^{\tau,t}_{T}}{P^{\tau,0}_{T}}\right)^{\frac{1}{T+1}}=P^{0,t}_{CCDI}. \tag{32}$$

4.3. The GEKS-QM index class

Let us define a general GEKS-type index family as follows:

$$P_{GEKS-QM}^{0,t}\left(r\right)=\prod_{\tau=0}^{T}\left(\frac{P_{QM}^{\tau,t}\left(r\right)}{P_{QM}^{\tau,0}\left(r\right)}\right)^{\frac{1}{T+1}}. \tag{33}$$

where $P_{QM}^{\tau,t}\left(r\right)\equiv P^{r}\left(p^{\tau},p^{t},q^{\tau},q^{t}\right)$.

The *QM* indices satisfy the *time reversal test*:

$$P_{QM}^{0,t}\left(r\right)P_{QM}^{t,0}\left(r\right)=\frac{\sqrt[r]{\sum_{i\in G_{0,t}}s_i^0\left(\frac{p_i^t}{p_i^0}\right)^{r/2}}}{\sqrt[r]{\sum_{i\in G_{0,t}}s_i^t\left(\frac{p_i^t}{p_i^0}\right)^{-r/2}}}\frac{\sqrt[r]{\sum_{i\in G_{0,t}}s_i^t\left(\frac{p_i^0}{p_i^t}\right)^{r/2}}}{\sqrt[r]{\sum_{i\in G_{0,t}}s_i^0\left(\frac{p_i^0}{p_i^t}\right)^{-r/2}}}=$$

$$=\frac{\sqrt[r]{\sum_{i\in G_{0,t}}s_i^0\left(\frac{p_i^t}{p_i^0}\right)^{r/2}}}{\sqrt[r]{\sum_{i\in G_{0,t}}s_i^t\left(\frac{p_i^0}{p_i^t}\right)^{r/2}}}\frac{\sqrt[r]{\sum_{i\in G_{0,t}}s_i^t\left(\frac{p_i^0}{p_i^t}\right)^{r/2}}}{\sqrt[r]{\sum_{i\in G_{0,t}}s_i^0\left(\frac{p_i^t}{p_i^0}\right)^{r/2}}}=1.$$

It can be proved that the following theorem holds (Białek, 2025):

Theorem 4. *For each $r\neq 0$ the multilateral price index $P_{GEKS-QM}^{0,t}\left(r\right)$ satisfies the following tests: transitivity, multi-period identity, responsiveness, continuity, positivity and normalization, weak price proportionality, homogeneity in prices and homogeneity in quantities, as well as commensurability.*

We may obtain some known particular cases of this index class:

$$P_{GEKS-QM}^{0,t}\left(r\to 0\right)=\prod_{\tau=0}^{T}\left(\frac{P_{T}^{\tau,t}}{P_{T}^{\tau,0}}\right)^{\frac{1}{T+1}}=P_{CCDI}^{0,t}, \tag{34}$$

$$P_{GEKS-QM}^{0,t}\left(1\right)\approx\prod_{\tau=0}^{T}\left(\frac{P_{W}^{\tau,t}}{P_{W}^{\tau,0}}\right)^{\frac{1}{T+1}}=P_{GEKS-W}^{0,t}, \tag{35}$$

$$P_{GEKS-QM}^{0,t}\left(2\right)=\prod_{\tau=0}^{T}\left(\frac{P_{F}^{\tau,t}}{P_{F}^{\tau,0}}\right)^{\frac{1}{T+1}}=P_{GEKS}^{0,t}. \tag{36}$$

4.4. The GEKS-LM index class

The Lloyd-Moulton price index (Loyd, 1975) can be written as follows:

$$P_{LM}^{\tau,t}\left(\sigma\right)=\left(\sum_{i\in G_{\tau,t}}s_i^{\tau}\left(\frac{p_i^t}{p_i^{\tau}}\right)^{1-\sigma}\right)^{\frac{1}{1-\sigma}}, \tag{37}$$

where the parameter σ denotes the elasticity of substitution. It can be shown (Loyd, 1975; Moulton, 1996) that under the assumption of cost-minimizing behavior, the index P_{LM} is exact for constant elasticity of substitution (CES) preferences. That is, for the fixed periods τ and t, it holds that

$$P_{LM}^{\tau,t}(\sigma) = \frac{c_\sigma\left(p^t\right)}{c_\sigma\left(p^\tau\right)}, \tag{38}$$

where the unit cost function $c_\sigma(p)$ is a CES aggregator function introduced in Arrow et al. (1961). Based on the Lloyd-Moulton index, let us define a general GEKS-type index family as follows:

$$P_{GEKS-LM}^{0,t}(\sigma) = \prod_{\tau=0}^{T}\left(\frac{P_{LM}^{\tau,t}(\sigma)}{P_{LM}^{\tau,0}(\sigma)}\right)^{\frac{1}{T+1}}, \tag{39}$$

with the following special cases:

$$P_{GEKS-LM}^{0,t}(0) = \prod_{\tau=0}^{T}\left(\frac{P_{La}^{\tau,t}}{P_{La}^{\tau,0}}\right)^{\frac{1}{T+1}} = P_{GEKS-L}^{0,t} \tag{40}$$

and

$$P_{GEKS-LM}^{0,t}(\sigma \to 1) = \prod_{\tau=0}^{T}\left(\frac{P_{GLa}^{\tau,t}}{P_{GLa}^{\tau,0}}\right)^{\frac{1}{T+1}} = P_{GEKS-GL}^{0,t}. \tag{41}$$

It can be proved that the following theorem holds (Białek, 2025):

Theorem 5. *For each $\sigma \neq 1$, the multilateral price index $P_{GEKS-LM}^{0,t}(\sigma)$ satisfies the following tests: transitivity, identity, multi-period identity, responsiveness, continuity, positivity and normalization, commensurability, strong price proportionality, as well as homogeneity in prices and homogeneity in quantities.*

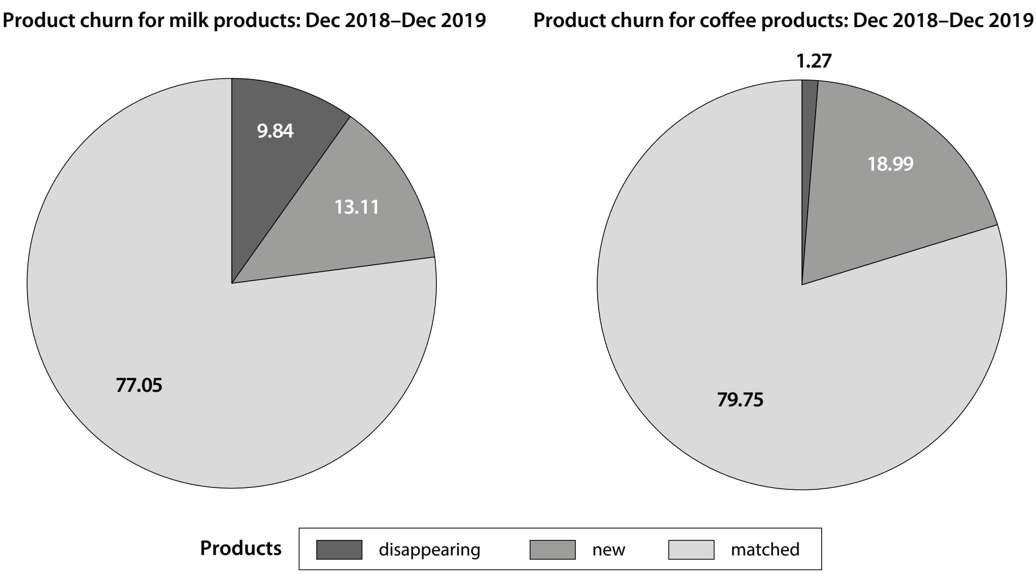
Remark 1. Please note that in this paper, we make a distinction between *strong* and *weak price proportionality tests*. To the best of the author's knowledge, this is the first distinction of this type in literature. However, this distinction is only relevant to multilateral indices, where the considerations involve many periods within a given time window. The *weak price proportionality* test assumes that price changes in all periods within the time interval are proportional to prices from the base period. This test states that in such a situation, the index comparing the current and base periods should depend only on the proportion of prices from the compared periods. In the case of a strong version of this test (i.e., the *strong proportionality test*), the same behavior of the index is expected under a less restrictive requirement because it assumes proportionality of prices in only two compared periods (prices from periods earlier than the current one do not have to be proportional to prices from the base period). It is easy to verify that a strong version of this test implies the occurrence of a weak version of the test. It can be shown that the GEKS, Geary-Khamis (Geary, 1958; Khamis, 1972), and TPD (de Haan and Krsinich, 2018) multilateral indices satisfy the *weak price proportionality test* but not the *strong price proportionality test*, while the SPQ (Diewert, 2020), GEKS-L, and GEKS-GL indices satisfy the *strong price proportionality test* (see an example in Białek (2025)).

Remark 2. In the author's opinion, there is no point in considering the *identity test, multi-period identity test* or *weak* or *strong proportionality tests* if we do not assume that the product universe is the same in the base and current periods. Of course, such an assumption does not exclude the phenomenon of product churn within the time interval under consideration.

5 EMPIRICAL ILLUSTRATION

An empirical illustration of the potential differences between the proposed indices (including representatives of general index classes) and the multilateral indices known from the literature (GEKS, CCDI, TPD, Geary-Khamis) is presented on the milk and coffee datasets. These datasets are anonymized sets of real data on coffee and milk sales in a retail chain operating in Poland. These datasets are available in the R package *PriceIndices* (Białek, 2021). The analysis was performed at the GTIN code level and covered the period Dec 2018–Dec 2019. The product rotation corresponding to these datasets and the analyzed period is presented in Fig. 1. Fig. 2 compares the indices presented in Sections 3.1, 3.2, 4.1 and 4.2 with popular multilateral indices. An analogous comparison, excluding indices from the GEKS-IQM class, is shown in Fig. 3. The next two figures (i.e., Fig. 4 and Fig. 5) consider the other two general classes of indices, i.e., the GEKS-QM and GEKS-LM.

Figure 1 Product churn in the scanner datasets considered.



Observing Figs. 2–5, one can see that, in general, the multilateral indices determined based on the *milk* and *coffee* datasets form three clusters of values. Specifically, the values of the GEKS-L, GEKS-GL, GEKS-AQI, and GEKS-AQU indices are close to each other (Cluster 1), further – the GEKS and CCDI indices approximate each other (cluster 2), and the TPD and Geary-Khamis indices have similar values (Cluster 3). In our study, it was found that values of indices from Cluster 3 are always between index values from Cluster 1 and Cluster 2. The similar values of the GEKS and CCDI index methods can be readily explained by the mutual approximation of superlative indices (Diewert, 1976), which serve as the underlying bilateral indices for these methods. However, it is particularly interesting that the indices satisfying the identity test form a distinct cluster (Cluster 1).

Figure 2 Comparison of the GEKS-L, GEKS-GL, GEKS-AQI, GEKS-AQU, and GEKS-GAQI indices with known multilateral indices.

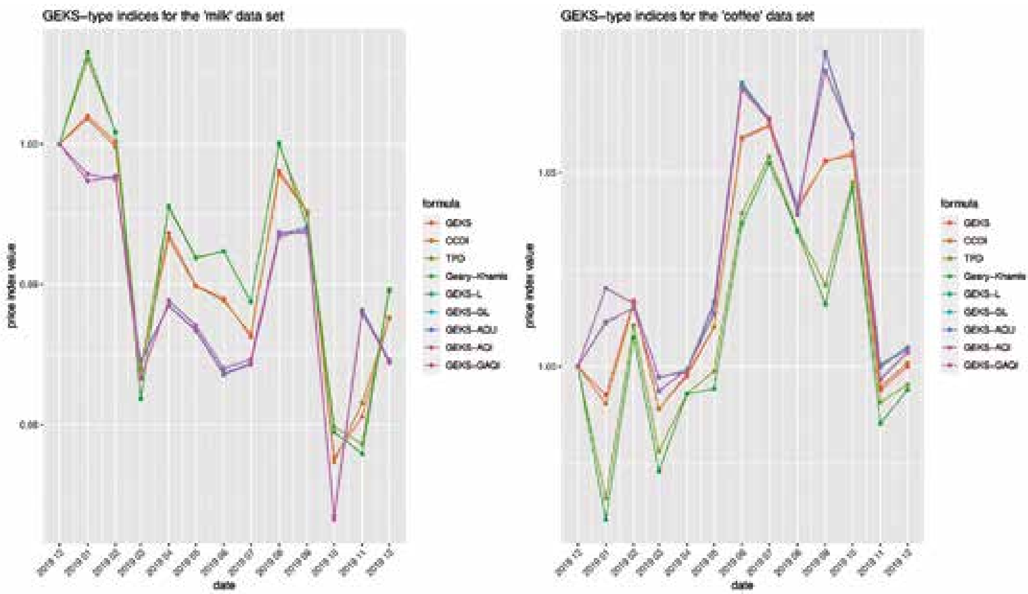


Figure 3 Comparison of the selected GEKS-IQM indices with known multilateral indices.

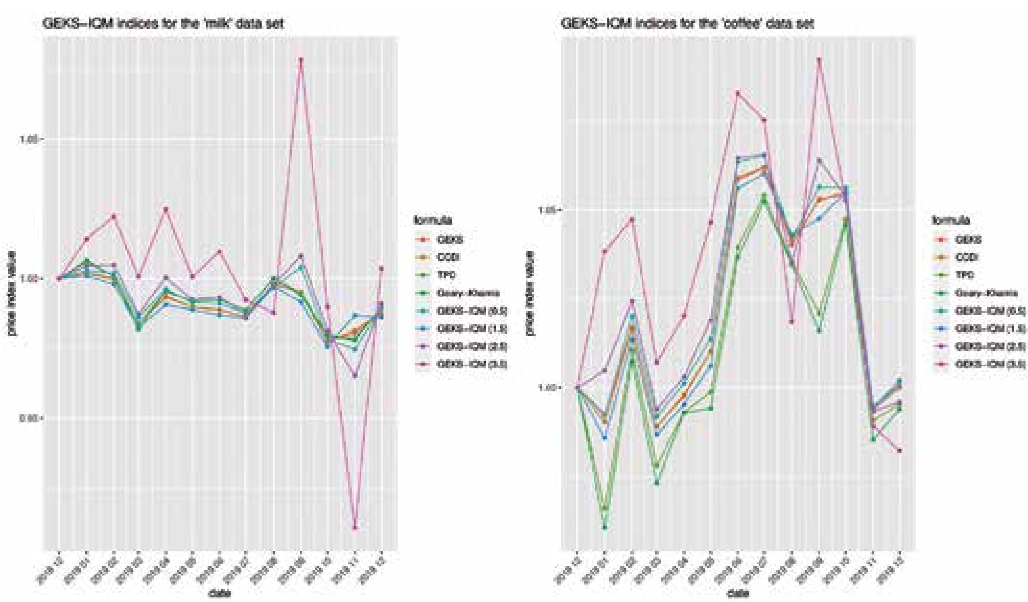


Figure 4 Comparison of the selected GEKS-QM indices with known multilateral indices.

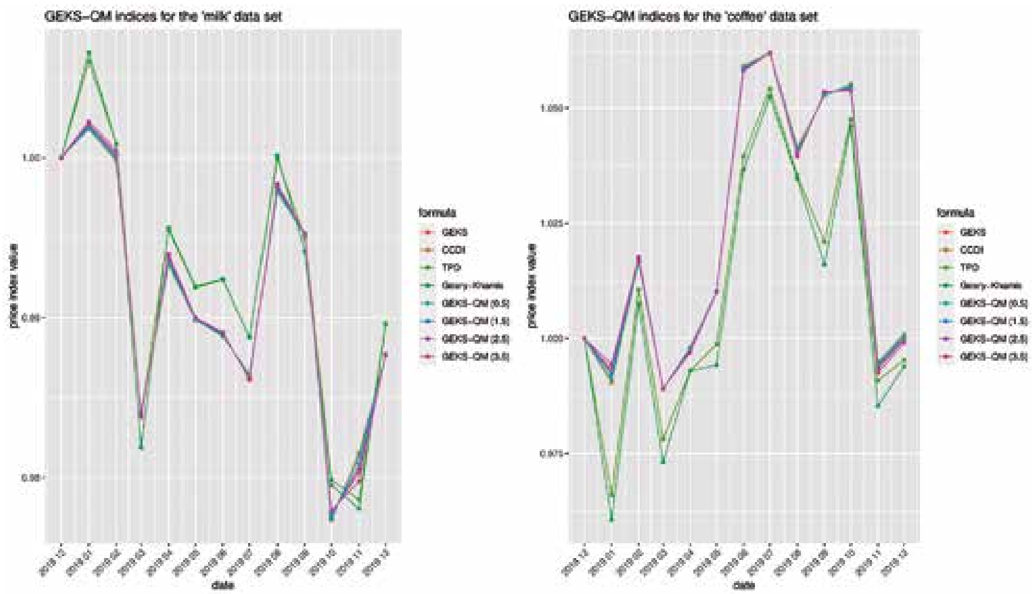
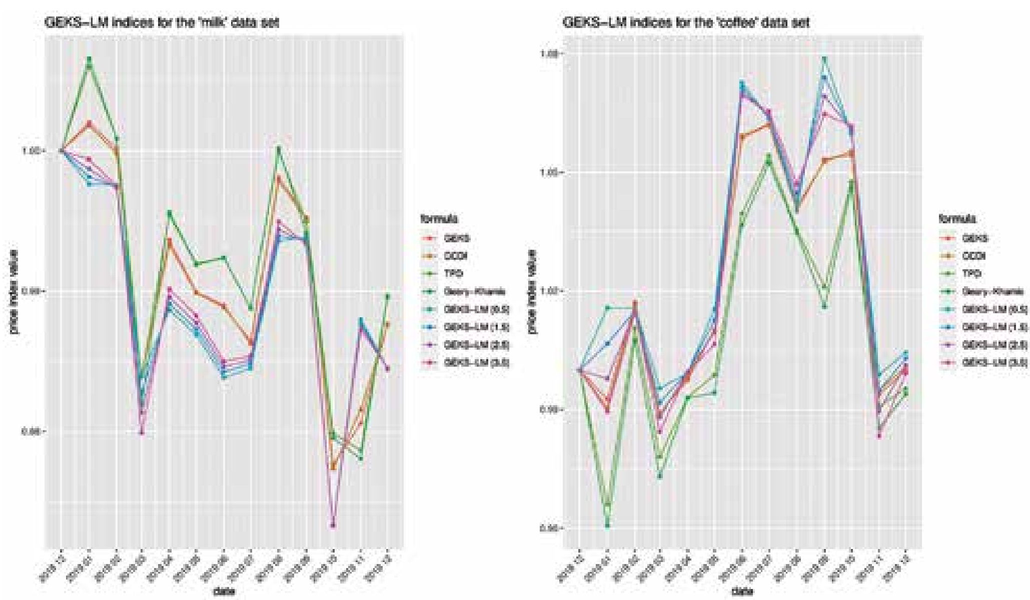


Figure 5 Comparison of the selected GEKS-LM indices with known multilateral indices.



6 SIMULATION STUDY

Our simulation study aims to investigate how price and quantity volatility affect the differences between GEKS-type indices. In the experiment, we generate price processes assuming that prices follow a geometric Brownian motion (GBM), which is commonly used for price modeling (see Ross (2014) or Hull (2018)) and whose parameters are relatively easy to estimate.

By discretizing the GBM price model, we effectively switch to a log-normal distribution, which is frequently used in price modeling, not only in finance but also in multilateral index number theory (see, for instance, Silver and Heravi (2007) or Białek (2020)). We assume that all prices at a given time (τ) are log-normally distributed, i.e.,

$$\log\left(p^{\tau}\right) \sim N\left(\mu_{\tau}, \sigma_{\tau}\right), \tag{42}$$

where $N\left(\mu_{\tau}, \sigma_{\tau}\right)$ denotes the normal distribution with mean μ_{τ} and standard deviation σ_{τ}.

We conducted four simulation experiments, each generating artificial datasets of $N = 100$ products observed over 13 months (Dec 2023–Dec 2024). Assuming that the prices and quantities were log-normally distributed, the following values of distribution parameters were used for each month:

a) for prices $\mu^{p} = \left\{\mu_{p}^{t} : t = 0,1,2,...,13\right\}$ and $\sigma^{p} = \left\{\sigma_{t}^{p} : t = 0,1,2,....,13\right\}$,
where $\mu_{p} = \{1, 1.01, 1.04, 1.05, 1.07, 1.11, 1.24, 1.13, 1.06, 1.03, 1.19, 1.12, 1.04\}$,
$\sigma^{p} = \gamma^{p} \mu^{p}$ $\gamma^{p} \in \{0.05, 0.5\}$,

b) for quantities $\mu^{q} = \left\{\mu_{q}^{t} : t = 0,1,2,...,13\right\}$ and $\sigma^{q} = \left\{\sigma_{t}^{q} : t = 0,1,2,....,13\right\}$, where $\mu_{t}^{1} = 6 / \mu_{t}^{p}$ and
$\sigma^{q} = \gamma^{q} \mu^{q}$ $\gamma^{q} \in \{0.05, 0.5\}$

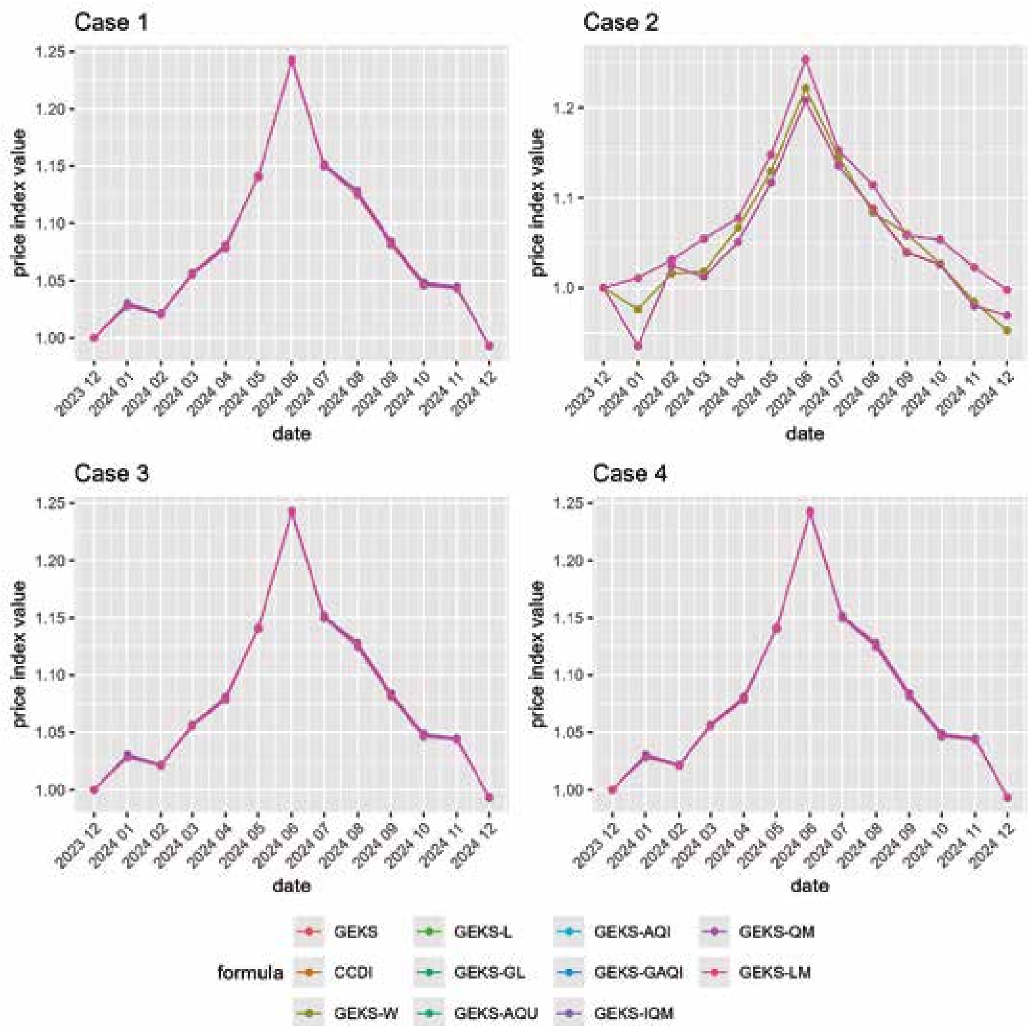
The parameters listed in (a) and (b) were selected to ensure that (1) two evident price peaks could be generated, e.g., there is a clear price increase in June 2023 and then in October 2023; (2) the price increase was accompanied by a fall in quantities (to reflect the natural negative correlation between prices and quantities); (3) it was possible to set varying levels of prices and quantities: low (Case 1: $\gamma^p = \gamma^q = 0.05$), high (Case 4: $\gamma^p = \gamma^q = 0.5$), and mixed (Case 2: $\gamma^p = 0.05$, $\gamma^q = 0.5$ and Case 3: $\gamma^p = 0.5$, $\gamma^q = 0.05$). Altogether, as already mentioned, four simulation experiments were performed.

Fig. 6 presents a comparison of the GEKS-type indices considered for Cases 1–4. The indices from the GEKS-QM and GEKS-IQM classes were determined for $r = 1.5$. The GEKS-LM index, on the other hand, was estimated for elasticity of substitution $\sigma = 0.7$. Analysis of Fig. 6 reveals that the differences between GEKS-type indices become noticeable only when there is high variation in quantities. In Case 2, where the variability of quantities is very large (and thus the levels of product consumption vary greatly), the differences between the indices considered exceeded as much as 10 percentage points (in other cases, they did not exceed 0.5 percentage points). If these indices begin to diverge, then, as in the empirical study, three clusters of GEKS-type index values emerge (see Section 5). Interestingly, even large price volatility does not generate substantial differences between the values of the GEKS-type indices (Case 3 in Fig. 6).

CONCLUSIONS

This paper expands on the results presented at the *Ottawa Group Meeting* held in Canada in 2024 (see Białek (2024) for details). The discussion of versions of the GEKS index was supplemented by a relatively new index, GEKS-GAQI, which was recently proposed by Białek and Pawelec (2025). An additional simulation study was conducted to determine the impact of price and quantity variability on potential differences between GEKS-type indices.

Figure 6 Comparison of the selected GEKS-type indices for Cases 1–4.



While the empirical illustration presented in Section 6 covers only two food groups and, thus, does not allow us to generalize conclusions, it provides some preliminary insights into the behavior of multilateral GEKS-type indices. Fig. 3 casts a bit of a shadow over the usefulness of indices from the GEKS-IQM class as they begin to lag significantly behind typical multilateral indices for the parameter . On the other hand, indices from the GEKS-QM class are stable due to the value of the parameter (Fig. 4). At the same time, differences between these indices and typical multilateral ones can be both small (no more than 0.5 p.p. as in the case of the *milk* set) and noticeable (as much as 2.5 p.p. for the *coffee* set). Fig. 5 shows that the indices from the GEKS-LM class also form a separate cluster of values and are moderately sensitive to the choice of parameter. Regarding the *coffee* data set, the indices from this class have substantially higher values than the other compared indices, with differences exceeding even 3 p.p. For the *milk* data set, indices from the GEKS-LM class generate the lowest values among the set of multilateral indices considered.

The simulation study conducted, presented in Section 6, also provides valuable practical conclusions. It was found that only large quantity volatility (and not price volatility) generates substantial differences in the values of GEKS-type indices. This conclusion differs from the observations of bilateral indices, where large price volatilities are primarily responsible for the resulting differences between the indices. In addition, the simulation study confirmed the formation of three clusters of GEKS-type index values when there are large differences between product consumption levels.

We can also draw some general conclusions about the analyzed family of multilateral indices, which may facilitate the final selection of one of the representatives of this family. From a theoretical (axiomatic) perspective, the GEKS-LM class of indices seems to be the most valuable (e.g., they satisfy the *identity test*). Important special cases of this class are the GEKS-L and GEKS-GL indices. To satisfy both the axiomatic and economic approaches simultaneously, it would be necessary to determine the elasticity of substitution parameter beforehand (e.g., by using a panel regression model that explains the behavior of quantities based on prices). This is the direction of the author's further research.

Please note that indices from the GEKS-AQU, GEKS-AQI and GEKS-GAQI classes also satisfy the *identity test*. However, a certain limitation of indices from these classes is the condition necessary to satisfy *homogeneity in prices* and *homogeneity in quantities*. This condition is a less realistic scenario, as it implies a constant set of products in the base and current periods being compared.

References

ARROW, K. J., CHENERY, H. B., MINHAS, B. S., SOLOW, M. (1961). Capital-Labor Substitution and Economic Efficiency. *The Review of Economics and Statistics* 43 (3): 225–50. http://www.jstor.org/stable/1927286.

AUSTRALIAN BUREAU OF STATISTICS. (2016). *Making Greater Use of Transactions Data to Compile the Consumer Price Index*. Canberra: Information Paper 6401.0.60.003.

BIAŁEK, J. (2020). Comparison of Elementary Price Indices. *Communications in Statistics-Theory and Methods* 49 (19): 4787–4803.

BIAŁEK, J., AND M. BERĘSEWICZ. (2021). Scanner Data in Inflation Measurement: From Raw Data to Price Indices. *The Statistical Journal of the IAOS* 37: 1315–36.

BIAŁEK, J. (2021). PriceIndices – a New R Package for Bilateral and Multilateral Price Index Calculations. *Statistika – Statistics and Economy Journal* 36 (2): 122–41.

BIAŁEK, J. (2022a). Improving Quality of the Scanner CPI: Proposition of New Multilateral Methods. *Quality and Quantity* 57: 2893–2921.

BIAŁEK, J. (2022b). The General Class of Multilateral Indices and Its Two Special Cases. In *Paper Presented at the 17th Meeting of the Ottawa Group on Price Indices, 6–10 June 2022, Rome, Italy*.

BIAŁEK, J. (2023). Quality Adjusted GEKS-Type Indices for Price Comparisons Based on Scanner Data. *Statistics in Transition New Series* 24 (June): 151–69. https://doi.org/10.59170/stattrans-2023-041.

BIAŁEK, J. (2024). Evolution of the GEKS Index. In: *Paper Presented at the 18th Meeting of the Ottawa Group, Ottawa, Canada, 13–15 May 2024*.

BIAŁEK, J. (2025). General Classes of GEKS-Type Price Indices With Application to Scanner Data. *Review of Income and Wealth* 71 (1): 1–21. https://doi.org/10.1111/roiw.12726.

BIAŁEK, J., PAWELEC, N. (2025). Proposition of the quality adjusted GEKS-type price index. *Statistika: Statistics and Economy Journal* 105 (3): 291–305. https://doi.org/10.54694/stat.2024.71.

CARLI, G. (1804). Del Valore e Della Proporzione de'metalli Monetati. In *Scrittori Classici Italiani Di Economia Politica*, 13:297–336.

CHESSA, A. (2015). Towards a Generic Price Index Method for Scanner Data in the Dutch CPI. In *14th Meeting of the Ottawa Group, Tokyo*, 20–22.

CHESSA, A. (2016). A New Methodology for Processing Scanner Data in the Dutch CPI. *Eurostat Review of National Accounts and Macroeconomic Indicators* 1: 49–69.

CHESSA, A., VERBURG, J., WILLENBORG, L. (2017). A Comparison of Price Index Methods for Scanner Data. In *15th Meeting of the Ottawa Group, Eltville*, 10–12.

DE HAAN, JAN, KRSINICH, F. (2014). *Scanner Data and the Treatment of Quality Change in Non-revisable Price Indexes, Journal of Business & Economic Statistics* 32(3): 341–358.

DE HAAN, J., KRSINICH, F. (2018). Time Dummy Hedonic and Quality-Adjusted Unit Value Indexes: Do They Really Differ? *Review of Income and Wealth* 64 (4): 757–76.
DIEWERT, W. E. (1976). Exact and Superlative Index Numbers. *Journal of Econometrics* 4 (2): 115–45.
DIEWERT, W. E. (2020). The Chain Drift Problem and Multilateral Indexes. Discussion Paper 20-07, Vancouver School of Economics.
DIEWERT, W. E, AND FOX, K. J. (2018). Substitution Bias in Multilateral Methods for CPI Construction Using Scanner Data. *UNSW Business School Research Paper* 2018-13.
ELTETÖ, O., KÖVES, P. (1964). On a Problem of Index Number Computation Relating to International Comparison. *Statisztikai Szemle* 42 (10): 507–18.
EUROSTAT (2022). *Guide on Multilateral Methods in the Harmonised Index of Consumer Prices*. Luxembourg: Publications Office of the European Union.
FISHER, I. (1922). *The Making of Index Numbers: A Study of Their Varieties, Tests, and Reliability*. Vol. xxxi. Houghton Mifflin.
GEARY, R. C. (1958). A Note on the Comparison of Exchange Rates and Purchasing Power Between Countries. *Journal of the Royal Statistical Society. Series A (General)* 121 (1): 97–99.
GINI, C. (1931). On the Circular Test of Index Numbers. *Metron* 9 (9): 3–24.
HULL, J. (2018). *Options, Futures, and Other Derivatives (10th Ed.)*. Pearson: Boston.
INKLAAR, R., DIEWERT, W. E. (2016). Measuring Industry Productivity and Cross-Country Convergence. *Journal of Econometrics* 191 (2): 426–33.
INTERNATIONAL LABOUR OFFICE. (2004). Consumer Price Index Manual: Theory and Practice. Geneva.
JEVONS, W. S. (1865). On the Variation of Prices and the Value of the Currency Since 1782. *Journal of the Statistical Society of London* 28 (2): 294–320.
KHAMIS, S. H. (1972). A New System of Index Numbers for National and International Purposes. *Journal of the Royal Statistical Society: Series A (General)* 135 (1): 96–121.
KONÜS, A. (1924). The Problem of the True Index of the Cost of Living. *The Economic Bulletin of the Institute of Economic Conjuncture* 9-10: 64–71.
LAMBORAY, C. (2017). The Geary Khamis Index and the Lehr Index: How Much Do They Differ. In *Paper to Be Presented at the 15th Meeting of the Ottawa Group*, 10–12.
LASPEYRES, K. (1871). IX. Die Berechnung Einer Mittleren Waarenpreissteigerung. *Jahrbücher für Nationalökonomie Und Statistik* 16 (1): 296–318.
LOYD, P. (1975). Substitution Effects and Biases in Nontrue Price Indices. *The American Economic Review* 65 (3): 301–13.
MOULTON, B. R. (1996). *Constant Elasticity Cost-of-Living Index in Share Relative Form*. Washington, D.C.: Bureau of Labor Statistics.
ROSS, S. (2014). *Variations on Brownian Motion. Introduction to Probability Models (11th Ed.)*. Elsevier: Amsterdam.
MICK, S., HERAVI, S. (2007). Why Elementary Price Index Number Formulas Differ: Evidence on Price Dispersion. *Journal of Econometrics* 140 (2): 874–83.
SZULC, B. (1964). Indices for Multiregional Comparisons. *Przeglad Statystyczny* 3: 239–54.
TÖRNQVIST, L. (1936). The Bank of Finland's Consumption Price Index. *Bank of Finland Monthly Bulletin*, 1–8.
VAN LOON, K. V., ROELS, D. (2018). Integrating Big Data in the Belgian CPI. In *Paper Presented at the Meeting of the Group of Experts on Consumer Price Indices, 8-9 May 2018, Geneva, Switzerland*.
VON AUER, L. (2019). The Nature of Chain Drift. In *Paper Presented at the 17th Meeting of the Ottawa Group on Price Indices, 8–10 May 2019, Rio de Janeiro, Brasil*.
WALSH, C. M. (1901). *The Measurement of General Exchange Value*. Macmillan; Co.

Causal relationships between inflation uncertainty, geopolitical risks, and inflation: New insights from Granger causality-in-quantiles tests

Zouheir Mighri[1] | *College of Business, University of Jeddah, Jeddah, Saudi Arabia*

Received 15.10.2025, **Accepted** (reviewed) 3.6.2026, **Published** 11.9.2026

Abstract

This study investigates the direction and magnitude of linear and nonlinear Granger causality-in-quantiles between inflation and inflation uncertainty, as well as from global geopolitical risks to inflation in Saudi Arabia over the period January 2008–December 2025. While no causality is detected in the mean, the results reveal significant quantile-dependent relationships. Linear Granger causality-in-quantiles indicates significant bidirectional causality between inflation and inflation uncertainty, with a predominantly positive and significant causal effect running from inflation to its uncertainty across most quantiles, thereby supporting the Friedman (1977) hypothesis. In contrast, nonlinear Granger causality-in-quantiles uncovers state-dependent effects, with inflation uncertainty exerting a significant negative influence on inflation at lower and middle quantiles, thereby supporting the Friedman-Ball hypothesis. Meanwhile, both linear and nonlinear Granger causality-in-quantiles reveal significant causal effects from global geopolitical risk measures to inflation across most quantiles; however, the sign analysis indicates that these effects are unstable and largely insignificant, providing no consistent or robust sign pattern linking external risk factors to inflation dynamics in Saudi Arabia. Robustness checks using the world uncertainty index confirm that global uncertainty has a limited and non-systematic influence on inflation dynamics. Overall, the findings emphasize strong state-dependent and asymmetric effects, with domestic uncertainty playing a central role, while external risk factors contribute marginally to inflation dynamics in Saudi Arabia.

Keywords	***DOI***	***JEL code***
Inflation, inflation uncertainty, geopolitical risk, Granger causality-in-quantiles.	*https://doi.org/10.54694/stat.2025.57*	*C11, C22, C52, D80, E31, H56*

[1] Department of Finance and Economics, College of Business, University of Jeddah, Jeddah, Saudi Arabia.
E-mail zmighri@uj.edu.sa. ORCID: <https://orcid.org/0000-0002-8850-7812>

INTRODUCTION

Geopolitical uncertainty and persistent inflation continue to weigh heavily on the global economy, weakening growth prospects and leading to downward revisions to global economic forecasts. Geopolitical risk, broadly defined as the likelihood of conflicts, terrorist attacks, or other disruptions to international relations that undermine stability, has intensified in recent years due to rising international tensions. Key risks include Russia-NATO frictions, the ongoing Russia-Ukraine conflict, cyberattacks, U.S.-China strategic competition, climate-related risks, energy security challenges, and the lingering effects of the COVID-19 pandemic. These developments have heightened market volatility and amplified uncertainty around inflation dynamics (Yang et al., 2023; Bouri et al., 2023; Lee et al., 2023; Adeosun et al., 2023).

The complex interplay between geopolitical shocks and inflation has therefore attracted increasing attention from academics, policymakers, and market participants (Lee et al., 2023; Yang et al., 2023; Caldara and Iacoviello, 2025). Geopolitical disruptions can directly transmit into the macroeconomy, creating inflationary pressures. For example, heightened uncertainty often induces consumer caution and delays in consumption, while businesses suspend investments and accumulate precautionary savings (Yang et al., 2023). Such dynamics exacerbate cyclical downturns by reducing the perceived value of future income streams and investment opportunities. Moreover, the magnitude of these effects varies according to the nature, scale, and duration of the geopolitical events (Noguera-Santaella, 2016).

Energy-dependent economies such as Saudi Arabia are particularly exposed to these risks, as fluctuations in energy markets are closely tied to geopolitical developments (Aliedan, 2022). Although the practice of anticipating geopolitical behavior has been part of international relations for centuries, its systematic integration into economic and political analysis has become especially pronounced since the early 20th century (Aliedan, 2022). This underscores the importance of viewing geopolitics not merely as a background condition, but as a central driver of both economic volatility and inflationary dynamics in the contemporary global landscape.

Understanding the impact of geopolitical risk on inflation is increasingly important, particularly in Saudi Arabia, given current global political tensions and economic challenges. Major events, such as the COVID-19 pandemic and the Russia-Ukraine conflict, have driven sharp increases in commodity and food prices, fueling inflation worldwide. While central banks have responded with interest rate hikes, tighter fiscal conditions have contributed to a slowdown in global growth. Although recent inflationary pressures are largely short-term, they may pose long-term economic challenges, necessitating both immediate and strategic policy responses.

This study addresses the following research question: Do adverse global geopolitical shocks increase or decrease inflation in Saudi Arabia? Theoretically, Caldara and Iacoviello (2025) argued that the effect of geopolitical tensions on inflation is ambiguous, as such shocks influence both supply and demand. On the supply side, conflicts can destroy capital, redirect resources inefficiently, disrupt trade and capital flows, and impair supply chains. On the demand side, uncertainty can reduce economic activity, delay investment and hiring, erode consumer confidence, and tighten financial conditions.

In addition to geopolitical risk, inflation uncertainty may be considered as an alternative determinant of inflation in Saudi Arabia. Historically, inflation has remained moderate, averaging 3.44% in 2020, 3.06% in 2021, 2.47% in 2022, and 2.33% in 2023, although the Consumer Price Index has shown a steady upward trend since 2010. This study further tries to answer the following second research question: Does inflation uncertainty raise or lower inflation in Saudi Arabia? Inflation uncertainty, that is, the unpredictable volatility in inflation, poses significant challenges for monetary policymakers (Barnett et al., 2020). It can arise from differences in international monetary regimes or political instability, both of which affect inflation expectations (Cukierman and Meltzer, 1986). Economic agents continuously update their expectations in response to new information, making inflation uncertainty dynamic and complex. Four main hypotheses describe the interplay between inflation and its uncertainty: positive and negative causal effects of inflation on uncertainty (Friedman, 1977; Ball, 1992; Pourgerami and Maskus, 1987; Ungar and Zilberfarb, 1993) and positive and negative effects of uncertainty on inflation (Cukierman and Meltzer, 1986; Holland, 1995).

This study aims to complement existing research on the causal effects of inflation uncertainty and geopolitical risk on inflation in Saudi Arabia and to inform policy recommendations for managing the economy during periods of crisis. Understanding how these factors influence the general price level is essential for monetary authorities seeking to contain inflation. The significance of this work lies in its contribution to the literature, addressing a gap in previous studies that have not explored the potential effects of these factors on Saudi Arabia's inflation dynamics.

The current study makes two main contributions. First, it examines the causal relationship between inflation and inflation uncertainty, as well as the causal impacts of global geopolitical risks on inflation in Saudi Arabia, within a quantile-based framework that captures distributional dynamics beyond conventional mean-based approaches. Investigating this relationship across quantiles is particularly important as it may identify the state-dependent effects that may differ between low- and high-inflation regimes, especially over the period spanning from January 2008 to December 2025, which encompasses major global and regional shocks. Moreover, evaluating the quantile-specific causal effects of global geopolitical risk, geopolitical threats, and geopolitical risk acts on inflation is crucial in the Saudi context, given its strong exposure to external shocks and oil market fluctuations. This approach provides deeper insights into how different dimensions of global geopolitical risk transmit to domestic inflation across varying economic conditions. Second, our study incorporates the sign of both linear and nonlinear Granger causality-in-quantiles, allowing for a more comprehensive understanding of not only whether causal relationships exist but also their direction and economic interpretation across different states of the distribution. This is particularly important because the sign of causality enables a direct assessment of competing theoretical hypotheses on the inflation-uncertainty relationship. To the best of our knowledge, this study is the first to examine the sign of nonlinear Granger causality-in-quantiles, thereby introducing a novel dimension to the empirical literature. By combining direction, magnitude, and sign within a distributional framework, the analysis uncovers hidden asymmetries and state-dependent dynamics that conventional approaches cannot detect.

The remainder of the paper is structured as follows. Section 1 presents the literature review regarding the nexus between geopolitical risks, inflation uncertainty, and inflation. Section 2 describes the data and methodology used in this study. Section 3 presents and discusses the main empirical results. Section 4 outlines the conclusion, policy implications, and limitations of the study.

1 LITERATURE REVIEW

1.1 Geopolitical risks and inflation

The empirical literature on the relationship between geopolitical risks and inflation remains relatively scarce, with only a handful of studies addressing this nexus. Caldara et al. (2025) identified two channels linking geopolitical risks to inflation: supply-side disruptions (disruptions to resources and trade, inefficient reallocation, and supply chain problems) and demand-side effects (decreased investment, employment, and consumer confidence due to uncertainty), with the net inflationary effect depending on the dominant channel. In a related study, Caldara and Iacoviello (2025) argued that geopolitical risks drive commodity price inflation through supply and demand effects in commodity markets. Kyriazis et al. (2023) also found that currency values directly affect inflation and indirectly affect exports and geopolitical risks, revealing a geopolitical risk-driven inflation channel, most evident in the U.S., Brazil, and South Korea (and less so in France and Germany). In Russia, inflation, exchange rates, and exports were found to constitute geopolitical risks, particularly during the Russo-Ukrainian War.

Meanwhile, Yang et al. (2023) showed that the links between geopolitical risks and oil price shocks and inflation in China, the U.S. and 27 European countries are time-varying, with post-COVID-19 inflation driven by shrinking oil reserves, higher oil prices and supply and demand imbalances, which are exacerbated by geopolitical risks. Similarly, Lee et al. (2023) found that geopolitical risks related to crude oil significantly affect the average and variance of core inflation in the United States and China, with the effect increasing during major geopolitical events and varying by country and type of causality. Alternatively, Adeosun et al. (2023) investigated the interplay among economic policy uncertainty, geopolitical risks, and inflation in the U.S., Canada, the UK, Japan, and China. They found that geopolitical risks in the U.S. and Canada display strong short- and medium-term coherence with inflation. In contrast, the UK and China exhibit strong short-term but weaker medium-term coherence, while Japan shows strong coherence only in the short term. Finally, Bouri et al. (2023) showed that the transmission of geopolitical risk and inflation across European and North American economies rose sharply during the Russian-Ukrainian war, exceeding even the levels of the 1970s energy crisis, and found that the global geopolitical risk (GPR) index positively predicts future inflation transmission, whereas the GPR Acts index only affects large transition changes.

1.2 Inflation uncertainty and inflation

The relationship between inflation and inflation uncertainty has been investigated in two main directions, each grounded in different economic perspectives. One strand focuses on the causal link running from inflation to inflation uncertainty. Within this strand, the literature is further divided according to the expected sign of the relationship. Friedman (1977) posited that inflation positively impacts uncertainty, since inconsistent monetary policy makes predicting future policy difficult—a view formally supported by Ball (1992). In contrast, Pourgerami and Maskus (1987) and Ungar and Zilberfarb (1993) argued for an inverse relationship, reasoning that higher inflation encourages greater effort in forecasting, thus reducing uncertainty. The second economic context addresses the opposite direction of causation, where uncertainty about inflation affects inflation itself. In this context, studies offer two opposing explanations, depending on the direction of the relationship. Cukierman and Meltzer (1986) argued for a positive causal effect, suggesting that increased uncertainty about inflation motivates policymakers to create a sudden inflationary surge, leading to higher inflation. In contrast, Holland (1995) proposed a negative causal effect, asserting that increased uncertainty prompts independent central banks to reduce inflation to minimize the real costs associated with that uncertainty.

The causal relationship between inflation and its uncertainty remains unclear and controversial (Barnett et al., 2020). Most empirical evidence supports Friedman's (1977) hypothesis (e.g., Berument and Dincer, 2005; Fountas et al., 2006, among others), although other studies have found evidence of a bidirectional causal relationship (e.g., Conrad and Karanasos, 2005; Albulescu et al., 2019, among others).

2 MATERIALS AND METHODS

2.1 Data and preliminary analysis

In this paper, we use monthly, seasonally adjusted Consumer Price Index (CPI) data from the International Monetary Fund (IMF) database to calculate the inflation rate (INFR).[2] We also collect monthly data on the global Geopolitical Risk (GPR) index, obtained from PolicyUncertainty.com. The GPR index, developed by Caldara and Iacoviello (2022), is calculated by counting the number of news articles related to adverse geopolitical events each month, relative to the total number of articles. This index provides a quantitative measure of geopolitical tensions and associated risks, with data tracing back to 1900. According to Caldara

[2] $INFR_t = 100 \times ((CPI_t - CPI_{t-12})/CPI_{t-12})$

and Iacoviello (2022), higher GPR values signal lower investment, stock prices, and employment, as well as an increased likelihood of economic crises and downside risks to the global economy. We supplement the baseline global GPR index with three alternative indicators. In particular, we employ the geopolitical risk threats (GPRT) index, which captures risks associated with rising tensions and the anticipation of conflict, and the geopolitical risk acts (GPRA) index, which reflects realized geopolitical events. Our dataset covers the period from January 2008 to December 2025, constrained by the availability of WUI data. All unadjusted series were seasonally adjusted using the X-13-ARIMA-SEATS program.

We further construct a new measure of inflation uncertainty (INFRU) based on the best-performing time-varying volatility model among several alternatives. Following Chan and Grant (2016), we estimate and compare variants of GARCH and SV models using Bayesian model comparison via Bayes factors to identify the most suitable proxy for inflation uncertainty.[3]

Table 1 Summary statistics

	INFR	INFRU	GPR	GPRT	GPRA	WUI
Minimum	–2.202	0.197	4.068	3.973	3.348	9.111
Median	2.469	1.082	4.522	4.578	4.453	9.943
Mean	2.975	1.880	4.580	4.639	4.465	10.024
Maximum	11.080	8.479	5.765	6.001	5.525	11.715
Std. Dev.	2.640	1.863	0.263	0.320	0.366	0.470
Skewness test	0.858***	1.579***	1.012***	0.866***	0.093	1.174***
Kurtosis test	4.063**	5.478***	4.570***	4.335***	3.433	4.768***
Jarque-Bera test	36.690***	144.998***	59.037***	43.053***	2.002	77.791***
ARCH-LM(10)	184.293***	188.703***	111.687***	113.308***	122.929***	140.606***
Q(10)	1034.105***	903.566***	452.152***	527.077***	626.826***	477.168***
$Q^2(10)$	999.530***	851.018***	436.654***	505.655***	591.237***	489.263***
Observation	216	216	216	216	216	216

Notes: The ARCH-LM test is performed for 10 lags. The superscripts ***, **, and * indicate rejection of the null hypothesis at the 1%, 5%, and 10% significance levels, respectively.
Source: Author's compilation

To evaluate the robustness of our results, we further incorporate the World Uncertainty Index (WUI) as a broader proxy for global uncertainty derived from country-level reports. The consistency of our results across this alternative proxy would reinforce the robustness and credibility of our empirical conclusions.

Table 1 presents preliminary summary statistics for the data series. Except for INFR and INFRU, all uncertainty indicators were converted to their logarithmic form. All series have positive mean and median values that are broadly similar. The WUI series exhibits the highest average, while INFR is the most volatile. Except for GPRA, all remaining series exhibit asymmetric distributions and positive

[3] The GARCH-class models include the GARCH (1,1), GARCH (2,1), GARCH-J, GARCH-M, GARCH-MA, and GARCH-*t* models as well as the GJR-GARCH model. For the SV-class models, we consider the standard SV, SV-2, SV-J, SV-M, SV-MA, SV with *t* innovations (SV-*t*), and SV-L models. For more details about these models as well as the estimation and comparison procedures, see Chan and Grant (2016).

kurtosis, indicating fat-tail behaviour. These characteristics violate the normality assumption, as confirmed by Skewness, Kurtosis, and Jarque-Bera tests, suggesting that all series, except GPRA, may respond differently to positive and negative shocks.

Table 2 Linearity test results

Variables	Harvey et al. (2008)'s test			Harvey et al. (2007)'s test			
	statistic W_λ	Decision		statistic $W^*_{10\%}$	statistic $W^*_{5\%}$	statistic $W^*_{1\%}$	Decision
INFR	2.62	Linear		3.24	3.76	3.97	Linear
INFRU	5.26*	Nonlinear		7.24	7.65	7.71	Linear
GPR	6.85**	Nonlinear		7.89*	8.95	9.66	Nonlinear
GPRT	6.87**	Nonlinear		7.97*	8.39	9.35	Nonlinear
GPRA	6.89**	Nonlinear		8.25*	8.49	9.79	Nonlinear
WUI	5.74*	Nonlinear		8.67*	8.43	8.56	Nonlinear
Critical Values							
1%	9.21					13.27	
5%	5.99				9.48		
10%	4.60			7.77			

Notes: ***, **, and * denote rejection of the null hypothesis of linearity at the 1%, 5%, and 10% levels, respectively. The W_λ statistic follows the chi-square distribution, and the relevant critical values are 9.21 (1%), 5.99 (5%), and 4.60 (10%). The critical values for W^* statistic are 13.27 (1%), 9.48 (5%), and 7.77 (10%).
Source: Author's compilation

All series also exhibit significant ARCH effects, suggesting that ARCH and GARCH-type models are suitable for capturing excess kurtosis and asymmetry. Additionally, Ljung-Box Q statistics for the levels and squared series indicate the presence of both linear and non-linear serial correlation. Given these characteristics, it is necessary to employ asymmetric modelling approaches.

Before conducting the empirical analysis, we perform a second set of preliminary tests to examine the linear properties of each series, using the tests of Harvey et al. (2007, 2008) with the null hypothesis of linearity against the alternative of non-linearity. Table 2 shows mixed results: the W_λ test (Harvey et al., 2008), which generally has better size and power than the W* test (Harvey et al., 2007), rejects the null of linearity for four cases at the 5% and 10% significance levels. However, linearity is not rejected for INFR, indicating that only uncertainty indices and INFRU exhibit non-linear patterns.

Table 3 Conventional unit root test results without structural breaks

Variables	ADF test				PP test		
	Level	FD	I(d)		Level	FD	I(d)
INFR	−2.350	−7.783***	I(1)		−2.919	−13.723***	I(1)
INFRU	−4.166***	−8.042***	I(0)		−3.629**	−12.239***	I(0)
GPR	−5.510***	−13.424***	I(0)		−7.259***	−21.660***	I(0)
GPRT	−6.260***	−13.408***	I(0)		−8.523***	−22.772***	I(0)
GPRA	−4.214***	−13.662***	I(0)		−5.210***	−21.421***	I(0)
WUI	−2.981	−14.137***	I(1)		−5.019***	−20.471***	I(0)

Notes: ***, and ** denote rejection of the null hypothesis at the 1% and 5% significance levels, respectively.
Source: Author's compilation

We further test the stationarity of the variables using the Augmented Dickey-Fuller (ADF), and Phillips-Perron (PP) tests, focusing on the null hypothesis with intercept and trend. Results in Table 3 show that INFR and WUI are generally I(1), while GPR, GPRT, and GPRA are mostly I(0). In addition, none of the series is I(2). In the presence of nonlinear structures and structural breaks, the conventional unit root tests may lose their explanatory power to reject the null of a unit root and thus provide misleading or biased results. These conventional tests may deceptively prove the variables to be I(1) or I(2) when there are one or more structural breaks in the series. To overcome these issues, we further apply the Zivot-Andrews (ZA) unit root test with a single unknown structural break, and the Lee-Strazicich (LS) unit root test with two unknown structural breaks. These tests have the advantage of accounting for one and two endogenous structural breaks in the intercept and both the intercept and trend terms, respectively. Using these tests allows us to determine evidence of structural break(s) in the data series and to prove the exclusion of I(2) variables.

Table 4 Unit root test results with structural breaks

Variables	ZA test			LS test		
	ADF-stat	TB		LM-stat	TB1	TB2
Level series						
INFR	–3.973	Feb–09		–5.560*	May–20	Aug–21
INFRU	–4.763**	Mar–09		–5.353*	Dec–11	Jun–19
GPR	–5.670***	Oct–20		–6.619***	Sep–15	Jul–21
GPRT	–6.003***	Jul–20		–5.942**	Nov–21	Dec–22
GPRA	–4.820**	Aug–20		–6.227**	Dec–14	Jul–21
WUI	–4.857**	Jul–24		–6.213**	Jul–20	Mar–23
First-Difference series						
ΔINFR	–8.754***	Jun–20		–8.745***	Nov–17	Jun–20
ΔINFRU	–7.541***	Jan–17		–7.872***	Jan–19	Apr–20
ΔGPR	–11.790***	Apr–25		–10.292***	Nov–19	Sep–23
ΔGPRT	–11.624***	May–25		–10.313***	Nov–19	Jul–21
ΔGPRA	–11.704***	Sep–23		–10.665***	Mar–17	Jun–17
ΔWUI	–11.590***	Aug–23		–10.405***	Jan–20	Jul–22

Notes: *** denotes rejection of the null hypothesis at the 1% significance level.
Source: Author's compilation

The results of ZA and LS tests were summarized in Table 4. They indicate that most variables, including INFR and WUI, are I(0) once structural breaks are accounted for. The findings also confirm the presence of time breaks in the data series and show that none of the variables in the model is I(2). The identified break dates can be attributed to global economic and financial crises, socio-economic turbulence, and several political and institutional factors in Saudi Arabia and worldwide.

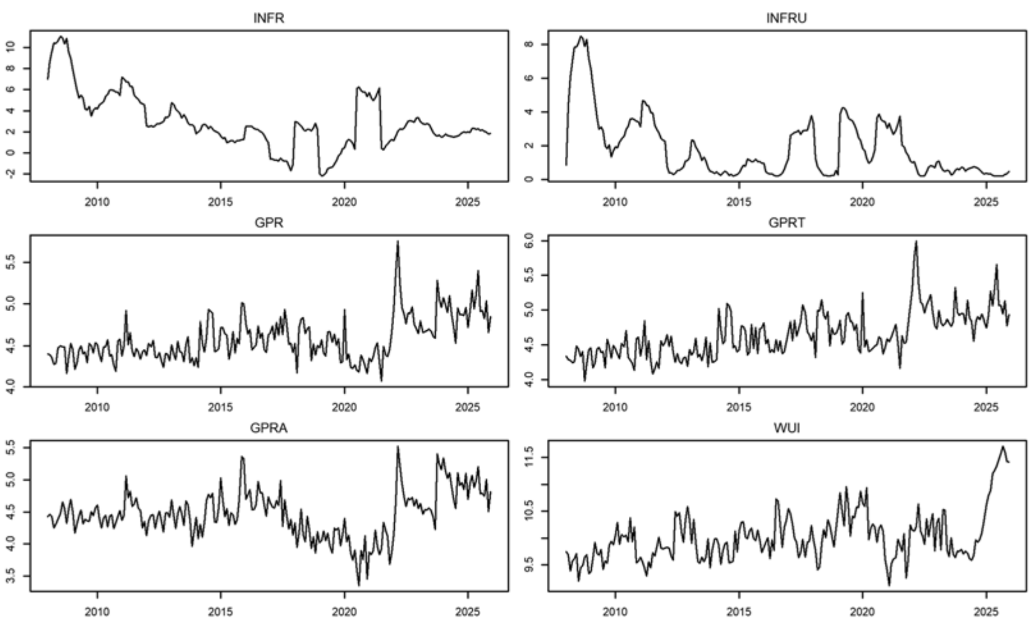
Table 5 LML of GARCH and SV models for INFR

GARCH	GARCH–2	GARCH–J	GARCH–M	GARCH–MA	GARCH–t	GARCH–GJR
–433.438 (0.094)	–409.498 (0.139)	–370.239 (0.185)	–402.978 (0.020)	–403.658 (0.062)	–407.325 (0.081)	–409.345 (0.194)
SV	SV–2	SV–J	SV–M	SV–MA	SV–t	SV–L
–417.419 (0.020)	–418.060 (0.062)	–419.655 (0.049)	–407.034 (0.032)	–408.547 (0.056)	–418.773 (0.028)	–419.045 (0.025)

Notes: The numbers between parentheses denote the numerical standard errors.
Source: Author's compilation

Given that unit root tests allowing for structural breaks indicate level stationarity for all variables, we proceed with level data for our subsequent empirical analyses. Interestingly, we model the inflation volatility and hence construct the INFRU series using fourteen time-varying volatility models, including seven GARCH models and their SV counterparts. Then, we compare model performance using Bayesian techniques (Chan and Grant, 2016; Barnett et al., 2020). Comparisons are conducted pairwise between GARCH and SV models, and between flexible and standard variants within each class. Model selection is based on log marginal likelihoods (LML) computed via the improved cross-entropy method (Chan and Eisenstat, 2015), with the model exhibiting the highest LML considered the best fit. Results, displayed in Table 5, indicate that GARCH-J is identified as the most suitable model for estimating INFRU.[4]

Figure 1 Dynamics of inflation, inflation uncertainty, and external risk measures



Source: Author's compilation

[4] Due to space considerations, we do not report the estimation results of the GARCH and SV models. These findings are available upon request from the corresponding author.

The summary statistics for the INFRU measure were reported in Table 1. INFRU is highly variable and right-skewed, indicating occasional large upward movements and a fat-tailed, non-normal distribution. Diagnostic tests show strong persistence and volatility clustering. Besides, the results in Tables 3–4 validate the stationarity assumption of the INFRU series at the level.

The plot of the INFRU series is shown in Figure 1. A comparison of INFR and INFRU plots across major crises shows a close but potentially nonlinear and positive relationship. INFR adjusts more gradually during events such as the global financial crisis, oil price shocks, the COVID-19 disruption, and subsequent geopolitical disturbances. At the same time, INFRU reacts with sharp, short-lived spikes that often precede or exceed inflation movements.

This pattern suggests that large shocks generate disproportionately strong uncertainty responses, indicating state-dependent and asymmetric dynamics in which uncertainty rises rapidly during stress periods and eases as conditions stabilize. Figure 1 also shows that geopolitical measures remain relatively stable over time but exhibit episodic surges, especially around 2022, reflecting heightened global tensions and their persistence at elevated levels thereafter. Meanwhile, WUI shows a more pronounced structural shift, remaining broadly stable before 2020 but increasing markedly in the post-COVID-19 pandemic period, indicating a transition toward a more uncertain global environment. Collectively, these patterns suggest that while inflationary pressures have become more contained over time, geopolitical developments and broader uncertainty have emerged as increasingly salient sources of macroeconomic risk.

2.2 Methods

2.2.1 Granger causality-in-mean tests

Let Y_t and X_t be two stationary time series. Let $\mathcal{F}_{t-1}^{Y} \equiv \left(Y_{t-1},\ldots,Y_{t-s}\right)' \in \mathbb{R}^s$ and $\mathcal{F}_{t-1}^{X} \equiv \left(X_{t-1},\ldots,X_{t-q}\right)' \in \mathbb{R}^q$ denote the past information sets of Y_t and X_t, respectively. The null hypothesis of Granger causality-in-distribution (Granger, 1969) from X_t to Y_t is given by:

$$F_Y\left(y \setminus \mathcal{F}_{t-1}^{Y}, \mathcal{F}_{t-1}^{X}\right) = F_Y\left(y \setminus \mathcal{F}_{t-1}^{Y}\right); \quad \forall y \in \mathbb{R} \tag{1}$$

where $F_Y(.)$ is the conditional distribution function operator.

In practice, the linear Granger causality-in-mean test is commonly used, with the null hypothesis given by:

$$\mathbb{E}\left(Y_t \setminus \mathcal{F}_{t-1}^{Y}, \mathcal{F}_{t-1}^{X}\right) = \mathbb{E}\left(Y_t \setminus \mathcal{F}_{t-1}^{Y}\right) \tag{2}$$

where $\mathbb{E}(.)$ is the conditional mean operator.

The null hypothesis given by Formula (2) is tested using the following VAR(p) model:

$$Y_t = \sum_{j=1}^{p} \alpha_j Y_{t-j} + \sum_{j=1}^{p} \beta_j X_{t-j} + \varepsilon_t \tag{3}$$

where p is selected using BIC.

We test for nonlinearity in VAR residuals using BDS and parameter stability tests (SupF, AveF, and ExpF) proposed by Andrews (1993), Andrews and Ploberger (1994), and Hansen (1997). If nonlinearity is detected, we apply nonlinear Granger causality-in-mean tests (Hiemstra and Jones, 1994; Diks and Panchenko, 2006 on the standardized VAR residuals.

2.2.2 Granger causality-in-quantiles tests

The Granger causality-in-mean tests overlook dependence in the tails of the conditional distribution. Therefore, we employ the Granger causality-in-quantiles test of Troster (2018), which allows detecting the possibly nonlinear or asymmetrical relationship between and across the conditional distribution[5]. This test is described by the following null hypothesis:

$$Q_{\tau}^{Y,X}\left(Y_t \setminus \mathcal{F}_{t-1}^{Y}, \mathcal{F}_{t-1}^{X}\right) = Q_{\tau}^{Y}\left(Y_t \setminus \mathcal{F}_{t-1}^{Y}\right); \quad \forall\, \tau \in (0,1) \tag{4}$$

where $Q_{\tau}^{Y,X}\left(. \setminus \mathcal{F}_{t-1}^{Y}, \mathcal{F}_{t-1}^{X}\right)$ and $Q_{\tau}^{Y}\left(Y_t \setminus \mathcal{F}_{t-1}^{Y}\right)$ denote the τ-quantiles of $F_Y\left(. \setminus I_t^Y, \setminus I_t^X\right)$ and $F_Y\left(. \setminus I_t^Y\right)$, respectively.

Under the null hypothesis, Troster (2018) proposed the following test statistic[6]

$$S_T = \frac{1}{Tn}\sum_{j=1}^{n}\left|\psi_{.j}^{'}\boldsymbol{W}\psi_{.j}\right| \tag{5}$$

where $\boldsymbol{W}$ is a weighting matrix, $\psi_{.j}^{'}$ is the jth column of the ψ matrix, with elements $\psi_{i,j} = \Psi_{\tau j}\left(Y_i - m\left(\mathcal{F}_i^{Y}, \theta_T\left(\tau_j\right)\right)\right)$. Here, the function $\Psi_{\tau j}(.)$ is defined as $\Psi_{\tau j}(\varepsilon) \equiv 1(\varepsilon \le 0) - \tau_j$ for all $\tau \in (0,1)$, and $m\left(\mathcal{F}_i^{Y}, \theta_T\left(\tau_j\right)\right)$ is a parametric quantile autoregressive (QAR) model for the conditional τ-quantile of Y_i.

We implement the S_T statistic defined in Formula (5) by specifying the following three linear parametric QAR models to model the conditional quantiles of Y_t for all $\tau \in (0,1)$:

$$QAR(1): m^1\left(\mathcal{F}_{t-1}^{Y}, \theta_T(\tau)\right) = \mu_0(\tau) + \mu_1(\tau)Y_{t-1} + \sigma_t\Phi_{\varepsilon}^{-1}(\tau) \tag{6}$$

$$QAR(2): m^2\left(\mathcal{F}_{t-1}^{Y}, \theta_T(\tau)\right) = \mu_0(\tau) + \mu_1(\tau)Y_{t-1} + \mu_2(\tau)Y_{t-2} + \sigma_t\Phi_{\varepsilon}^{-1}(\tau) \tag{7}$$

$$QAR(3): m^3\left(\mathcal{F}_{t-1}^{Y}, \theta_T(\tau)\right) = \mu_0(\tau) + \mu_1(\tau)Y_{t-1} + \mu_2(\tau)Y_{t-2} + \mu_3(\tau)Y_{t-3} + \sigma_t\Phi_{\varepsilon}^{-1}(\tau) \tag{8}$$

where the parameters $\theta(\tau) = \left(\mu_0(\tau), \mu_1(\tau), \mu_2(\tau), \mu_3(\tau), \sigma_t\right)^{'}$ are estimated using the maximum likelihood method for all $\tau \in (0,1)$, and $\Phi_{\varepsilon}^{-1}(\tau)$ denotes the inverse of the standard normal distribution function.

To check the sign of the causal relationship between the series, we estimate the QAR models in Formula (6)- Formula (8) including lagged variables of another variable. We provide the results using the following extended QAR models:

$$QAR(1): Q_{\tau}^{Y}\left(Y_t \setminus \mathcal{F}_{t-1}^{Y}, \mathcal{F}_{t-1}^{Z}\right) = \mu_1(\tau) + \mu_2(\tau)Y_{t-1} + \beta(\tau)Z_{t-1} + \sigma_t\Phi_{\varepsilon}^{-1}(\tau) \tag{9}$$

[5] For more details, see Troster et al. (2018), Troster et al. (2019), Ahmed et al. (2022), Mighri et al. (2022), Mighri and Sarkodie (2024), and Mighri et al. (2026), among others.

[6] The critical values for S_T are obtained via subsampling.

$$QAR(2): Q_\tau^Y\left(Y_t \setminus \mathcal{F}_{t-1}^Y, \mathcal{F}_{t-1}^Z\right) = \mu_1(\tau) + \mu_2(\tau) Y_{t-1} + \mu_3(\tau) Y_{t-2} + \beta(\tau) Z_{t-1} + \sigma_t \Phi_\varepsilon^{-1}(\tau) \quad (10)$$

$$QAR(3): Q_\tau^Y\left(Y_t \setminus \mathcal{F}_{t-1}^Y, \mathcal{F}_{t-1}^Z\right) = \mu_1(\tau) + \mu_2(\tau) Y_{t-1} + \mu_3(\tau) Y_{t-2} + \mu_4(\tau) Y_{t-3} + \beta(\tau) Z_{t-1} + \sigma_t \Phi_\varepsilon^{-1}(\tau) \quad (11)$$

where the parameters $\mu_i(\tau)$ for $i = 1,2,3$, $\beta(\tau)$, and σ_t are estimated by maximum likelihood for all $\tau \in (0,1)$.

In Formula (9)–Formula (11), we focus on the sign of the estimated parameter $\beta(\tau)$. If $\beta(\tau) > 0$, then increases (decreases) in Z will lead to increases (decreases) in Y. On the contrary, if $\beta(\tau) < 0$, then increases (decreases) in Z will result in decreases (increases) in Y.

The modeling framework advanced by Troster (2018) allows for any $\sqrt{T}$-consistent estimator of $\theta_T(\tau)$ such as the CAViaR estimator by Engle and Manganelli (2004). Troster et al., 2019 introduced nonlinear QAR models using CAViaR specifications, which specify an AR process of conditional quantiles. Interestingly, we employ two specifications of the nonlinear CAViaR models, namely the symmetric absolute value (SAV) and the asymmetric slope (AS). Formally, both SAV and AS are specified as follows:

$$SAV: Q_\tau^Y\left(Y_t \setminus \mathcal{F}_{t-1}^Y\right) = \alpha(\tau) + \beta(\tau) Q_\tau^Y\left(Y_{t-1} \setminus \mathcal{F}_{t-2}^Y\right) + \gamma(\tau)\left|Y_{t-1}\right| \quad (12)$$

$$AS: Q_\tau^Y\left(Y_t \setminus \mathcal{F}_{t-1}^Y\right) = \alpha(\tau) + \beta(\tau) Q_\tau^Y\left(Y_{t-1} \setminus \mathcal{F}_{t-2}^Y\right) + \gamma(\tau)\left(Y_{t-1}\right)^+ + \delta(\tau)\left(Y_{t-1}\right)^- \quad (13)$$

where $(Y_{t-1})^+ = \max(Y_{t-1}, 0)$ and $(Y_{t-1})^- = -\min(Y_{t-1}, 0)$.

To evaluate the sign of nonlinear Granger causality-in-quantiles, we extend the nonlinear CAViaR models by including lagged values of Z_t, i.e.,

$$SAV: Q_\tau^Y\left(Y_t \setminus \mathcal{F}_{t-1}^Y\right) = \alpha(\tau) + \beta(\tau) Q_\tau^Y\left(Y_{t-1} \setminus \mathcal{F}_{t-2}^Y\right) + \gamma(\tau)\left|Y_{t-1}\right| + \theta(\tau) Z_{t-1} \quad (14)$$

$$AS: Q_\tau^Y\left(Y_t \setminus \mathcal{F}_{t-1}^Y\right) = \alpha(\tau) + \beta(\tau) Q_\tau^Y\left(Y_{t-1} \setminus \mathcal{F}_{t-2}^Y\right) + \gamma(\tau)\left(Y_{t-1}\right)^+ + \delta(\tau)\left(Y_{t-1}\right)^- + \theta(\tau) Z_{t-1} \quad (15)$$

The sign of nonlinear Granger causality-in-quantiles is determined by the estimated coefficient $\theta(\tau)$. Specifically, if $\theta(\tau) > 0$, this indicates positive nonlinear causality, implying that increases (decreases) in Z_{t-1} lead to increases (decreases) in the conditional quantile of Y_t. Conversely, if $\theta(\tau) < 0$, this indicates negative nonlinear causality, implying that increases (decreases) in Z_{t-1} lead to decreases (increases) in the conditional quantile of Y_t.

3 RESULTS AND DISCUSSION

3.1 Preliminary test results

We begin the causality analysis with the conventional linear Granger causality test based on VAR(p) models. Consistent with the unit root test results, we use the variables in levels rather than their first differences. The findings, reported in Table 6, indicate no strong evidence of linear Granger causality between uncertainty indicators and inflation. The only weak result is a marginal effect from INFR to GPRT at the 10% significance level. This lack of significance may reflect nonlinearity or non-normality in the data. Additionally, the BG (Breusch, 1978; Godfrey, 1978) and ES (Edgerton and Shukur, 1999) tests indicate no serial correlation in the VAR residuals, confirming that the models are well specified.

Overall, linear relationships between inflation and uncertainty measures are absent, suggesting that standard linear Granger causality tests in the mean may be misspecified.

Table 6 Tests for linear Granger causality in mean and serial correlation in VAR residuals

	p-value (F-test)	*p*	P-Value (BG-test)	p-value (ES-test)
INFRU to INFR	0.417	2	0.300	0.317
GPR to INFR	0.784	1	0.219	0.238
GPRT to INFR	0.327	2	0.727	0.765
GPRA to INFR	0.997	2	0.216	0.241
WUI to INFR	0.147	1	0.185	0.199
INFR to INFRU	0.623	2	0.300	0.317

Note: The optimal lag length p of VAR model is selected based on BIC, with a maximum lag length set to 12, such that VAR residuals are serially uncorrelated under the null hypothesis. To address potential heteroskedasticity in the residuals, we employ a robust heteroskedasticity-consistent covariance matrix estimator for VAR residuals (MacKinnon and White, 1985). The superscript ** indicates statistical significance at the 5% level.
Source: Authors' compilation

We further examine potential nonlinearity in the causal relationships between inflation and uncertainty measures using the BDS test for nonlinearity in VAR residuals (see Table 7). Significant results appear only for GPRT → INFR, showing strong evidence of nonlinearity. However, most remaining pairs of variables show insignificant nonlinearity. Overall, there is evidence of nonlinear dynamics, particularly involving geopolitical risk threats and inflation, suggesting that linear models may be insufficient and thus supporting the use of nonlinear models.

Table 7 Tests for nonlinearity in VAR residuals, parameter constancy, and parameter stability

	BDS (m=2)	BDS (m=3)	BDS (m=4)	SupF	AveF	ExpF
INFR, INFRU	0.617	0.542	0.464	0.000***	0.056*	0.000***
INFR, GPR	0.119	0.212	0.348	0.000***	0.050*	0.000***
INFR, GPRT	0.007***	0.011**	0.007***	0.001***	0.123	0.002***
INFR, GPRA	0.778	0.673	0.569	0.001***	0.108	0.001***
INFR, WUI	0.116	0.210	0.344	0.000***	0.047**	0.001***
INFRU, INFR	0.986	0.895	0.866	0.000***	0.000***	0.000***

Note: The superscripts ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.
Source: Authors' compilation

In addition, we assess parameter constancy and stability using the SupF, AveF, and ExpF tests. The results in Table 7 indicate that all SupF and ExpF statistics are statistically significant, implying parameter instability and structural changes in the relationships between inflation and uncertainty variables. Only limited evidence of stability appears in some AveF results. Overall, the results suggest that model parameters are generally not constant over time, supporting the presence of structural changes and reinforcing the need for nonlinear or time-varying modelling approaches.

Based on the BDS test results, we apply nonlinear Granger causality tests in the mean. Table 8 reports test results for nonlinear Granger causality in the mean using the HJ (Hiemstra and Jones, 1994) and DP (Diks and Panchenko, 2006) tests. The findings indicate no evidence of nonlinear causality from uncertainty indicators to inflation. There is only evidence of a nonlinear causality from inflation to its uncertainty using the DP test. Overall, nonlinear effects mainly run from inflation to inflation uncertainty, while the remaining uncertainty indicators do not significantly predict inflation in a nonlinear framework.

Table 8 Tests for nonlinear Granger causality-in-mean

	L	**P-value (HJ)**	**P-value (DP)**
INFRU to INFR	2	0.734	0.640
GPR to INFR	1	0.899	0.912
GPRT to INFR	2	0.633	0.316
GPRA to INFR	2	0.609	0.337
WUI to INFR	1	0.945	0.931
INFR to INFRU	2	0.978	0.041**

Note: The optimal lag length ***L*** of bivariate VAR models is selected based on BIC such that VAR residuals are serially uncorrelated under the null hypothesis. The superscript ** indicates statistical significance at the 5% level.
Source: Authors' compilation

3.2 Linear Granger causality-in-quantiles results

To apply the linear Granger causality-in-quantiles test, we use a grid of 9 quantiles, i.e., $T_n = \{\tau_j\}_{j=1}^{n} =$.[0.10,0.90].[7] We classify the distribution into lower ([0.10, 0.30]), middle ([0.40, 0.60]), and upper ([0.70, 0.90]) quantile intervals. Interestingly, we examine the Granger causality at each quantile τ using the S_T test statistic, while considering only stationary variables. The S_T statistic is calculated using a subsample size of $b = \left[kT^{2/5} \right] \approx 43$, with $k = 5$ and T = 216.[8]

The results in Table 9 suggest that inflation uncertainty and geopolitical risk play an important role in shaping inflation dynamics in Saudi Arabia, particularly during extreme inflation conditions. The significant causality from inflation uncertainty and geopolitical risk measures to inflation at the lower and upper quantiles indicates that external uncertainty and global risk shocks affect Saudi inflation mainly during periods of unusually low or high inflation, rather than during normal conditions. This reflects Saudi Arabia's strong integration with global markets, especially through oil prices, trade and financial channels, making domestic inflation sensitive to global uncertainty shocks in stressed environments.

Meanwhile, the strong causal effect from inflation to inflation uncertainty across most quantiles implies that higher or more volatile inflation increases uncertainty about future price developments, highlighting feedback effects in domestic price dynamics. Besides, the findings are highly consistent across the two- and three-lag specifications of the QAR model, suggesting that the detected causal relationships are robust and not sensitive to lag length. Overall, the results indicate that Saudi inflation is driven by both internal uncertainty and global geopolitical risk conditions, with effects that are asymmetric and state-dependent.

7 To conserve space, we do not report the results for a grid of 19 quantiles as they do not alter our conclusion. These results are available from the corresponding author upon request.
8 Our results remain similar for other values of the constant parameter *k*.

Table 9 Linear Granger causality-in-quantiles test results

τ	INFRU to INFR			INFR to INFRU			GPR to INFR		
	L **= 1**	***L*** **= 2**	***L*** **= 3**	***L*** **= 1**	***L*** **= 2**	***L*** **= 3**	***L*** **= 1**	***L*** **= 2**	***L*** **= 3**
0.10	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***
0.20	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***
0.30	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***
0.40	0.006***	0.017**	0.006***	0.128	0.070*	0.093*	0.006***	0.006***	0.006***
0.50	0.576	0.500	0.517	0.105	0.006***	0.006***	0.523	0.488	0.500
0.60	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***
0.70	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***
0.80	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***
0.90	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***
$\forall\tau$	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***
τ	**GPRT to INFR**			**GPRA to INFR**			**WUI to INFR**		
	L **= 1**	***L*** **= 2**	***L*** **= 3**	***L*** **= 1**	***L*** **= 2**	***L*** **= 3**	***L*** **= 1**	***L*** **= 2**	***L*** **= 3**
0.10	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***
0.20	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***
0.30	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***
0.40	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***
0.50	0.529	0.488	0.494	0.523	0.488	0.500	0.576	0.547	0.523
0.60	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***
0.70	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***
0.80	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***
0.90	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***
$\forall\tau$	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***

Note: This table reports the subsampling p-values for the $\boldsymbol{S_T}$ statistic. $\boldsymbol{L}$ is the lag order of the QAR model. The superscripts ***, **, and * indicate rejection of the null hypothesis at the 1%, 5%, and 10% significance levels, respectively.
Source: Authors' compilation

The linear Granger causality-in-quantiles analysis reveals the lead-lag relationships between uncertainty measures and inflation. However, understanding these relationships requires not only establishing causality but also determining the direction and sign of the causal effects. To address this, we employ quantile regression (QR) estimates, which complement Granger causality-in-quantiles by allowing interpretation of the sign of causal relationships. More importantly, we assess the sign of linear Granger causality-in-quantiles by estimating three QAR models as described in Formula (9)–Formula (11).

Table 10 reports the sign and magnitude of the linear Granger causality-in-quantiles effects using QAR models, showing whether the causal impact is positive or negative across quantiles. The results indicate that global geopolitical risk and inflation uncertainty generally exert positive effects on Saudi inflation, although their impacts are not consistently significant across all quantiles. In contrast, inflation strongly and positively influences inflation uncertainty across most quantiles, indicating a reinforcing feedback effect in domestic price dynamics and hence supporting the Friedman-Ball hypothesis. The findings also show that geopolitical threats generally exert a positive but weak effect on Saudi inflation, with significance mainly at lower quantiles, while geopolitical acts have no systematic impact on inflation.

Table 10 Sign of linear Granger causality-in-quantiles

τ	INFRU to INFR				INFR to INFRU				GPR to INFR		
	QAR(1)	QAR(2)	QAR(3)		QAR(1)	QAR(2)	QAR(3)		QAR(1)	QAR(2)	QAR(3)
0.10	0.130	0.121	0.136		0.263***	0.286***	0.289***		0.246	0.342	0.212
0.20	0.033	0.046	0.052		0.333***	0.333***	0.331***		0.325	0.224	0.164
0.30	0.055	0.015	0.023		0.322***	0.322***	0.323***		−0.003	−0.007	−0.130
0.40	0.047	−0.004	−0.031		0.310***	0.309***	0.312***		−0.002	−0.063	−0.047
0.50	0.029	−0.025	−0.030		0.312***	0.314***	0.309***		0.041	0.039	0.048
0.60	0.080	−0.041	−0.046		0.321***	0.322***	0.322***		0.022	−0.034	0.001
0.70	0.108	−0.024	−0.047		0.332***	0.328***	0.324***		0.053	0.047	0.047
0.80	0.107	0.090	−0.050		0.347***	0.332***	0.284***		0.309	0.340	0.172
0.90	0.092	0.143	0.101		0.246***	0.250***	0.253***		0.146	0.199	0.207
τ	GPRT to INFR				GPRA to INFR				WUI to INFR		
	QAR(1)	QAR(2)	QAR(3)		QAR(1)	QAR(2)	QAR(3)		QAR(1)	QAR(2)	QAR(3)
0.10	0.400	0.445	0.211		0.101	0.111	0.004		−0.048	−0.046	−0.053
0.20	0.310*	0.328*	0.328*		0.000	−0.009	−0.081		0.093	0.077	0.092
0.30	0.004	0.018	−0.001		−0.019	−0.087	−0.105		0.123	0.119	0.130
0.40	−0.002	−0.051	−0.041		0.016	0.009	0.004		0.041	0.047	0.049
0.50	0.006	−0.004	−0.010		0.058	0.054	0.058		0.021	0.032	0.030
0.60	0.010	−0.036	−0.042		0.010	0.005	0.001		0.007	−0.001	−0.003
0.70	0.047	0.045	0.046		0.026	0.037	0.031		−0.048	−0.055	−0.040
0.80	0.251	0.275	0.261		0.192	0.100	0.067		−0.077	−0.102	−0.103
0.90	0.145	0.159	0.183		−0.027	−0.028	−0.044		−0.057	−0.081	−0.093

Note: The superscripts ***, **, and * denote rejection of the null hypothesis at the 1%, 5%, and 10% significance levels, respectively.
Source: Authors' compilation

3.3 Nonlinear Granger causality-in-quantiles results

We now turn to the nonlinear Granger causality-in-quantiles test results using nonlinear CAViaR models (SAV and AS). To motivate this choice, we first compare quantile forecasts for INFR using linear and nonlinear QR models. Specifically, we compare the root mean squared error (RMSE) of 120 recursive monthly out-of-sample quantile forecasts of all selected quantiles of the distribution. The first forecast was performed on January 1, 2016. For each forecast, the in-sample period was expanded by one month to generate updated predictions.

Table 11 RMSE of out-of-sample quantile forecasts

τ	QAR(1)	QAR(2)	QAR(3)	SAV	AS
0.10	1.0298	**1.0072**	1.0240	1.7306	1.0605
0.20	0.9721	**0.9672**	0.9708	1.4103	0.9902
0.30	0.9572	**0.9534**	0.9577	1.3497	0.9671
0.40	0.9538	**0.9506**	0.9516	1.3421	0.9608
0.50	0.9542	0.9552	0.9563	1.3486	0.9539
0.60	0.9570	0.9557	0.9574	1.3642	0.9541
0.70	**0.9615**	0.9635	0.9647	1.3714	0.9696
0.80	0.9908	0.9992	1.0027	1.3447	0.9809
0.90	**1.0398**	1.0494	1.0604	1.3657	1.1633

Note: The model with the lowest RMSE value is in bold.
Source: Authors' compilation

Table 11 compares the forecasting performance of linear QAR and nonlinear CAViaR models (SAV, AS) for predicting changes in inflation. The results indicate that forecasting performance varies across inflation regimes. Linear QAR models yield the lowest RMSE values, particularly at the lower and upper quantiles, thereby outperforming and offering better specification than their competing nonlinear CAViaR models. This finding suggests that inflation dynamics are relatively stable and linear during low- and high-inflation periods. However, nonlinear CAViaR models become more competitive or outperform at the middle and upper quantiles, indicating that inflation exhibits stronger nonlinear and asymmetric behaviour in normal and high-inflation regimes.

Table 12 Nonlinear Granger causality-in-quantiles test results

τ	INFRU to INFR			INFR to INFRU			GPR to INFR	
	SAV	AS		SAV	AS		SAV	AS
∀τ	0.007***	0.355		0.014**	0.022**		0.007***	0.341
0.10	0.007***	0.551		0.029**	0.022**		0.007***	0.348
0.20	0.007***	0.464		0.007***	0.007***		0.007***	0.326
0.30	0.007***	0.486		0.029**	0.036**		0.007***	0.696
0.40	0.007***	0.203		0.159	0.180		0.080*	0.572
0.50	0.268	0.558		0.094*	0.138		0.290	0.638
0.60	0.377	0.377		0.145	0.145		0.399	0.239
0.70	0.449	0.384		0.159	0.087*		0.391	0.188
0.80	0.333	0.181		0.065*	0.043**		0.225	0.167
0.90	0.493	0.058*		0.065*	0.130		0.384	0.036**
τ	GPRT to INFR			GPRA to INFR			WUI to INFR	
	SAV	AS		SAV	AS		SAV	AS
∀τ	0.007***	0.312		0.007***	0.275		0.007	0.268
0.10	0.007***	0.406		0.007***	0.283		0.007***	0.420
0.20	0.007***	0.261		0.007***	0.500		0.007***	0.275
0.30	0.007***	0.652		0.007***	0.725		0.007***	0.283
0.40	0.080*	0.594		0.087*	0.601		0.058*	0.457
0.50	0.297	0.645		0.297	0.630		0.261	0.623
0.60	0.391	0.246		0.399	0.210		0.362	0.254
0.70	0.384	0.181		0.362	0.167		0.362	0.232
0.80	0.210	0.116		0.210	0.087*		0.348	0.507
0.90	0.326	0.036**		0.341	0.036**		0.341	0.065*

Note: This table reports the subsampling p-values for the $\boldsymbol{S_T}$ statistic. The superscripts ***, **, and * indicate rejection of the null hypothesis at the 1%, 5%, and 10% significance levels, respectively.
Source: Authors' compilation

Table 12 displays the nonlinear Granger causality-in-quantiles test results for INFR using nonlinear quantile specifications under H_0 in Formula (3). The results show evidence of nonlinear, quantile-dependent causality between inflation and uncertainty measures. This suggests that Saudi inflation responds asymmetrically to global uncertainty and domestic inflation dynamics, particularly during specific economic conditions. The SAV specification reveals strong nonlinear causal effects from inflation uncertainty and geopolitical risk indicators (GPR, GPRT, GPRA) to Saudi inflation, mainly at lower quantiles, indicating that uncertainty and external risk shocks influence domestic inflation primarily during low-inflation periods. This may reflect a stronger sensitivity of domestic prices to global developments when inflationary pressures are subdued, possibly through oil prices, import costs, and global market conditions. The strong nonlinear feedback from inflation to its uncertainty, particularly at lower and (extreme) upper quantiles, implies that inflation fluctuations increase uncertainty about future price movements, reinforcing instability in domestic price expectations. Overall, the findings highlight the dominant role of external uncertainty in shaping Saudi inflation and the presence of asymmetric state-dependent inflation dynamics.

Table 13 reports the sign of nonlinear Granger causality-in-quantiles, allowing us to identify not only whether causal relationships exist but also whether their effects are positive or negative across different quantiles of the distribution. The results indicate that the only robust nonlinear causal relationship is from inflation uncertainty to inflation, which is significantly negative at lower and middle quantiles, implying that higher uncertainty dampens inflation, particularly in low-inflation regimes. This effect gradually weakens and becomes insignificant toward the upper quantiles, highlighting strong state-dependent behaviour. In contrast, there is no evidence of a reverse causal effect (INFR to INFRU), and geopolitical risk measures exhibit sporadic and mostly insignificant effects, with only limited significance at specific quantiles. Overall, the direction of causality is predominantly negative and driven by domestic uncertainty, while external risk factors play a minor and inconsistent role across different market conditions.

Table 13 Sign of nonlinear Granger causality-in-quantiles

τ	INFRU to INFR		INFR to INFRU		GPR to INFR	
	SAV	AS	SAV	AS	SAV	AS
0.10	−0.773***	−0.021	0.006	0.002	−0.323	−0.074
0.20	−0.520***	0.014	0.015	−0.004	0.618	−0.054
0.30	−0.422***	0.02	0.01	−0.004	0.331	0.087
0.40	−0.345***	0.026	0.021	0.019	0.22	0.042
0.50	−0.198*	0.006	0.003	0.003	0.128	−0.092
0.60	−0.082	0.047	0.006	0.007	0.046	−0.076
0.70	−0.05	0.057	−0.008	−0.009	−0.001	−0.08
0.80	−0.026	0.077	0.004	0.031	0.037	−0.09
0.90	0.012	0.065	−0.002	−0.003	−0.028	−0.068

τ	GPRT to INFR		GPRA to INFR		WUI to INFR	
	SAV	AS	SAV	AS	SAV	AS
0.10	1.183	0.188	0.183	−0.039	−0.094	0.009
0.20	0.45	−0.06	0.322	0.066	0.122	0.008
0.30	0.269	0.081	0.212	−0.054	0.038	0.008
0.40	0.193	0.072	0.108	0.043	0.062	0.001
0.50	0.152	0.021	0.084	−0.061	0.081	0.001
0.60	0.064	0.035	−0.0003	−0.072	0.006	0.008
0.70	0.12	−0.033	−0.02	−0.076	0.018	0.007
0.80	0.087	−0.038	−0.123	−0.158	0.054	0.039
0.90	0.195	−0.006	−0.157	−0.151	−0.009	0.099

Note: The superscripts ***, **, and * denote rejection of the null hypothesis at the 1%, 5%, and 10% significance levels, respectively.
Source: Authors' compilation

3.4 Robustness checks

As a robustness check, the inclusion of WUI as an alternative global risk measure confirms the limited role of external uncertainty in driving inflation dynamics in Saudi Arabia. While the nonlinear Granger causality-in-quantiles results indicate some evidence of causality from WUI to inflation at specific quantiles, particularly in the lower and upper tails, these effects are neither consistent nor stable across the distribution. Moreover, the sign analysis reveals that the impact of WUI is weak, mixed, and statistically insignificant across most quantiles, with no clear directional pattern. Overall, these findings suggest that global uncertainty does not exert a robust or systematic influence on inflation, reinforcing the conclusion that domestic factors, rather than external risk measures, play a more dominant role in shaping inflationary dynamics in Saudi Arabia.

4 CONCLUSION, POLICY IMPLICATIONS, AND LIMITATIONS

The findings of this study underscore the importance of nonlinear and state-dependent dynamics in understanding inflation dynamics in Saudi Arabia. While linear Granger causality-in-quantiles confirms bidirectional linkages between inflation and inflation uncertainty, the sign analysis reveals that inflation tends to increase its uncertainty, whereas nonlinear results show that inflation uncertainty can reduce inflation in lower and middle regimes, supporting the Friedman and Friedman-Ball hypotheses, respectively. In contrast, although global geopolitical risk measures exhibit significant causal effects, their direction is unstable and lacks consistency, indicating a limited and non-systematic role in driving inflation.

These results carry several important policy implications. First, the Saudi central bank should strengthen policy credibility and communication strategies to anchor inflation expectations and reduce uncertainty, particularly given its measurable impact on inflation dynamics. Second, the negative effect of uncertainty on inflation in low-inflation regimes suggests that policymakers should avoid excessive tightening during such periods, as uncertainty itself may already dampen inflationary pressures. Third, given the weak and unstable influence of global geopolitical risks, monetary policy should remain primarily focused on domestic conditions, while external shocks should be managed through targeted fiscal and structural policies, especially those linked to oil revenue stabilization. Finally, the presence of quantile-dependent effects highlights the need for state-contingent policy frameworks, where policy responses are adapted to prevailing inflation regimes rather than relying on a uniform approach, thereby enhancing the effectiveness of macroeconomic stabilization in Saudi Arabia.

This study is not without limitations and offers avenues for future research. Future studies may extend the analysis by integrating time-frequency coherence and quantile causality frameworks, employing wavelet transforms to capture dynamic interactions and lead-lag relationships across different frequencies. Moreover, future research could expand the analysis to include all Gulf Cooperation Council countries, enabling a comparative assessment of inflation-uncertainty dynamics across oil-dependent economies. Finally, given the increasing political tensions between the U.S. and Iran, which have recently intensified and affected regional stability, energy infrastructure, and economic conditions in the Gulf, further investigation is warranted to better understand how such evolving geopolitical risks influence inflation dynamics in Saudi Arabia and the broader region.

ACKNOWLEDGMENT

This work was funded by the University of Jeddah, Jeddah, Saudi Arabia, under grant No. (UJ-22-DR-37). The authors, therefore, acknowledge with thanks the University of Jeddah for its technical and financial support.

References

ADEOSUN, O. A., TABASH, M. I., VO, X. V., ANAGREH, S. (2023). Uncertainty measures and inflation dynamics in selected global players: a wavelet approach. *Quality and Quantity* 57, 3389-3424. https://doi.org/10.1007/s11135-022-01513-7.

AHMED, A., GRANBERG, M., TROSTER, V., UDDIN, G. S. (2022). Asymmetric dynamics between uncertainty and unemployment flows in the United States. *Studies in Nonlinear Dynamics and Econometrics*, 26(1), 155–172. https://doi.org/10.1515/snde-2019-0058.

ALBULESCU, C. T., TIWARI, A. K., MILLER, S. M., GUPTA, R. (2019). Time–frequency relationship between US inflation and inflation uncertainty: evidence from historical data. *Scottish Journal of Political Economy*, 66(5), 673–702. https://doi.org/10.1111/sjpe.12207.

ALIEDAN, M. (2022). The Geopolitics of International Trade in Saudi Arabia: Saudi Vision 2030. *Cuadernos de Economia*, 45(127), 11–19. https://doi.org/10.32826/cude.v1i127.601.

ANDREWS, D. W. K. (1993). Tests for Parameter Instability and Structural Change With Unknown Change Point. *Econometrica*, 61(4), 821. https://doi.org/10.2307/2951764.

ANDREWS, D. W. K., PLOBERGER, W. (1994). Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative. *Econometrica*, 62(6), 1383. https://doi.org/10.2307/2951753.

BALL, L. (1992). Why does high inflation raise inflation uncertainty? *Journal of Monetary Economics*, 29(3), 371–388. https://doi.org/10.1016/0304-3932(92)90032-W.

BARNETT, W. A., JAWADI, F., FTITI, Z. (2020). Causal relationships between inflation and inflation uncertainty. *Studies in Nonlinear Dynamics and Econometrics* 24(5). https://doi.org/10.1515/snde-2019-0094.

BERUMENT, H., DINCER, N. (2005). Inflation and inflation uncertainty in the G-7 countries. *Physica A: Statistical Mechanics and Its Applications*, 348, 371–379. https://doi.org/10.1016/j.physa.2004.09.003.

BOURI, E., GABAUER, D., GUPTA, R., KINATEDER, H. (2023). Global geopolitical risk and inflation spillovers across European and North American economies. *Research in International Business and Finance*, 66, 102048. https://doi.org/10.1016/j.ribaf.2023.102048.

BREUSCH, T. S. (1978). Testing for Autocorrelation in Dynamic Linear Models. *Australian Economic Papers*, 17(31), 334–355. https://doi.org/10.1111/j.1467-8454.1978.tb00635.x.

CALDARA, D., IACOVIELLO, M. (2022). Measuring Geopolitical Risk. *American Economic Review*, 112(4), 1194-1225. https://doi.org/10.1257/aer.20191823.

CALDARA, D., CONLISK, S., IACOVIELLO, M., PENN, M. Do Geopolitical Risks Raise or Lower Inflation?. *Journal of International Economics*, 159, 104188. https://doi.org/10.1016/j.jinteco.2025.104188.

CHAN, J. C. C., EISENSTAT, E. (2015). Marginal Likelihood Estimation with the Cross-Entropy Method. *Econometric Reviews*, 34(3), 256–285. https://doi.org/10.1080/07474938.2014.944474.

CHAN, J. C. C., GRANT, A. L. (2016). Modeling energy price dynamics: GARCH versus stochastic volatility. *Energy Economics*, 54, 182–189. https://doi.org/10.1016/j.eneco.2015.12.003.

CONRAD, C., KARANASOS, M. (2005). On the inflation-uncertainty hypothesis in the USA, Japan and the UK: A dual long memory approach. *Japan and the World Economy*, 17(3), 327–343. https://doi.org/10.1016/j.japwor.2004.03.002.

CUKIERMAN, A., MELTZER, A. H. (1986). A Theory of Ambiguity, Credibility, and Inflation under Discretion and Asymmetric Information. *Econometrica*, 54(5), 1099. https://doi.org/10.2307/1912324.

DIKS, C., PANCHENKO, V. (2006). A new statistic and practical guidelines for nonparametric Granger causality testing. *Journal of Economic Dynamics and Control*, 30(9–10), 1647–1669. https://doi.org/10.1016/j.jedc.2005.08.008.

EDGERTON, D., SHUKUR, G. (1999). Testing autocorrelation in a system perspective testing autocorrelation. *Econometric Reviews*, 18(4), 343-386. https://doi.org/10.1080/07474939908800351.

ENGLE, R. F., MANGANELLI, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics*, 22(4), 367–381. https://doi.org/10.1198/073500104000000370.

FOUNTAS, S., KARANASOS, M., KIM, J. (2006). Inflation uncertainty, output growth uncertainty and macroeconomic performance. *Oxford Bulletin of Economics and Statistics*, 68(3), 319–343. https://doi.org/10.1111/j.1468-0084.2006.00164.x.

FRIEDMAN, M. (1977). Nobel Lecture: Inflation and Unemployment. *Journal of Political Economy*, 85(3), 451–472. https://doi.org/10.1086/260579.

GRANGER, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 37(3), 424-438. https://doi.org/10.2307/1912791.

HARVEY, D. I., LEYBOURNE, S. J., TAYLOR, A. M. R. (2007). A simple, robust and powerful test of the trend hypothesis. Journal of Econometrics, 141(2), 1302–1330. https://doi.org/10.1016/j.jeconom.2007.02.005.

HARVEY, D. I., LEYBOURNE, S. J., XIAO, B. (2008). A powerful test for linearity when the order of integration is unknown. Studies in Nonlinear Dynamics & Econometrics 12(3), 1-24. https://doi.org/10.2202/1558-3708.1582.

HOLLAND, A. S. (1995). Inflation and Uncertainty: Tests for Temporal Ordering. Journal of *Money, Credit and Banking*, 27(3), 827-837. https://doi.org/10.2307/2077753.

LEE, C. C., OLASEHINDE-WILLIAMS, G., ÖZKAN, O. (2023). Geopolitical oil price uncertainty transmission into core inflation: Evidence from two of the biggest global players. *Energy Economics*, 126. https://doi.org/10.1016/j.eneco.2023.106983.

MIGHRI, Z., RAGOUBI, H., SARWAR, S., WANG, Y. (2022). Quantile Granger causality between US stock market indices and precious metal prices. *Resources Policy*, 76, 102595. https://doi.org/10.1016/j.resourpol.2022.102595.

MIGHRI, Z., SARKODIE, S. A. (2024). Interdependency and causality between green technology innovation and consumption-based carbon emissions in Saudi Arabia: fresh insights from quantile-on-quantile and causality-in-quantiles approaches. *Environmental Science and Pollution Research*, 31(6), 9288–9316. https://doi.org/10.1007/s11356-023-31571-9.

MIGHRI, Z., RAGOUBI, H., SARWAR, S. (2026). Do Stock Market Return and Volatility Respond to U.S. Partisan Conflict Uncertainty? *Scottish Journal of Political Economy*, e70075. https://doi.org/ 10.1111/sjpe.70075.

NOGUERA-SANTAELLA, J. (2016). Geopolitics and the oil price. *Economic Modelling*, 52(January 1973), 301–309. https://doi.org/10.1016/j.econmod.2015.08.018.

POURGERAMI, A., MASKUS, K. E. (1987). The effects of inflation on the predictability of price changes in Latin America: Some estimates and policy implications. *World Development*, 15(2), 287–290. https://doi.org/10.1016/0305-750X(87)90083-0.

TROSTER, V. (2018). Testing for Granger-causality in quantiles. *Econometric Reviews*, 37(8), 850–866. https://doi.org/10.1080/07474938.2016.1172400.

TROSTER, V., BOURI, E., ROUBAUD, D. (2019). A quantile regression analysis of flights-to-safety with implied volatilities. *Resources Policy*, 62, 482–495. https://doi.org/10.1016/j.resourpol.2018.10.004.

TROSTER, V., SHAHBAZ, M., UDDIN, G. S. (2018). Renewable energy, oil prices, and economic activity: A Granger-causality in quantiles analysis. *Energy Economics*, 70, 440–452. https://doi.org/10.1016/j.eneco.2018.01.029.

YANG, T., DONG, Q., DU, M., DU, Q. (2023). Geopolitical risks , oil price shocks and inflation : Evidence from a TVP–SV–VAR approach. *Energy Economics*, 127(PB), 107099. https://doi.org/10.1016/j.eneco.202.

Cluster Analysis of the EU Countries during the Covid-19 Period (2019–2021)

Andrii Hrabariev[1] | *Kyiv National Economic University named after Vadym Hetman, Kyiv, Ukraine*
Vasyl Derbentsev[2] | *Kyiv National Economic University named after Vadym Hetman, Kyiv, Ukraine*
Mykhailo Baraniuk[3] | *Kyiv National Economic University named after Vadym Hetman, Kyiv, Ukraine*

Received 11.8.2025 (revision received 26.11.2025), **Accepted** (reviewed) 11.12.2025, **Published** 11.9.2026

Abstract

This study investigates structural shifts in European Union countries banking system typologies during the COVID-19 period from 2019 to 2021. Using the k-means clustering method (compared with hierarchical agglomerative clustering and Gaussian mixture models) on 10 financial and macroeconomic indicators (e.g., non-performing loans, z-score, GDP growth) at the country level for 25 European Union countries, we identified four distinct clusters with stable typological definitions but evolving country compositions. Comparison of cluster profiles reveals the pandemic period as a structural catalyst, reinforcing the resilience of Nordic Resilience Systems) and Core Stability Systems clusters. Significant compositional shifts occurred: Dynamic Growth Systems absorbed advanced economies, balancing growth and stability, while High Intermediation Systems transformed into a vulnerable typology, marked by high fiscal risk and low stability buffers. These findings highlight structural shifts associated with the pandemic period and differentiated policy responses, and inform regulatory strategies for enhancing EU banking resilience.

Keywords	***DOI***	***JEL code***
EU banking systems, COVID-19, cluster analysis, financial resilience, banking and macroeconomic indicators, structural shifts, temporal analysis.	*https://doi.org/10.54694/stat.2025.40*	*G21, G28, C38*

INTRODUCTION

The stability and efficiency of banking systems are paramount to the economic health and resilience of any region, particularly within a highly integrated economic bloc like the European Union (EU).

[1] Andrii Hrabariev, PhD, Associated Professor, Department of Informatics and Systemology, Kyiv National Economic University named after Vadym Hetman, Prospekt Beresteiskyi 54/1, Kyiv, 03057, Ukraine. E-mail: andrgrab@gmail.com, phone: +380 68 158 8558. ORCID: https://orcid.org/0000-0001-6165-0996.

[2] Vasyl Derbentsev, PhD, Professor, Department of Informatics and Systemology, Kyiv National Economic University named after Vadym Hetman, Prospekt Beresteiskyi 54/1, Kyiv, 03057, Ukraine. Corresponding author: e-mail: derbv@kneu.edu.ua, phone: +380 67 407 6179. ORCID: https://orcid.org/0000-0002-8988-2526.

[3] Mykhailo Baraniuk, PhD student, Department of Informatics and Systemology, Kyiv National Economic University named after Vadym Hetman, Prospekt Beresteiskyi 54/1, Kyiv, 03057, Ukraine.

Defined by a complex interplay of macroeconomic conditions, regulatory frameworks, and inherent financial sector characteristics, the performance and robustness of national banking sectors significantly influence overall financial stability (Demirgüç-Kunt et al., 2021).

The period between 2019 and 2021 presents a unique and critical juncture for examining these dynamics. The year 2019 serves as a baseline, representing the pre-pandemic financial landscape, while 2021 captures the immediate aftermath and initial recovery phase from the unprecedented COVID-19 shock. This period was also characterized by massive monetary and fiscal policy interventions, including the European Central Bank's Pandemic Emergency Purchase Programme (PEPP) and extensive loan moratoriums, designed to buffer banking systems from the pandemic's direct impacts (ECB, 2021; EBA, 2020). Understanding the shifts and resilience within EU banking systems during this turbulent period is crucial for policymakers, regulators, and market participants alike (EBA, 2020).

Traditional assessments of financial stability often rely on aggregate indicators or country--specific analyses. However, a more granular understanding can be achieved through unsupervised machine learning techniques, specifically clustering, which group countries based on their inherent financial similarities (Mihalech & Košíková, 2021; Poznyak & Kolyada, 2023). Such methods, applied to comprehensive financial and macro-financial indicators, can reveal distinct banking system typologies and underlying structural similarities not immediately apparent through conventional analysis.

This study aims to investigate the stability and evolution of banking system typologies across EU countries (excluding Greece and Luxembourg, identified as consistent outliers) over the 2019–2021 period. Utilizing the k-means clustering method (K-Means), compared with hierarchical agglomerative clustering (HAC) and Gaussian mixture models (GMM), we identify four robust typologies based on ten key financial and macro-financial indicators. Our approach examines how these typologies, defined by characteristic profiles of asset quality, capitalization, and macroeconomic exposure, transformed in country composition and risk characteristics during the pandemic period. Country-level analysis captures systemic patterns informing macroprudential policy, though it necessarily aggregates over within-country heterogeneity.

Considering these objectives, we propose the following *working hypotheses:*

H1: Resilience of Core Clusters. Despite the exogenous shock of the COVID-19 pandemic and associated policy responses, core clusters of financially stable and mature Western European banking systems, as well as highly resilient Nordic banking systems, will demonstrate a significant stability in their composition from 2019 to 2021, exhibiting robust financial health indicators throughout the period.

H2: Emergence of Differentiated Vulnerabilities. The pandemic period will be associated with re--clustering of countries with pre-existing or newly exposed vulnerabilities, particularly among Central and Eastern European (CEE) nations and smaller, tourism-dependent economies. These newly formed or significantly altered clusters will exhibit higher levels of non-performing loans (NPLs), increased government debt, and potentially lower z-scores in 2021 compared to 2019.

H3: Dynamic Adaptation of Specialized Systems. Countries previously identified as having "specialized" or "transitional" banking systems will exhibit diverse adaptive responses to the pandemic period, leading to their re-alignment into either more stable core groups or new clusters with altered risk profiles defined by specific emerging risks or opportunities. This will manifest in notable shifts in their cluster membership and indicators mean values.

By empirically testing these hypotheses through a comparative analysis of clustering results and the dynamics of key financial indicators, this research aims to contribute significantly to the understanding of banking system resilience and the evolving financial architecture of the European Union during periods of major economic shocks combined with unprecedented policy interventions.

1 LITERATURE SURVEY

The COVID-19 pandemic significantly impacted EU banking systems, prompting extensive research on their resilience, typologies, and policy responses during 2019–2021.

Robust pre-crisis capital buffers (CET1 ~15% in 2019 vs. ~7% in 2008) mitigated losses during the pandemic (EBA, 2020). Dursun-de Neef and Schandlbauer (2021) found that worse-capitalized banks increased lending by ~10% to avoid loan loss recognition, while better-capitalized banks faced a ~15% rise in delinquent loans. Cao and Chou (2022) confirmed that banks with higher regulatory capital (CAR >12%) lent more resiliently, particularly in Nordic and Core Western European clusters, maintaining NPL ratios below 2% (Demirgüç-Kunt et al., 2021). Clichici and Zeldea (2021) highlighted the resilience of Central and Eastern European (CEE) banks, with low NPL ratios (~2%) and high capitalization.

Alessi et al. (2022) noted challenges in NPL provisioning in Southern banks due to GDPG declines (~6.1% in 2020), while Ari et al. (2021) reported a ~20% increase in delinquent loans in vulnerable systems. Kozak (2021) emphasized equity ratios' role in the CEE resilience, and Salazar et al. (2023) linked IFRS 9 adoption to improved z-score stability in Western European banks. Foglia et al. (2022) underscored high volatility connectedness in Eurozone banks.

Clustering methodologies reveal EU banking heterogeneity. Mihalech and Košíková (2021) applied clustering to EBA risk indicators, identifying groups (eastern, southern, northern, core). Danko and Suchý (2021) used clustering to assess financial integration, identifying leaders and lagging systems, supporting our typology.

Kaminskyi and Nehrey (2023) applied clustering to assess US banks' resilience to COVID-19 shocks, using indicators like total assets and recovery rates, reinforcing clustering applicability for identifying resilience patterns, despite focusing on a non-EU context. Our study extends these by using a comprehensive set of financial and macroeconomic indicators (see Table A1), capturing longitudinal shifts across 2019–2021.

Monetary and fiscal interventions stabilized EU banks. The European Central Bank's Pandemic Emergency Purchase Programme (PEPP, €1.85 trillion) reduced funding pressures (ECB, 2021). EBA (2020) reported that loan moratoriums increased CAR by ~1.5% and liquidity by ~10%, stabilizing NPL at ~2.5% by Q4 2020 (EBA, 2021). Feyen et al. (2021) highlighted PEPP's role in mitigating GDEBT pressures (>100% in Southern Europe), though IMF (2021) noted that €2 trillion in loan guarantees raised fiscal risks. Demir and Danisman (2021) emphasized fiscal support's role in liquidity stability. Colak and Öztekin (2021) found ECB policies boosted lending by ~15% in resilient banks.

The macroeconomic environment shaped banking responses. The World Bank (2022) reported a 6.1% GDP contraction in 2020, with unemployment rate (UNEMP) peaking at 7.8% in 2020. Chodnicka-Jaworska (2022) found that GDPG declines and UNEMP spikes delayed bank rating downgrades by 6–12 months. Plikas et al. (2024) showed that lockdown strictness increased SME NPLs by ~30% in Southern Europe.

Existing literature provides insights but has gaps. Studies like Dursun-de Neef and Schandlbauer (2021), Alessi et al. (2022), and Foglia et al. (2022) focus on bank performance but overlook structural shifts. Mihalech and Košíková (2021), Danko and Suchý (2021), and Kaminskyi and Nehrey (2023) apply clustering but use fewer indicators and different contexts.

Our study addresses these gaps by clustering the EU-25 (excluding Greece and Luxembourg) using 10 indicators from TheGlobalEconomy.com (2025) and the World Bank's Global Financial Development Database (2025). We applied K-Means compared with the results obtained by other methods to identify dynamic typologies and provide a basis for developing regulatory strategies.

2 METHODS

This study applies a multistage methodology designed to explore the structural characteristics and resilience of the banking systems of the EU member states under the influence of the COVID-19 pandemic. The methodological approach is focused on multivariate statistical analysis, including feature selection techniques and unsupervised clustering algorithms, with the goal of identifying stable groupings of countries and detecting any systemic shifts across selected time periods (Hastie et al., 2009).

2.1 Research Design and Periodization

The analytical design is structured around two reference years: 2019 (pre-COVID baseline) and 2021 (post-pandemic adaptation). This temporal segmentation allows for comparison of structural changes in national banking systems across distinct phases of external pressure. By employing a uniform set of indicators and methodological procedures across these years, we ensure internal consistency and comparability of the clustering results.

The overall process consists of the following key stages: (1) selection and grouping of relevant indicators; (2) data collection and pre-processing; (3) dimensionality reduction and indicators filtering; and (4) cluster analysis and validation.

2.2 Data Sources and Pre-processing

Data were sourced from TheGlobalEconomy.com, an aggregator providing access to harmonized socio-economic indicators from official databases, including the World Bank's Global Financial Development Database, the European Central Bank's Statistical Data Warehouse, and the IMF DataMapper.

The analysis covers 25 EU member states, excluding Greece and Luxembourg. Preliminary clustering of all EU27 countries revealed that Luxembourg consistently acted as a standalone outlier cluster across all three methods (K-Means, HAC, and GMM). Similarly, Greece was identified as an outlier by both the K-Means and HAC algorithms. These exclusions were methodologically justified, as Luxembourg's exceptionally large banking sector relative to its domestic economy (deposits at 413.21% of GDP) and Greece's unique post-sovereign debt crisis financial profile (high non-performing loans at 38.07% in 2019) would disproportionately skew cluster formation and distort the analysis of the remaining EU member states.

Missing values (e.g., bank assets for Cyprus, bank concentration for Germany) were imputed using regional medians. Outliers were identified via the interquartile range (IQR) method and winsorized to preserve variance.

All indicators x_{ij} were normalized to the range [0,1] using the min-max normalization:

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

where $\max(x_j)$, $\min(x_j)$ – are the minimum and maximum values of the indicators across all countries n – number of countries, m – number of indicators.

2.3 Indicator Selection

This study initially considered 15 indicators to capture the resilience, efficiency, and macroeconomic context of EU banking systems, grouped into three analytical blocks: (1) banking system resilience and stability (e.g., z-score, non-performing loans ratio (NPL)), (2) access to and efficiency of financial intermediation (e.g., bank credit to private sector), and (3) macroeconomic and institutional environment (e.g., GDP growth, government debt). These indicators were selected based on their theoretical significance, empirical use in prior financial stability studies (Alessi et al., 2022; Chodnicka-Jaworska, 2022; Colak & Öztekin, 2021; Demirgüç-Kunt et al., 2021), and data availability across EU members.

The selection of the final set of indicators was a critical step to ensure the reliability and interpretability of the cluster analysis, as excessive correlation (multicollinearity) between indicators can distort clustering results (Hastie et al., 2009). To mitigate multicollinearity and enhance clustering performance, a hybrid feature selection approach was applied:

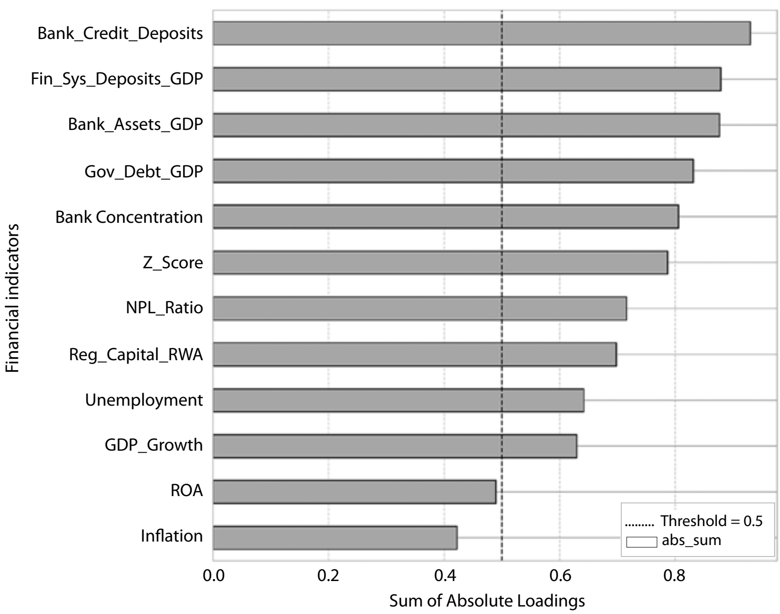
Stage 1: Correlation Analysis and Indicator Elimination. We conducted a detailed Pearson correlation analysis among all 15 indicators for both periods (2019 and 2021) to identify the redundant. Indicators with a correlation coefficient exceeding $|\rho| > 0.7$ with others were identified as candidates for exclusion. In such cases, we retained the indicator with greater theoretical relevance or a higher contribution to the variance. Through this process, three indicators were excluded:

1. Bank Credit to Private Sector (% GDP) – high correlation with Bank Assets to GDP ($\rho = 0.83$). We retained BAGDP as it more directly captures overall banking sector size relative to the economy.
2. Current Account Balance (% GDP) – high correlation with Government Debt ($\rho = -0.68$) and FDGDP ($\rho = 0.71$), reflecting fiscal-financial linkages. We retained GDEBT as it more directly captures sovereign risk affecting banking systems.
3. Foreign Direct Investment (% GDP) – moderate correlations with multiple indicators (GDPG: $\rho = 0.65$, BAGDP: $\rho = 0.58$) and less direct impact on banking system resilience compared to domestic credit dynamics.

Following correlation-based elimination, 12 indicators remained for further analysis.

Stage 2: Principal Component Analysis (PCA) Validation. Second, PCA assessed the variance explained by the indicators (Figure 1).

Figure 1 Sum of Absolute PCA Loadings for 12 Indicators



Source: Own construction

The effectiveness of PCA for dimensionality reduction in economic data has been empirically confirmed by recent comparative studies. Poznyak and Kolyada (2023) evaluated 11 dimensionality reduction methods and found PCA to demonstrate superior performance in preserving data structure while reducing complexity, particularly for macroeconomic and financial indicators. This supports our choice of PCA for feature selection validation in the present study.

PCA was applied to the remaining 12 indicators to assess the variance explained (Figure 1). The first four PCA components accounted for approximately 80% of the total variance, with key contributions from CAR, financial system deposits to GDP ratio (FDGDP), bank concentration (BCONC), and confirming the relevance of the selected indicators.

Two indicators with the lowest cumulative loadings – Return on Assets (ROA) and Inflation Rate (CPI) – were excluded. ROA's exclusion is justified as it is already incorporated as a component of the z-score, while CPI showed less direct relevance to banking system stability during the COVID-19 period characterized by supply shocks. This yielded a final set of 10 indicators (see Table A1 in Appendix).

This two-stage approach ensures that each of the final 10 indicators contributes to unique information while maintaining low multicollinearity (all pairwise $|\rho| < 0.7$), providing accurate and interpretable cluster demarcation.

2.4 Cluster Analysis and Validation of Results

The core stage of the analysis involved the application of unsupervised clustering algorithms to the normalized data (Jain, 2010). To ensure robustness of the identified country groupings, we conducted experimental validation using three clustering methods: K-Means, Hierarchical agglomerative clustering (HAC) with Ward's linkage, and Gaussian Mixture Models (GMM).

Each method embodies different assumptions, serving as validation where no ground truth exists (Hastie et al., 2009):

- *K-Means*, a centroid-based method, minimizes the sum of squared distances between data points and cluster centroids. The algorithm used the Euclidean distance and random centroid initialization.
- *HAC* constructs a dendrogram with Ward's linkage and squared Euclidean distance, minimizing within-cluster variance (Rousseeuw & Kaufman, 1990). The dendrogram was cut at $k = 4$.
- *GMM* – modelled data as a mixture of multivariate normal distributions, capturing potential elliptical clusters (McLachlan & Peel, 2000). A full covariance matrix was used, with $k = 4$ selected based on the Bayesian Information Criterion (BIC).

  Each method embodies different assumptions: the k-means method assumes spherical clusters, hierarchical agglomerative clustering assumes hierarchical structure, and Gaussian mixture models assume elliptical, probabilistic clusters.

Each country is represented as a data point – a vector of normalized indicator values. Cluster centroids represent mean indicator values for countries within each cluster, defining typical banking system profiles. Preliminary experiments demonstrated that all three methods produced largely consistent country groupings, with only minor variations at cluster boundaries. When $k = 4$ clusters were specified for both 2019 and 2021 periods, the methods achieved over 85% agreement in country assignments. This cross--method convergence confirms strong evidence that the identified clusters represent genuine structural patterns in EU banking systems rather than methodological artifacts.

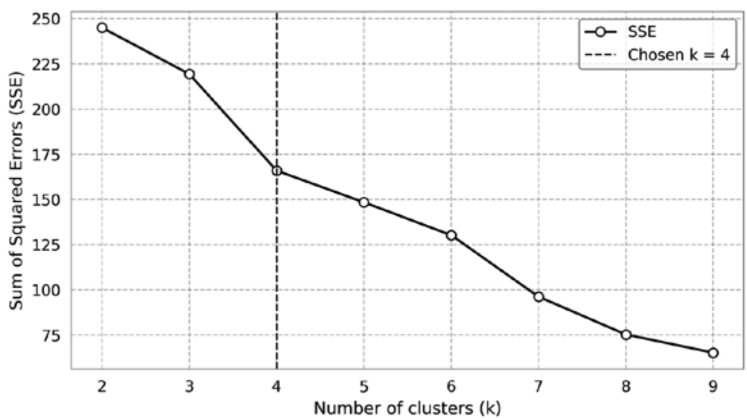
K-Means was selected as the principal method for: (1) Clear and interpretable partition through distance-based logic and defined centroids representing "typical" banking system profiles (Jain, 2010); (2) Computational efficiency and reproducibility, allowing iterative comparison across periods (Hastie et al., 2009); (3) Superior empirical performance: (Poznyak & Kolyada, 2023; Osadcha et al., 2025) found K-Means the most effective, with agglomerative clustering ranking second; (4) Euclidean distance and normalized indicators satisfied K-Means assumptions; (5) Nearly identical results to GMM and HAC, confirming robustness; (6) Established precedent in banking sector research (Mihalech & Košíková, 2021; Kaminskyi & Nehrey, 2023).; (7) The final typological classification is based on K-Means, which combines interpretability, computational simplicity, and empirical robustness.

We chose $k = 4$ as the optimal number of clusters despite the silhouette score suggesting higher values ($k = 8$) because the elbow method indicated a significant reduction in the sum of squared errors (SSE) at this point (Figure 2), marking a clear "elbow" in the curve (Rousseeuw & Kaufman, 1990).

Additionally, considering the specific context of clustering EU countries, a smaller number of clusters offers a more interpretable and actionable segmentation, balancing model simplicity with meaningful differentiation among groups. This decision aligns with the practical goal of deriving clear insights for policy or economic analysis, where overly granular clusters might obscure broader patterns.

Clustering results were validated by: (1) Cluster profiles: average indicator values characterized financial profiles (Table 1, Table 2); (2) Temporal mapping: country transitions from 2019 to 2021 assessed stability and reconfiguration during the pandemic period (Figure 5, detailed country migrations in Table A4); (3) method consistency: high agreement (>85%) among K-Means, HAC, and GMM were confirmed robust typologies (Jain, 2010).

Figure 2 Results of the elbow method for determining the optimal number of clusters based on the k-means method



Source: Own construction

3 EMPIRICAL RESULTS

This section presents the clustering analysis results for EU banking systems in 2019 (pre-pandemic baseline) and 2021 (post-pandemic), followed by a comprehensive temporal analysis of structural transformations induced by the COVID-19 crisis.

Data processing and analysis were performed in Google Colab using Python 3.10 with libraries: *pandas* and *numpy* for data manipulation; *scikit-learn* for normalization, PCA, K-Means, HAC, GMM, and validation metrics; *matplotlib* and *seaborn* for visualization; *statsmodels* and *scipy* for correlation and statistical validation.

3.1 Baseline Clusters Analysis (2019)

The application of K-Means to the 2019 data consistently reveals four distinct banking system typologies within the EU. The results of K-Means are shown in Figure 3 (projection on the two principal components), and Table 1 presents the cluster composition and key characteristic profiles for each identified banking system typology.

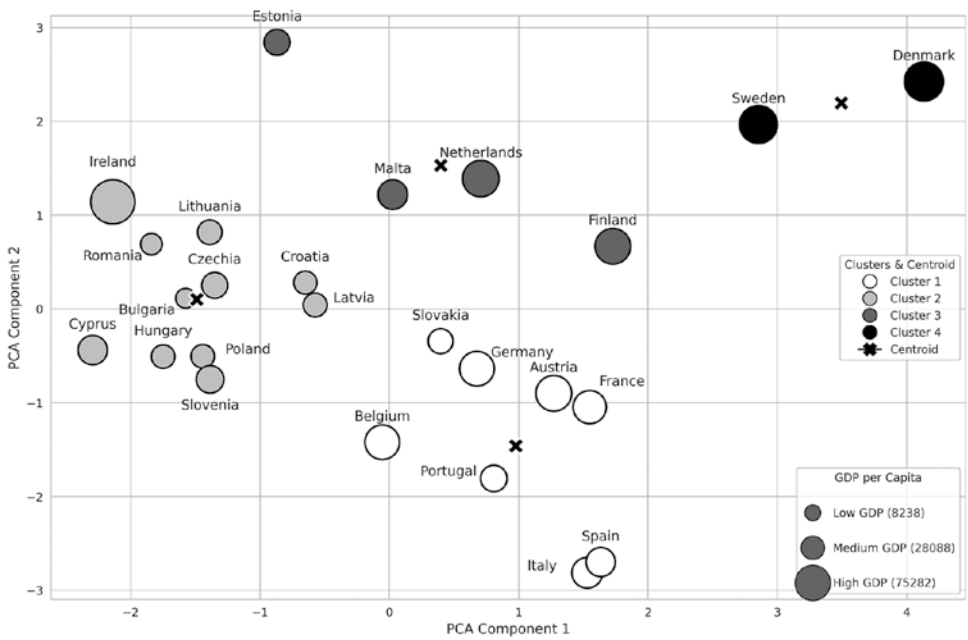
The Core Stability System (CSS) cluster (Cluster 1), comprising major Eurozone economies (Germany, France, Italy), represents the stable backbone of the EU financial system. This cluster demonstrates exceptional macroeconomic stability with the lowest unemployment (4.28%) and moderate government debt (50.44%). Their banking systems exhibit healthy asset quality (NPL: 2.96%), robust capital adequacy (CAR: 20.38%), and solid stability buffers (z-score: 14.44), establishing them as the benchmark for mature, well-regulated financial intermediation.

Table 1 EU Banking System Cluster Composition and Profiles (2019)

Name/Abbreviation		Countries	Profile
Cluster 1	Core Stability Systems (CSS)	Austria, Belgium, Germany, Spain, France, Italy, Portugal, Slovakia	Mature, stable: NPL 2.96%, CAR 20.38%, z-score 14.44, GDEBT 50.44%
Cluster 2	Dynamic Growth Systems (DGS)	Bulgaria, Cyprus, Czechia, Croatia, Hungary, Ireland, Lithuania, Latvia, Poland, Romania, Slovenia	Growth-oriented: GDPG 3.53%, NPL 5.44%, CAR 19.47%, z-score 13.56
Cluster 3	High Intermediation Systems (HIS)	Estonia, Finland, Malta, Netherlands	Deep markets: BAGDP 100.45%, LADR 114.00%, NPL 2.42%, z-score 14.62
Cluster 4	Resilience Systems (RS)	Denmark, Sweden	Exceptional: z-score 31.54, NPL 0.98%, GDEBT 34.65%

Source: Own construction. Note: All indicators represent cluster means. For detailed values of all 10 indicators, see Table A2 in ANNEX,

Figure 3 K-Means Cluster Visualization on PCA Components (Excluding Luxembourg and Greece as outliers, Point Size by GDP per Capita, 2019)



Source: Own construction

The Dynamic Growth System (DGS, Cluster 2) reflects economies balancing growth dynamism with elevated credit risk (Cluster 2). This group, including most Central and Eastern European countries plus

Ireland, demonstrates the highest GDP growth (3.53%) but also the highest NPL ratio (5.44%), slightly exceeding optimal ranges. Despite strong capital adequacy (CAR: 19.47%), their lower z-score (13.56) indicates reduced stability buffers. Ireland's consistent inclusion highlights its unique financial system characteristics that align with developmental banking patterns despite high GDP per capita.

The High Intermediation Systems (HIS, Cluster 3) features highly sophisticated financial sectors with deep market penetration. Comprising Estonia, Finland, Malta, and Netherlands, this cluster is characterized by significantly elevated banking sector depth relative to GDP. These economies show significantly higher Bank Assets to GDP (100.45%) and elevated loan-to-deposit ratios (114.00%), indicating extensive financial intermediation, sophisticated non-deposit funding sources, and significant cross-border banking operations. Despite sector complexity, they maintain excellent asset quality (NPL: 2.42%) and strong capital positions (CAR: 20.40%), resulting in healthy z-scores (14.62). The combination of deep financial markets, low credit risk, and robust capitalization distinguishes this cluster from other banking system typologies.

The Resilience Systems (NRS, Cluster 4), consisting solely of Denmark and Sweden, represent the pinnacle of financial resilience. This archetype exhibits exceptional characteristics: the lowest government debt (34.65%), minimal NPLs (0.98%), and an outstanding z-score (31.54) indicating virtually no insolvency risk. Their massive banking sectors (BAGDP: 154.93%, LADR: 227.56%) reflect sophisticated international operations managed under superior prudential frameworks, establishing them as the gold standard for banking system stability.

3.2 Post-Pandemic Cluster Evolution (2021)

To ensure robustness and comparability, the clustering analysis for 2021 was conducted using the same set of indicators and methodology (K-Means) as the 2019 baseline. Furthermore, Greece and Luxembourg were again excluded from the primary analysis due to their persistent status as outliers, often forming singleton clusters or significantly skewing the results.

It is important to emphasize that cluster numbering and naming remain consistent across both periods (2019 and 2021), as presented in Tables 1 and 2. This approach allows us to trace structural transformations in the EU banking systems by observing changes in country composition within fixed cluster typologies. The movement of countries between clusters reveals the pandemic's role as a structural catalyst, reshaping banking system alignments rather than merely causing temporary disruptions.

By comparing the clustering solutions for 2019 and 2021, we can assess the stability of the identified clusters and pinpoint which country groupings remained resilient and which were reshuffled due to the crisis. This analysis provides key insights into the differential effects of the pandemic on various banking systems.

The results of K-Means for 2021 are presented in Figure 4, and the full results of the cluster assignments by method are presented in ANNEX (Table A3). For the main analysis, however, we focus on the consolidated profiles and their temporal evolution, with detailed cluster composition and profiles in Table 2.

The 2021 analysis, employing identical methodology, reveals significant structural reconfiguration following the COVID-19 shock. While maintaining four clusters, country compositions and characteristic profiles transformed substantially, reflecting differentiated pandemic impacts and recovery trajectories.

The Core Stability Systems (CSS, Cluster 1) expanded to include, demonstrating the core's attractive stability during crisis recovery. This group achieved a strong economic rebound (GDPG: 5.96%) while maintaining sound banking fundamentals. Despite increased fiscal burden (GDEBT rose from 50.44% to 66.61% due to pandemic support measures), they improved asset quality (NPL decreased to 2.14%) and strengthened capital positions (CAR: 21.94%). The modest z-score decline (from 14.44 to 13.68) reflects controlled absorption of pandemic shock.

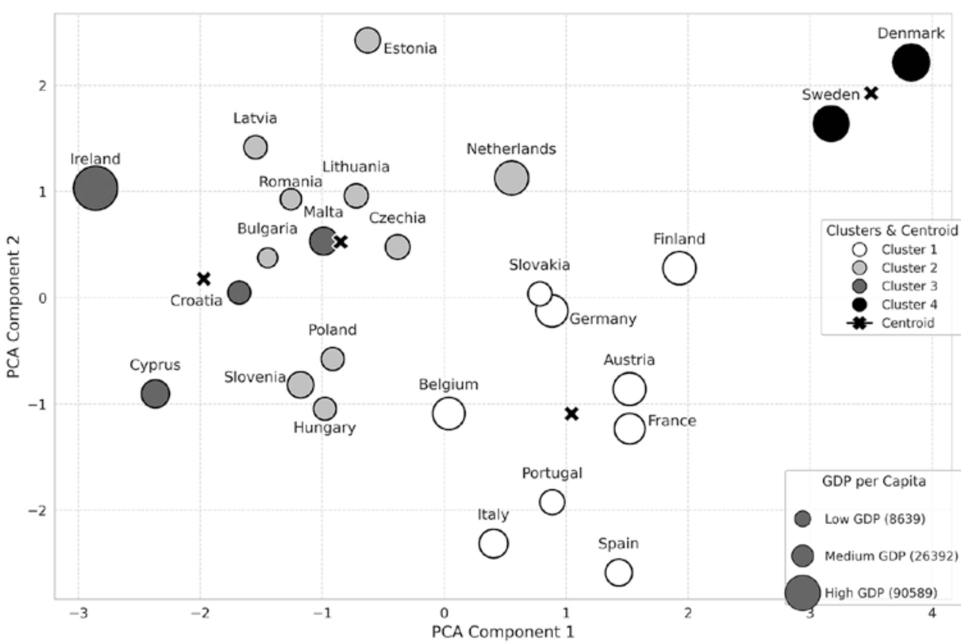
Table 2 EU Banking System Cluster Composition and Profiles (2021)

Name/Abbreviation		Countries	Profile
Cluster 1	Core Stability Systems (CSS)	Austria, Belgium, Germany, Spain, Finland, France, Italy, Portugal, Slovakia	Recovery: GDPG 5.96%, NPL 2.14%, CAR 21.94%, z-score 13.68, GDEBT 66.61%
Cluster 2	Dynamic Growth Systems (DGS)	Bulgaria, Czechia, Estonia, Hungary, Lithuania, Latvia, Netherlands, Poland, Romania, Slovenia	High growth: GDPG 8.27%, NPL 3.71%, CAR 20.68%, z-score 12.46
Cluster 3	High Intermediation Systems (HIS)	Cyprus, Croatia, Ireland, Malta	Vulnerable: GDPG 8.76%, GDEBT 88.25%, z-score 10.17, NPL 3.59%
Cluster 4	Nordic Resilience Systems (NRS)	Denmark, Sweden	Strengthened: z-score 32.89, NPL 0.83%, GDEBT 36.35%

Source: Own construction. Note: Italic indicates new cluster members in 2021 compared to 2019. All indicators represent cluster means. For detailed values of all 10 indicators, see Table A3 in ANNEX.

A critical reconfiguration occurred within the Dynamic Growth Systems (DGS, Cluster 2) – it began to contain countries from the former 2019 DGS cluster alongside Estonia and Netherlands, which migrated from the 2019 HIS cluster. This convergence was driven by shared rapid recovery dynamics (GDPG: 8.27%), and represents a fundamental typological shift where advanced economies aligned with high-growth markets. However, this dynamism carries costs: elevated NPLs (3.71%) and the lowest z-score among stable clusters (12.46%) indicate reduced stability buffers despite strong growth momentum.

Figure 4 K-Means Cluster Visualization on PCA Components (Excluding Luxembourg and Greece as outliers, Point Size by GDP per Capita, 2021)



Source: Own construction

The High Intermediation Systems (HIS, Cluster 3) experienced the most substantial reconfiguration. In 2019, it consisted of highly integrated and low-risk banking systems (Estonia, Finland, Malta, Netherlands). By 2021, its composition shifted almost entirely toward peripheral and island economies (Cyprus, Croatia, Ireland, Malta), with only Malta remaining from the original group. The 2021 HIS cluster exhibits high government debt levels (88.25%) and markedly weaker stability buffers (z-score: 10.17). This shift reflects not a change in the cluster's structural definition – high financial intermediation relative to economic size – but rather uneven pandemic-era impact on countries with greater exposure to financial-sector fragilities.

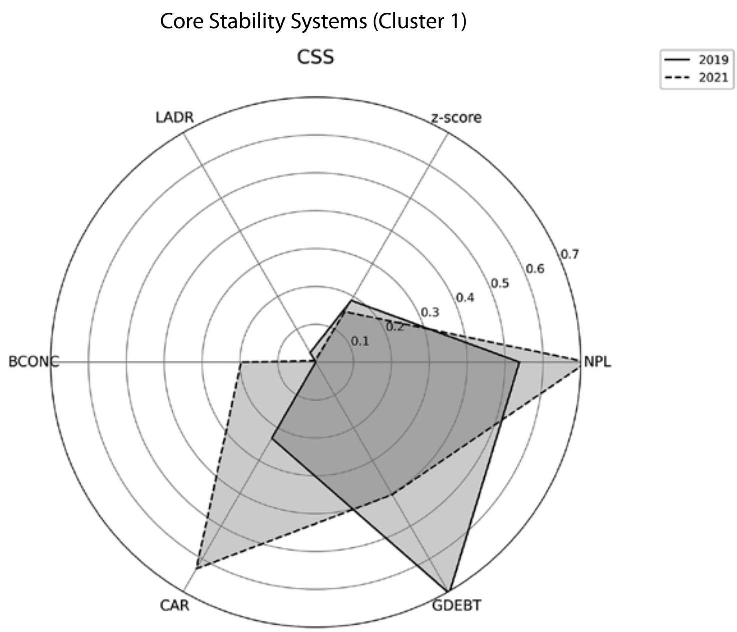
The Nordic Resilience Systems (NRS, Cluster 4) remained unchanged and strengthened its exceptional position. Post-pandemic metrics show further improvement: z-score increased to 32.89, NPLs decreased to 0.83%, while maintaining minimal fiscal burden (GDEBT: 36.35%). This paradoxical strengthening during global crisis confirms their unique resilience and superior risk management capabilities.

3.3 Temporal Analysis of Cluster Profile Changes (2019 vs. 2021)

The comparative analysis of 2019–2021 cluster profiles reveals the COVID-19 pandemic's role as a structural catalyst, creating both reinforcement of existing strengths and emergence of new vulnerabilities. Figure 5 visualizes these transformations across six key indicators, demonstrating how different banking system archetypes absorbed and adapted to the exogenous shock.

As established in Section 3.2, cluster numbering and naming remain fixed across both periods, enabling direct comparison of how identical cluster typologies evolved in their characteristic profiles and country composition.

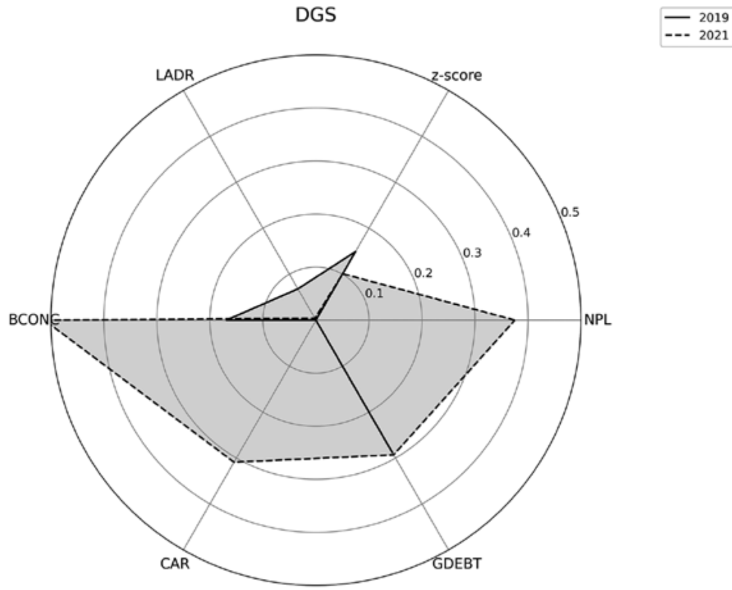
Figure 5 Temporal profile changes of EU Banking Clusters (2019 vs. 2021, Normalized values with inverted NPL and GDEBT ratios)



Source: Own construction

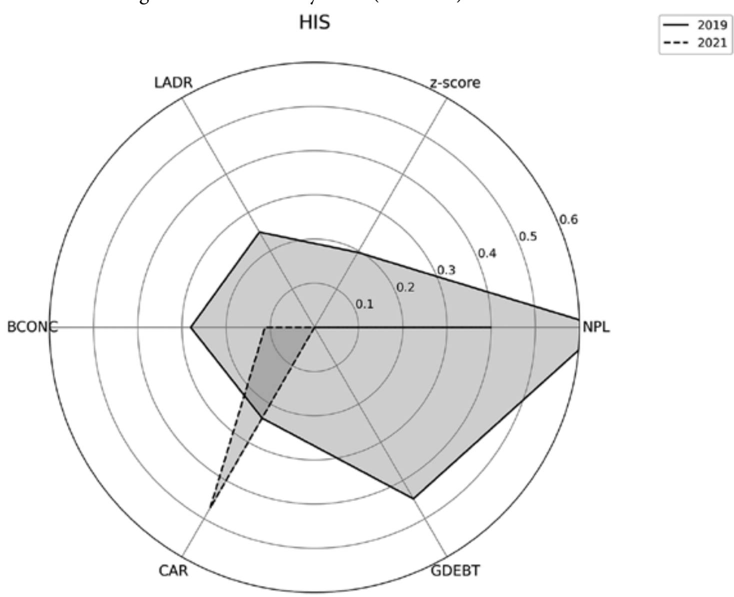
Figure 5 (continuation)

Dynamic Growth Systems (Cluster 2)



Source: Own construction

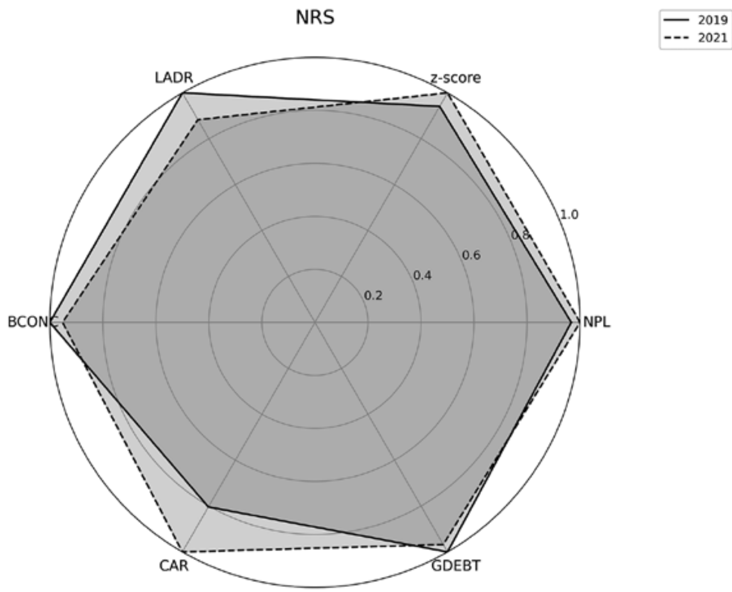
High Intermediation Systems (Cluster 3)



Source: Own construction

Figure 5 (continuation)

Nordic Resilience Systems (Cluster 4)



Source: Own construction

The radar chart presents normalized profiles for each of the four clusters, allowing direct visual comparison of structural changes. The most striking transformations occurred in Clusters 3 (HIS) and 2 (DGS), while Clusters 1 (CSS) and 4 (NRS) demonstrated remarkable stability.

The Core Stability Systems evolution (Cluster 1, top-left chart) demonstrates controlled shock absorption by mature economies. While z-score declined moderately by 5.3% (14.44→13.68), reflecting reduced stability buffers, these systems successfully managed credit risk with NPL improvements of 27.7% (2.96%→2.14%). The substantial GDEBT increase of 32.1% (50.44%→66.61%) reflects massive fiscal intervention, highlighting the trade-off between immediate economic support and long-term fiscal health. Decreased concentration and deleveraging (LADR decline) suggest strategic market adjustments.

The Dynamic Growth Systems (Cluster 2) retained its core CEE composition while absorbing Estonia and the Netherlands from the former HIS group. This configuration shows the strongest growth (GDPG: 8.27%) but with reduced stability buffers (z-score: 13.56→12.46), indicating that rapid recovery was accompanied by increased systemic sensitivity. Asset quality improved (NPLs: 5.44%→3.71%), suggesting genuine economic rebound rather than excessive credit expansion. The convergence of advanced financial systems (Estonia, Netherlands) with CEE economies represents a structural realignment driven by post-crisis growth dynamics.

The High Intermediation Systems (Cluster 3) experienced the most substantial transformation. All original 2019 members except Malta migrated to other clusters, while Cyprus, Croatia, and Ireland joined. In 2021, government debt rose to 88.25%, stability buffers weakened sharply (z-score: 10.17), and NPLs increased (2.42%→3.59%). The reduction in loan-to-deposit ratios (114.00%→76.44%) signals forced deleveraging and diminished intermediation capacity. The cluster shifted from a low-risk profile to a high-risk, fiscally constrained configuration requiring closer supervisory attention, while maintaining high financial intermediation (BAGDP: 88.88%) (see Table A4).

The Nordic Resilience Systems (Cluster 4, top-left chart) defies conventional crisis expectations. Denmark and Sweden not only maintained their exceptional position but paradoxically strengthened it: z-score increased by 4.2% (31.54→32.89), NPLs decreased by 15.3% (0.98%→0.83%), while fiscal position remained virtually unchanged (+4.9% GDEBT increase). This counter-cyclical performance establishes Nordic systems as uniquely resilient, capable of strengthening during the global financial stress through superior institutional frameworks and risk management.

These temporal dynamics confirm all three research hypotheses: Nordic and Core Stability clusters demonstrated resilience (H1), new vulnerabilities emerged in reconfigured groupings (H2), and specialized systems underwent dynamic adaptation through cluster realignment (H3). The pandemic acted not merely as a temporary shock but as a structural transformation catalyst, creating new banking system typologies that reflect post-crisis realities of differentiated resilience, fiscal burden, and growth-stability trade-offs across the European Union.

4 DISCUSSION

This study identifies clear structural shifts in the EU banking system typologies between 2019 and 2021. These changes must be interpreted considering the unprecedented fiscal and monetary interventions implemented during the COVID-19 period, including the ECB's PEPP programme, loan moratoria, and extensive guarantee schemes. As these measures mitigated direct balance-sheet deterioration, the transformations captured in our analysis represent the *combined* effect of pandemic pressures and policy support rather than the pandemic in isolation.

Country-level aggregation allows us to detect systemic patterns and cross-national typologies, though it necessarily masks the within-country heterogeneity. For our purpose, identifying structural banking system configurations, this macro-level approach is appropriate, but bank-level analyses would complement our findings.

The results provide strong support for *Hypothesis 1:* the Nordic Resilience Systems (NRS) and Core Stability Systems (CSS) clusters maintained robust stability buffers, with improved asset quality and limited deterioration in intermediation metrics. These systems demonstrate the capacity of mature institutional environments to absorb large shocks when combined with substantial policy support.

Hypothesis 2 is confirmed by the transformation of the High Intermediation Systems (HIS). In 2021, this cluster (Cyprus, Croatia, Ireland, Malta) exhibited the highest sovereign debt levels and weakest stability indicators. Countries that migrated into HIS were characterized by a strong sovereign–bank nexus and elevated exposure to pandemic-related fiscal expansion, making the cluster the most risk-sensitive typology.

Hypothesis 3 is supported by the adaptive evolution of the Dynamic Growth Systems (DGS), which retained its CEE core while integrating Estonia and the Netherlands. This convergence produced a fast-growing but moderately fragile typology, combining high post-pandemic expansion with slightly reduced stability buffers.

Limitations. Several limitations must be acknowledged: (1) Aggregate country data obscure within-country banking heterogeneity; (2) a two-period observational design prevents causal attribution, as structural changes may also reflect pre-existing trends or heterogeneous policy capacities; (3) Greece and Luxembourg were excluded due to outliers' extreme characteristics, limiting generalizability; (4) the period 2019–2021 captures only immediate adjustments, with longer-term effects uncertain; and (5) qualitative factors (supervisory intensity, regulatory quality) are not incorporated, though they likely influence resilience.

CONCLUSION

This study provides an empirical assessment of structural shifts in the EU banking system typologies during 2019–2021. Using K-Means clustering, we identify four stable archetypes whose characteristics persist, though country composition changed during the pandemic period.

The findings confirm that external shocks, can act as structural catalysts reshaping banking system configurations. The pandemic reinforced resilient systems, exposed latent vulnerabilities in high--intermediation structures, and facilitated convergence patterns transcending traditional geographic boundaries.

These results have implications for macroprudential policy. They support a cluster-based regulatory approach tailoring supervisory intensity to risk characteristics of each typology rather than applying uniform requirements across the EU. Specifically, the vulnerable HIS cluster requires stricter capital buffers and enhanced sovereign-bank risk monitoring, while the dynamic DGS cluster warrants close credit expansion oversight to prevent asset quality deterioration.

ACKNOWLEDGMENT

We would like to thank TheGlobalEconomy.com, and, in particular, the team leader Neven Valev, for providing easy access to harmonized socioeconomic data, which facilitated the collection and analysis of indicators for this study. We also acknowledge compliance with the publication ethics rules of the journal Statistika: Statistics and Economy Journal (available at: https://www.csu.gov.cz/statistika_journal).

References

ALESSI, L., BRUNO, B., CARLETTI, E., NEUGEBAUER, K., & WOLFSKEIL, I. (2021). Cover your assets: Non-performing loans and coverage ratios in Europe. *Economic Policy*, 36(108): 685–733. <https://doi.org/10.1093/epolic/eiab013>.

ARI, A., CHEN, S., & RATNOVSKI, L. (2021). The dynamics of non-performing loans during banking crises: A new database with post-COVID-19 implications. *Journal of Banking and Finance*, 133: 106–140. <https://doi.org/10.1016/j.jbankfin.2021.106140>.

CAO, YIFEI & CHOU, JEN-YU. (2022). Bank resilience over the COVID-19 crisis: The role of regulatory capital. *Finance Research Letters*, 48(C). <https://doi.org/1 10.1016/j.frl.2022.102891>.

CHODNICKA-JAWORSKA, P. (2022). Impact of COVID-19 on European banks' credit ratings. *Economic Research-Ekonomska Istraživanja*, 36(3): 1–17. <https://doi.org/10.1080/1331677X.2022.2153717>.

CLICHICI, D., & ZELDEA, C.-G. (2021). The resilience of central and eastern european banking systems during the covid-19 crisis. *Economy and Sociology*, 2: 20–29. <https://doi.org/10.36004/nier.es.2021.2–02>.

COLAK, G., & ÖZTEKIN, O. (2021). The impact of COVID-19 pandemic on bank lending around the world. *Journal of Banking and Finance*, 133: 106207. <https://doi.org/10.1016/j.jbankfin.2021.106207>.

DANKO, J., & SUCHÝ, E. (2021). The Financial Integration in the European Capital Market Using a Clustering Approach on Financial Data. *Economies*, *9*(2), 89. <https://doi.org/10.3390/economies9020089>.

DEMIR, E., & DANISMAN, G. O. (2021). Banking sector reactions to COVID-19: The role of bank-specific factors and government policy responses. *Research in International Business and Finance*, 58: 101508. <https://doi.org/10.1016/j.ribaf.2021.101508>.

DEMIRGÜÇ-KUNT, A., PEDRAZA, A., & RUIZ-ORTEGA, C. (2021). Banking sector performance during the COVID-19 crisis. *Journal of Banking and Finance*, 133: 106305. <https://doi.org/10.1016/j.jbankfin.2021.106305>.

DURSUN-DE NEEF, H. Ö., & SCHANDLBAUER, A. (2021). COVID-19 and lending responses of European banks. *Journal of Banking and Finance*, 133: 106236. <https://doi.org/10.1016/j.jbankfin.2021.106236>.

EUROPEAN BANKING AUTHORITY (EBA). (2020). *The EU banking sector: First insights into the COVID-19 impacts* (EBA/REP/2020/17) [online]. <https://service.betterregulation.com/document/442177>.

EUROPEAN BANKING AUTHORITY (EBA). (2021). *Risk dashboard: Q4 2020* [online]. <https://www.eba.europa.eu/risk-analysis-and-data/risk-dashboard>.

EUROPEAN CENTRAL BANK (ECB). (2021). *Financial stability review: May 2021* [online]. <https://www.ecb.europa.eu/pub/financial-stability/fsr/html/ecb.fsr202105~c9f6a47f42.en.html>.

FEYEN, E., ALONSO GISPERT, T., KLIATSKOVA, T., & MARE, D. S. (2021). Financial sector policy response to COVID-19 in emerging markets and developing economies. *Journal of Banking and Finance*, 133: 106184. https://doi.org/10.1016/j.jbankfin.2021.106184.

FOGLIA, M., ADDI, A., & ANGELINI, E. (2022). The Eurozone banking sector in the time of COVID-19: Measuring volatility connectedness. *Global Finance Journal*, 51(C). <https://doi.org/ 10.1016/j.gfj.2021.100677>.

HASTIE, T., TIBSHIRANI, R., & FRIEDMAN, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer. <https://doi.org/10.1007/978-0-387-84858-7>.

INTERNATIONAL MONETARY FUND (IMF). (2021). Departmental Papers 2021, 008: COVID-19: How will European banks fare? [online]. <https://doi.org/10.5089/9781513572772.087>.

JAIN, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8): 651–666. <https://doi.org/10.1016/j.patrec.2009.09.011>.

KAMINSKYI, A. & NEHREY, M. (2023). Cluster Method Applying to Covid-19 Event Study for the Largest USA Banks. CEUR, 3641. <https://ceur-ws.org/Vol-3641/paper4.pdf>

KOZAK, S. (2021). The impact of COVID-19 on bank equity and performance: The case of Central Eastern South European countries. *Sustainability*, 13(19): 11036. <https://doi.org/10.3390/su131911036>.

MCLACHLAN, G. J., & PEEL, D. (2000). *Finite Mixture Models*. Wiley. <https://doi.org/10.1002/0471721182>.

MIHALECH, P., & KOŠÍKOVÁ, M. (2021). Cluster analysis of the EU banking sector based on EBA risk indicators. In: *EDAMBA 2021: 24th International Scientific Conference for Doctoral Students and Young Researchers*. pp. 306–316. <https://doi.org/10.53465/EDAMBA.2021.9788022549301.306–316>.

OSADCHA, N., ZHYTKEVYCH, O., & MATVIYCHUK, A. (2025). Comprehensive comparative analysis of clustering methods in the context of modeling nations' decarbonization potentials. SSRN. http://dx.doi.org/10.2139/ssrn.5598455.

PLIKAS, J. H., KENOURGIOS, D., & SAVVAKIS, G. A. (2024). COVID-19 and non-performing loans in Europe. *Journal of Risk and Financial Management*, 17(7): 271. <https://doi.org/10.3390/jrfm17070271>.

POZNYAK, S., & KOLYADA, Y. (2023). Comparative analysis of the effectiveness of dimensionality reduction algorithms and clustering methods on the problem of modelling economic growth. Neuro-Fuzzy Modelling Techniques in Economics, 12, 67–110. http://doi.org/10.33111/nfmte.2023.067

ROUSSEEUW, P. J., & KAUFMAN, L. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley. <https://doi.org/10.1002/9780470316801>.

SALAZAR, Y., MERELLO, P., & ZORIO-GRIMA, A. (2023). IFRS 9, banking risk and COVID-19: Evidence from Europe. *Finance Research Letters*, 56: 104130. <https://doi.org/10.1016/j.frl.2023.104130>.

THEGLOBALECONOMY.COM. (2025). *Economic and financial indicators database* [online]. https://www.theglobaleconomy.com.

WORLD BANK. (2022). *World Development Report 2022: Finance for an equitable recovery* [online]. <https://doi.org/10.1596/978-1-4648-1730-4>.

WORLD BANK'S GLOBAL FINANCIAL DEVELOPMENT DATABASE. (2025). [online]. <https://www.worldbank.org/en/publication/gfdr/data/global-financial-development-database>

ANNEX/ APPENDIX

Table A1 The complete list financial and macro-financial indicators used in the study

	Abbr.	Full Indicator Name	Brief Description	Motivation for Inclusion	Preferred Range / Interpretation
1	GDPG	GDP Growth Rate	Annual percentage change of Gross Domestic Product	It reflects overall economic health and potential for loan growth/repayment	Higher positive values (e.g., >2–3%) indicate robust economic activity
2	UNEMP	Unemployment Rate	Percentage of the total labor force that is unemployed	It directly affects household income and ability to service debt, impacting NPLs	Lower values (<6–7%) indicate a healthy labor market
3	GDEBT	Government Debt to GDP	General government gross debt as a percentage of GDP	It indicates sovereign risk and potential fiscal constraints affecting banking sector support	Lower to moderate (<60–90%) suggests fiscal prudence and capacity

Table A1 (continuation)

	Abbr.	Full Indicator Name	Brief Description	Motivation for Inclusion	Preferred Range / Interpretation
4	BAGDP	Bank Assets to GDP	Total banking sector assets as a percentage of GDP	It reflects the size of the banking sector relative to the economy, indicating systemic importance.	Moderate to high (e.g., 70–200%). Extremely high values may signal systemic risk.
5	FDGDP	Financial System Deposits to GDP	Total deposits in the financial system as a percentage of GDP	It indicates the stability and size of the funding base for the financial sector	Moderate to high (e.g., 60–100%) suggests a stable funding source
6	BCONC	Bank Concentration	Assets of the three largest banks as a share of total banking sector assets.	It measures market power and potential "too-big-to-fail" implications; can influence competition	Moderate (e.g., 50–80%). Very high values (>90%) may signal systemic risk or lack of competition
7	NPL	Non-Performing Loans Ratio	Non-performing loans to total gross loans	It is a key indicator of asset quality and credit risk for banks	Lower is better (<3–5%), reflecting healthy loan portfolios
8	LADR	Bank Credit to Deposits Ratio	Total bank loans to total bank deposits	It measures liquidity risk; indicates reliance on non-deposit funding	Around 70–100% indicates efficient deposit utilization; >100% signals higher reliance on other funding
9	CAR	Regulatory Capital to RWA	Regulatory Tier 1 Capital to Risk-Weighted Assets.	It is primary measure of a bank's capital adequacy and ability to absorb losses	Higher is better (>15–20%) for strong capital buffers (Basel III minimum is 8%)
10	z-score	Banking System Z-score	Composite index: (ROA + capital/asset ratio) / std. dev of ROA. Indicates banking system stability	It indicates distance to insolvency (number of standard deviations); higher values imply greater stability and lower insolvency risk	Higher is better (>10–15) for lower insolvency risk

Source: Own construction. All data sourced from TheGlobalEconomy.com (2025), aggregating harmonized indicators from Eurostat, World Bank, World Bank Global Financial Development Database (GFDD), European Central Bank (ECB), and International Monetary Fund (IMF)
Note: Preferred ranges for indicators in Table A1 are defined based on regulatory standards (e.g., Basel III, Maastricht criteria) and empirical studies [Alessi et al., 2022; Cao & Chou, 2022; Chodnicka-Jaworska, 2022; Colak & Öztekin, 2021; Demirgüç-Kunt et al., 2021; Foglia et al., 2022; IMF, 2021; Salazar et al., 2023; World Bank, 2022].

Table A2 Consolidated Average Factor Values for EU Banking System Clusters (2019 Baseline)

Cluster Name (ID)	GDPG	UNEMP	GDEBT	BAGDP	FDGDP	BCONC	NPL	LADR	CAR	z-score
CSS (Cluster 1)	2.57	4.28	50.44	64.51	69.17	67.52	2.96	80.81	20.38	14.44
DGS (Cluster 2)	3.53	6.59	72.82	70.43	69.25	70.75	5.44	86.76	19.47	13.56
HIS (Cluster 3)	2.56	5.44	64.25	100.45	76.35	72.86	2.42	114.00	20.40	14.62
NRS (Cluster 4)	2.13	5.92	34.65	154.93	65.64	86.62	0.98	227.56	22.62	31.54
Desired/Optimal Range	>2–3%	<6–7%	<60–90%	70–200%	60–100%	50–80%	<3–5%	70–100%	>15–20%	>10–15

Source: Own construction.
Note: Cluster abbreviations: Core Stability Systems (CSS), Dynamic Growth Systems (DGS), High Intermediation Systems (HIS), Nordic Resilience Systems NRS.

Table A3 Consolidated Average Factor Values for EU Banking System Clusters (2021 Post-Pandemic)

Cluster Name (ID)	GDPG	UNEMP	GDEBT	BAGDP	FDGDP	BCONC	NPL	LADR	CAR	z-score
CSS (Cluster 1)	5.96	5.67	66.61	74.90	81.22	71.29	2.14	77.00	21.94	13.68
DGS (Cluster 2)	8.27	6.88	72.52	73.15	86.52	77.29	3.71	77.04	20.68	12.46
HIS (Cluster 3)	8.76	6.34	88.25	88.88	85.35	69.66	3.59	76.44	21.32	10.17
NRS (Cluster 4)	6.66	6.88	36.35	157.86	73.72	85.70	0.83	209.86	23.39	32.89
Desired/Optimal Range	>2–3%	<6–7%	<60–90%	70–200%	60–100%	50–80%	<3–5%	70–100%	>15–20%	>10–15

Source: Own construction

Nowcasting Cost of Living Using Google Trends: The Case of Selected Asian Countries

Yee Sheng Mak[1] | *School of Mathematical Sciences, Universiti Sains Malaysia, Minden, Penang, Malaysia*
Siok Kun Sek[2] | *School of Mathematical Sciences, Universiti Sains Malaysia, Minden, Penang, Malaysia*
Mohsen Ayyash[3] | *School of Mathematical Sciences, Universiti Sains Malaysia, Minden, Penang, Malaysia*

Received 14.10.2025 (revisions received 13.3.2026, 30.3.2026), **Accepted** (reviewed) 5.4.2026, **Published** 11.9.2026

Abstract

This paper examines the cost of living (COL) in six selected Asian countries: Indonesia, Japan, Korea, Malaysia, Singapore, and Thailand. Motivated by rising inflationary pressures and limitations of traditional economic indicators – such as reporting delays and limited granularity – this study explores whether real-time data from Google Trends searches can improve the tracking and prediction of COL patterns in several country-specific settings. The analysis covers the period from 2010 to 2024 and employs Mixed Data Sampling (MIDAS) regression. The results demonstrate that MIDAS models provide superior flexibility and responsiveness by accommodating high-frequency Google Trends data alongside lower-frequency macroeconomic variables. Notably, MIDAS models yield improved predictive accuracy and timeliness in capturing COL dynamics. The findings also highlight the significance of search terms like debt, inflation, and unemployment as behavioural proxies for public economic concerns. This study underscores the potential of integrating search engine data into COL analysis to improve the timeliness and sensitivity of economic monitoring.

Keywords	***DOI***	***JEL code***
Cost of living, Google Trends, macroeconomic indicators, Mixed Data Sampling (MIDAS) regression, real time data	*https://doi.org/10.54694/stat.2025.56*	*C82, C22, R20, C53, E31, D12, D31*

INTRODUCTION

The cost of living (COL) denotes the resources required to sustain a standard of living, based on essential expenditures such as housing, food, transport, and healthcare. It is a key measure of household welfare

[1] School of Mathematical Sciences, Universiti Sains Malaysia, 118 00 USM Penang, Malaysia.
[2] School of Mathematical Sciences, Universiti Sains Malaysia, 118 00 USM Penang, Malaysia. Corresponding author: e-mail: sksek@usm.my. <ORCID ID: https://orcid.org/0000-0002-3616-9815>.
[3] School of Mathematical Sciences, Universiti Sains Malaysia, 118 00 USM Penang, Malaysia. <ORCID: https://orcid.org/0000-0002-5618-6313>.

and an important input into wage and policy decisions (ILO, 2022; OECD, 2011). In Asia, COL has risen sharply in recent decades, shaped by rapid urbanisation, structural transformation, and persistent inflation. These pressures have been exacerbated by external shocks: the 2008 Global Financial Crisis triggered food and fuel price surges that hit low-income households hardest, while the COVID-19 pandemic caused severe income losses and drove an estimated 90 million people into extreme poverty worldwide, many in Asia (FAO, 2021).

Asia's diversity makes it a compelling setting for COL analysis. High-income economies such as Singapore, Japan, and South Korea consistently rank among the world's most expensive countries, driven by housing, urban costs, and currency strength. By contrast, South and Central Asian nations such as Pakistan, Bangladesh, and India remain among the least expensive globally, reflecting lower wages, rural consumption patterns, and modest price levels (Clissitt, 2023). Southeast Asia lies between: Malaysia and Thailand record mid-range living costs but show steeper upward trends compared to advanced peers. This heterogeneity underscores how development stages and economic structures shape COL dynamics.

Traditional COL measures, including the consumer price index (CPI) and household surveys, remain central but face well-known shortcomings: delayed publication, lack of regional granularity, and reliance on static consumption baskets that fail to capture behavioural shifts during shocks (Cavallo, 2013; ILO, 2022; OECD, 2011). These limitations highlight the value of complementary, higher-frequency indicators. Digital data offers such potential. Real-time information sources enhance responsiveness and allow for timelier detection of economic changes (Polash, 2026). Google Trends, in particular, provides a cost-free measure of public search activity and has been widely used for nowcasting unemployment, inflation, and consumer sentiment (Choi and Varian, 2012; D'amuri and Marcucci, 2017; Vosen and Schmidt, 2011). Its applications have broadened significantly across disciplines (Jun et al., 2018), though concerns remain about reliability for low-frequency terms (Gummer and Oehrlein, 2024). Despite this, its potential for monitoring COL specifically remains underexplored, especially in emerging Asian contexts. To date, no study has applied Google Trends search data in analyzing the cost of living. Hence, utilizing the real-time data using Google Trends search in studying COL could contribute to new insights in understanding the behaviour of households, forecasting analysis, and policy planning or decision-making.

This study investigates COL dynamics in six Asian economies – Indonesia, Japan, South Korea, Malaysia, Singapore, and Thailand – between 2010 and 2024. By comparing Malaysia with both regional peers and advanced economies, the analysis highlights differences in COL trajectories across development levels. The study pursues three objectives: (i) to evaluate the relationship between macroeconomic indicators and COL, (ii) to assess the explanatory power of Google Trends data, individually and as a composite index, and (iii) to test whether hybrid models integrating digital and traditional data, particularly through MIDAS methods, yield improved insights into COL dynamics.

1 BACKGROUND STUDY – GOOGLE TRENDS

Google Trends is a tool that provides data on the relative popularity of search queries over time. It enables users to explore patterns in search behaviour, with data normalized so that the highest search volume in a selected period is assigned a value of 100, and all other values are scaled accordingly. This tool is widely used in economic research for forecasting key indicators, analyzing consumer behaviour, and predicting market movements (Choi and Varian, 2012). For instance, researchers have employed Google Trends data to anticipate changes in unemployment rates, housing activity, and even financial markets.

Google Trends offers a powerful example of real-time data in practice. As a freely accessible platform, it provides continuous updates on search activity, reflecting the interests, concerns, and behaviours of populations around the world. Its applications span across business, policymaking, and research. For businesses, Google Trends is invaluable for market research, consumer insights, and forecasting.

It helps identify seasonal demand patterns, detect emerging product categories, and benchmark brand visibility against competitors. In the field of digital marketing, it guides search engine optimization (SEO) strategies by revealing rising keywords and enabling micro-targeted content creation (Polash, 2026). Geographically, Google Trends allows for precise regional analysis of search activity, supporting localized marketing campaigns and enabling researchers to uncover cultural or regional differences in public concerns (Hölzl, 2025). From a policy perspective, it has been increasingly used to nowcast economic indicators such as unemployment, inflation, and consumer sentiment, offering governments a low-cost supplement to traditional surveys, particularly in data-scarce environments (Hölzl, 2025). Its role in crisis detection and public health monitoring has also been well documented, with search spikes serving as early warning signals for events such as influenza outbreaks and the COVID-19 pandemic. In academia and journalism, Google Trends contributes to the study of human behaviour, opinion dynamics, and media coverage by providing a rich source of digital behavioural data (Hölzl, 2025). Importantly, it is not only versatile but also cost-effective, offering free, continuously updated insights that can be integrated into decision-making at multiple levels.

The practical applications of these insights are wide-ranging. In economics, Google Trends data has been incorporated into forecasting models for unemployment, inflation, and housing market dynamics, allowing policymakers and analysts to generate faster estimates than traditional surveys permit. In public health, search patterns have been used to detect outbreaks of influenza and, more recently, to track public concern and behavioural responses during the COVID-19 pandemic. In business, companies use real-time analytics and Google Trends to monitor consumer demand, identify emerging product categories, and adapt their marketing strategies to shifting public interests. These applications illustrate the versatility of Google Trends as both a predictive and diagnostic tool, particularly in contexts where conventional data sources are delayed, costly, or incomplete.

Beyond these practical uses, methodological innovation represents another crucial area of application. The integration of Google Trends with other big data sources, such as social media feeds, economic indicators, or mobility data, has opened new avenues for interdisciplinary research. This approach not only mitigates the limitations of search-only data but also enhances the robustness of findings by triangulating across multiple streams of real-time information. Studies such as those by Jun et al. (2018) and Gummer & Oehrlein (2024) collectively demonstrate that while Google Trends is an accessible and cost-efficient source of data, its full potential is realized when applied with careful attention to methodological considerations and combined with complementary datasets.

While Google Trends is a powerful, free tool for analyzing search behaviour and identifying public interest, it is not a direct measure of absolute search volume. Users must be cautious in utilizing and interpreting the information. Google Trends data are not a raw data, but it is a normalized, relative data (Jun et al., 2018). Such data is useful for comparing and analyzing changes in trends over time, but it cannot provide precise and absolute measurements. In addition, online search might be reliable to reflect personal situation, but it is less reliable when the search variables are unknown to the individual (Niesert et al., 2020). Google Trends search is also claimed to lack sentiment. It shows the volume of search for a topic, but it does not tell why or what sentiment the search has. High frequency data, such as daily data, is only available for short periods. Google Trends cannot provide for long-term trend analysis (over several years) without losing precision.

There are many cases of improper application and analysis using data from Google Trends by researchers. Hölzl et al. (2025) conducted a systematic review on the misuse of Google Trends applications in the social sciences. They revealed many improper sampling and estimation procedures or mistakes by researchers in the literature. The majority of studies did not perform an internal validity test, did not concern about data reliability, and did not concern about the generalizability of results. Failing to address these issues leads to misleading and inaccurate results collected from Google Trends.

Together, these applications underline the dual nature of Google Trends as both an opportunity and a challenge. Its broad accessibility and near real-time availability make it a powerful instrument for academia, business, and public policy. The literature demonstrates that when applied appropriately, Google Trends offers significant value as a tool for forecasting, behavioural analysis, and decision-making in fast-changing environments (Gummer and Oehrlein, 2024; Jun et al., 2018).

2 LITERATURE REVIEW

2.1 Theoretical Review

The cost of living is underpinned by several core economic theories that collectively explain how prices, consumption, and policy interact over time. Index number theory provides the statistical foundation for measuring changes in economic variables, particularly prices, through the construction of index numbers relative to a base period. Central to this is the cost of living index, which quantifies the expenditure required to maintain a constant standard of living over time. The concept originated with Konüs (1939) and was advanced by Frisch (1936). Traditional price indices such as the Laspeyres and Paasche indices, have notable limitations in accounting for shifts in consumer behaviour when relative prices change (Triplett, 1992). To address these limitations, superlative indices, such as the Fisher Ideal and Törnqvist indices were developed. Theoretical justification for these was provided by Diewert (1983), who demonstrated that such indices offer accurate approximations of the "true" cost-of-living index. These indices are widely used for inflation measurement and cost of living analysis due to their ability to reflect substitution effects and real-world price dynamics.

Complementing this, consumer choice theory explains how individuals allocate limited resources to maximize utility amidst changing prices and income. Two key mechanisms are the substitution effect and the income effect (Keat and Young, 2009; Latimaha et al., 2018). Rising living costs tend to restrict substitution, especially for inelastic necessities such as healthcare and housing. This theory highlights how price changes disproportionately affect lower-income households, emphasizing the importance of public policy in mitigating inflation's impact (Latimaha et al., 2018).

Adding a global dimension, trade openness theory explores how liberalization of trade and capital flows influences domestic prices and consumer welfare. Greater openness can reduce inflationary pressures through access to cheaper imports and productivity gains (Latimaha et al., 2018; Squalli and Wilson, 2006). However, it also increases exposure to external shocks such as global commodity price fluctuations and exchange rate volatility. For open economies like Malaysia, currency depreciation raises import costs, thereby exacerbating household financial burdens (Lafrance and Schembri, 2000; Latimaha et al., 2018). While trade openness can enhance consumer welfare, it necessitates policy responses to manage distributional impacts and external vulnerabilities.

Finally, macroeconomic theory provides a broader context for understanding COL dynamics. The Phillips Curve illustrates the trade-off between inflation and unemployment, with both influencing household welfare (Cebula & Toma, 2008; Squalli & Wilson, 2006). Monetary and fiscal policies – including interest rate adjustments and government spending – play crucial roles in controlling inflation and shaping household affordability (International Monetary Fund, 2015; Latimaha et al., 2018; United States Congress, 1996). Wage dynamics are also pivotal; while rising wages can help preserve real incomes, they may contribute to inflation via a wage–price spiral (Blanciforti and Kranner, 1997; Cebula, 1980). Effective policymaking requires balancing growth, price stability, and protection of real incomes, especially for households vulnerable to economic shocks.

2.2 Empirical Review

Consumer Price Index (CPI), representing overall inflation, consistently shows a positive relationship with the cost of living. Bosire and Omwenga (2022) found that inflation increased household living expenses in Kajiado North, Kenya. Similarly, Najaf and Najaf (2016) observed that oil price volatility and exchange rate fluctuations heightened the cost of living in Pakistan through inflationary effects. Latimaha et al. (2018) reported similar findings for Malaysia, identifying inflation as a key driver of rising living costs. Overall, CPI emerges as a consistently significant and positive contributor to living costs across diverse countries and methodologies.

Gross Domestic Product (GDP), commonly used as a proxy for national income or economic growth, typically influences the cost of living through rising income levels and increased demand for goods and services. Jenkins (2000) found that temporary increases in household income in the UK led to short-term rises in living costs. Cebula and Todd (2004) similarly reported a positive association between economic activity – linked to GDP – and the cost of living in Florida counties. Although many studies focus on income or wages rather than GDP directly, the evidence consistently supports a positive relationship between GDP and cost of living.

Population size, especially in urban areas, is also positively associated with living costs. Haworth and Rasmussen (1973) observed that larger U.S. metropolitan populations incurred higher living expenses due to increased demand for housing and public services. Cebula and Toma (2008) confirmed these findings, noting that population density was positively correlated with the cost of living in U.S. counties. In Malaysia, Latimaha et al. (2018) identified urbanization as a significant factor contributing to elevated living costs. Collectively, these studies indicate that larger and denser populations drive up living expenses, particularly in urban settings.

Exchange rate fluctuations, especially currency depreciation, are widely found to increase the cost of living by raising the price of imported goods. Hobdari and Uppal (2017) showed that depreciation led to food price shocks and higher living costs in Sub-Saharan Africa. Similarly, Najaf and Najaf (2016) found that exchange rate instability contributed to inflation and increased living expenses in Pakistan. Lafrance and Schembri (2000) confirmed this relationship in Canada, where currency depreciation was linked to higher living costs. Overall, the literature consistently identifies a positive relationship between exchange rate depreciation and the cost of living.

Housing and property costs are frequently identified as key contributors to living expenses. Kurre (2000) reported that housing prices significantly increased the cost of living in Pennsylvania counties. Cebula (1980) and Cebula and Toma (2008) also found that housing costs and urban amenities were positively associated with living costs. In Malaysia, Latimaha et al. (2018) highlighted housing expenses as a major determinant of living costs in capital cities. Additionally, Bridel and Lontoh (2014) showed that the removal of fuel subsidies raised transportation costs, which could indirectly impact housing affordability through commuting expenses. Across all reviewed studies, housing and property prices are consistently linked to higher living costs.

Unemployment shows a more complex and indirect relationship with the cost of living. Latimaha et al. (2019) found that unemployment in Malaysian states indirectly contributed to higher living costs by exacerbating urban crime issues. However, unemployment is not commonly analyzed as a direct determinant of living costs. The limited evidence suggests a positive but indirect link, indicating a need for further research in this area.

2.3 Research Gaps

A thorough review of existing literature reveals significant gaps in methodology, data usage, and analytical scope that this study seeks to address. A critical shortcoming in existing research is its reliance on conventional macroeconomic indicators – such as inflation, GDP growth, and employment – that are typically delayed in reporting, aggregated at national levels, and insufficient for capturing individuals' and communities' lived economic realities. These top-down measures lack the granularity required to reflect public sentiment and the day-to-day experiences of economic pressure, thereby limiting their relevance for timely and localized policy decisions. As economies face increasingly rapid shifts, especially during crises, the need for real-time, high-frequency data has grown. Google Trends, by reflecting what people search for in real time, offers an immediate and cost-effective complement to traditional statistics.

Previous studies also have some shortcomings in terms of methodology, scope, and data used. A major methodological limitation is the predominance of conventional linear regression models, such as ordinary least squares (OLS). For instance, studies by Blanciforti and Kranner (1997) and Cebula and Todd (2004) used linear techniques to analyse cost-of-living trends in the U.S., often neglecting complex economic behaviours such as nonlinearities, structural changes, and asymmetric responses. These approaches may oversimplify interactions and fail to capture the nuanced behaviour of cost-related variables across different countries and time periods.

Geographical and temporal constraints further weaken the generalizability of prior research. Many studies are confined to specific regions or policy events – for example, Bridel and Lontoh's (2014) study of Malaysia's 2013 fuel reform or Bosire and Omwenga's (2022) work in a Kenyan sub-county. Such narrow scopes hinder comparative and longitudinal analysis. Next, data limitations also constrain the utility of existing studies. Most rely on low-frequency, aggregated macroeconomic indicators from official sources. While reliable, these data suffer from reporting lags, lack household-level detail, and fail to capture real-time public sentiment. This disconnect between delayed statistics and immediate lived experience weakens the policy relevance of cost-of-living research. As a result, there is a growing recognition of the importance of real-time data sources that can complement and enhance traditional measures. In recent years, Google Trends has gained prominence in economic research, particularly in forecasting unemployment, consumer behaviour, and financial markets. However, its application to COL analysis remains limited, leaving an important gap that this study seeks to fill.

To overcome these issues, researchers have begun exploring advanced methodologies and novel data sources. Mixed Data Sampling (MIDAS) models offer a robust framework for integrating high- and low-frequency data without sacrificing information richness. Rather than relying on macroeconomic indicators, which are slow and delay in capturing economic behaviours, this study applies a new approach by integrating traditional macroeconomic indicators with alternative digital data – specifically, Google Trends search behaviour. Real-time search queries related to inflation, housing, healthcare, and government subsidies are analyzed to construct thematic indices that act as proxies for public attention and sentiment. These digital indicators complement official statistics by capturing emerging concerns as they unfold, allowing for earlier detection of economic stress compared to lagged macroeconomic measures.

3 DATA AND METHODOLOGY

3.1 Data Collection and Processing

This study is focused on six selected Asian countries: Indonesia (IDN), Japan (JAP), Korea (KOR), Malaysia (MYS), Singapore (SGP), and Thailand (THA). These countries represent two main clusters, the developed Asia (JAP, KOR, and SGP) versus the developing Asia (IDN, MYS, and THA). This study seeks to compare the cost of living (COL) in these two clusters in Asia, reflecting the COL situation of very high COL versus those that are experiencing increasing COL in Asia. The results might not fully represent the conditions of all countries in Asia, but they could tell the conditions of COL in many countries in Asia.

The data are in different frequencies, covering the range of 2010 to 2024. This study uses COL as the main variable of interest, measured annually using data from Numbeo, a global database that compiles user-reported consumer prices and living costs. Numbeo, one of the most comprehensive databases of cost-of-living data, calculates its index based on crowdsourced prices for core goods and services such as food, rent, utilities, and transportation. The index is benchmarked against New York, which is assigned a value of 100. For example, Kuala Lumpur, Malaysia, has a cost of living index of 33.3, implying that the overall expenses there are approximately 33.3% of those in New York. An index value below 100 indicates lower relative costs, while a value above 100 reflects higher expenses. This index plays a key role in wage benchmarking, relocation planning, business site selection, and policy formulation. Nonetheless, its accuracy may be affected by inconsistencies in data collection, currency volatility, and differing consumption patterns across regions.

The data collection process of Numbeo involves user-generated input and manually gathered information from reputable sources. The reliance on such crowdsourced data can be prone to human error and seasonal fluctuations (Numbeo, 2026). The cost coverage also overlooks local, non-consumer costs such as taxes, healthcare, and childcare, and it does not account for individual lifestyle differences. The index is calculated from a fixed „basket of goods" for a hypothetical four-person family (Numbeo, 2026). Such a rigid lifestyle weighting fails to account for individual spending habits. Furthermore, the use of New York City as a baseline may not well reflect the consumption habits of residents in other locations. Despite these limitations, Numbeo offers a platform and free access to its extensive dataset for COL analyses and comparisons, covering thousands of cities, providing various specific metrics for international comparisons

In addition to COL, several macroeconomic indicators were included to capture broader economic and price dynamics: consumer price index (CPI), gross domestic product (GDP), population size (POP), real effective exchange rate (RER), real residential property price index (RPI), and unemployment rate (URT). These variables reflect key dimensions of economic performance, labour market conditions, inflationary pressure, and housing affordability. Data for these indicators were obtained from multiple sources, including Numbeo, the International Monetary Fund (IMF), the Federal Reserve Economic Data (FRED), the Department of Statistics Malaysia (DOSM), and the Singapore Department of Statistics (SINGSTAT).

Table 1 summarizes the variables used, including their descriptions, frequencies, sources, and relevant notes. For GDP, a natural logarithm transformation was applied to account for its large scale; the transformed variable is referred to as LNGDP in the analysis. Variables reported annually, namely COL and POP, were converted to quarterly frequency using linear interpolation. Linear interpolation is a mathematical technique used to estimate unknown values that fall between two known data points by assuming a straight line exists between them. It provides simple and easy implementation. It is very effective when the data follows a smooth trend as linear interpolation can consistently yield the lowest error rates and perform better than much more complex machine learning models (Zuhairi et al., 2025).

Table 1 List of variables used in the analysis

Variable	Description	Frequency	Source	Remarks
COL	Cost of living index	Annual	Numbeo	
CPI	Consumer price index	Monthly	IMF	IDN data obtained from FRED
GDP	Gross domestic product	Quarterly	IMF	MYS data from DOSM; log- transformed as LNGDP
POP	Population (millions)	Annual	IMF	
RER	Real effective exchange rate	Monthly	FRED	
RPI	Real residential property price index	Quarterly	FRED	
URT	Unemployment rate (%)	Quarterly	IMF	MYS data from DOSM; SGP data from SINGSTAT

Source: Authors' computation

Complementing macroeconomic indicators, the study incorporates monthly Google Trends data for six search terms linked to economic concerns: *debt*, *inflation*, *loan*, *recession*, *tax*, *and unemployment*. These search indices, retrieved in monthly frequency without aggregation, serve as proxies for public sentiment and offer high-frequency insights into cost-of-living pressures.

The following four model specifications were estimated:

Model 1: COL= F (CPI, LNGDP, POP, RER, RPI, URT)

Model 2: COL= F (*debt*, *inflation*, *loan*, *recession*, *tax*, *unemployment*)

Model 3: COL= F(PC1)

Model 4: COL= F (CPI, LNGDP, POP, RER, RPI, URT, PC1)

In the above, PC1 refers to the first principal component obtained from Principal Component Analysis (PCA) of the six Google Trends indices. The difference between Model 2 and Model 3 is, Model 2 is based on keywords search using Google Trends, each keyword is associated with a single search volume index. Then, in Model 3, all six indices from Model 2 are integrated to generate a composite index using PCA method. As a summary, Model 1 is based on macroeconomic data alone, Model 2 and Model 3 are regressed on Google Trends data while Model 4 consists of macroeconomic data and Google Trends data. Each model was estimated using two approaches.

3.2 Methodology – MIDAS

Mixed Data Sampling (MIDAS) regression is an econometric technique designed to incorporate regressors observed at higher frequencies into a regression with a lower-frequency dependent variable, without resorting to temporal aggregation. The method allows for flexible and parsimonious modeling of lag effects from high-frequency predictors by using weighting schemes that restrict or parameterize the lag structure.

The general form of the MIDAS regression is:

$$y_t = \mathbf{X}_t^{(L)'}\beta + f\left(\{\mathbf{X}_t^{(H)}\} : \nu, \ell\right) + \varepsilon_t$$

where y_t is the dependent variable observed at low frequency, $\mathbf{X}_t^{(L)}$ is the vector of low-frequency regressors with coefficients β; $\mathbf{X}_t^{(H)}$ is a set of high-frequency regressors, $f(\cdot)$ is a lag weighting function parameterized by ν over ℓ high-frequency lags, and ε_t is the error term.

Several forms of MIDAS regressions differ in how they specify the weighting function for the high-frequency lags. In the step-weighted approach, the high-frequency lags are grouped into steps of equal length, with each step sharing the same coefficient. Another approach uses polynomial distributed lag (PDL) weighting, also known as Almon lags, which constrains the lag coefficients to lie on a polynomial curve. This method yields smoothly varying weights that can capture gradual decay, rise, or hump-shaped lag effects.

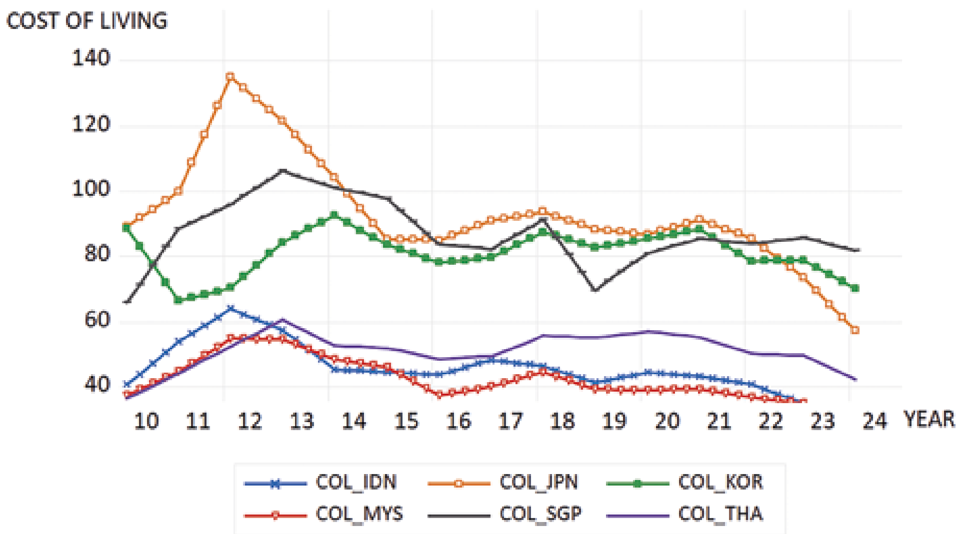
In contrast to these structured models, the unrestricted MIDAS (U-MIDAS) formulation allows each high-frequency lag to enter the model with its own freely estimated coefficient. To improve parsimony while retaining the flexibility of U-MIDAS, the general-to-specific (GETS-MIDAS) procedure begins with the full set of unrestricted lags and applies a model selection algorithm to eliminate those that are statistically insignificant. The selection is typically based on information criteria such as the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC), resulting in a sparser model that retains only the most informative lags.

4 RESULTS AND DISCUSSION

This section presents the key empirical findings of the study, focusing on the comparative performance of different estimation models and techniques in modelling and predicting the cost of living (COL) across selected Asian countries. The primary aim is to evaluate whether incorporating Google Trends data enhances model performance relative to relying solely on traditional macroeconomic variables.

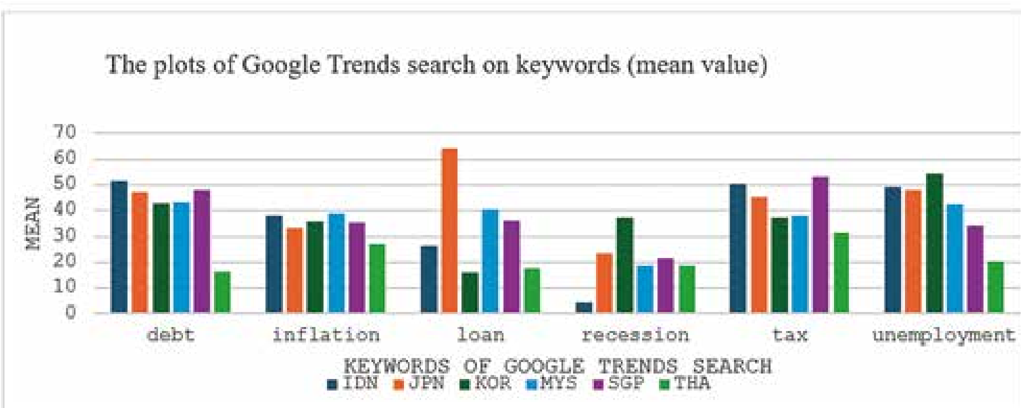
Below are the plots of COL and regressors (macroeconomics versus Google Trends keywords indices).

Figure 1 Cost of living (2010–2024)



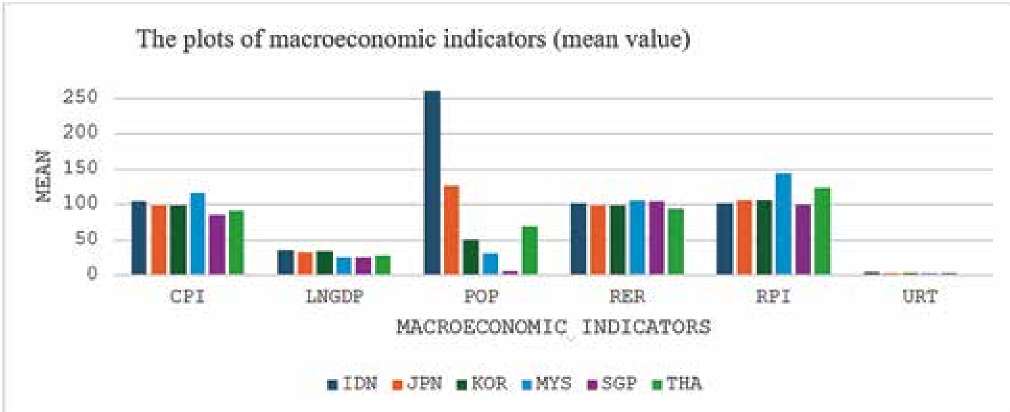
Source: Authors' computation

Figure 2 Mean values of Google Trends keywords indices



Source: Authors' computation

Figure 3 Mean values of macroeconomic indicators



Source: Authors' computation

The cost of living plots (Figure 1) show that there are two clusters of COL where Japan, Korea, and Singapore are under the same cluster with higher COL; Indonesia, Malaysia, and Thailand are in the lower COL cluster. The cluster with higher COL includes countries with higher economic development compared to the lower cluster with lower economic development. All countries show a declining trend in the 2020s relative to New York, U.S.

In terms of Google Trends search (Figure 2), each keyword search is accompanied by a core index. These indices reflect the main concern of individuals. For instance, Japan is more concerned about loans; while Singapore and Indonesia are concerned about taxes and debt, and Korea is concerned about unemployment. Malaysia is more concerned about inflation. In general, keywords debt, tax, and unemployment show higher values; these keywords are searched more frequently, while the word recession has a much lower search volume.

In terms of macroeconomic indicators (Figure 3), Malaysia exhibited higher mean values (2010–2024) in inflation, real exchange rate, and real property index than other countries. Japan and Indonesia are leading in terms of population size. Thailand also showed a relatively higher mean in the real property index. These macroeconomic factors might contribute to higher COL in respective countries, but they may not track the behaviour of households in time.

4.1 Model Performance by Country

The analysis involves the following steps. First of all, preliminary tests (unit-root and cointegration) are performed to examine the properties of the data. Unit-root tests reveal that all variables are not stationary at the level but are stationary in first differences. As all variables are stationary in first differences, the cointegration test is performed and reveals the existence of a long-run relationship. Hence, all models (Model 1, 2, 3, and 4) are estimated at the level form (long-run relationship) by comparing MIDAS (PDL-MIDAS, STEP-MIDAS, U-MIDAS, and GETS-MIDAS) with the benchmark OLS. Results are evaluated using five metrics. Finally, forecast averaging analysis is performed on the best model to proxy COL for each country.

Model evaluation is based on standard metrics: Adjusted R-squared, Loglikelihood, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). Higher Adjusted R-squared and less negative Loglikelihood values indicate better model fit, while lower RMSE and MAE suggest improved forecast precision. Across the four specified models, estimations were conducted using Ordinary Least Squares

(OLS) and four variants of Mixed Data Sampling (MIDAS) regression – PDL-MIDAS, STEP-MIDAS, U-MIDAS, and GETS-MIDAS.

Tables 2 to 5 present the detailed evaluation metrics for Models 1 to 4, respectively, across all six countries in the study. Each table reports the values of Adjusted R-squared, Loglikelihood, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) for the five estimation methods: OLS, PDL-MIDAS, STEP-MIDAS, U-MIDAS, and GETS-MIDAS. These metrics allow for a comprehensive comparison of model fit and predictive accuracy, highlighting the strengths and weaknesses of each estimation approach in different country-model combinations. For each country and metric, the best-performing value is highlighted in bold and italics to facilitate identification of the most accurate method. Additionally, the final column in each table indicates the best overall model for each country, based on a holistic evaluation of all metrics.

Table 2 Evaluation metrics for Model 1

Country	Method	R-SQ	LOGLIKE	RMSE	MAE	Best Model
IDN	OLS	***0.9872***	***–60.2265***	0.7705	***0.6323***	OLS
	PDL-MIDAS	0.9824	–67.1751	0.8594	0.6925	
	STEP-MIDAS	0.9826	–68.0871	0.8743	0.7296	
	U-MIDAS	0.9837	–60.9400	***0.7640***	0.6336	
	GETS-MIDAS	0.9845	–64.4392	0.8162	0.6578	
JPN	OLS	***0.9833***	–112.6004	2.0251	***1.4550***	OLS
	PDL-MIDAS	0.9799	–116.8464	2.1062	1.5416	
	STEP-MIDAS	0.9800	–117.9158	2.1483	1.5637	
	U-MIDAS	0.9797	–112.9325	***1.9589***	1.4696	
	GETS-MIDAS	0.9818	***–99.8482***	11.6083	8.2500	
KOR	OLS	0.9182	–101.3504	1.6378	1.3984	U-MIDAS
	PDL-MIDAS	0.9229	–99.8482	1.5374	1.2535	
	STEP-MIDAS	0.9208	–101.8469	1.5954	1.2863	
	U-MIDAS	0.9030	***–92.9654***	***1.3534***	***1.0785***	
	GETS-MIDAS	***0.9329***	–96.1009	1.7897	1.4996	
MYS	OLS	0.9815	–64.4726	0.8167	0.6698	U-MIDAS
	PDL-MIDAS	0.9819	–60.9549	0.7482	0.6103	
	STEP-MIDAS	0.9828	–61.4998	0.7557	0.6065	
	U-MIDAS	0.9873	***–44.5608***	***0.5523***	***0.4312***	
	GETS-MIDAS	***0.9901***	–46.0247	0.5674	0.4388	
SGP	OLS	0.9365	–112.6106	2.0254	1.4560	U-MIDAS
	PDL-MIDAS	0.9391	–110.9186	1.8872	1.4027	
	STEP-MIDAS	0.9407	–111.4155	1.9047	1.4240	
	U-MIDAS	0.9317	***–109.8249***	***1.8494***	***1.3992***	
	GETS-MIDAS	***0.9424***	–112.4420	1.9412	1.4481	
THA	OLS	0.9347	–71.5218	0.9329	0.7742	U-MIDAS
	PDL-MIDAS	0.9470	–67.4159	0.8432	0.7012	
	STEP-MIDAS	0.9467	–68.8133	0.8654	0.7142	
	U-MIDAS	***0.9487***	***–62.3800***	***0.7682***	***0.6191***	
	GETS-MIDAS	0.9468	–70.5475	0.8936	0.6775	

Source: Authors' computation

Table 3 Evaluation metrics for Model 2

Country	Method	R-SQ	LOGLIKE	RMSE	MAE	Best Model
IDN	OLS	***0.9958***	–106.5034	4.7127	3.4624	U-MIDAS
	PDL-MIDAS	0.9697	–78.9397	1.0438	0.7963	
	STEP-MIDAS	0.9716	–81.5466	1.0955	0.8692	
	U-MIDAS	0.9810	***–45.9353***	***0.5665***	***0.4494***	
	GETS-MIDAS	0.9864	–62.0820	15.5234	13.9179	
JPN	OLS	***0.9969***	–214.8034	7.4933	5.6438	U-MIDAS
	PDL-MIDAS	0.9784	–117.8481	2.0481	1.6732	
	STEP-MIDAS	0.9793	–119.4305	2.1113	1.7425	
	U-MIDAS	0.9901	***–73.3924***	***1.0637***	***0.8632***	
	GETS-MIDAS	0.9833	–140.7901	2.7096	2.2567	
KOR	OLS	***0.9972***	–154.9581	5.2479	4.0543	U-MIDAS
	PDL-MIDAS	0.9803	–83.1075	1.2652	1.0338	
	STEP-MIDAS	0.9806	–83.8267	1.2908	1.0549	
	U-MIDAS	0.9910	***–48.5823***	***0.6561***	***0.5195***	
	GETS-MIDAS	0.9893	–69.1165	1.0057	0.8807	
MYS	OLS	***0.9911***	–125.3293	6.2892	4.4566	U-MIDAS
	PDL-MIDAS	0.9700	–84.2465	2.2122	1.8016	
	STEP-MIDAS	0.9713	–85.6197	2.2753	1.8533	
	U-MIDAS	0.9807	***–57.3927***	***1.2815***	***1.0614***	
	GETS-MIDAS	0.9787	–70.0708	1.6547	1.4036	
SGP	OLS	***0.9963***	–137.4049	4.6192	3.5945	U-MIDAS
	PDL-MIDAS	0.9744	–73.9705	1.0722	0.8770	
	STEP-MIDAS	0.9744	–73.9803	1.0726	0.8773	
	U-MIDAS	0.9870	***–40.7327***	***0.5705***	***0.4581***	
	GETS-MIDAS	0.9833	–56.0417	0.8561	0.6995	
THA	OLS	***0.9986***	–115.1474	2.8051	2.1862	U-MIDAS
	PDL-MIDAS	0.9807	–67.6515	0.9093	0.7461	
	STEP-MIDAS	0.9809	–67.9927	0.9207	0.7549	
	U-MIDAS	0.9891	***–30.9165***	***0.4290***	***0.2847***	
	GETS-MIDAS	0.9879	–54.2076	0.8252	0.6733	

Source: Authors' computation

Table 4 Evaluation metrics for Model 3

Country	Method	R-SQ	LOGLIKE	RMSE	MAE	Best Model
IDN	OLS	***0.9900***	***−58.6022***	***0.7311***	***0.4845***	OLS
	PDL-MIDAS	0.9587	−97.1553	1.4626	1.0704	
	STEP-MIDAS	0.9596	−97.0758	1.4605	1.0720	
	U-MIDAS	0.9568	−96.6511	1.4490	1.0785	
	GETS-MIDAS	0.9608	−97.2928	1.4664	1.0602	
JPN	OLS	0.9588	−139.2548	3.3485	2.6539	U-MIDAS
	PDL-MIDAS	0.9579	−140.9368	3.2904	2.5678	
	STEP-MIDAS	0.9588	−140.9193	3.2893	2.5714	
	U-MIDAS	***0.9851***	***−110.5077***	***1.8729***	***1.0235***	
	GETS-MIDAS	0.9597	−140.8451	3.2848	2.5856	
KOR	OLS	***0.9145***	***−105.3153***	1.7650	1.4938	U-MIDAS
	PDL-MIDAS	0.9056	−109.4711	1.8373	1.5563	
	STEP-MIDAS	0.9079	−109.3660	1.8337	1.5455	
	U-MIDAS	0.9080	−107.0721	***1.7575***	***1.4601***	
	GETS-MIDAS	0.9085	−109.7051	1.8453	1.5547	
MYS	OLS	***0.9658***	−83.6015	1.1717	0.9583	U-MIDAS
	PDL-MIDAS	0.9718	−78.4188	1.0338	0.8567	
	STEP-MIDAS	0.9721	−78.6740	1.0387	0.8679	
	U-MIDAS	0.9706	***−77.7849***	***1.0218***	***0.8342***	
	GETS-MIDAS	0.9711	−80.1721	1.0679	0.8568	
SGP	OLS	***0.9196***	***−121.6273***	***2.4011***	***1.8327***	OLS
	PDL-MIDAS	0.9098	−125.6759	2.4803	1.8731	
	STEP-MIDAS	0.9114	−125.7198	2.4824	1.8703	
	U-MIDAS	0.9074	−124.6798	2.4350	1.8495	
	GETS-MIDAS	0.9148	−125.7297	2.4828	1.8646	
THA	OLS	0.9143	***−81.5154***	***1.1265***	***0.9433***	OLS
	PDL-MIDAS	0.9134	−84.8369	1.1643	1.0210	
	STEP-MIDAS	0.9155	−84.7220	1.1618	1.0150	
	U-MIDAS	0.9130	−83.2799	1.1312	0.9525	
	GETS-MIDAS	***0.9169***	−84.8085	1.1637	1.0056	

Source: Authors' computation

Table 5 Evaluation metrics for Model 4

Country	Method	R-SQ	LOGLIKE	RMSE	MAE	Best Model
IDN	OLS	***0.9883***	–57.3356	0.7288	0.6101	U-MIDAS
	PDL-MIDAS	0.9832	–63.9516	0.8087	0.6446	
	STEP-MIDAS	0.9840	–64.7096	0.8204	0.6634	
	U-MIDAS	0.9841	***–55.2170***	***0.6858***	***0.5626***	
	GETS-MIDAS	0.9850	–62.8451	13.1594	10.8055	
JPN	OLS	***0.9839***	–111.0176	1.9655	***1.3930***	U-MIDAS
	PDL-MIDAS	0.9798	–114.9708	2.0343	1.5244	
	STEP-MIDAS	0.9810	–115.3348	2.0481	1.5403	
	U-MIDAS	0.9783	***–109.7732***	***1.8476***	1.3938	
	GETS-MIDAS	0.9812	–110.8418	94.5502	93.3076	
KOR	OLS	0.9232	–99.0843	1.5692	1.3386	U-MIDAS
	PDL-MIDAS	0.9441	–89.1627	1.2614	1.0180	
	STEP-MIDAS	0.9466	–89.9252	1.2793	1.0397	
	U-MIDAS	0.9390	***–84.4279***	***1.1555***	***0.9634***	
	GETS-MIDAS	***0.9488***	–90.6846	1.2975	1.0498	
MYS	OLS	0.9822	–62.9381	0.7934	0.6383	U-MIDAS
	PDL-MIDAS	0.9845	–54.6518	0.6657	0.5435	
	STEP-MIDAS	0.9846	–57.1992	0.6979	0.5679	
	U-MIDAS	0.9896	***–33.2582***	***0.4480***	***0.3506***	
	GETS-MIDAS	**0.9910**	–42.6560	0.5331	0.4257	
SGP	OLS	0.9365	–111.9823	2.0016	1.4639	U-MIDAS
	PDL-MIDAS	0.9388	–109.0439	1.8228	1.3988	
	STEP-MIDAS	***0.9418***	–109.6673	1.8440	1.4379	
	U-MIDAS	0.9333	***–104.2763***	***1.6688***	***1.2850***	
	GETS-MIDAS	0.9374	–115.2292	2.0441	1.4498	
THA	OLS	0.9376	–69.6867	0.9011	0.7450	U-MIDAS
	PDL-MIDAS	0.9443	–66.7720	0.8333	0.6880	
	STEP-MIDAS	0.9451	–68.4000	0.8588	0.6945	
	U-MIDAS	0.9441	***–59.7833***	***0.7321***	***0.5644***	
	GETS-MIDAS	***0.9468***	–79.5475	0.8936	0.6775	

Source: Authors' computation

In Model 1, which includes only macroeconomic variables, performance differed by country, with both OLS and U-MIDAS achieving strong results. In Indonesia and Japan, OLS outperformed the MIDAS variants, particularly in terms of Adjusted R-squared and MAE, suggesting that in these contexts, the macroeconomic indicators captured most of the relevant variation in COL. However, in Korea, Malaysia, Singapore, and Thailand, U-MIDAS consistently outshone OLS, delivering lower RMSE and higher loglikelihood values. This pattern suggests that in these countries, even for macroeconomic data, the flexible structure of U-MIDAS allowed for better exploitation of temporal patterns. Notably, GETS- MIDAS occasionally recorded high R-squared values but was penalized by its higher forecast errors, confirming that good in-sample fit does not necessarily translate to better predictive performance.

Model 2, which relies exclusively on individual Google Trends indices, demonstrated the strongest predictive capacity overall. U-MIDAS was the clear leader in every country, achieving the best loglikelihood, RMSE, and MAE. Although OLS reported slightly higher R-squared values in some cases, the difference was marginal and did not compensate for its inferior forecast accuracy. The superiority of U-MIDAS in this model highlights the model's ability to exploit the real-time nature of search trends data, which may reflect emerging concerns about the cost of living more promptly than traditional indicators. PDL-MIDAS and STEP-MIDAS were moderately effective but failed to challenge U-MIDAS in any country. GETS-MIDAS again underperformed, particularly in Indonesia, where its forecast errors were significantly higher.

For Model 3, which utilises a composite Google Trends index generated via Principal Component Analysis (PCA), performance varied more widely. In Indonesia and Singapore, OLS outperformed all MIDAS methods, possibly due to the strong explanatory power of the composite index in these contexts. Conversely, U-MIDAS was the top performer in Korea and Malaysia, achieving the lowest RMSE and MAE, suggesting that in these cases, the flexible dynamic weighting of search data within U-MIDAS was beneficial. Japan presented a mixed outcome: while OLS had a higher R-squared, U-MIDAS achieved the best RMSE and loglikelihood, tipping the balance in favour of U-MIDAS. Thailand's results were relatively balanced, though OLS narrowly led in most performance metrics. The mixed findings for Model 3 suggest that a composite index may dilute specific signals present in individual search series, resulting in performance differences depending on country-specific dynamics.

In Model 4, which combines macroeconomic variables with the composite Google Trends index, U- MIDAS emerged as the most robust and reliable method in five out of six countries. It consistently achieved the lowest RMSE and highest loglikelihood, reflecting strong predictive ability and good model fit. Japan was the only exception, where OLS surpassed U-MIDAS in R-squared and MAE, although U-MIDAS still had better RMSE and loglikelihood. PDL-MIDAS, STEP-MIDAS, and GETS-MIDAS occasionally ranked highly in specific metrics but could not match U-MIDAS's overall consistency. These results underscore the added value of integrating behavioural data with economic indicators, particularly when processed through a flexible modelling framework such as U-MIDAS.

4.2 Comparative Insights

Table 6 reports the estimation performance metrics for all four models across the six selected countries. The metrics include R-squared (R-SQ), log-likelihood (LOGLIKE), root mean square error (RMSE), and mean absolute error (MAE). For each country, the best value for each metric is highlighted in bold and italics. The best overall model for each country – determined by the collective performance across all metrics – is indicated in the final column. This comparative analysis provides a comprehensive evaluation of model fit and forecasting accuracy across different model specifications and countries.

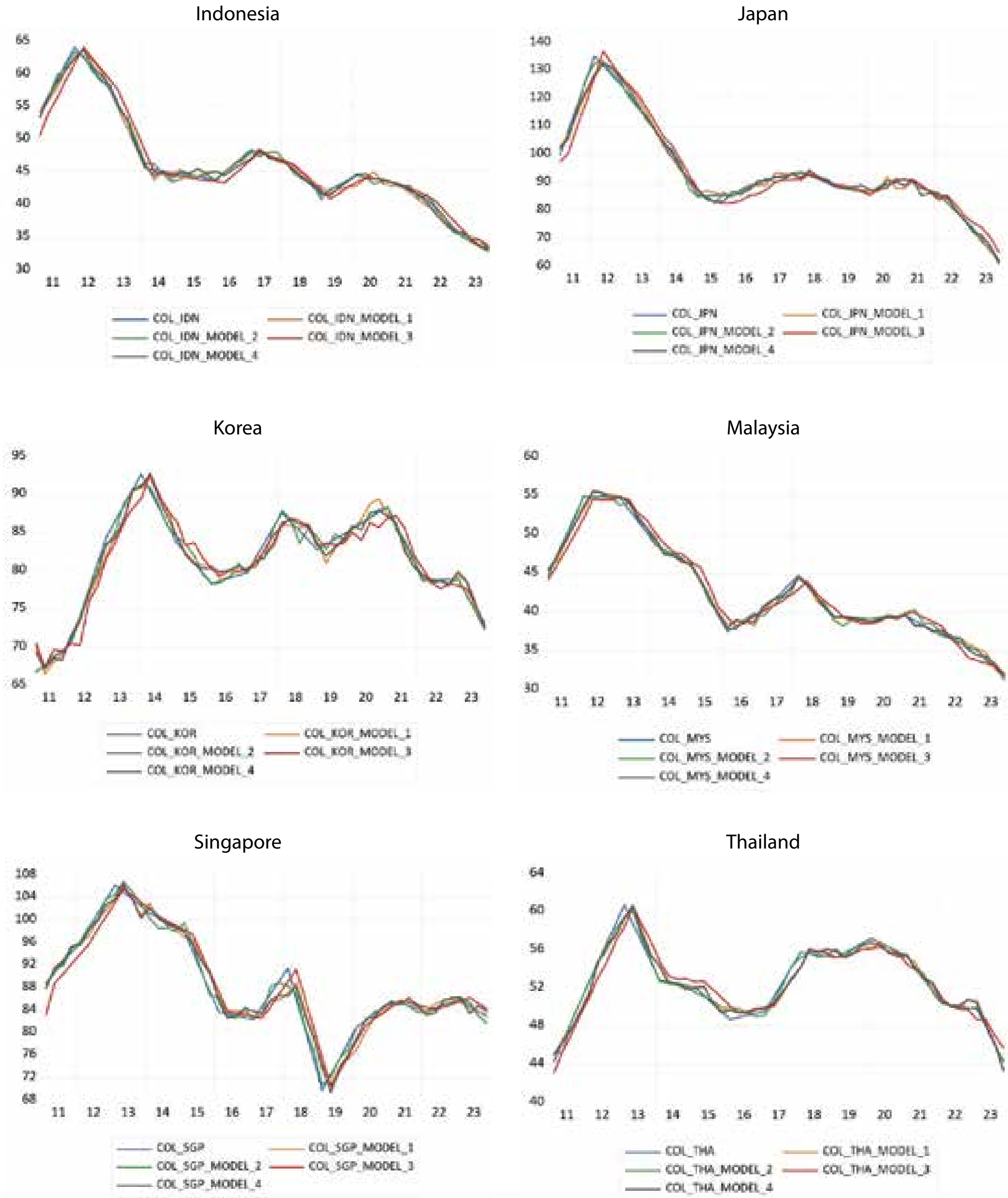
Table 6 Estimation Performance Metrics and Best Model Selection

Country	Model	R-SQ	LOGLIKE	RMSE	MAPE	Best Model
IDN	1	***0.9872***	-60.2265	0.7705	1.3803	2
	2	0.9810	***-45.9353***	***0.5665***	***0.9672***	
	3	0.9900	-58.6022	0.7311	1.0074	
	4	0.9841	-55.2170	0.6858	1.2178	
JPN	1	0.9833	-112.6004	2.0251	1.4507	2
	2	***0.9851***	***-82.6423***	***1.1179***	***0.9515***	
	3	0.9851	-110.5077	1.8729	1.0555	
	4	0.9783	-109.7732	1.8476	1.4610	
KOR	1	0.9030	-92.9654	1.3534	1.3312	2
	2	***0.9569***	***-58.0789***	***0.7094***	***0.5946***	
	3	0.9080	-107.0721	1.7575	1.7959	
	4	0.9390	-84.4279	1.1555	1.1908	
MYS	1	0.9873	-44.5608	0.5523	0.9951	4
	2	0.9812	-37.1707	0.4816	0.8535	
	3	0.9706	-77.7849	1.0218	1.9427	
	4	***0.9896***	***-33.2582***	***0.4480***	***0.8144***	
SGP	1	0.9317	-109.8249	1.8494	1.6331	2
	2	***0.9532***	***-78.2403***	***1.0304***	***0.8819***	
	3	0.9074	-124.6798	2.4350	2.1621	
	4	0.9333	-104.2763	1.6688	1.5077	
THA	1	0.9487	-62.3800	0.7682	1.1739	2
	2	***0.9891***	***-30.9165***	***0.4290***	***0.5426***	
	3	0.9143	-81.5154	1.1265	1.8053	
	4	0.9441	-59.7833	0.7321	1.0685	

Source: Authors' computation

Figure 4 provides the plots of real COL data (in dark blue line) versus the estimated COL under different models. In most cases, Model 2 outperforms the other models as COL estimated using Model 2 is the closest to the real COL data.

Figure 4 Estimated COL versus real COL data



Source: Authors' computation

The comparative analysis yields several important insights. First, models that incorporate Google Trends data – especially Model 2 and Model 4 – generally outperform the baseline macroeconomic model (Model 1). Specifically, Model 2 was the best-performing specification in five out of six countries, while Model 4 led only in Malaysia, where Model 2 was a close second. This suggests that search data, either on its own or in combination with traditional indicators, provides a richer and more timely signal for modelling the cost of living. Model 3, which uses a composite index, was the weakest performer across all countries, reinforcing the idea that disaggregated search data retains more useful information than a single aggregated measure.

Secondly, the results emphasize the importance of using appropriate estimation techniques. MIDAS models – particularly U-MIDAS – consistently outperformed OLS in terms of forecast accuracy, regardless of the model specification. This confirms the methodological advantage of MIDAS in handling mixed-frequency data, allowing for better exploitation of real-time search trends.

Finally, these findings have broader implications. They demonstrate that integrating behavioural data, such as Google search trends, with traditional economic analysis can enhance the monitoring and forecasting of COL pressures. This is relevant for both developed and developing economies, as policymakers seek more responsive tools to detect and manage economic challenges. The effectiveness of MIDAS models in this context highlights their value in real-time economic analytics, supporting data-driven policy interventions that are timely and adaptive.

4.3 Discussion

Google Trends keywords search reveals that people from different countries show their specific concerns, reflecting their cost of living particulars. Relatively, Japan is highly concerned about 'loan', Singapore is more concerned about 'debt' and 'tax', Korea is relatively more concerned about 'debt', 'recession' and 'unemployment'; Indonesia is concerned about 'tax' and 'unemployment', while Malaysia is concerned about 'inflation'. These keywords are the main search and could predict the movement of COL better than the macroeconomic factors, as they reflect the real-time needs and behaviour shifts. These keywords searches can be classified into financial stress (such as debt, loan), economic uncertainty or shock (recession, inflation), and policy stance (tax). These factors are recognized to be influential in nowcasting COL across countries, but with different magnitudes. Comparing the results between the developed versus developing clusters, there is no clear-cut on which factor may dominate COL according to the country development cluster.

In general, loans, debt, and taxes are more relevant or influential, while recession is less influential. These three determinants are more influential as they could induce higher financial burden and risk, lower purchasing power, and lower living standard. As discussed in Chai et al. (2020), indebtedness may expose financial stress and financial risk, and other economic implications, including a reduction of future consumption. On the other hand, the impose of taxes on goods and services induces higher prices. An increase in one key product triggers a cyclical chain reaction of price increases in many more products across the economy, leading to cost-push inflation. The increase of income tax causes a lower spending ability. All these lead to higher COL. According to Maina et al. (2024), Direct taxation is more influential than indirect taxation in affecting consumer burden and the house price index. The loan also affects COL in the same way by boosting higher financial stress and lowering spending ability. As COL is defined as the amount needed to sustain a minimal living standard, the increase of debt, loans, and taxes will lead to a drop in money available to sustain the same living standard, leading to a spike in COL.

The level of development does not directly determine the factors leading to the increase of COL, but country-specific factors (e.g., economic structure, economic stability, changes of policies) and living environment/styles might have more direct impacts on COL. Hence, the results might not fully represent the two clusters within Asia or outside Asia. Rather, it reflects the relative influence of each factor across countries by comparing countries with different characteristics.

Next, the nowcasting results show that all six Asian countries are expected to show a lower cost of living compared to the benchmark of the U.S., with Korea and Thailand show deeper declining trend than the others and Singapore might remain high in COL. The results provide useful information to the policymaker to focus on the keywords and macroeconomic variables that show greater magnitude in each country, as these are the main determinants of COL. Such integration of macroeconomic data and Google Trends data could be utilized by policymakers in other countries in analyzing the dynamics of COL.

CONCLUSIONS

This study investigated the potential of Google Trends data as a real-time, high-frequency indicator for modelling and forecasting the cost of living (COL) in six Asian countries: Indonesia, Japan, Korea, Malaysia, Singapore, and Thailand. Traditional macroeconomic data often suffer from reporting delays and limited granularity. In contrast, search engine data provide a timely and novel approach to capturing public interest and sentiment related to economic conditions. Four model specifications were estimated using both Ordinary Least Squares (OLS) and several MIDAS regression techniques. These models incorporated macroeconomic variables, Google Trends data, or a combination of both. Across all countries, MIDAS regression methods outperformed OLS, underscoring the advantages of mixed-frequency modelling for tracking dynamic economic phenomena.

Key findings indicate that Model 2 – based solely on Google Trends indices – achieved the highest predictive accuracy in Indonesia, Japan, Korea, Singapore, and Thailand. In Malaysia, Model 4, which combines macroeconomic indicators with a composite Google index, performed best; however, Model 2 was a close second. This highlights the strong predictive value of search behaviour data even in countries with moderate digital infrastructure. Furthermore, individual Google search indices proved more informative than aggregated indices, reinforcing the usefulness of specific search terms for economic modelling.

These findings carry important policy implications. The integration of real-time digital behavioural data with traditional macroeconomic indicators can greatly enhance the responsiveness and precision of cost-of-living monitoring systems. Policymakers, particularly in developing and digitally engaged economies, should consider incorporating search engine data to complement official statistics. This can support more timely and targeted interventions during inflationary periods or economic shocks. Researchers could also explore the use of other digital data sources, such as social media trends or e-commerce data, and apply alternative modelling approaches, including machine learning techniques. Additionally, investigating the causal mechanisms behind search behaviour and its relationship with economic indicators would deepen the understanding of how digital activity reflects real-world conditions.

Although Google Trends is a powerful tool for data search behaviour, it also has some downsides. Issues of reliability, particularly for low-frequency search terms, necessitate careful validation and methodological refinement. Researchers should pay caution in utilizing and interpreting the information. In the era of data analytics and big data, data is valuable. Google Trends is a very useful tool if it is applied with careful attention to methodological considerations and combined with complementary datasets.

While Google Trends is highly utilized to make data-driven decisions, including anticipating shifts in consumer demand/market trends, optimizing marketing strategies, and content planning, its applications in policy decision and planning are lacking behind. The policymakers, governmental organizations and research institutions could apply Google Trends to monitor public interest/reactions, nowcast economic trends, and observe real-time and fast-moving trends to provide better forecasts on economic scenarios. This tool is also useful to detect the geological disasters (the outbreak of diseases, floods condition, etc.) and provide early warning on the change of symptoms or hotspot detection. In terms of decision-making, the real-time data is feasible in making forecasts, revealing fast-moving trends to gauge public concerns and responses, detecting outbreaks of diseases, so that immediate action can be taken in emergency situations. It also provides additional information in policy-decision, cost-effective planning and development.

In conclusion, this study provides compelling evidence that Google Trends data can enhance the modelling and forecasting of cost-of-living dynamics. The successful integration of real-time behavioural data with traditional economic indicators offers valuable tools for researchers and policymakers to improve economic analysis, surveillance, and policy response in an increasingly digitalised world.

DECLARATION OF GENERATIVE AI IN THE WRITING

We did not apply any AI tool in the writing process.

References

BLANCIFORTI, L. A., KRANNER, E. (1997). *Estimating county cost of living indexes: The issue of urban versus rural.* [online]. Regional Research Institute Working Paper No. 9718, National Institute for Occupational Safety and Health, Morgantown. [cit. 20.8.2025]. <https://researchrepository.wvu.edu/rri_pubs/186/>.

BOSIRE, W. M., OMWENGA, J. Q. (2022). Determinants of cost of living of the households in Kajiado North Sub-County. *International Journal of Social Science and Humanities Research*, 10(4): 148–167. <https://doi.org/10.5281/zenodo.7183190>.

BRIDEL, A., LONTOH, L. (2014). *Lessons Learned: Malaysia's 2013 Fuel Subsidy Reform* [online]. Geneva: International Institute for Sustainable Development. [cit. 20.8.2025].<https://www.iisd.org/sites/default/files/gsi/ffs_malaysia_lessonslearned.pdf>

CAVALLO, A. (2013). Online and official price indexes: Measuring Argentina's inflation. *Journal of Monetary Economics*, 60(2): 152–165. < https://doi.org/10.1016/j.jmoneco.2012.10.002>.

CEBULA, R. J. (1980). Determinants of geographic living-cost differentials in the United States: An empirical note. *Land Economics*, 56(4): 477–481. < https://doi.org/10.2307/3146225>.

CEBULA, R. J., TODD, S. (2004). An empirical note on determinants of geographic living-cost differentials for counties in the state of Florida, 2003. *The Review of Regional Studies*, 34(1): 112–119. <https://doi.org/10.52324/001c.8372.>.

CEBULA, R. J., TOMA, M. (2008). An empirical analysis of determinants of interstate living-cost differentials, 2005. *Journal of Regional Analysis and Policy*, 38(3): 1–7. <https://doi.org/ 10.22004/ag.econ.133000>.

CHAI, Y., ABDULLAH YUSOF, S., MOHD AMIN, R., MOHD ARSHAD, M. N. (2020). Household Debt and Household Spending Behaviour: Evidence from Malaysia. *Jurnal Ekonomi Malaysia*, 54(1): 111–120. <http://dx.doi.org/10.17576/JEM-2020-5401-8>.

CHOI, H., VARIAN, H. (2012). Predicting the present with Google Trends. *Economic Record*, 88(S1): 2–9. <https://doi.org/10.1111/j.1475-4932.2012.00809.x>

CLISSITT, C. (2023). *Cost of living around the world in 2023* [online]. London: MoveHub. [cit. 20.8.2025]. <https://www.movehub.com/advice/cost-of-living-around-the-world/>.

D'AMURI, F., MARCUCCI, J. (2017). The predictive power of Google searches in forecasting US unemployment. *International Journal of Forecasting*, 33(4): 801–816. <https://doi.org/10.1016/j.ijforecast.2017.03.004>.

DEPARTMENT OF STATISTICS MALAYSIA (DOSM). (2024). *Gross domestic product and unemployment rate statistics* [online]. Putrajaya: Department of Statistics Malaysia. [cit. 20.8.2025]. <https://open.dosm.gov.my/dashboard>.

DIEWERT, W. E. (1983). The theory of the cost-of-living index and the measurement of welfare change. In: DIEWERT, W. E., MONTMARQUETTE, C (eds.) *Price Level Measurement*. Ottawa: Statistics Canada.

FEDERAL RESERVE BANK OF ST. LOUIS. (2024). *FRED economic data* [online]. St. Louis: Federal Reserve Bank of St. Louis. [cit. 20.8.2025]. < https://fred.stlouisfed.org/>.

FOOD AND AGRICULTURE ORGANIZATION (FAO). (2021). FAO Statistical Yearbook 2021 [online]. Rome: FAO. [cit. 20.8.2025]. <https://openknowledge.fao.org/server/api/core/bitstreams/522c9fe3-0fe2-47ea-8aac-f85bb6507776/content>.

FRISCH, R. (1936). The problem of index numbers. *Econometrica*, 4: 1–38. <https://doi.org/10.2307/1907119>.

GUMMER, T., OEHRLEIN, A.-S. (2024). Using Google Trends data to study high-frequency search terms: Evidence for a reliability-frequency continuum. *Social Science Computer Review*, 43(4): 814–826. <https://doi.org/10.1177/08944393241279421>.

HAWORTH, C. T., RASMUSSEN, D. W. (1973). Determinants of metropolitan cost of living variations. *Southern Economic Journal*, 40(2): 183–192. <https://doi.org/10.2307/1056869>.

HOBDARI, M. N. A., UPPAL, A. (2017). *Food Inflation in Sub-Saharan Africa: Causes and Policy Implications* [online]. International Monetary Fund. [cit. 20.8.2025]. <https://www.imf.org/external/pubs/ft/wp/2016/wp16247.pdf>.

HÖLZL, J., KEUSCH, F., SAJONS, C. (2025). The (mis)use of Google Trends data in the social sciences - A systematic review, critique, and recommendations. *Social science research,* 126, 103099. <https://doi.org/10.1016/j.ssresearch.2024.103099>.

INTERNATIONAL MONETARY FUND (IMF). (2015). *Malaysia: 2015 Article IV consultation – staff report*. Washington, D.C.: International Monetary Fund. [cit. 20.8.2025]. <https://www.elibrary.imf.org/view/journals/002/2015/058/002.2015.issue-058-en.xml>.

INTERNATIONAL LABOUR ORGANIZATION (ILO). (2022). *Understanding Cost of Living Indicators* [online]. Geneva: ILO. [cit. 20.8.2025]. <https://share.google/gJ9Pf8r2KLtX72KNd>.

JENKINS, S. P. (2000). Modelling household income dynamics. *Journal of Population Economics*, 13(4): 529–567. <https://doi.org/10.1007/s001480050151>.

JUN, S.-P., YOO, H. S., CHOI, S. (2018). Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. *Technological Forecasting and Social Change*, 130: 69–87. <https://doi.org/10.1016/j.techfore.2017.11.009>.

KEAT, P. G., YOUNG, P. K. Y. (2009). *Managerial Economics: Economic Tools for Today's Decision Makers*. Pearson Education.

KONÜS, A. A. (1939). The problem of the true index of the cost of living. *Econometrica*, 7(1): 10–29. <https://doi.org/10.2307/1906997>.

KURRE, J. A. (2000). *Differences in the Cost of Living Across Pennsylvania's 67 Counties*. Harrisburg: The Center for Rural Pennsylvania. [cit. 20.8.2025]. <https://www.rural.pa.gov>.

LAFRANCE, R., SCHEMBRI, L. L. (2000). *The exchange rate, productivity, and the standard of living*. [online]. Bank of Canada Review. [cit. 20.8.2025]. <https://www.bankofcanada.ca/wp-content/uploads/2010/06/r001b-e.pdf>.

LATIMAHA, R., BAHARI, Z., ISMAIL, N. A. (2018). A descriptive analysis on the basic needs budget by the middle income earners to identify the most expensive city to live: An analysis involving capital cities in Malaysia. *Malaysian Journal of Business and Economics*, 5(2): 1–18. <https://doi.org/10.51200/mjbe.v0i0.1606>.

LATIMAHA, R., BAHARI, Z., ISMAIL, N. A. (2019). Examining the linkages between street crime and selected state economic variables in Malaysia: A panel data analysis. *Malaysian Journal of Economics*, 53(1): 59–72. <http://dx.doi.org/10.17576/JEM-2019-5301-6>.

MAINAL, S. A., ABU HANIPAH, F. A., ZUHAIMI, N. L., KAMARUDIN, N. N. S. (2024). The Impact of Taxation on the Cost of Living: A Comprehensive Analysis. *Information Management and Business Review*, 16(3(I)S), 628–643. <https://doi.org/10.22610/imbr.v16i3(I)S.4089N>.

NUMBEO. (2024). *Cost of Living Index Database* [online]. Belgrade: Numbeo. [cit. 20.8.2025]. <https://www.numbeo.com/cost-of-living/>.

NAJAF, R., NAJAF, K. (2016). Impact of oil prices on the stock exchange of Pakistan and Malaysia. *International Journal of Research* -GRANTHAALAYAH, 4(8): 84-91. <https://doi.org/10.5281/zenodo.61161>.

NUMBEO. (2026). *Methodology and motivation*. [online]. Belgrade: Numbeo. [cit. 20.8.2025]. <https://www.numbeo.com/common/motivation_and_methodology.jsp>.

NIESERT, R. F., OORSCHOT, J. A., VELDHUISEN, C. P., BRONS, K., LANGE, R. J. (2020). Can Google search data help predict macroeconomic series? *International Journal of Forecasting*, 36(3): 1163–1172. <https://doi.org/10.1016/j.ijforecast.2018.12.006.>.

POLASH, A. (2026). *How to use Google Trends for research and stay ahead of your competitors*. [online]. Florida: Hosting.com. [cit. 20.8.2025]. <https://hosting.com/blog/how-to-use-google-trends-for-research/>.

SINGAPORE DEPARTMENT OF STATISTICS (SINGSTAT). (2024). *Unemployment Rate Statistics* [online]. Singapore: Singapore Department of Statistics. [cit. 20.8.2025]. < https://www.singstat.gov.sg/>.

SQUALLI, J., WILSON, K. (2006). A new approach to measuring trade openness. Economic and Policy Research Unit Working Paper 06–07, Zayed University, Dubai, UAE. <https://www.zu.ac.ae/epru/documents/06-07-web.pdf>.

TRIPLETT, J. E. (1992). Economic theory and BEA's alternative quantity and price indexes. *Survey of Current Business*, 72: 49–52.

UNITED STATES CONGRESS. (1996). Toward a more accurate measure of the cost of living: final report to the Senate Finance Committee. Dec. 4, 1996 updated version. [Washington, D.C.]: Advisory Commission to Study the Consumer Price Index. <https://catalog.hathitrust.org/Record/003239902>.

VOSEN, S., SCHMIDT, T. (2011). Forecasting private consumption: Survey-based indicators vs. Google trends. *Journal of Forecasting*, 30(6): 565–578. <https://doi.org/10.1002/for.1213>.

ZUHAIRI, A. H., YAKUB, F., HARUM, A. N., OMAR, M., SHARIFUDDIN, M., FARUQ, A., SINHA, V., RAZAK, K. A., & ABDULLAH, S. (2025). Comparative analysis of missing data imputation methods for flood features from Langat river in Selangor, Malaysia. *Malaysian Journal of Computer Science*, 38(3): 224–237. <https://doi.org/10.22452/mjcs.vol38no3.2>.

Composite Index of Health Indicators and Typology of the Mexican States Using and Extending the Performance Interval Approach

Jesus A. Trevino-Cantu[1] | *Universidad Autonoma de Nuevo Leon, Mexico*

Received 27.10.2025 (revision received 23.2.2026), **Accepted** (reviewed) 18.3.2026, **Published** 11.9.2026

Abstract

This study employs a performance-interval approach to assess health performance across Mexico's 32 states, addressing limitations of traditional single-value indices. Using OECD-selected indicators covering resources, health status, and risks, the methodology calculates a midpoint (MP) for average performance and an amplitude (A) for the balance of internal indicators. Extending this approach, the research develops a typology classifying states based on MP and A values to guide policy. Results reveal significant regional inequalities and institutional fragmentation; for instance, Mexico City exhibits high resource density yet high mortality rates due to its fragmented institutions. The analysis identifies four state groups, ranging from robust systems to those facing severe risks and uncertainty. By capturing internal disparities often obscured by single-value indices, the extended approach incorporates health system fragmentation into a composite index and a typology, as identified analytically in the current literature. The findings suggest that effective policy requires addressing resource scarcity and institutional coordination to reduce health disparities and improve outcomes across regions.

Keywords	***DOI***	***JEL code***
Composite index; Performance interval; Health indicators; Health fragmentation; Mexico.	*https://doi.org/10.54694/stat.2025.59*	*I10, R11, C43, I18*

INTRODUCTION

Health is a key aspect of human development and a universal right, reflecting a population's social, economic, and territorial conditions. In recent decades, population health has become strategically important in development agendas at both the international and national levels. In Mexico, access to quality

[1] Urban Studies Graduate Program, School of Architecture, Universidad Autonoma de Nuevo Leon, San Nicolas de Los Garza, Nuevo Leon, 66455, Mexico. Email: jtrevino41@hotmail.com. ORCID: <https://orcid.org/0000-0002-0211-2140>.

services, along with the availability of human resources and infrastructure, is essential for understanding the health situation across the country's states (OECD, 2023). However, these conditions are not evenly distributed geographically (Bhattacharya, Hyde, Tu, 2014): health indicators in Mexico show significant inequalities between states, driven by structural factors, unequal resources, and varying risk dynamics across territories (Gómez-Dantés et al., 2024).

The availability of various tools and methods for measuring health enables comparisons and a better understanding of regional health system performance. The OECD health analytical framework categorizes indicators into health status, risk factors, and healthcare resources. This classification into meaningful groups provides a comprehensive, comparable, and policy-relevant basis for assessing health system performance and population health in OECD countries and their subregions (Joumard, André, Nico, 2008; OECD, 2023; Tolonen et al., 2021). The collected data enable consistent benchmarking across countries through uniform definitions and methods, enhancing comparability and facilitating tracking over time (Özen and Özen, 2024). The indicators cover multiple dimensions, including life expectancy, avoidable mortality, smoking, obesity, and health infrastructure, offering a holistic view of health system performance (OECD, 2023). This analytical foundation supports actionable, policy-relevant evaluations tailored to various health system needs while maintaining standardization (OECD, 2023). It encourages effective communication among policymakers by establishing a common language and promoting knowledge sharing across sectors (Özen and Özen, 2024). The indicators align with health determinants and cross-sectoral policies, fostering coherence in public health efforts (Tolonen et al., 2021). Well-organized methodological data ensures reliability and support analysis at both local and national levels (Özen and Özen, 2024; Braithwaite et al., 2017). Their adoption by many countries promotes collaboration and sets standards for health system evaluation (Özen and Özen, 2024). In summary, the OECD analytical framework effectively balances comprehensiveness, data quality, and global relevance in assessing state-level health performance (OECD, 2023; Tolonen et al., 2021).

Despite progress in recent decades, Mexico still faces growing fragmentation within its regional health system. Inequities in service access, medical infrastructure, health outcomes, and exposure to socio-environmental risks create a complex, uneven, and multifaceted landscape (Gómez-Dantés et al., 2023; Meneses-Navarro et al., 2022). This work provides a multidimensional analysis of health across Mexico's 32 states. Composite indicators of state-level health performance offer a concise summary that helps identify issues and guide public policy. Traditional assessments that rely solely on single averages or indicators often obscure differences between states, making it harder to develop effective policies and prioritize resources. The lack of tools that combine resources, status, and risk with variability criteria limits the effectiveness of public health strategies. The methodological approach shows that the performance interval index (Mazziotta and Pareto, 2020) is a practical option for capturing variability within each state's health system.

The research presents an analytical and methodological approach to developing a composite indicator of state-level health performance in Mexico. The study examines key concepts, the rationale for the issue, and its social importance. It also establishes a systematic, methodological process and a logical framework for combining indicators into an interval index with three levels of compensation: full compensation, partial compensation, and full non-compensation. The aim is to enrich academic literature and inform public policies to reduce health disparities and improve social well-being.

Accurate measurement of state-level health performance is essential for developing effective and equitable public policies. Mexico's health issues, including chronic diseases and resource disparities, require innovative methodological approaches to capture their complexity. The performance interval-based method for composite indices not only identifies states with low and high performance but also pinpoints those with wide or narrow internal variations in their indicators (Mazziotta and Pareto, 2020; Drago and Gatto, 2023). This information is essential for designing strategies tailored to specific regions.

Furthermore, this approach enables the comparison and visualization of regional patterns that are often overlooked by traditional techniques, thereby supporting the creation of regionally focused and fair public policies (Drago, 2021).

Objectives. To analyze the performance and hierarchy of a composite health index for Mexico's states and to recommend a policy framework. Achieving this involves five specific tasks:

1. Use the OECD conceptual framework of resources, status, and risk to select key health indicators that answer the question: What are representative indicators of the health system, as defined in the database?
2. Develop a composite health index for Mexican states using an adapted version of the performance interval method by Mazziotta and Pareto (2020), referred to as the original version in this study. This task aims to determine: What is the performance and ranking of each state based on its interval midpoint (MP), and how balanced the indicators are in each case based on the interval amplitude (A)?
3. Identify the key indicators in the MP and the geometric mean and verify the relationship between these two partially compensatory indices. This activity addresses the question: How can we confirm that MP is a partially compensatory index, and which indicators are most important?
4. Propose a regional classification of state-level health performance based on the MP and A of the states, using an arithmetic interval (interval average). The key question is: How can the MP and A be combined to create a typology that informs the design of policy frameworks?
5. Recommend tailored courses of action based on the previous performance typology. The relevant question is: Which tasks and actions are appropriate given the performance typology?

The main purpose of these general and specific objectives is to guide the narrative in this document rather than address each question individually in the case study.

1 CONCEPTUAL APPROACHES

1.1 Dimensions and indicators of health

The concept of health is multidimensional, encompassing structural, individual, and territorial factors and determinants. According to the WHO (2024), health is not merely the absence of disease but also includes physical, mental, and social well-being. Building on this idea, the OECD (2023) selects, compiles, and organizes indicators into three analytical dimensions related to health system performance.

a. *Resources* (inputs). These indicators reflect the capacity of health systems, including infrastructure, personnel, and equipment. Resources represent the state's operational capacity and define the system's potential to respond to health needs. The database provides the following resource indicators (per 1 000 inhabitants): Active physicians, Nurse practitioners, and Beds. The OECD (2023) also includes hospital discharges in this dimension. This indicator is excluded from this analysis because it does not directly measure health resources. High or low rates of hospital discharges do not necessarily indicate better health care resources; they should be interpreted alongside results and system efficiency.

b. *Status* (results) are indicators that measure the overall outcome of the system in terms of population health. Health status reflects the population's health condition and shows accumulated achievements or deficits in longevity, mortality, and morbidity. The most representative indicators of a population's health status are life expectancy and mortality (Bhattacharya, Hyde, Tu, 2014). The database includes life expectancy at birth, crude mortality rate, age-adjusted mortality rate, and infant mortality rate.

c. *Risk* (determinants). These indicators represent individual or societal factors that increase the likelihood of illness or early death. They are social and behavioral determinants of health. Risks serve as probabilistic predictors of changes in health status, especially when combined. The database includes the Obesity rate, Population exposure to fine particulate matter (PM2.5), Mortality rate from respiratory diseases, and Mortality rate from circulatory diseases.

The operational nature of the OECD approach is flexible enough to accommodate various theoretical perspectives, but it does not fully align with all of them. For example, Sen's (1985) capabilities methodological procedure and OECD health resources (physicians, nurses, and beds) agree that a well-funded health system is essential not only for preventing disease but also for enabling people to develop their abilities and lead fulfilling lives.

The OECD approach also connects to Marmot's social determinants of health framework (Marmot, 2015; Marmot et al., 2020). Both the OECD and Marmot aim to incorporate social and structural factors into health analysis. However, they differ in focus and scope: the OECD emphasizes measuring resources, status, and risks for comparison and policy development, whereas Marmot promotes an equity and social justice agenda focused on changing the structural conditions that create the social gradient in health.

1.2 Recent literature on the case study and the advantages of the performance interval

Mexico's health system exhibits profound subnational heterogeneity, necessitating analytical frameworks that capture disparities and guide social policy. Composite indices are promising tools for empirical assessment, yet methodological limitations constrain their diagnostic utility in addressing regional inequities. Gómez-Dantés et al. (2016) used Global Burden of Disease data to analyze the geographic pattern of cancer. They found a discordant health transition across Mexican states from 1990 to 2013, thereby demonstrating geographic disparities and producing regional rankings showing southern states lagging their northern counterparts in mortality and disease burden.

Empirical measurements of health systems are vital for identifying specific areas that require intervention. However, these measurements also present additional complexities that demand careful interpretation. Urquieta-Salomon and Villarreal (2016) documented the evolution of health coverage, showing progress while highlighting structural challenges in the Mexican health system.

Socioeconomic factors influence performance metrics, thereby affecting the overall assessment of the system's effectiveness. Smith and Goldman (2007) identified socioeconomic disparities in health among older adults. Barraza-Llorens et al. (2002) examined income-related inequalities from 2000 to 2006 and found that health care utilization varied significantly by socioeconomic status.

Together, these studies show that although coverage has expanded, equity remains elusive. The current literature indicates that composite indices fail to capture the social segmentation and institutional fragmentation of Mexico's health system (Meneses-Navarro, 2025). Segmentation refers to the classification of the population, primarily by formal or informal labor market status, whereas fragmentation refers to the diversity of institutions serving the segmented population (e.g., IMSS, ISSSTE, SEDENA, MARINA, PEMEX, health institutions associated with unions, such as universities) (Becerril-Montekio et al., 2024). The combination of segmentation and fragmentation creates a "sanitary apartheid" characterized not only by the divide between those who have and those who lack access to the health system but also by disparities in health resources and in the quality of medical care among those who already have health services (Knaul et al., 2012).

Critical methodological gaps persist in the composite index literature: single-value indices blend uniformly low performance with highly variable performance, leading to mismatched interventions; compensatory aggregation obscures indicator substitution, thereby masking fragmentation; and existing typologies stratify states by average scores alone, ignoring organizational styles (institutional fragmentation across different sectors of the population). The performance interval approach addresses these gaps by quantifying and combining midpoint and amplitude to suggest a regional health typology. This two-dimensional diagnosis enables precision in public policy and prevents the uniform allocation of resources across and within states for indicators.

2 METHODOLOGY

The current study uses health indicators from the OECD-Explorer (Health Statistics – Regions) for Mexico's 32 federal entities. The database groups these indicators into three dimensions: resources, status, and risk. The methodology combines these indicators using the performance interval approach in five steps. A final sixth step extends the interval approach to developing a typology that supports the creation of territorial strategies for decision-making and policy development.

a. *Normalization*. This research employs vector normalization, the most stable among the normalization options used in decision-making methods, such as TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) (Bouhedja, Bouhedja, Benselhoub, 2024), which aligns with the typology described in this methodology:

$$z = \frac{x_{ij}}{\sqrt{\sum_{j=1}^{m} x_{ij}^2}}, \qquad \text{Eq. (1)}$$

Where x_{ij} is the value of indicator j in observation i in the input data matrix; the sum of squares is calculated for each indicator (per column). This normalization does not change the shape of the distribution of the original values (i.e., the same skewness and kurtosis).

b. *Negative polarity*. Indicators with an inverse relationship (negative correlation) with the composite index have ~~a~~ negative polarity, meaning they need to be inverted to align with it. Negative polarity can be reversed by applying any standardization or normalization procedure for negative indicators, linear operations (subtracting from the maximum value), or non-linear procedures (reciprocal transformation), as detailed in Alaimo (2023). In this research, indicators with negative polarity are reversed following Mulliner, Malys, Maliene (2016):

$$x_i' = max + min - x_i. \qquad \text{Eq. (2)}$$

c. *Weighting*. There are objective and subjective weightings. Objective weighting is based on data-driven criteria such as variability, statistical properties, or optimization procedures. Subjective weighting involves value judgments or preferences by experts, stakeholders, or index designers. This research assigns equal weights based on the number of indicators in each dimension, as explained in the case study.

d. *Aggregation*. All indicators are synthesized using the performance interval approach (Mazziotta and Pareto, 2020). This method builds a composite index that considers both the midpoint and the range of each interval, offering a synthetic and reliable view of the health situation by state. This approach identifies for each territorial unit the minimum (LB) and maximum (UB) of the normalized and weighted values that define the performance interval. In the case study, the UB (upper bound) of the interval is the arithmetic mean of the normalized, equally weighted values for each row (Mexican state).

The amplitude of the interval (A) is the difference between the maximum and minimum values of the indicators in each observation:

$$A = UB - LB \qquad \text{Eq. (3)}$$

And the midpoint (MP) of this interval is:

$$MP = (LB + UB)/2 \qquad \text{Eq. (4)}$$

The midpoint (MP) of this interval is a composite index that considers both the average value and its relative stability (variability). The MP is a partially compensatory index, like other single-value indices (e.g., geometric mean), that determines the hierarchy of performance in each state. The interval approach simultaneously accounts for both the midpoint performance value and the internal variation of each state, offering a more detailed view than a traditional index based solely on averages.

e. *Importance of normalized indicators, Z_i*. It is assessed through an iterative process that involves repeatedly calculating the MP by removing one indicator at a time and adjusting the weights w_i for the remaining indicators. This method is based on the idea that the indicator whose removal results in a lower correlation is the most important, as its absence has a greater effect on the observation hierarchy. This same simulation process is applied to the weighted geometric mean for normalized data. The study uses the multiplicative version, denoted as GM_w^{Π} (there is also an additive version that produces the same result):

$$GM_w^{\Pi} = \left(\prod_{i=1}^{j=7} z_i^{w_i} \right)^{\frac{1}{\sum_{i=1}^{j=7} w_i}} \qquad \text{Eq. (5)}$$

The number seven in Eq. (5) represents the count of indicators in this research. When the sum of the weights equals one, the formula simplifies to multiplication inside parentheses:

$$GM_w^{\Pi} = \prod_{i=1}^{j=7} z_i^{w_i} = z_1^{w_1} * z_2^{w_2} * z_3^{w_3} \dots z_7^{w_7} \qquad \text{Eq. (6)}$$

f. *Typology*. The combination of MP and A creates a typology of health performance for the states. The key to developing this typology is to calculate an interval arithmetic (interval average) using the mean of all minimums (LB) on one side and the mean of all maximums (UB) on the other, considering all individual intervals (Drago and Gatto, 2023; Drago, 2021). The significant points of the interval average are the midpoint ($\overline{MP}$) and the amplitude (Ā):

$\overline{MP} = \left(\overline{LB} + \overline{UB} \right) / 2$; and $\bar{A} = \overline{UB} - \overline{LB}$. These values allow the states to be classified into four groups:

a. High performance, low variability (MP > $\overline{MP}$ & A < Ā): States with effective healthcare systems and stable indicators. Policy: Continue current measures and enhance innovation.
b. High performance, high variability (MP > $\overline{MP}$ & A ≥ Ā): Successful states with internal imbalance in indicators. Policy: focus interventions in areas with unbalanced indicators.
c. Low performance, low variability (MP ≤ $\overline{MP}$ & A < Ā): States with poor results but internally balanced indicators. Policy: restructure general policy and increase resources.
d. Low performance, high variability (MP ≤ $\overline{MP}$ & A ≥ Ā): States with low performance and unbalanced indicators. Policy: implement institutional reform and emphasize territorial equity with balanced indicators.

This typology emphasizes the unique challenges of each group of states, allowing for the creation of policies tailored to each one. The approach supports the principles of territorial equity and institutional efficiency, which are vital for reaching health goals across various geographical settings (Kruk et al., 2015).

MP and A can be divided into any number of strata to develop more detailed typologies. Different stratification methods can be used for this purpose, such as natural breaks or intervals around the mean or median. However, as the number of categories or types increases, interpretation becomes more complicated.

3 HEALTH INDICATORS IN THE MEXICAN STATES

Data are represented by variables that, when combined, form indicators. Health indicators in the OECD are categorized into three dimensions, as mentioned earlier: resources, status, and risk (Table 1). The number of indicators varies by dimension and may sometimes include missing data. These characteristics lead to the following decisions:

a. Indicators with missing information for specific states were retrieved from alternative sources instead of being omitted. For example, obesity data for Mexico does not appear in the OECD database. The data show the percentage of the population with overweight and obesity by state in 2018, available on the Mexican website for the 2030 Agenda.
b. The Hospital discharges indicator has been removed from the analysis because it does not effectively measure resources.
c. The data for the indicators are averages from recent years to avoid short-term effects and to update lagging data. The exceptions are the PM2.5 and Obesity indicators, as they are available only for 2020 and 2018, respectively. The remaining indicators are averages for the periods 2020-2022, 2023, or 2024.
d. Indicators with a coefficient of variation less than 10% are excluded from the study because they lack sufficient discriminatory power between states. This includes the indicators Infant mortality and Life expectancy in the Status dimension, and Obesity in the Risk dimension.

Once the previous operational decisions were made, the input matrix was ready to be used in the performance interval procedure. The final matrix has seven indicators grouped into three dimensions, covering the 32 states of Mexico (Table 1 and Table 2).

Table 1 Health indicators, by dimension, polarity, and description.

Dimensions (polarity)	Abbreviation and description
Resources	
Physicians (+). Per 1 000 habts.	*Physicians.* General practitioners who assume responsibility for providing ongoing care to individuals and families, or specialists such as pediatricians, obstetricians-gynecologists, psychiatrists, medical specialists, and specialist surgeons.
Nurses (+). Per 1 000 habts.	*Nurses.* They provide direct patient care services, including professional nurses, nurse practitioners, and foreign nurses licensed to practice and actively practicing in the country. Midwives are excluded, unless they work primarily as nurses.
Beds (+). Per 1 000 habts.	*Beds*. They are hospital beds with regular maintenance and staffing, immediately available for the care of admitted patients. They include beds in all hospitals, including general hospitals, mental health and substance abuse hospitals, and other specialty hospitals; occupied and unoccupied beds.
Status	
Crude mortality rate. (-). Per 1 000 habts.	*MortR*. It is the ratio of deaths during the year to the average population of that year.
Risk	
Population exposure to fine particles PM2.5	*PM2.5*. Population exposure to PM2.5
Respiratory system diseases mortality rate (-)*	*RespiraM*. Number of deaths from diseases of the respiratory system, for 100 000 habts.
Circulatory system diseases mortality rate (-)*	*CirculaM*. Number of deaths from diseases of the circulatory system, for 100 000 habts.

Source: Based on OECD data explorer (https://data-explorer.oecd.org/).
* The OECD, in its country-level analysis, examines circulatory and respiratory diseases under Health status (OECD Health at a Glance, 2023, Chapter 3). However, in its database, both indicators are reported in the health risk classification at the level of regions (Data Explorer-Health statistics - Regions). Since the main data source in this research is at the regional level (TL2), the classification of health indicators is the same as that used in the OECD Data Explorer.

Table 2 Selected health indicators. Input data (decision matrix)

States	Resources			Status	Risk		
	Physicians	Nurses	Beds	MortR	RespiraM	CirculaM	PM2.5
Aguascalientes (Ags)	1.35	1.8	1.0	5.97	57.4	125.27	13.8
Baja California (BC)	0.90	1.2	0.9	7.08	57.4	163.13	16.4
Baja California Sur (BCS)	1.78	1.4	1.1	5.87	46.6	133.88	10.9
Campeche (Cam)	1.50	1.6	1.1	7.04	52.8	177.69	15.6
Coahuila (Coa)	1.28	1.4	1.2	7.28	55.8	190.69	13.5
Colima (Col)	1.70	1.7	1.2	8.27	65.5	171.58	12.3
Chiapas (Chis)	0.83	1.3	0.6	6.30	49.0	166.80	18.0
Chihuahua (Chih)	1.20	1.8	1.1	8.45	70.7	213.85	13.2
Ciudad de Mexico (CdMx)	2.45	3.2	2.3	9.28	76.9	234.64	17.2
Durango (Dgo)	1.18	1.3	1.0	6.71	60.7	209.09	13.1
Guanajuato (Gto)	1.00	1.5	1.0	7.66	57.3	203.31	14.3
Guerrero (Gro)	1.15	1.4	0.8	6.62	42.6	158.77	13.1
Hidalgo (Hgo)	1.03	1.2	0.8	6.71	45.1	187.43	16.0
Jalisco (Jal)	1.03	1.5	1.2	7.37	79.0	179.28	12.9
Mexico (Mex)	0.88	1.1	0.8	7.32	58.6	169.84	17.3
Michoacan (Mich)	1.00	1.1	1.0	7.66	60.3	192.25	13.2
Morelos (Mor)	1.15	1.2	0.8	9.04	65.5	224.37	14.8
Nayarit (Nay)	1.35	1.2	0.7	6.79	62.0	162.14	11.5
Nuevo Leon (NL)	0.95	1.6	1.1	6.84	64.8	189.81	14.1
Oaxaca (Oax)	0.88	1.0	0.7	7.90	51.3	213.59	14.6
Puebla (Pue)	0.83	1.0	0.8	7.82	61.9	220.14	15.7
Queretaro (Qro)	0.95	1.1	0.8	6.09	48.8	145.65	15.3
Quintana Roo (QRoo)	1.03	1.4	0.7	5.26	36.1	94.67	13.8
San Luis Potosi (SLP)	0.80	1.2	1.0	7.78	68.9	217.33	14.8
Sinaloa (Sin)	1.35	1.5	1.0	6.73	64.8	180.23	11.9
Sonora (Son)	1.48	1.5	1.2	7.97	75.6	211.23	13.2
Tabasco (Tab)	1.48	1.6	0.9	7.31	51.4	195.83	17.4
Tamaulipas (Tam)	1.23	1.8	1.3	6.91	56.6	188.82	14.2
Tlaxcala (Tla)	1.15	1.2	0.7	7.44	89.2	185.92	16.2
Veracruz (Ver)	1.00	1.1	0.8	8.51	61.4	250.32	15.7
Yucatan (Yuc)	1.25	1.6	1.1	7.15	66.5	211.21	14.1
Zacatecas (Zac)	1.33	1.6	0.9	8.03	71.3	192.13	13.0
Mean	1.200	1.439	0.986	7.285	60.371	186.277	14.409
Stdev	0.339	0.403	0.295	0.907	11.341	32.713	1.787
CV	0.283	0.280	0.299	0.125	0.188	0.176	0.124

Source: OECD-explorer. Data averaged for the recent period (2020, onwards)—indicator abbreviations in Table 1.

The performance interval approach differs from studies on value intervals in the input matrix because it produces both compensatory and non-compensatory measurements from a single continuous-value matrix. It is not tied to indices based on input ranges, underscoring the difficulty of accurately representing aspects such as individual preferences through value ranges (Jahanshahloo, Lotfi, Izadikhah, 2006).

The original performance interval procedure is not applied in its original form; instead, it is applied with four variations and two extensions, indicated when appropriate. Variations are common because the original version's main purpose is to develop composite indices using the performance interval method. It is important, however, to mention them to avoid claims of distorting or falsifying the original procedure. Extensions are needed for a more effective graphical representation of the interval method and to develop the typology proposed in this research. Among the variations and extensions, the study incorporates observations from the original approach into the graphical representation of the interval. The study lists all steps, variations, observations, and extensions in order. The first four steps involve adaptations or variations:

1. *Vector normalization instead of rescaled standardization in the original version* (Table 2). This step is essential to make indicators expressed in different units of measurement comparable. There are several standardization procedures, with varying effects on the composite index (OECD, 2008; Talukder, Hipel, Vanloon, 2017). The original version uses the z standardization multiplied by 10, and ± 100. The version applied in this research employs vector normalization, as explained in the methodology.

2. *Revert indicators with negative polarity* so that higher values indicate better performance (see Table 3). The methodology details this procedure.

Table 3 Normalized matrix with reversed negative indicators

States	Physicians	Nurses	Beds	Polarity reversal= (max + min) - xi			
				MortR	RespiraM	CirculaM	PM2.5
Ags	0.1916	0.2093	0.1719	0.2064	0.1953	0.2055	0.1839
BC	0.1277	0.1382	0.1605	0.1794	0.1954	0.1701	0.1522
BCS	0.2519	0.1658	0.1834	0.2087	0.2265	0.1974	0.2192
Cam	0.2129	0.1935	0.1834	0.1806	0.2086	0.1565	0.1620
Coa	0.1810	0.1619	0.2063	0.1747	0.1999	0.1443	0.1875
Col	0.2413	0.2014	0.2063	0.1508	0.1722	0.1622	0.2022
Chis	0.1171	0.1501	0.1032	0.1982	0.2197	0.1666	0.1327
Chih	0.1703	0.2172	0.1948	0.1464	0.1571	0.1226	0.1912
CdMx	0.3477	0.3791	0.3897	0.1266	0.1393	0.1032	0.1425
Dgo	0.1668	0.1580	0.1719	0.1885	0.1858	0.1271	0.1924
Gto	0.1419	0.1777	0.1719	0.1655	0.1956	0.1325	0.1778
Gro	0.1632	0.1658	0.1375	0.1906	0.2381	0.1741	0.1924
Hgo	0.1455	0.1422	0.1375	0.1885	0.2310	0.1473	0.1571
Jal	0.1455	0.1737	0.2063	0.1725	0.1333	0.1550	0.1948
Mex	0.1242	0.1303	0.1318	0.1738	0.1919	0.1638	0.1413
Mich	0.1419	0.1303	0.1719	0.1655	0.1872	0.1428	0.1912
Mor	0.1632	0.1382	0.1318	0.1323	0.1722	0.1128	0.1717
Nay	0.1916	0.1382	0.1261	0.1865	0.1822	0.1710	0.2119
NL	0.1348	0.1856	0.1891	0.1853	0.1741	0.1451	0.1802
Oax	0.1242	0.1185	0.1203	0.1597	0.2129	0.1229	0.1741
Pue	0.1171	0.1145	0.1375	0.1617	0.1824	0.1167	0.1607
Qro	0.1348	0.1343	0.1375	0.2033	0.2201	0.1864	0.1656
QRoo	0.1455	0.1658	0.1261	0.2234	0.2567	0.2341	0.1839
SLP	0.1135	0.1461	0.1719	0.1626	0.1624	0.1194	0.1717
Sin	0.1916	0.1816	0.1776	0.1880	0.1741	0.1541	0.2070
Son	0.2093	0.1816	0.2063	0.1581	0.1431	0.1251	0.1912
Tab	0.2093	0.1935	0.1547	0.1739	0.2126	0.1395	0.1400
Tam	0.1739	0.2132	0.2235	0.1837	0.1977	0.1460	0.1790
Tla	0.1632	0.1422	0.1261	0.1709	0.1040	0.1488	0.1547
Ver	0.1419	0.1264	0.1375	0.1451	0.1839	0.0885	0.1607
Yuc	0.1774	0.1895	0.1834	0.1778	0.1691	0.1251	0.1802
Zac	0.1881	0.1895	0.1490	0.1567	0.1555	0.1429	0.1936

Source: Abbreviations and calculations based on Table 2.

3. *Equal weighting by dimension* (Table 4). Equal weighting is used when there is no strong reason to consider one dimension or indicator more important than others (the principle of insufficient reason or the principle of indifference). Segmenting the weights helps prevent dominance by dimensions with the most indicators. The indicators in this research represent the social right to health. There are no convincing arguments that one social right is more important than another, nor that one indicator of the social right to health is more important than others. In the absence of a firm argument for differential weighting, these indicators are weighted equally, as suggested by the literature (Babbie, 2021; Mothupi et al., 2021; Nguefack-Tsague, Klasen, Zuchini, 2011; UNICEF, n.d.). This weighting strategy also prioritizes conceptual clarity and policy transparency. Unlike expert-weighting (vulnerable to stakeholder capture) or statistical weighting (opaque to non-technical audiences), equal weighting by dimension provides a defensible, auditable baseline that highlights trade-offs rather than obscuring them.

Table 4 Normalized and equally weighted matrix according to the number of indicators by dimension

States	Resources (1/3)			Status (1/3)	Risk (1/3)		
	Physicians	Nurses	Beds	MortR	RespiraM	CirculaM	PM2.5
Weighting	0.11111	0.1111111	0.11111	0.33333	0.11111	0.11111	0.11111
Ags	0.0213	0.0233	**0.0191**	0.0688	0.0217	0.0228	0.0204
BC	**0.0142**	0.0154	0.0178	0.0598	0.0217	0.0189	0.0169
BCS	0.0280	**0.0184**	0.0204	0.0696	0.0252	0.0219	0.0244
Cam	0.0237	0.0215	0.0204	0.0602	0.0232	**0.0174**	0.0180
Coa	0.0201	0.0180	0.0229	0.0582	0.0222	**0.0160**	0.0208
Col	0.0268	0.0224	0.0229	0.0503	0.0191	**0.0180**	0.0225
Chis	0.0130	0.0167	**0.0115**	0.0661	0.0244	0.0185	0.0147
Chih	0.0189	0.0241	0.0216	0.0488	0.0175	**0.0136**	0.0212
CdMx	0.0386	0.0421	0.0433	0.0422	0.0155	**0.0115**	0.0158
Dgo	0.0185	0.0176	0.0191	0.0628	0.0206	**0.0141**	0.0214
Gto	0.0158	0.0197	0.0191	0.0552	0.0217	**0.0147**	0.0198
Gro	0.0181	0.0184	**0.0153**	0.0635	0.0265	0.0193	0.0214
Hgo	0.0162	0.0158	**0.0153**	0.0628	0.0257	0.0164	0.0175
Jal	0.0162	0.0193	0.0229	0.0575	**0.0148**	0.0172	0.0216
Mex	**0.0138**	0.0145	0.0146	0.0579	0.0213	0.0182	0.0157
Mich	0.0158	**0.0145**	0.0191	0.0552	0.0208	0.0159	0.0212
Mor	0.0181	0.0154	0.0146	0.0441	0.0191	**0.0125**	0.0191
Nay	0.0213	0.0154	**0.0140**	0.0622	0.0202	0.0190	0.0235
NL	**0.0150**	0.0206	0.0210	0.0618	0.0193	0.0161	0.0200
Oax	0.0138	**0.0132**	0.0134	0.0532	0.0237	0.0137	0.0193
Pue	0.0130	**0.0127**	0.0153	0.0539	0.0203	0.0130	0.0179
Qro	0.0150	**0.0149**	0.0153	0.0678	0.0245	0.0207	0.0184
QRoo	0.0162	0.0184	**0.0140**	0.0745	0.0285	0.0260	0.0204
SLP	**0.0126**	0.0162	0.0191	0.0542	0.0180	0.0133	0.0191
Sin	0.0213	0.0202	0.0197	0.0627	0.0193	**0.0171**	0.0230
Son	0.0233	0.0202	0.0229	0.0527	0.0159	**0.0139**	0.0212
Tab	0.0233	0.0215	0.0172	0.0580	0.0236	**0.0155**	0.0156
Tam	0.0193	0.0237	0.0248	0.0612	0.0220	**0.0162**	0.0199
Tla	0.0181	0.0158	0.0140	0.0570	**0.0116**	0.0165	0.0172
Ver	0.0158	0.0140	0.0153	0.0484	0.0204	**0.0098**	0.0179
Yuc	0.0197	0.0211	0.0204	0.0593	0.0188	**0.0139**	0.0200
Zac	0.0209	0.0211	0.0166	0.0522	0.0173	**0.0159**	0.0215

Source: Calculations based on Table 3. Numbers in bold indicate the lowest value in each row. Abbreviations for the states correspond to Table 2.

4. *High correlation between MP and geometric mean for weighted data (GMw)*. The original version by Mazziotta and Pareto (2020) states that MP and GM are highly correlated. The difference in this research is the use of weighted GM for standardized data. This change is necessary because the original version does not group the indicators by subcomponents. In contrast, the calculations in this study use equally weighted data based on the number of dimensions and indicators in each.

The MP is calculated as the average of the lower limit (LB) and the upper limit (UB) of the interval. LB represents the minimum value of each observation (state), while UB is the average of the entire row (Table 5). Additionally, the methodology provides a formula for determining the geometric mean of weighted values (GMw). The case study confirms the original claim of a high rank correlation between MP and GMw (0.96, $p < 0.001$). This strong correlation not only demonstrates that the MP is a partially compensatory index but also, unlike GMw, is preferable because it includes information about the interval's amplitude and the balance of indicators in each observation.

Table 5 Performance intervals for health indicators in the states

State	Min (LB)	MP	Mean (UB)	Amplitude (A)	GMw	Rank MP	Rank GMw	MP-GM	MP	Interval
Ags	0.0191	0.0237	0.0282	0.0091	0.1970	2	2	0	High	Narrow
BC	0.0142	0.0189	0.0235	0.0093	0.1633	22	23	–1	Low	Narrow
BCS	0.0184	0.0241	0.0297	0.0113	0.2065	1	1	0	High	Wide
Cam	0.0174	0.0219	0.0263	0.0089	0.1834	4	5	–1	High	Narrow
Coa	0.0160	0.0208	0.0255	0.0094	0.1774	8	10	–2	High	Narrow
Col	0.0180	0.0220	0.0260	0.0080	0.1796	3	9	–6	High	Narrow
Chis	0.0115	0.0175	0.0236	0.0121	0.1600	26	25	1	Low	Wide
Chih	0.0136	0.0187	0.0237	0.0101	0.1635	24	22	2	Low	Wide
CdMx	0.0115	0.0207	0.0299	0.0184	0.1815	9	7	2	High	Wide
Dgo	0.0141	0.0195	0.0249	0.0108	0.1729	16	13	3	Low	Wide
Gto	0.0147	0.0192	0.0237	0.0090	0.1650	20	20	0	Low	Narrow
Gro	0.0153	0.0207	0.0261	0.0108	0.1807	9	8	1	High	Wide
Hgo	0.0153	0.0198	0.0242	0.0089	0.1671	14	19	–5	High	Narrow
Jal	0.0148	0.0195	0.0242	0.0094	0.1682	16	17	–1	Low	Narrow
Mex	0.0138	0.0180	0.0223	0.0085	0.1544	25	26	–1	Low	Narrow
Mich	0.0145	0.0188	0.0232	0.0087	0.1612	23	24	–1	Low	Narrow
Mor	0.0125	0.0165	0.0204	0.0079	0.1416	30	31	–1	Low	Narrow
Nay	0.0140	0.0195	0.0251	0.0111	0.1736	16	11	5	High	Wide
NL	0.0150	0.0199	0.0248	0.0099	0.1728	13	14	–1	High	Narrow
Oax	0.0132	0.0173	0.0215	0.0083	0.1474	27	29	–2	Low	Narrow
Pue	0.0127	0.0168	0.0209	0.0081	0.1441	29	30	–1	Low	Narrow
Qro	0.0149	0.0201	0.0252	0.0103	0.1735	12	12	0	High	Wide
QRoo	0.0140	0.0211	0.0283	0.0143	0.1932	7	3	4	High	Wide
SLP	0.0126	0.0172	0.0218	0.0092	0.1510	28	27	1	Low	Narrow
Sin	0.0171	0.0217	0.0262	0.0091	0.1828	5	6	–1	High	Narrow
Son	0.0139	0.0191	0.0243	0.0104	0.1679	21	18	3	Low	Wide
Tab	0.0155	0.0202	0.0249	0.0094	0.1727	11	15	–4	High	Narrow
Tam	0.0162	0.0215	0.0267	0.0105	0.1859	6	4	2	High	Wide
Tla	0.0116	0.0165	0.0215	0.0099	0.1484	30	28	2	Low	Narrow
Ver	0.0098	0.0150	0.0202	0.0104	0.1393	32	32	0	Low	Wide
Yuc	0.0139	0.0193	0.0247	0.0108	0.1721	19	16	3	Low	Wide
Zac	0.0159	0.0198	0.0236	0.0077	0.1644	14	21	–7	High	Narrow
Int. avg	**0.0145**	**0.0195**	**0.0245**	**0.0100**						

Source: Calculations based on Table 4. Abbreviations for the states correspond to Table 2.

The case study verifies the following details listed in the original version:

a. The MP values are lower than those of the GMw because the minimum value penalty is higher in the interval version. The case study confirms this in the original version (Table 5).

b. The absolute difference highlights the significant impact of the minimum value penalty in the MP-GMw ranking. The MP-GMw difference shows that the states with the highest penalties, due to their lower standardized and weighted minimum values, are Colima (Col), Hidalgo (Hgo), and Zacatecas (Zac) (Table 5).

5. *Identification of the most important indicators at the midpoint (MP) and the weighted geometric mean (GMw)*. This task is not addressed in the original version of the performance interval. The rank correlation shows that the two most important indicators (lowest correlation) in the MP simulation are respiratory system diseases mortality rate (RespiraM) and circulatory system diseases mortality rate (CirculaM), with $r = 0.847$, $p < 0.001$, and $r = 0.842$, $p < 0.001$, respectively. The simulation for the GMw confirms the higher relevance of these indicators but adds PM2.5. Another interesting fact is that omitting an indicator has a greater impact on the GMw than on the MP (Table 6).

Table 6 Identification of the most important variables in MP and GMw by rank correlation

Simulation for MP								
All indicators		**Without**						
		Physicians	**Nurses**	**Beds**	**MortR**	**RespiraM**	**CirculaM**	**PM2.5**
All indicators	1.000	0.932	0.927	0.929	0.942	**0.847**	**0.842**	0.901
Simulation for GMw								
All indicators	1.000	0.724	0.724	0.724	0.773	**0.597**	**0.597**	**0.597**

* All rank correlation coefficients are significant for $p < 0.001$.

6. *Graph of the intervals and interpretation*. In the interval graph, the states are displayed in descending order of performance summarized in the MP (Figure 1). The gray rectangle, which is not included in the original version, represents the interval average, and the dotted line shows its MP. The original version emphasizes the following aspects in the interval chart:

a. The graph illustrates the hierarchy across three levels of compensation: LB and UB, representing the full non-compensatory and full compensatory extremes, respectively. MP is partially compensatory. The full non-compensatory approach prevents strong performance in one indicator from concealing poor results in others. The MP, like the GMw, is partially compensatory because it allows some degree of trade-off between indicators. UB is full compensatory (it is an average) because an excellent performance in one indicator can offset low performance in another, and it treats essential indicators as if they were completely interchangeable. Full compensatory indices are recommended for public reports or communications due to their simplicity. Full non-compensatory and partially compensatory indices are intended for specialists supporting decision-making, as they require greater technical expertise (El Gibari, Cabello, Gómez, and Ruiz, 2021).

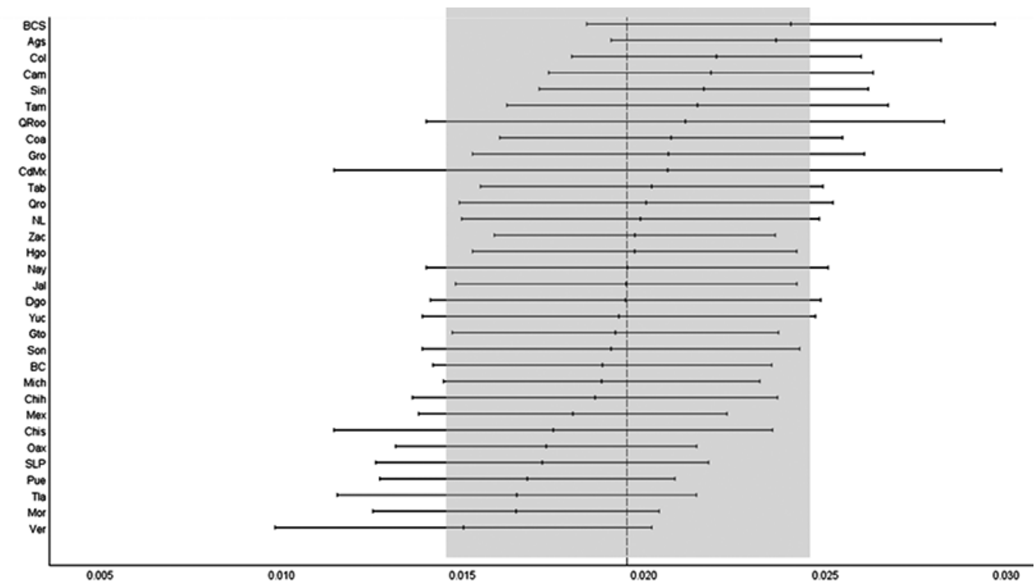
b. One state may perform better than another from a non-compensatory perspective (LB), but do poorly from a full compensatory viewpoint (UB lower). This is the case, for example, with Colima (Col) and Campeche (Cam). The interval of Col is entirely within the interval of Cam.

c. A state may be considered a better performer if its LB is higher than the UB of another, regardless of its degree of compensation. There are no cases related to this health situation in the states of Mexico.
d. The narrower the amplitude, the more stable the performance of a state. For example, Mexico City (CdMx) performs better than Tabasco (Tab), but it is less balanced, as shown by its wider amplitude.

7. *Typology based on MP and interval amplitude,* as outlined in the study methodology. This task is not part of the original interval approach but complements it and provides greater clarity. Figure 1 illustrates the components of the individual intervals and their average. The fundamental elements of the interval average—lower limit, upper limit, midpoint, and amplitude—serve as references for evaluating the MP and A of each observation. States to the right of the dotted line perform at higher levels. Segments outside the gray rectangle fall above or below the average rectangle's boundaries. One way to integrate both criteria, MP and A, is to use the interval average to develop a typology for decision-making, as recommended in the methodology. The gray rectangle in Figure 1 helps locate the MP and A of each characterized state within the typology cells.

The average interval allows for the stratification of MP and A, as outlined in the methodology. Combining these stratifications into a cross-tabulation categorizes states into four main groups, each with specific policy recommendations (Table 7).

Figure 1 Performance interval of health indicators.



Source: Table 5. The gray area indicates the interval average, and the dotted line shows its midpoint (MP). Each horizontal line represents a state's interval, with the central mark indicating the MP. The states are ordered by MP in descending order. Abbreviations for the states correspond to Table 2.

The four-group typology in Table 7, derived from performance (MP) and amplitude (A) metrics, provides a starting point for designing targeted interventions to address institutional fragmentation in Mexico's federal health system. Rather than merely redistributing resources, the amplitude metric diagnoses coordination failures across subsystems, including IMSS, ISSSTE, and state *Secretarías de Salud*, enabling precision equity by aligning intervention design with each state's institutional reality.

Table 7 Typology and public policy suggestions for states

Typology	State (minimum value)	Public policy suggested
Robust state with consolidated performance (Midpoint high + narrow interval)	Ags (Beds) Cam (CirculaM) Coa (CirculaM) Col (CirculaM) Hgo (Beds) NL (Physicians) Sin (CirculaM) Tab (CirculaM) Zac (CirculaM)	*Maintain and enhance current performance*: Continue supporting existing programs and consider expanding successful policies to additional states. *Invest in innovation:* Use stability to test advanced programs (e.g., digital health services). *Prepare for future risks:* Strengthen resilience against unexpected shocks, since current performance remains stable.
State with good performance, but unstable or uneven indicators (Midpoint high + wide interval)	BCS (Nurses) CdMx (CirculaM) Gro (Beds) Nay (Beds) Qro (Nurses) QRoo (Beds) Tam (CirculaM)	*Addressing inequality*: Interventions should target intra-state disparities, supporting weaker indicators. *Monitoring and evaluation*: Enhancing data collection and skills to better understand the causes of internal variability. *Responsive programs*: Create flexible policies customized to varied local needs (e.g., health programs).
State with weak but consistent system (Low midpoint + narrow interval)	BC (Physicians) Gto (CirculaM) Jal (RespiraM) Mex (Physicians) Mich (Nurses) Mor (CirculaM) Oax (Nurses) Pue (Nurses) SLP (Physicians) Tla (RespiraM)	*Basic support*: Broad investments in health. *Incremental improvement*: Interventions can be implemented and scaled consistently thanks to uniformity of conditions. *Capacity building*: Strengthen administrative and institutional ability to adopt and carry out reforms.
State with severe risks and a high degree of uncertainty (Low midpoint + wide interval)	Chis (Beds) Chih (CirculaM) Dgo (CirculaM) Son (CirculaM) Ver (CirculaM) Yuc (CirculaM)	*High priority for intervention*: These are the most vulnerable regions that require immediate and adequately resourced actions. *Tailor-made approaches*: There is no one-size-fits-all solution; craft local policies. *Local involvement*: Engage community members to identify particular issues and encourage participatory growth.

Group 1: Robust states with consolidated performance (Aguascalientes, Zacatecas, Nuevo León) exhibit high MP and narrow amplitude, signaling systemic coherence in which resources, status, and risks align. This reflects successful cross-institutional coordination. The operational pathway positions these states as integration laboratories: formalize cross-institutional referral protocols through state *Comisiones de Salud*, building on Aguascalientes' integrated electronic records system; target municipal peripheries where amplitude begins to widen by deploying mobile non-communicable disease (NCD) clinics funded by the federal catastrophic expenditure fund; and monitor amplitude stability over time, using widening A as an early warning signal for preemptive coordination even when overall performance remains steady.

Group 2: High-performing but unstable states (Mexico City, Quintana Roo, Baja California Sur) show high MP but wide amplitude, indicating institutional decoupling, in which resource concentration fails to reduce mortality due to coverage gaps. Mexico City's paradox (high physician density yet the worst circulatory mortality rates) stems from residents' reliance on underfunded state services despite their proximity to inaccessible IMSS facilities. The pathway implements outcome-linked coordination: redirect a portion of IMSS-Bienestar transfers to fund a shared NCD registry jointly managed by the federal SSA and the city government, with facility bonuses tied to cross-institutional verification of patient outcomes; send promotores to high-pollution, high-informality *colonias populares* to conduct screenings and escort patients to the nearest facilities regardless of affiliation, supported by transportation vouchers; and apply *Fondo de Aportaciones para los Servicios de Salud* (FASSA) flexibility clauses to permit spending on cross-institutional coordination rather than facility-specific inputs, directly addressing the resource-outcome decoupling.

Group 3: Weak but consistent states (Oaxaca, Puebla, Guerrero) show low MP and narrow amplitude, indicating uniform underdevelopment (all indicators move together at low levels) and signaling systemic capacity deficits rather than misalignment. The pathway prioritizes limited infrastructure: channel *Programa de Fortalecimiento para la Salud* (PROSA) funds toward integrated primary care units that combine physician staffing, NCD screening equipment, and air quality monitors to address all three dimensions simultaneously and prevent new resources from replicating fragmentation; focus on indigenous municipalities where narrow amplitude masks ethnic inequities, embedding bilingual community health workers within PROSA units to ensure culturally competent care without creating parallel systems; and avoid quick-win mortality interventions that could widen amplitude by improving outcomes without strengthening local resources, thereby creating dependency and future volatility.

Group 4: States with severe risks and high uncertainty (Veracruz, Chiapas, Chihuahua) show low MP and wide amplitude, signaling a structural crisis in which indicators move in opposing directions due to institutional collapse in specific domains. The pathway implements emergency triage with accountability: activate the federal *Early Warning Mechanism* to declare health system emergencies in critical municipalities, enabling direct SSA management while rebuilding coordination capacity; use amplitude as a targeting metric, prioritizing municipalities where A exceeds the state average (e.g., by over 30%), and deploy mobile surgical units for acute mortality spikes while negotiating ad hoc institutional protocols; and establish a clear exit strategy, transitioning back to state management only after amplitude narrows (e.g., below 0.010) for two consecutive years, ensuring stability precedes autonomy.

This operationalization shows that amplitude is not merely diagnostic but also a tool for intervention design. Traditional rankings would misguide policy: deprioritizing Mexico City despite its fragmentation crisis, or treating Oaxaca and Veracruz as identical low performers, even though they require opposite interventions. Critically, all pathways rely on existing Mexican institutional mechanisms (FASSA flexibility, *Comisiones de Salud*, PROSA) rather than proposing politically infeasible system-wide integration. This transforms the analysis from a methodological exercise into an applied, actionable tool for SSA policymakers addressing Mexico's documented segmentation crisis, advancing both theoretical innovation and practical implementation to strengthen equitable health systems.

4 DISCUSSION

The performance interval approach reveals critical dimensions of Mexico's health system that conventional single-value indices obscure, particularly the internal heterogeneity of state-level health systems and the policy implications of that heterogeneity. Performance intervals bound composite indices between non-compensatory (e.g., minimum value) and fully compensatory (e.g., arithmetic mean) extremes, with the midpoint approximating partial compensability, as with geometric means. Averages or single indices assume fixed substitutability, which can over-penalize or over-compensate unbalanced units; intervals quantify this range explicitly. This approach also reveals how conventional composite indices obscure the critical heterogeneity of indicators within jurisdictions.

Key findings reveal severe resource-outcome decoupling in Mexico's health system, exemplified by Mexico City. Ranking 9th by midpoint, Mexico City exhibits the nation's widest amplitude, indicating extreme disparity between abundant resources (physicians, nurses) and the worst mortality outcomes. This paradox highlights institutional fragmentation rather than scarcity. Among low-performing states, Veracruz shows chaotic imbalance (wide amplitude), requiring targeted rebalancing, while Oaxaca displays uniformly weak performance (narrow amplitude), demanding broad capacity-building. These distinctions are lost in conventional rankings. Sensitivity analysis further shows that Mexico's imbalance stems from inefficient resource-to-outcome conversion rather than input scarcity. By diagnosing why systems underperform, the approach enables precision equity: directing institutional reform to volatile systems (Veracruz), innovation investments to stable, high-performing systems (Aguascalientes), and emergency triage to states with both low performance and high uncertainty (Chiapas). This moves health economics beyond ranking-based resource allocation toward balanced, indicators-aware policy.

A possible explanation for wide intervals in high-resource states is institutional decoupling rather than resource scarcity. This paradox reflects institutional fragmentation within an urban epidemiological transition, not resource scarcity. This dissociation arises from four interacting mechanisms:

a. *Segmented coverage*: Mexico City illustrates the consequences of institutional fragmentation and social segmentation. Despite its exceptional health infrastructure, it has the nation's highest circulatory mortality (Table 2). About 38% of Mexico City's population relies on underfunded state-level services (Secretaría de Salud capitalina), while federal regimes (IMSS/ISSSTE) serve formal-sector workers, creating access gaps even with near resources (Gómez-Dantés et al., 2023).

b. *Urban risk concentration*: Daily PM2.5 exposure (17.2 μg/m^3), above the WHO 24-hour limit (15 μg/m^3).

c. *Cost-sharing barriers and geographic segregation within the metropolis*: obesogenic environments exacerbate the non-communicable disease (NCD) burden among informal workers and marginalized colonias, where proximity to resources does not translate into use.

d. *Outcome accountability gaps*: Mexico's historical focus on input metrics (beds, personnel) in performance evaluation emphasizes infrastructure investment without linking it to mortality outcomes, precisely why amplitude exposes what averages conceal. Quintana Roo exhibits a mirror pattern: tourism-driven resource concentration (hospital infrastructure for visitors) coexists with population-level vulnerability to respiratory mortality (36.1/100 000), revealing how economic specialization fragments health systems within states.

Why does circulatory/respiratory mortality dominate composite performance? The sensitivity analysis (Table 6) shows that respiratory ($r = 0.847$) and circulatory ($r = 0.842$) mortality are the most influential indicators, reflecting Mexico's advanced epidemiological transition but uneven institutional adaptation across and within states. Unlike resource indicators (physicians, beds) that measure system capacity, these mortality metrics capture system effectiveness, which is fragmented across Mexico's institutional system. This explains why amplitude becomes policy-relevant: it quantifies the gap between capacity and effectiveness created by fragmentation, making visible the institutional failures that single-value indices cannot capture.

Policy implications. As detailed in the typology, wide intervals in resource-rich states signal not a need for more inputs but for integration mechanisms, such as shared electronic records across national and local facilities, unified referral protocols for NCDs, and outcome-based financing that penalizes institutional fragmentation. Conversely, narrow intervals in low-performing states (Oaxaca, Puebla) reflect uniformly weak systems, where input expansion remains necessary but insufficient without parallel governance reforms to prevent new resources from merely replicating existing fragmentation in wide-interval states. The amplitude metric thus shifts the policy question from 'how much should be invested?' to 'which institutional arrangements are most effective at converting resources into mortality reduction? and 'which institutional system will convert resources into mortality reduction?' This diagnostic precision is essential for addressing Mexico's documented segmentation crisis.

CONCLUSION

The performance interval approach used for the indicators selected for Mexican states is an innovative and statistically reliable tool for assessing regional health. This method combines average performance with the indicators' internal variability. This dual focus offers a more comprehensive view of health, highlighting states with consistent, balanced performance; those with good conditions but internal disparities; and those with weak systems and high uncertainty.

The method used shows that it is important to consider both average results and the balance of internal indicators for a full diagnosis. The performance range measures the overall effectiveness and consistency of the health system, highlighting aspects that traditional methods, such as averages or rankings, might overlook.

The methodology goes beyond traditional approaches by analyzing all three compensation levels simultaneously within the performance interval. This improves the technical interpretation and clear communication of the results, making it suitable for application in other public policy sectors and for public health officials in Mexico.

The study also emphasizes that mortality rates from respiratory and circulatory diseases are most critical for estimating the partially compensatory composite index (MP). This finding underscores the need to target these indicators to improve population health without neglecting institutional fragmentation.

Moreover, the strong correlation between the midpoint and the weighted geometric mean confirms the index's consistency. At the same time, the interval's width highlights the states with the greatest internal imbalance, a key consideration for fair resource distribution.

This study advances the health economics literature by reframing regional health system inequality not only as a gradient of resource scarcity but also as an internal imbalance among indicators. Interval-based findings, such as performance intervals in composite indices, reveal imbalances across indicators that single-value averages or rankings often mask, leading to more cautious policy decisions. A policy focused on the internal heterogeneity of indicators within each state targets imbalances, such as low life expectancy despite high income, whereas average-based rankings might uncritically place high-income units at the top. Analytically, the interval methodology reframes health policy to prioritize equilibrium (short intervals) over interventions that inadvertently reinforce spatial imbalances by overcompensating weak indicators with strong ones.

Rather than generic prescriptions to address inequality, the interval approach shows that Mexico's state-level indicator imbalances may stem not only from resource scarcity but also from institutional decoupling across parallel health subsystems, including IMSS (formal-sector workers), ISSSTE (public employees), and *Secretaría de Salud* (uninsured/residual population).

Operationalization pathway. The federal Ministry of Health (SSA) and the states should co-design an outcome-based referral protocol to address current imbalances in indicators. First, implement the 2023 General Health Law reforms by establishing joint coordination councils to create a unified non-communicable disease (NCD) registry across public health institutions (e.g., IMSS, ISSSTE, and SSA) to ensure continuity of care. Second, redirect a percentage (e.g., 15%) of IMSS-Bienestar transfers toward outcome-linked payments, rewarding facilities for achieving 80% disease control rather than consultation volume, focusing on mortality reduction over inputs. Third, deploy public community health workers to marginalized areas with high levels of informal employment to conduct door-to-door screenings and escort patients to the nearest facilities using state-funded transport vouchers, regardless of institutional affiliation. Finally, measure success by the reduction in amplitude, narrowing the gap between resource indicators and catastrophic disease outcomes. A target of a 20% reduction in amplitudes over three years would indicate effective institutional integration. This strategy prioritizes impact over indices, incentivizing results rather than system saturation while ensuring excluded populations have access to care.

This intervention fundamentally differs from single-value approaches. For example, traditional single-value indices might classify Mexico City as moderately performing (MP rank 9) and deprioritize it for resource allocation. In contrast, amplitude diagnostics highlight its institutional fragmentation, which calls for coordination mechanisms instead of additional resources. Notably, the proposed approach suggests using Mexico's existing legal framework, such as coordination councils and IMSS-Bienestar transfers, rather than pursuing impractical system-wide integration due to current federalism and political limitations.

Future research should examine how the sensitivity of the performance interval varies across different normalization or weighting criteria for indicators. It is also advisable to investigate the included and excluded indicators separately, examine the relationship between the composite index and economic variables such as GDP, and relate the index results to the types of health programs implemented in each state.

AUTHOR STATEMENT

Jesus A. Trevino-Cantu. The author was solely responsible for the conception and design of the study, data acquisition and analysis, adaptation and extension of the composite index methodology, interpretation of results, and preparation and revision of the manuscript, including intellectual content.

No funding acquisition or external resources were involved. The author approves the final version and takes responsibility for the accuracy and integrity of the work.

References

ALAIMO, L. S. (2023). The Complexity of Social Phenomena and the Construction of Indicators. In: DI BELLA, E., FACHELLI, S., LÓPEZ-ROLDÁN, P., SUTER, C. (eds.), *Measuring Gender Equality: A multidisciplinary analysis of some EU countries*. Social Indicators Research Series 87. Cham, Switzerland: Springer. <https://doi.org/10.1007/978-3-031-41486-2_2>.

BABBIE, E. R. (2021). *The practice of social research* (15th ed.). Cengage AU.

BARRAZA-LLORENS, M., BERTOZZI, S., GONZALEZ-PIER, E., GUTIÉRREZ, J. P. (2002). Addressing inequity in health and health care in Mexico. *Health Affairs, 21*, 47–56. <10.1377/hlthaff.21.3.47>.

BECERRIL-MONTEKIO, V., MENESES-NAVARRO, S., PELCASTRE-VILLAFUERTE, B. E., SERVÁN-MORI, E. (2024). Segmentation and fragmentation of health systems and the quest for universal health coverage: Conceptual clarifications from the Mexican case. *Journal of Public Health Policy*, *45*(1), 164-174. <https://doi.org/10.1057/s41271-024-00470-9>.

BHATTACHARYA, J., HYDE, T., TU, P. (2014). *Health Economics*. London: Palgrave Macmillan.

BRAITHWAITE, J., HIBBERT, P., BLAKELY, B., PLUMB, J., HANNAFORD, N., LONG, J. C., MARKS, D. (2017). Health system frameworks and performance indicators in eight countries: A comparative international analysis. *Sage Open Medicine*. 2017;5. <10.1177/2050312116686516>.

BOUHEDJA, M., BOUHEDJA, S., BENSELHOUB, A. (2024). Testing the suitability of vector normalization procedure in topsis method: Application to wheel loader selection. *Technology audit and production reserves*, *2*(2(76)), 52-62. <https://doi.org/10.15587/2706-5448.2024.301207>.

DRAGO, C. (2021). The analysis and the measurement of poverty: An interval-based composite indicator approach. *Economies*, 9(4): 145. <https://doi.org/10.3390/economies9040145>.

DRAGO, C., GATTO, A. (2023). Gauging energy poverty in developing countries with a composite metric of electricity access. *Utilities Policy*, 81: 101486. <https://doi.org/10.1016/j.jup.2022.101486>.

EL GIBARI, S., CABELLO, J. M., GÓMEZ, T., RUIZ, F. (2021). Composite indicators as decision-making tools: The joint use of compensatory and Noncompensatory schemes. *International Journal of Information Technology & Decision Making*, 20(3): 847–879. <https://doi.org/10.1142/s0219622021500231>.

GÓMEZ-DANTÉS, O., SERVÁN-MORI, E., CERECERO, D., FLAMAND, L., MOHAR, A. (2024). Mexico's health system, 2023. *Salud Pública de México*, 67(1): 91–105. <https://doi.org/10.21149/15802>.

GÓMEZ-DANTÉS, O., SERVÁN-MORI, E., CERECERO-GARCÍA, D., BERTOZZI, S. M. (2023). Origin, impacts, and potential solutions to the fragmentation of the Mexican health system: A consultation with key actors. *Health Research Policy and Systems*, *21*(1), 80. <https://doi.org/10.1186/s12961-023-01025-2>.

GÓMEZ-DANTÉS, H., LAMADRID-FIGUEROA, H., CAHUANA-HURTADO, L., SILVERMAN-RETANA, O., MONTERO, P., GONZÁLEZ-ROBLEDO, M. C., FITZMAURICE, C., PAIN, A., ALLEN, C., DICKER, D. J., HAMAVID, H., LÓPEZ, A., MURRAY, C., NAGHAVI, M., LOZANO, R. (2016). The burden of cancer in Mexico, 1990-2013. *Salud Pública de México*, 118-131. <https://doi.org/10.21149/spm.v58i2.7780>.

JAHANSHAHLOO, G., LOTFI, F. H., IZADIKHAH, M. (2006). An algorithmic method to extend TOPSIS for decision-making problems with interval data. *Applied Mathematics and Computation*, 175(2): 1375–1384. <https://doi.org/10.1016/j.amc.2005.08.048>.

JOUMARD, I., ANDRÉ, C., NICQ, C. (2008). Health Status Determinants: Lifestyle, Environment, Health Care Resources and Efficiency. *OECD Economics Department Working Papers*, No. 627. Paris: OECD Publishing. <https://doi.org/10.1787/240858500130>.

KNAUL, F. M., GONZÁLEZ-PIER, E., GÓMEZ-DANTÉS, O., GARCÍA-JUNCO, D., ARREOLA-ORNELAS, H., BARRAZA-LLORÉNS, M., SANDOVAL, R., CABALLERO, F., HERNÁNDEZ-AVILA, M., JUAN, M., KERSHENOBICH, D., NIGENDA, G., RUELAS, E., SEPÚLVEDA, J., TAPIA, R., SOBERÓN, G., CHERTORIVSKI, S., FRENK, J. (2012). The quest for universal health coverage: Achieving social protection for all in Mexico. *The Lancet*, *380*(9849), 1259-1279. <https://doi.org/10.1016/s0140-6736(12)61068-x>.

KRUK, M. E., MYERS, M., VARPILAH, S. T., DAHN, B. T. (2015). What is a resilient health system? Lessons from Ebola. *The Lancet*, 385(9980): 1910–1912. <https://doi.org/10.1016/S0140-6736(15)60755-3>.

MARMOT, M. (2015). The health gap: The challenge of an unequal world. *The Lancet*, 386(10011): 2442–2444. <https://doi.org/10.1016/s0140-6736(15)00150-6>.

MARMOT, M., ALLEN, J., BOYCE, T., GOLDBLATT, P., MORRISON, J. (2020). *Health Equity in England: The Marmot Review 10 Years On*. Institute of Health Equity. <https://www.health.org.uk/publications/reports/the-marmot-review-10-years-on>

MAZZIOTTA, M., PARETO, A. (2020). Composite indices construction: The performance interval approach. *Social Indicators Research*, 161(2–3): 541–551. <https://doi.org/10.1007/s11205-020-02336-5>.

MENESES-NAVARRO, S. (2025). Mexico frente a la universalidad en salud: Desigualdad, segmentación y reformas contradictorias. *Gaceta Sanitaria*, 102533. <https://doi.org/10.1016/j.gaceta.2025.102533>.

MENESES-NAVARRO, S., PELCASTRE-VILLAFUERTE, B. E., BECERRIL-MONTEKIO, V., SERVÁN-MORI, E. (2022). Overcoming the health systems' segmentation to achieve universal health coverage in Mexico. *The International Journal of Health Planning and Management*, *37*(6), 3357-3364. <https://doi.org/10.1002/hpm.3538>.

Mexico | SIODS | Sistema de Información de los Objetivos de Desarrollo Sostenible. (n.d.). <https://agenda2030.mx/#/home>

MOTHUPI, M. C., DE MAN, J., TABANA, H., KNIGHT, L. (2021). Development and testing of a composite index to monitor the continuum of maternal health service delivery at provincial and district level in South Africa. PLoS ONE, 16(5), e0252182. <https://doi.org/10.1371/journal.pone.0252182>.

MULLINER, E., MALYS, N., MALIENE, V. (2016). Comparative analysis of MCDM methods for the assessment of sustainable housing affordability. *Omega*, 59: 146–156. <https://doi.org/10.1016/j.omega.2015.05.013>.

NGUEFACK-TSAGUE, G., KLASEN, S., ZUCCHINI, W. (2011). On weighting the components of the human development index: A statistical justification. *Journal of Human Development and Capabilities*, *12*(2), 183-202. <https://doi.org/10.1080/19452829.2011.571077>.

OECD. (2023). *Health at a Glance 2023: OECD Indicators*. OECD Publishing. <https://doi.org/10.1787/7a7afb35-en>.

OECD. (2008). *Handbook on Constructing Composite Indicators: Methodology and User Guide*. OECD Publishing, Paris. <https://doi.org/10.1787/9789264043466-en>.

ÖZEN, H., ÖZEN, D. (2024). Latent profile analysis for the classification of OECD countries with health indicators. *Gulhane Medical Journal*: 62–67. <https://doi.org/10.4274/gulhane.galenos.2023.69926>.

WHO. ORGANIZACIÓN PANAMERICANA DE LA SALUD (PAHO). (2024). *Perfil de país - México: Salud en las Américas*. <https://hia.paho.org/es/perfiles-de-pais/mexico>.

SEN, A. (1985). *Commodities and Capabilities*. India: North-Holland.

SMITH, K. V., GOLDMAN, N. (2007). Socioeconomic differences in health among older adults in Mexico. *Social Science & Medicine, 65*, 1372–1385. <10.1016/j.socscimed.2007.05.023>.

TALUKDER, B., HIPEL, K. W., VANLOON, G. W. (2017). Developing composite indicators for agricultural sustainability assessment: Effect of normalization and aggregation techniques. *Resources*, 6(4): 66. <https://doi.org/10.3390/resources6040066>.

TOLONEN, H., REINIKAINEN, J., KOPONEN, P., ELONHEIMO, H., PALMIERI, L., TIJHUIS, M. J. (2021). Cross-national comparisons of health indicators require standardized definitions and common data sources. *Arch Public Health* 79: 208. <https://doi.org/10.1186/s13690-021-00734-w>.

UNICEF. (n.d.). Technical briefing note 3: Seven reasons for equal weighting of dimensions in child poverty measurement. <https://data.unicef.org/wp-content/uploads/2021/12/Technical-Briefing-Note-3.pdf>.

URQUIETA-SALOMÓN, J. E., VILLARREAL, H. J. (2016). Evolution of health coverage in Mexico: Evidence of progress and challenges in the Mexican health system. *Health Policy and Planning*, 31(1), 28-36. <https://doi.org/10.1093/heapol/czv015>.

New Developments in the System of National Accounts

Stanislava Hronová [1] | *Prague University of Economics and Business, Prague, Czech Republic*

On 11–12 June 2026, the nineteenth conference of the *Association de Comptabilité Nationale* (ACN) was held in Paris at the OECD premises in Boulogne-Billancourt. The international conference was organised under the auspices of the Institut National de la Statistique et des Etudes Economiques (INSEE) and with the support of the OECD. The conference was attended by around one hundred experts from more than ten countries representing national statistical offices, universities, research institutes, and other national and international institutions, including Eurostat, the International Monetary Fund, and the OECD.

The conference programme was divided into five relatively independent thematic sessions. The topics selected by the organisers generated considerable interest among all participants, both because of the wide range of issues addressed and because of the richness of the subsequent discussions.

The conference proceedings were, once again this year, highly interesting and valuable. The following overview presents some of the most significant contributions.

The conference was officially opened by Monica Brezzi, Director of the Statistics Division at the OECD, who emphasised the importance of bringing together experts in the field of national accounting, particularly at a time of economic turbulence when it is essential to provide a consistent view of economic developments in European countries based on the harmonised methodology of national accounts. At present, as work is underway to complete the ESA 2030 standard, the OECD has identified three main areas of activity: data preparation and processing, technical assistance, and methodology. The implementation of the new ESA 2030 standard will require very close cooperation between national accounts experts and other statisticians in order to ensure that phenomena such as the environment, the digital economy, income inequalities, and other emerging issues are properly reflected in official statistics.

The second invited speaker was Fabrice Lenglart, Director-General of INSEE. In his presentation, he focused mainly on new challenges arising from recent developments in society – including digitalisation, artificial intelligence, the emergence of new products and new professions – which official statistics and national accounts must capture. Another important development will undoubtedly be the introduction of the digital euro, which is expected to play an increasingly important role in cashless payments.

The conference was chaired by François Lequiller, President of ACN and former Director of the National Accounts Department at INSEE and of the OECD National Accounts Division. The first thematic session, devoted to *New standards for national accounts*, focused on the changes that will result from the forthcoming United Nations and European Union standards.

In his contribution *Presentation of SNA 2025 and ESA 2030*, John Verrinder (Eurostat) discussed the main parameters of the changes introduced by the new standards. He recalled the continuity between

[1] Department of Economic Statistics, Faculty of Informatics and Statistics, Prague University of Economics and Business, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic. E-mail: hronova@vse.cz. ORCID: <0000-0002-3568-9755>.

the latest United Nations and European Union standards and stressed the need to harmonise the frameworks applied in European and non-European countries.

The ESA 2030 standard will introduce changes affecting both the basic framework and the values of certain macroeconomic aggregates, as well as changes that will extend the scope of the existing framework. The first group of conceptual changes includes, in particular, the inclusion of the depletion of natural resources as a production cost (which will affect the measurement of net domestic product), and the introduction of new categories of renewable energy resources and natural resources as non-produced assets. Another significant change will be the inclusion of data (not only databases) among produced assets. This means that data will be incorporated into the output, with an impact on the value of gross domestic product. Similarly, crypto-assets will be included, with their mining and validation activities recognised as output, also affecting the value of gross domestic product.

The second group of changes concerns, in particular, more detailed monitoring of household income and expenditure, as well as the development of information on household assets and liabilities.

In conclusion, Verrinder stressed that ESA 2030 will retain a stable structure, with revisions incorporated both into the main framework and in the annexes. The implementation of the key changes in national statistical systems is expected to take place in 2030, while partial modifications and extensions will be introduced gradually.

The presentation was followed by a lively discussion that focused mainly on the proposal that "net" aggregates should become more prominent at the expense of "gross" aggregates. This approach was questioned in view of the long-standing tradition, practical experience and substantive arguments concerning the international comparability of estimates of the consumption of fixed capital. In addition, "gross" indicators, particularly GDP, are widely used in the construction of relative indicators, including those required by European legislation.

A particularly interesting contribution in this session was the presentation *Cryptocurrencies and National Accounts* by Jeremy Fouliard (INSEE). Cryptocurrencies represent a new phenomenon that undoubtedly poses significant challenges, especially regarding their definition as production activities (mining and validation), their recording in terms of stocks and changes in assets and liabilities, and their international comparability. Given the volatility of this type of asset, Fouliard called for a cautious and gradual approach to methodological changes. Crypto-assets have their own underlying technology, their own operating mechanisms and their own specific "culture" (codes and anonymity). A fundamental question is for example what cryptocurrencies actually represent. Should they be classified as financial or non-financial assets? As produced or non-produced assets? If they are financial assets, do they have a corresponding liability? If crypto-assets have no counterpart liability, they can only be classified as non-financial, non-produced assets. A further and broader challenge for official statistics will be the classification and recording of the various transactions involving these crypto-assets.

The second thematic session was devoted to the topic of *Military Expenditure*. In his presentation *Military Expenditure – Its Recording in the French National Accounts*, Jean-Cyprien Héam (INSEE) addressed not only the specific features of France but also placed them in the broader context of EU and NATO requirements.

He emphasised that, although exceptions exist regarding the inclusion of military expenditure in the government deficit under the Maastricht rules, neither the definition of military expenditure nor the calculation of the government balance (net lending/net borrowing) is affected by these exceptions. This creates additional requirements for official statistics in terms of the accuracy of definitions, transparency and international comparability of the various categories of military expenditure.

Particular difficulties arise in the area of investment. These include how to record deliveries of complex weapons systems (such as ships, submarines and aircraft), which are often produced in several different countries; how to record deliveries of ammunition to storage facilities abroad; and how to address issues related to confidentiality and security requirements.

Héam also recalled the differences between the classification of defence expenditure according to COFOG and according to NATO. Unlike COFOG, the NATO classification does not include civil defence within military expenditure. On the other hand, it applies a broader concept to expenditure on customs services, police forces and border protection where these activities have a paramilitary role. It also adopts a broader approach to expenditure on research and development infrastructure and facilities with dual military and civilian use.

Pierre Muller (ACN) reviewed the history of recording military expenditure in the French national accounts (*Military Expenditure and National Defence Service: Evolution of Approaches in National Accounts*). He recalled that, until the beginning of the 1970s, military expenditure was not included in the output in French national accounts, because non-market output itself was not taken into account at that time (similarly to the situation in the Material Product System used in socialist countries).

Subsequently, in accordance with the international standards then in force, all military expenditure in France (with the exception of military accommodation buildings) was included in intermediate consumption. The adoption of the SNA 1993 and ESA 1995 standards brought a major change: military expenditure on infrastructure, construction, weapons systems, transport equipment and other tangible and intangible fixed assets was reclassified from intermediate consumption to gross fixed capital formation, while ammunition and light weapons were recorded as inventories.

Muller also noted that, with the adoption of the SNA 2025 standard, "military" data will also be included within military expenditure. Given the specific characteristics of defence expenditure, he considered the establishment of a Defence Satellite Account to be a useful development. Such an account would enable more precise definitions of activities whose outputs are included in production measures and military investment and would ultimately provide a more accurate assessment of national defence expenditure. However, Muller expressed concern that such efforts may encounter difficulties because of confidentiality requirements.

The third thematic session focused on *Satellite Accounts*. John Verrinder (Eurostat) presented the position, importance and role of satellite accounts within the System of National Accounts. He emphasised their broad thematic scope, including health, education, social protection, tourism, and research and development, while noting that many other satellite accounts remain relatively uncommon in EU countries. These include, for example, the Social Accounting Matrix, the Household Production Satellite Accounts, the Social Economy Satellite Accounts and the Non-profit Institutions Satellite Accounts. In the ESA 2030 standard, these thematic extensions and elements expanding the basic framework will be addressed in Chapter 20.

The OECD approach to satellite accounts was presented by Michael Mueller (OECD), who used the example of the *Health Satellite Account for OECD countries*, while Viktoria Kis (OECD) presented the *Education Satellite Account for OECD countries*. The complex issues involved in the statistical description of healthcare systems were further illustrated through French experience by Clément Delecourt and Jeremy Fouliard (both INSEE).

The thematic session on *Environmental Accounts* included four presentations. The most stimulating contribution, in my view, was the paper *How Can Climate Be Taken into Account in National Accounts?* presented by Sylvain Larrieu and Sébastien Roux (both INSEE).

The System of National Accounts provides information on the actual state of the economy, including aspects of sustainability, and also takes into account the costs associated with greenhouse gas emissions. The authors' fundamental question was how the current SNA framework could be modified in order to capture the consequences of climate change. They proposed modifications to several aggregates. Since greenhouse gas emissions contribute to the degradation of climate-related assets, they should be considered as a form of capital consumption. This approach would make it possible to create indicators such as Adjusted Net Domestic Product or Adjusted Net Saving, which would incorporate the damage caused by emissions into the consumption of fixed capital. The authors nevertheless stressed that such aggregates could not, under any circumstances, have the same status as conventional national accounting indicators. They should instead be regarded as the results of experimental approaches.

The final session focused on issues related to the separate measurement of *Price and volume* developments in aggregate values. Three interesting papers were presented: *Inequalities in the Face of Inflation: The Issue of Real Estate Prices* (Didier Blanchet, Marc Fleurbaey, Craig Pesme, Come Zegna-Rata, INSEE and Paris School of Economics), *Measuring the Volume of Non-market Education Output* (Jeremy Fouliard, INSEE), and *Values, Volumes, Surplus and Consumer Utility: What Are the Relationships?* (Didier Blanchet, Marc Fleurbaey, Craig Pesme, INSEE and Paris School of Economics).

The contribution addressing the impact of rising real estate prices on inequalities in the household sector was particularly interesting and highly relevant also for the current situation in the Czech Republic. In general, increases in the prices of different commodities do not affect all households equally. From the perspective of widening inequalities, this is most clearly reflected in developments in real estate prices.

In general, INSEE uses three main price indices: the Laspeyres consumer price index, which serves as the basis for measuring inflation; an analogous price index used as the basis for the harmonised inflation measure; and the deflator of household final consumption expenditure (the Paasche price index). According to the authors, it would therefore be appropriate to develop different deflators for different socio-economic categories of households. They compared the structure of household final consumption expenditure across socio-economic categories and attempted to estimate the corresponding inflation rates for these groups. They concluded that households with lower saving rates are the groups most adversely affected by price developments.

The conference concluded with the *General Assembly of ACN*, during which the report on the Association's activities and financial management was approved and further possibilities for the development of this international organisation were discussed.

In this context, it should be noted that membership in ACN, hosted by INSEE, is voluntary and free of charge. Individuals interested in joining the Association may find further information on the INSEE website. All presentations delivered at the conference are available on the INSEE website. Available at: https://www.insee.fr/fr/information/8992587.

More Than Just a Data Guide to the Historical Development of the Press, Film, and Radio in the Czech Lands – Book Review

Stanislava Hronová[1] | *Prague University of Economics and Business, Prague, Czechia*

PIŠTORA, L. *Dějiny masových médií v číslech. Od počátků do konce první republiky.* (History of the Mass Media in Numbers: From the Beginnings to the End of the First Czechoslovak Republic.) 1st Ed., Prague: Karolinum, 2025, 308 p., ISBN 978-80-246-5845-2

At the beginning of the year, a new publication History of the Mass Media in Numbers with the subtitle From the Beginnings to the End of the First Czechoslovak Republic, which is the work of Ladislav Pištora, an author known to all those who are interested in data, appeared in the offer of new titles of professional booksellers. It was published by the Karolinum Press of Charles University.

The publication is a data guide to the development of the media (the press, film, and radio) gradually in the territory of Bohemia, today's Czech Republic, and interwar Czechoslovakia. In the introduction to the publication, the author outlines all the pitfalls he had to face when drafting the publication, however, which for him (and perhaps precisely due to that) represented both a challenge and the actual reason for such an undertaking. Statistical data on the development of the media were not systematically monitored and published for a long time, because they did not represent a key, especially economic, interest. If any data appeared at the beginning of the development of periodicals, it was rather sporadically, mainly

1 Department of Economic Statistics, Faculty of Informatics and Statistics, Prague University of Economics and Business, Czechia. E-mail: hronova@vse.cz. ORCID: <https://orcid.org/0000-0002-3568-9755>.

to support certain claims, and their publication was more or less random. The situation was better in the case of film (since the beginning of the 20th century) and later regarding radio, too (since the 1920's).

When listing the pitfalls involved in the compilation of such a data publication, the author highlighted issues related to the temporal, spatial, and factual comparability of the data. These issues and their possible solutions are addressed in a comprehensive accompanying text commenting on the presented data. The aim of the publication is thus not to analyse the development of these media, but to describe them using statistical data. In this respect, it is necessary to appreciate the author's careful approach to solving all the issues related to the availability, quality, and comparability of the data. The publication has an extensive note-taking apparatus and bibliography and data sources.

Leaving aside the introduction, the publication is divided into three thematic chapters devoted to the development of the press, film, and radio. The text part of the chapters is accompanied by a number of detailed tables illustrating the described period. At the end of each of the three chapters, there are also summary tables. Detailed (and summary) tables are accompanied by charts in a number of places. Unfortunately, orientation in them is often difficult for the reader (if not impossible, in the case of pie charts), given that the charts, originally designed for colour printing, were adopted for use in this publication without being adapted for the capabilities of black and white printing.

The author places the beginning of the development of periodical press (newspapers and magazines), traceable in the data, at the turn of the 18th and 19th centuries, which is a period considered by historians to be the period of the origin of Czech journalism. Data on periodicals in that initial period are sporadic and the author admits that he had to rely solely on secondary sources. Statistical data sources can only be traced back to approximately the 1860's, when, in connection with the publication of the Press Act, periodicals were monitored for the purposes of registration and authorisation (by repressive authorities).

The author divided the development of periodicals in data according to historical events as well as data availability into the period of the first half of the 19th century, then the period around the revolutionary year 1848, then the period after the issuance of the Press Act (1862) up until the 1890's, then the period until the end of the First World War, and, finally, the period of the so-called First Czechoslovak Republic. For each of these periods, the author describes and documents available data sources, the specifics of print runs of individual titles in terms of their number, thematic focus, the language of publication, etc. However, the text of the chapters is not a mere accompaniment to the presented tables, but rather a detailed description and commentary on the situation concerning periodicals at each of the monitored stages of development. Logically, the longest sub-chapter is devoted to the period of the so-called First Czechoslovak Republic, in which the author drew from rather rich and trustworthy data sources and documents. I appreciate that in the period after the establishment of Czechoslovakia, the author deals in detail with the first statistical survey of periodicals in 1920, which was one of the first on this topic on a European scale.

The beginning of the historical stage of the development of film in data, the author situates to the very beginnings of cinema (the end of the 19th century), although he adds that the period until the end of the First World War is not data rich in this respect, i.e. it is not possible to talk about a systematic monitoring and reporting of the number of films, titles, distributors, cinemas, viewers, etc. In the Czech Lands, we have the first figures on the number of licences granted to operate cinematographic enterprises as early as the year 1896, i.e. only one year after the first presentation of the Lumière brothers' cinematograph in Paris. The development of technology, the accessibility of film screenings for a wide audience led not only to their rapid spread, but also to the possibility of statistics to monitor all these phenomena. Given a sufficient database on the development of cinematography, the author focuses (in terms of both data and text) on the situation in the Czech Lands and Czechoslovakia as well as in developed European countries (France, Germany, the United Kingdom of Great Britain and Northern Ireland) and the United States of America. Rich data sources, easier data comparability (including international comparisons), and high-quality accompanying text make this chapter very interesting and readable.

The development of radio broadcasting in Czechoslovakia dates to 1923, when the first regular broadcasting from a transmitter in Kbely took place. An important moment in its further development in terms of data was the nationalisation of radio in 1925, i.e. the fact that, under the then applicable law, the state became the only radio broadcaster. It ensured regular monitoring of a range of quantitative data on all aspects of radio broadcasting (licences, transmitters, receivers, listeners, licence fee payers, broadcasting genres, etc.), including regional disaggregation. Given the availability of sufficient data sources, the author deals with international comparisons (as in the case of film).

The *History of the Mass Media in Numbers* is a publication intended for all those who are interested in the historical development of the media from a data perspective. The author dealt with a number of pitfalls in the form of difficult comparability and low availability of data from the 19th century and, on the contrary, showed what interesting quantitative and qualitative information can be obtained especially from the period of the so-called First Czechoslovak Republic. It will undoubtedly find its readers among experts and students in the field of the media, but also among the general public interested in the history of the press, film, and radio.

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