Volume 26, Number 3, September 2025 **ISSN 1234-7655**

STATISTICS IN TRANSITION

new series

An International Journal of the Polish Statistical Association and Statistics Poland

IN THIS ISSUE:

ISSN 1234-7655

Head of Editorial Office/Secretary
Patryk Barszcz, *Statistics Poland, Warsaw, Poland, e-mail: p.barszcz@stat.gov.pl, phone number +48 22 – 608 33 66*

Managing Editor
Adriana Nowakowska, *Statistics Poland, Warsaw, Poland, e-mail: a.nowakowska3@stat.gov.pl*

Technical Assistant
Rajmund Litkowiec, *Statistical Office in Rzeszów, Rzeszów, Poland, e-mail: r.litkowiec@stat.gov.pl*

Address for correspondence
Statistics Poland, al. Niepodległości 208, 00-925 Warsaw, Poland, tel./fax: +48 22 – 825 03 95

CONTENTS

STATISTICS IN TRANSITION new series, *September 2025*
Vol. 26, No. 3, pp. III

Submission information for Authors

Statistics in Transition new series (SiTns) is an international journal published jointly by the Polish Statistical Association (PTS) and Statistics Poland, on a quarterly basis (during 1993–2006 it was issued twice and since 2006 three times a year). Also, it has extended its scope of interest beyond its originally primary focus on statistical issues pertinent to transition from centrally planned to a market-oriented economy through embracing questions related to systemic transformations of and within the national statistical systems, world-wide.

The SiT*ns* seeks contributors that address the full range of problems involved in data production, data dissemination and utilization, providing international community of statisticians and users – including researchers, teachers, policy makers and the general public – with a platform for exchange of ideas and for sharing best practices in all areas of the development of statistics.

Accordingly, articles dealing with any topics of statistics and its advancement – as either a scientific domain (new research and data analysis methods) or as a domain of informational infrastructure of the economy, society and the state – are appropriate for *Statistics in Transition new series*.

Demonstration of the role played by statistical research and data in economic growth and social progress (both locally and globally), including better-informed decisions and greater participation of citizens, are of particular interest.

Each paper submitted by prospective authors are peer reviewed by internationally recognized experts, who are guided in their decisions about the publication by criteria of originality and overall quality, including its content and form, and of potential interest to readers (esp. professionals).

Manuscript should be submitted electronically to the Editor:
sit@stat.gov.pl,
GUS/Statistics Poland,
Al. Niepodległości 208, R. 296, 00-925 Warsaw, Poland

It is assumed, that the submitted manuscript has not been published previously and that it is not under review elsewhere. It should include an abstract (of not more than 1600 characters, including spaces). Inquiries concerning the submitted manuscript, its current status etc., should be directed to the Editor by email, address above, or w.okrasa@stat.gov.pl.

For other aspects of editorial policies and procedures see the SiT Guidelines on its Web site: https://sit.stat.gov.pl/ForAuthors.

STATISTICS IN TRANSITION new series, *September 2025*
Vol. 26, No. 3, pp. V–VI

Policy Statement

The broad objective of *Statistics in Transition new series* is to advance the statistical and associated methods used primarily by statistical agencies and other research institutions. To meet that objective, the journal encompasses a wide range of topics in statistical design and analysis, including survey methodology and survey sampling, census methodology, statistical uses of administrative data sources, estimation methods, economic and demographic studies, and novel methods of analysis of socio-economic and population data. With its focus on innovative methods that address practical problems, the journal favours papers that report new methods accompanied by real-life applications. Authoritative review papers on important problems faced by statisticians in agencies and academia also fall within the journal's scope.

***

Abstracting and Indexing Databases

Statistics in Transition new series is currently covered in:

BASE – Bielefeld Academic Search Engine

CEEOL – Central and Eastern European Online Library

CEJSH (The Central European Journal of Social Sciences and Humanities)

CNKI Scholar (China National Knowledge Infrastructure)

CNPIEC – cnpLINKer

CORE

Current Index to Statistics

Dimensions

DOAJ (Directory of Open Access Journals)

EconPapers

EconStore

Electronic Journals Library

Elsevier – Scopus

Genamics JournalSeek

Google Scholar

Index Copernicus

J-Gate

JournalGuide

JournalTOCs

Keepers Registry

MIAR

Microsoft Academic

OpenAIRE

ProQuest – Summon

Publons

QOAM (Quality Open Access Market)

ReadCube

RePec

SCImago Journal & Country Rank

TDNet

Technische Informationsbibliothek (TIB) – German National Library of Science and Technology

Ulrichsweb & Ulrich's Periodicals Directory

WanFang Data

WorldCat (OCLC)

Zenodo

STATISTICS IN TRANSITION new series, September 2025
Vol. 26, No. 3, pp. VII–XI

From the Editor

It is with great pleasure that we present our readers with the September issue consisting of 12 articles arranged in three sections: *Original Research Papers, Conference Papers, and Research Communicates and Letters*. Altogether 24 authors from a large set of countries – Australia, Poland, Jordan, India, Nigeria, Indonesia, and Italy – present results of their investigations in a wide spectrum of research areas.

Original Research Papers

The section of original research papers starts with the article by **Ravi Dutta-Powell** entitled ***The perils of premature evaluation: reassessing the application of Benford's Law to the USA's COVID-19 data.*** In follow-up of a review of earlier applications of Benford's Law to the COVID-19 data in the United States – that claimed these data's non-conformity with Benford's Law – more recent granular data are used to demonstrate that this was likely due to the earlier data being unsuitable for such applications. It also demonstrates that the same dataset, analyzed in different ways, can show vastly different levels of conformity with Benford's Law. Specifically, most US states show high degrees of conformity for the COVID-19 cases and cumulative deaths when the Robust Order of Magnitude (ROM) is over 3 and data at the county level is used to analyze state outcomes. Conversely, when the county data is aggregated to the state level and analyzed (i.e. case totals for all counties are summed to create a single state figure for each day of the pandemic), every state shows non-conformity.

Leszek Morawski's paper ***Education expansion and income inequality: evidence from Poland (2005–2019*)** shows that educational change reduced poverty and income inequality incidence and depth, and describes the consequences of this change using a microsimulation decomposition based on a tax-benefit microsimulation model. Using a microsimulation approach, the author estimates that the impact of the abovementioned educational change on changing material poverty risk corresponds to 40% of the policy effect associated with changes in tax and benefit rules. For the Gini index, the educational effect amounted to 91% of the policy effect. The results show that educational changes in Poland between 2005 and 2019 significantly impacted income inequality and the risk of material poverty.

In the article ***Area-biased one-parameter exponential distribution with financial applications*** **Abdullah Hardan** and **Loai Alzoubi** propose the area-biased one-parameter linear exponential distribution the main properties of which – such as the

moments and the related measures, the harmonic mean and the mode – are derived and analyzed using the reliability analysis functions along with the pdfs of the minimum, maximum and the kth order statistics. Additionally, employed are the quantile function, the mean absolute deviations from the mean and the median jointly with the mean waiting and residual lifetime. A simulation study using the MLE, OLS, WLS, MPS, CVM and AD methods of estimating parameters is conducted showing that the estimators are unbiased and consistent. Three real financial data applications prove the goodness of fit for this distribution. They show that the suggested distribution fits the real data better than the competence distributions.

The next paper, ***Bayesian estimation of two-parameter power Rayleigh distribution and its application*** by **Mohd Irfan** and **Anup Kumar Sharma** explores classical and Bayesian approaches to the estimation of unknown parameters and reliability functions for the power Rayleigh distribution. The maximum likelihood estimator (MLE) method is considered in classical estimation. The Bayesian estimation, on the other hand, uses several loss functions under informative and non-informative prior distributions, utilizing the Lindley technique and Markov Chain Monte Carlo (MCMC) method for Bayesian computations. Approximate confidence intervals are established based on the MLEs using the delta technique, while Bayes credible intervals are determined using the MCMC method. A simulation study is conducted to compare the performance of these methods in terms of biases and mean square errors, showing that Bayesian estimators outperform their classical counterparts. Additionally, two real datasets are presented for illustrative purposes.

Krzysztof Brania's and **Henryk Gurgul's** article ***The impact of the COVID-19 pandemic on forecast uncertainty of macroeconomic data releases*** focuses on the uncertainty associated with macroeconomic data forecasts measured by the surprise indicator (SI). Moreover, the authors examine whether the distribution of SI depends on the economy, category of indicator or time, considering pre-pandemic, pandemic and post-pandemic periods in the context of the COVID-19 crisis. The construction of a sentiment indicator that is intended to aggregate all information that is jointly released through macroeconomic indicators was also proposed, Macroeconomic data releases are very important benchmarks of the economy. Therefore, the vast majority of financial market analysts and traders closely monitor both the projected estimates and the intuitively more impactful actual values.

The paper by **Nureni Olawale Adeboye** and **Olumide Sunday Adesina** discusses ***Bayesian and frequentist modelling of West African economic growth: a dynamic panel approach.*** As the empirical outcomes of previous studies examining the relationship between economic growth and socio-economic indicators have been inconclusive and contradictory the current research employed an alternative strategy. A dynamic panel model is estimated *via* three robust dynamic panel data estimators of

the generalized method of moment (GMM), frequentist instrumental variable (IV) and the Bayesian IV on real and simulated data. Various model performance criteria such as Wald statistics, leave-out-one cross-validation and the Pareto checks were used for validity verification. The results of the robust diagnostics checks and a model strength metric showed that the family of IV models outperformed the GMM. Thus, the estimation provided by the Bayesian IV is upheld and recommended in modelling dynamic panel data as it provides robust estimates of the parameters of interest.

In the work entitled ***Exploring the stochastic production frontier in the presence of outliers: a simulation study,*** **Anik Djuraidah** and **Ismail Pranata** present the results of a simulation conducted to compare five SPF models: Normal-half Normal, Normal-Gamma, Normal-Weibull, Normal-Rayleigh, and Student's-t-half Normal. Applying simulated data across nine scenarios with varying data amounts and outlier percentages, the findings prove that the SPF Student's t-half Normal model provides the most accurate prediction of technical efficiency. Using a heavy-tailed distribution, such as the Student's t distribution, for the disturbance component is more effective in handling outliers in the response variable than modifying the inefficiency of the component distribution.

The paper by **Alina Jędrzejczak**, **Małgorzata Misztal**, and **Dorota Pekasiewicz** entitled ***Regional differentiation of income distributions in Poland*** examines the regional differences in the total income distribution in Poland. Both average income levels and income inequality and poverty parameters are included in the analyzes. The study, based on individual data from the Household Budget Survey, used parametric and non-parametric methods for estimating inequality and poverty measures, as well as cluster analysis methods. In the parametric approach, the empirical income distributions in Poland were approximated using the theoretical Dagum distribution. This enabled the segmentation of voivodships in terms of the estimated characteristics of the equivalent household income distribution. The results confirmed anticipations that income distributions in Poland differ significantly across regions. The obtained clusters allowed identifying groups of regions that may require separate social policies aimed at upholding household income or at reducing income inequality.

Błażej Suproń's paper ***Modelling the asymmetric relationship between energy and CO_2 emissions in the Visegrad group: empirical evidence from a panel NARDL approach*** presents an attempt to assess the impact of renewable energy consumption, non-renewable energy consumption and economic growth on the volume of carbon dioxide (CO_2) emissions in the Visegrad countries between 1991 and 2021. Using a Nonlinear Autoregressive Distributed Lag (NARDL) model for panel data, the research captures both long-term dependencies and short-term dynamics. The results show that a reduction in CO_2 emissions yielded by a significant long-term decrease

in non-renewable energy consumption is proportionally larger than the increase in the emissions caused by the growth in the consumption of such energy. GDP growth in the V4 countries increases CO_2 emissions, but GDP decline contributes significantly more to the reduction in emissions. On the contrary, renewable energy consumption consistently reduces CO_2 emissions over the long term, with no significant asymmetry detected. In the short term, both economic growth and non-renewable energy consumption increase CO_2 emissions.

Conference Papers

XXXXII Multivariate Statistical Analysis 2024, Lodz, Poland

Stefano Bonnini's and **Michela Borghesi's** paper entitled ***Multivariate two-sample permutation test with directional alternative for categorical data*** presents a distribution-free test, based on the permutation approach, on treatment effects with a multivariate categorical response variable. It refers to a typical case-control biomedical study, performed to investigate the effect of the treatment called "assisted motor activity" (AMA) on the health of comorbid patients affected by "low back pain" (LBP), "hypertension" and "diabetes". Specifically, the goal was to test whether the AMA determines an improvement in the functionality and the perceived health status of patients. Two independent samples (treated and control group) were compared according to 13 different binary or ordinal outcomes. The null hypothesis of the test assumes the equality in the distribution of the multivariate responses of the two groups, while under the alternative hypothesis, the health status of the treated patients is better. The approach proposed in this work is based on the Combined Permutation Test (CPT) method, which is suitable for analyzing multivariate categorical data in the presence of confounding factors.

XV Scientific Conference MASEP 2024 – Measurement and Assessment of Social and Economic Phenomena, Warsaw, Poland

Abu Feyo Bantu, **Andrzej Kozyra**, and **Józef Wiora** discuss ***Normality tests for transformed large measured data: a comprehensive analysis.*** The study emphasizes the importance of assessing and transforming large datasets, such as GNSS measurements, to ensure normality for the validation of parametric statistical tests. The untransformed GNSS latitude data were identified as non-normal using various visual an statistical tests, including histograms, Q-Q plots, skewness, kurtosis, and Statistical tests: KS, AD, DA, SW, JB, CVM, Chi2, and LF. From among the transformation techniques, the rank-based Inverse Normal Transformation (INT) shown relatively higher effectiveness in enhancing data normality, as validated by various testing methods. The efficiency of Statistical tests was assessed using ROC and AUC analysis, which successfully categorized untransformed data as non-normal and transformed data as normal.

These findings underscore the necessity of using tailored transformation methods in large-scale data applications, particularly in geospatial and industrial fields, to enhance the reliability and applicability of parametric statistical methods.

Research Communicates and Letters

Archana Panigrahi, **Priyaranjan Dash**, and **Gopabandhu Mishra** present ***A minimum variance unbiased estimator of finite population variance using auxiliary information***. A class of estimators of finite population variance (S_y^2) using auxiliary information has been developed under simple random sampling without replacement (SRSWOR) scheme. An attempt has been made to derive the minimum variance unbiased estimator of finite population variance from the proposed class of unbiased estimators. The efficiency of the class of estimators under optimality is compared with the usual unbiased estimator, ratio type estimator, product type estimator, regression type estimator, exponential ratio type estimator, exponential product type estimator, and ratio-in-regression estimator, both theoretically and empirically under general conditions and under bivariate normality. The proposed class of estimator performs better than these estimators under certain realistic conditions. The proposed class of estimators is generalized for the case of multi-auxiliary variables.

Włodzimierz Okrasa
Editor

STATISTICS IN TRANSITION new series, September 2025
Vol. 26, No. 3, pp. 1–13, https://doi.org/10.59139/stattrans-2025-024
Received – 04.02.2025; accepted – 30.04.2025

The perils of premature evaluation: reassessing the application of Benford's Law to the USA's COVID-19 data

Ravi Dutta-Powell[1]

Abstract

This paper reviews earlier applications of Benford's Law to the COVID-19 data in the United States that claimed these data's non-conformity with Benford's Law, and uses later and more granular data to demonstrate that this was likely due to the earlier data being unsuitable for such applications. It also demonstrates that the same dataset, analyzed in different ways, can show vastly different levels of conformity with Benford's Law. Specifically, most US states show high degrees of conformity for the COVID-19 cases and cumulative deaths when the Robust Order of Magnitude (ROM) is over 3 and data at the county level is used to analyze state outcomes. Conversely, when the county data is aggregated to the state level and analyzed (i.e. case totals for all counties are summed to create a single state figure for each day of the pandemic), every state shows non-conformity. Only new deaths showed the reverse pattern - this is likely because new deaths at the county level do not span sufficient orders of magnitude, and aggregation to the state level overcomes this. This suggests that some instances of non-conformity with Benford's Law in the literature may be caused by its applications to inappropriate datasets or methodological issues.

Key words: Benford's Law, COVID-19 data.

1. Introduction

Since 2020, the COVID-19 pandemic has impacted nearly every aspect of human existence, with over three quarters of a billion cases and over 7 million deaths reported to the World Health Organisation (World Health Organisation, 2024). COVID-19 data is key for understanding the pandemic, both in terms of measuring its impact but also in terms of developing strategies to mitigate its spread. However, since the early days of the pandemic, questions have been raised with respect to the quality of the data that is being reported (Campolieti, 2022; Sambridge & Jackson, 2020). There have been suggestions of undercounting and misreporting of data, which has persisted through to the present day (Neumayer & Plümper, 2022). One common tool used to identify potential anomalies is Benford's Law (Benford, 1938; Nigrini, 2012), which posits that

[1] Independent researcher. L12 309 Kent St Sydney NSW 2000, Australia. E-mail: ravi.dutta.powell@gmail.com. ORCID: https://orcid.org/0000-0003-4197-3477.

the first digits of a sufficiently large and naturally generated dataset will conform to a particular frequency.

Benford's Law has previously been applied to the H1N1 influenza epidemic (Idrovo et al., 2011), and has seen renewed interest during the COVID-19 pandemic. Early analysis using Benford's Law yielded mixed results – for example, some analysis suggested broad conformity (Sambridge & Jackson, 2020), whilst others highlighted nonconformity, but noted this was consistent across a range of countries, and did not necessarily suggest data manipulation (Koch & Okamura, 2020). Farhadi (2021) conducted an analysis covering a period up to September 2020, and found that only a minority of countries adhered well to Benford's Law, whilst Isea (2020) found that Italy, Portugal, Netherlands, United Kingdom, Denmark, Belgium and Chile all did not conform to Benford's Law based on data up until April 2020. Campolieti (2022) found that nearly every US state up until June 2020, as well as New York City and the District of Columbia, deviated substantially from Benford's Law with respect to new daily cases, although as will be demonstrated this is likely due to the data not providing a sufficiently large sample size and not covering sufficient orders of magnitude – two key prerequisites for conformity with Benford's Law. Similarly, early data for countries in the European Union found a wide variance in conformity, with the perhaps counterintuitive result that countries with higher vaccination rates saw deviations from Benford's law (Kolias, 2022). Neumayer and Plumper (2022) found that autocratic regimes appear to have lower COVID-19 mortality rates on the surface, but in fact an analysis of excess mortality suggests that this is largely due to data manipulation.

Later work has shown more widespread conformity, as more data has emerged. Farhadi and Lahooti (2021) showed that as time had progressed, more countries showed greater conformity to Benford's Law, and two years in most countries showed conformity with Benford's Law (Farhadi & Lahooti, 2022a). Similarly, Campanelli (2023) found that a majority of countries showed conformity with Benford's Law when using the Euclidean Distance statistic, with only a minority showing significant non-conformity. Others have shown that conformity with Benford's Law is often positively correlated with indices of development (Balashov et al., 2021), or with indices of democratic freedom (Kilani, 2021). In the US, research has suggested that partisan bias may have led to data manipulation and misreporting of data at the county level (Eutsler et al., 2023).

In this paper, we review data for the United States of America (henceforth the US), covering the full period of the pandemic (January 2020 – March 2024), and revisit previous analyses. Section 2 below outlines the data sources used and the methodology, where we identify the key requirements of Benford's Law and appropriate measures. Section 3 covers how this approach is then applied to the various sources of US data, where we identify potential shortcomings of previous applications. We also demon-

strate multiple ways of analyzing the same dataset using Benford's Law, and demonstrate how different approaches can lead to radically different results. Section 4 provides a discussion of the results and our conclusions.

2. Data and methods

This paper uses two datasets to assess cases and deaths due to COVID-19 – the Centre for Disease Control's (CDC) dataset on daily cases and deaths at a state level, and the New York Times (NYT) COVID-19 cases and deaths tracker, available at https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties.csv. CDC data for the period 6 March 2020 to 5 August 2020 was originally used by Campolieti (2022), and a more recent update of the data has also been included. The New York Times data covers slightly over three years, and has been used by several authors more recently (Eutsler et al., 2023; Rocha Filho et al., 2023). The datasets differ in terms of their coverage and figures – the CDC has its own collection methods, provides only state level data, and provides data on new cases and deaths, whilst the New York Times sources data directly from states and counties, provides county level data, and provides the cumulative tally of cases and deaths. All data and analysis code is available at https://osf.io/j2s9d/.

One potential shortcoming of cumulative data, particularly with respect to the spread of an infectious disease, is that the data can have "plateaus" where there are few new deaths or cases. The cumulative total can then be "stuck" in a particular numerical range, which may lead to certain leading digits being overrepresented and thus lead to nonconformity with Benford's Law. An alternative to cumulative data is thus to generate daily changes, which may show greater conformity (daily change data can easily be calculated from cumulative data, and vice-versa). Copies of the datasets used, and all analysis code, are available at https://osf.io/592bu/.

To assess whether data complies with Benford's Law, there are several criteria that must be met: the data should have a large sample size, cover several orders of magnitude, have a positively skewed distribution, and be the result of some natural or semi-random process (Goodman, 2016). The first two criteria are particularly notable, as whilst there is agreement on the overall principles, there are a range of definitions for acceptable thresholds. A common threshold for minimum sample size of 1,000 is often cited (Farhadi, 2021; Nigrini, 2012), but others have suggested lower thresholds of 500 (Cerqueti & Provenzano, 2023). Regardless, it is likely that there is no hard and fast rule, and that some smaller samples will show conformity whilst some other, larger samples will not. Similarly, there is no definitive answer for how many orders of magnitude (OOM) a dataset must span to ensure compliance, although it is generally agreed that there should be at least 3 (Farhadi & Lahooti, 2022b; Fewster, 2009; Kossovsky, 2021). Notably, Kossovsky (2021) suggests using the "Robust Order of Magnitude" (ROM), which has the advantage of excluding any outliers, and requiring

the ROM to exceed at least 3 to deem a dataset suitable for Benford analysis (though it may be possible to observe conformity at values below 3). The ROM is defined as:

$$\text{ROM} = Log_{10}(P_{99\%} \,/\, P_{1\%})$$

where $P_{99\%}$ and $P_{1\%}$ represent the 99th percentile and 1st percentile of the data, respectively.

There are a range of methods for assessing compliance with Benford's Law, but research has identified issues with some. For example, some research has used Pearson's x^2 test as a prominent method for evaluating the conformity of data with Benford's Law (Campolieti, 2022). However, the x^2 test is limited, in ways that are particularly problematic for use with data expected to conform with Benford's Law – larger datasets are more likely to show conformity with Benford's Law, whereas the χ^2 test is more likely to show non-conformity as the sample size increases, making it almost paradoxically unsuitable for evaluating conformity with Benford's Law (Cerqueti & Lupi, 2023; Koch & Okamura, 2020; Kossovsky, 2021; Nigrini, 2012). Notably, the χ^2 test also typically has a minimum sample size requirement, with each potential "cell" (in this case, the 9 potential values for the frequency of each leading digit) required to have a value of 5 or more (McHugh, 2013). For a test of Benford's Law, where there are 9 potential cells, this implies a minimum sample size of 45.

A more common alternative is the Mean Absolute Deviation or MAD (Kolias, 2022; Kossovsky, 2021; Nigrini, 2012), which is not sensitive to sample sizes, although recent research suggests that at very large sizes it can also have limitations (Cerqueti & Lupi, 2023). Instead of significance testing, a series of thresholds have been to classify the levels of conformity: close conformity (MAD < 0.006), acceptable conformity (0.006 < MAD < 0.012), marginally acceptable conformity (0.012 < MAD < 0.015), and nonconformity (MAD > 0.015) (Nigrini, 2012). Note these are not the only thresholds for MAD – earlier research has used lower thresholds (Drake & Nigrini, 2000), however more recently the consensus in the literature has coalesced around the thresholds listed (see, for example, Campolieti, 2022; Cerqueti & Provenzano, 2023; Kolias, 2022, Kossovsky, 2021). As such, to ensure consistency with the broader literature and specific works cited, we used these thresholds. MAD is defined as:

$$\text{MAD} = \frac{1}{9}\sum_{i=1}^{9} |Ex - Obs|$$

where *Ex* is the expected proportion for any given leading digit, and *Obs* is the actual observed proportion for that leading digit.

More recently, the Sum of Squared Deviations (SSD) has emerged as a more scale-agnostic alternative to MAD (Kossovsky, 2021). Compared to MAD, it also treats fewer, larger deviations from Benford's Law more severely than several smaller deviations - the latter of which is much more likely to occur naturally due to random chance (Kossovsky, 2021). SSD is defined as

$$\text{SSD} = \sum_{i=1}^{9} (Ex - Obs)^2 \times 10^4$$

Hence this paper will apply both SSD and MAD, but with preference given to the findings of SSD, given its simplicity of not dividing deviations by the number of digits involved, especially in cases of a large base in a number system. Similarly, datasets will note whether they have sufficient sample size (N > 1,000), and have sufficient robust orders of magnitude (i.e. ROM > 3).

3. Results

3.1. Early analysis

We first review the data on daily deaths at the state level in the US, initially analyzed in the early period of the COVID-19 pandemic (Campolieti, 2022) to evaluate whether it meets the minimum requirements for potential conformity with Benford's Law. Table S1 in the supplementary materials provides a summary of each state, as well as Washington DC and New York City, along with brief summary statistics of the total sample, range, and OOM/ROM measures.

Concerningly, none of the states even comes close to the minimum N of 1,000 – every state has an N < 150 (full details supplementary materials). In fact, 5 states have N < 45, making the use of x^2 inappropriate. Similarly, no state has a ROM over 3 – only 3 states exceed even a ROM of 2.5. In fact, several states have data that *does not even cover a single order of magnitude.* Thus, it is perhaps unsurprising that Campolieti (2022) previously found widespread nonconformity with Benford's Law – in many cases, the data was such that it was not even mathematically possible for conformity to exist. It is noteworthy that other research around this time also found deviations from Benford's Law in the US for some time periods (Koch & Okamura, 2020).

3.2. Later data

Subsequently, as the pandemic progressed, more data has become available. We now turn to an analysis of CDC data covering a longer period (until October 2022). For deaths, no states had a N > 1,000 (although some came close, with over N > 900), and no states had a ROM of over 3 - just three states and New York City had a ROM above even 2.5, making deaths likely unsuitable for Benford's Law (see supplementary materials for full details).

Similarly, no state had cases where N > 1,000, but 22 had a ROM of over 3. Table 1 shows a summary of the results, with full results for all states in the supplementary materials. For states with a ROM over 3, 12 showed nonconformity on MAD, but just two of them showed nonconformity on SSD. For the 30 states with a ROM below 3, the results were different – 5 states showed nonconformity on SSD and 18 states showed nonconformity on MAD. Of note, states that showed nonconformity on SSD were a subset of those who showed nonconformity on MAD.

Table 1: Summary of results for states from CDC data (cases)

Specification	MAD		SSD	
Subset	**Conforming**	**Nonconforming**	**Conforming**	**Nonconforming**
ROM > 3	10	12	20	2
ROM < 3	12	18	25	5
Total	22	30	45	7

MAD nonconformity: MAD > 0.015. SSD nonconformity: SSD > 100. Full data are contained in supplementary materials.

However, when combining all states and analyzing the US as a whole, leading to N = 42,889 and ROM = 3.879, the results show very close conformity with Benford's Law (MAD = 0.002, SSD = 0.608).

3.3. County level data

Finally, we consider the NYT county-level data, covering a period up to March 2023. This allows for data from each county in a state to be pooled together, resulting in substantially larger sample sizes for each state of well over N > 1,000, with N > 10,000 for the vast majority of states. All states also have full coverage of first digits in the lowest order of magnitude, meaning that any non-conformity is not due to a lack of coverage of digits in the lowest order of magnitude (Goodman, 2023). The data is provided as a cumulative daily tally of all cases and deaths, by county.

3.3.1. Couty level data – deaths

Table 2: Summary of results for states from NYT data (cumulative daily deaths, by county)

Specification	MAD		SSD	
Subset	**Conforming**	**Nonconforming**	**Conforming**	**Nonconforming**
ROM > 3	23	8	28	3
ROM < 3	7	12	13	6
Total	30	20	41	9

MAD nonconformity: MAD > 0.015. SSD nonconformity: SSD > 100. Full data is contained in supplementary materials.

Table 2 summarizes the results for cumulative daily deaths by county, analyzed at the state level (i.e. each county's daily figure is treated as a separate observation, and combined with other county-day observations within the state to create the total dataset). Thirty-one states show a ROM of over 3. Of these, just 3 states show nonconformity on SSD and MAD (Connecticut, Massachusetts and Hawaii), with 5 further

states (Nevada, Oregon, Louisiana, Ohio and Kansas) showing nonconformity on MAD only. Of note, Connecticut, Massachusetts, and Hawaii are the three smallest states by N in this group.

For completeness, we also review the 19 states with a ROM below 3. Of these, 6 show non-conformity on SSD and MAD, whilst a further 6 show nonconformity on MAD only. Notably, just one state (Kentucky) shows close conformity on MAD (none do so on SSD) - interestingly, it has the second highest N of this group.

Notably, the US as a whole (N = 3,162,163, ROM = 3.436) shows strong conformity with Benford's Law (MAD = 0.003, close conformity; SSD = 2.35, acceptable).

Another option for analyzing this information is to aggregate the daily county-level observations into a single, state-level observation. That is, for each given date, we can sum all observations across every county in each state, to get a single figure for cumulative deaths by state. This does reduce the overall sample size for each state, but each state still has N > 1,000. Strikingly, every single state shows nonconformity on MAD and SSD using this analysis, suggesting that this approach may not be suitable. However, once again, the US as a whole (N = 54,976, ROM = 4.052) shows acceptable conformity (MAD = 0.0197, SSD = 16.97).

Alternatively, instead of cumulative deaths, we can analyze new daily deaths, again using individual county-level and then aggregated state-level data. Perhaps unsurprisingly, county-level new death data does not show conformity for any states - this is to be expected, however, as despite most states have sizeable N, they all have limited ROMs. Not a single state had a ROM over 3 (few states even had a raw OOM of over 3), and several states barely covered even a single order of magnitude. Hence in this situation, it may be appropriate to aggregate county-level data to state level, as it will increase the ROM for most states, even though it will reduce the N substantially.

Table 3: Summary of results for states from New York Times data (new daily deaths, aggregated)

Specification	MAD		SSD	
	Conforming	**Nonconforming**	**Conforming**	**Nonconforming**
All states	28	22	44	6

MAD nonconformity: MAD > 0.015. SSD nonconformity: SSD > 100. Full data is contained in supplementary materials.

Surprisingly, despite this leading to N < 1,000 for nearly every state, and no state having a ROM over 3, there was a moderate degree of conformity (Table 3). It should be noted that the ROMs for states were generally higher in aggregated new deaths scenario, as compared to county-level new deaths, which may have contributed to this result.

3.3.2. Cases

For cumulative cases, all states have very high *Ns* and high ROMs. This analysis shows just 2 states with a non-conforming SSD for cases – Connecticut and Hawaii, although Delaware is also very close to the threshold (Table 4). Notably, these three states represent the three of the four smallest by *N*, suggesting that again a very large sample size can be required to find a high degree of compliance with Benford's Law. Under MAD, these three states are also non-conforming, along with a further 8 - Rhode Island, New Hampshire, Vermont, Arizona, Massachusetts, Nevada, New Jersey, and Iowa. However, 10 of these states are in the bottom 11 by *N*, again suggesting that even large *Ns* may not always show conformity. The outlier is Iowa. One potential explanation for Iowa's nonconformity is the fact that they are a particularly rural state, with a large number of sparsely populated counties. Hence, these individual counties are likely to have a large number of relatively small numbers of cases, leading to fairly stable cumulative case counts and potentially causing non-compliance with Benford's Law.

Table 4: Summary of results for states from New York Times data (cumulative daily cases, county-level)

Specification	MAD		SSD	
	Conforming	Nonconforming	Conforming	Nonconforming
All states	39	11	48	2

MAD nonconformity: MAD > 0.015. SSD nonconformity: SSD > 100. Full data is contained in supplementary materials.

As with cumulative deaths above, we also try aggregating cumulative case data at the county-level to generate daily state-level observations. Strikingly, once again every single state shows nonconformity, despite all states having $N > 1{,}000$ and ROM of well over 3 (most are well over 3.5). However, despite this, the US as a whole ($N = 55{,}876$, ROM = 5.82) shows acceptable conformity with Benford's Law (MAD = 0.0063, SSD = 6.244).

Table 5: Summary of results for states from NYT data (new daily cases, by county)

Specification	MAD		SSD	
Subset	Conforming	Nonconforming	Conforming	Nonconforming
ROM > 3	14	1	15	0
ROM < 3	27	8	33	2
Total	41	9	48	2

MAD nonconformity: MAD > 0.015. SSD nonconformity: SSD > 100. Full data is contained in supplementary materials.

When considering new cases, 15 states have a ROM above 3, with all showing conformity on SSD and just one (Texas) showing nonconformity on MAD (Table 5). Even states with a ROM below 3 show fairly high levels of conformity - just two states show nonconformity on SSD, and 8 on MAD. And, as with other results, the US as a whole (N = 1,764,406, ROM = 2.914) shows acceptable conformity (MAD = 0.0105. SSD = 19.39).

Finally, new cases can also be aggregated at the state level. In line with most of the other results, states generally have greater nonconformity when considering aggregated data, as compared to when county-level data remains disaggregated (Table 6).

Table 6: Summary of results for states from NYT data (new cases, aggregated)

Specification	MAD		SSD	
Subset	**Conforming**	**Nonconforming**	**Conforming**	**Nonconforming**
ROM > 3	15	17	32	2
ROM < 3	11	7	16	2
Total	26	24	48	2

MAD nonconformity: MAD > 0.015. SSD nonconformity: SSD > 100. Full data is contained in supplementary materials.

4. Discussion and Conclusion

The results of this work show how different views of the same data can lead to very different results, and the importance of selecting an appropriate sample for analysis of conformity with Benford's Law. Based on the results, it seems that if data is available at different levels (e.g., state and national, or state and county), then the level of data below the level of interest should be used if it provides sufficient orders of magnitude (ideally 3, but it is possible to show conformity below this). That is, state level data should be used to calculate national level results, and county level data should be used to calculate state level results. Data should not be aggregated and summed together to a higher level (for example, summing daily county-level observations to get a single state level observation for each day), unless absolutely necessary. In general, when each county-day observation was treated as its own datapoint, most states showed more conformity with Benford's Law. In contrast, when observations for counties were summed each day to produce a single, aggregated daily figure for states, conformity was much lower. Perhaps most strikingly, cumulative deaths and cumulative cases saw the majority of states showing conformity when using county-level data, but literally every state showed nonconformity on both cases and deaths when analyzing aggregate data (despite the US as a whole also showing conformity).

However, there was one exception to this, namely the NYT data on new deaths. In this case, county level data does not cover sufficient orders of magnitude – as such,

aggregating several observations together can make sense and may in fact be necessary. When county-day observations for new deaths (which spanned fewer orders of magnitude than new cases, or cumulative cases/deaths) were treated as individual data points, no state showed conformity. This was expected, as no state had a ROM of over 3 and in fact several had a ROM below 1. However, when new deaths per county were summed to provide a single, daily state figure, many states actually showed some degree of conformity, despite nearly every state having $N < 1{,}000$ and ROM < 3.

Overall, this highlights the importance of large sample sizes and orders of magnitude, particularly the ROM measure. Some datasets with very large *Ns* (e.g., new daily deaths at the county level) but low orders of magnitude show high nonconformity, whilst others with high ROMs but relatively lower *Ns* (e.g., cumulative cases, aggregated to the state level) also showed nonconformity. Whilst there were some cases where despite a low ROM and a low N, some states did show conformity with Benford's Law, these were in the minority. Indeed, just because some data with low *N* and limited ROM complies with Benford's Law does not mean that we should expect all data to comply (Goodman, 2023). Of note, whenever states were divided into groups based on a ROM above or below 3, there was typically greater conformity in the group with ROMs above 3, suggesting that the thresholds for *N* and ROM, whilst important, are more in the nature of guidelines than hard and fast rules.

This highlights the importance of a comparative approach, using multiple, similar datasets to test conformity, and using the most appropriate level of analysis where multiple levels are available. The results of this work confirm more recent work that has found broad conformity with Benford's Law for the US as a whole and for most states (Rocha Filho et al., 2023). It fits with the broader trend of early indications of non-conformity, but later data showing clear conformity (Farhadi & Lahooti, 2021, 2022a). It reinforces the need to ensure that data has the potential to actually be compliant with Benford's Law, before testing for compliance (Ausloos et al., 2021), and indeed helps to identify why earlier analysis was more prone to finding deviations from Benford's Law. It can be tempting to try and find sources of non-compliance, and forget the underlying prerequisites of Benford's Law. It is likely that a thorough review of previous work of Benford's Law, particularly from early in the COVID-19 pandemic but potentially in a range of fields, would likely find many such examples of its application to datasets that were not at all suitable, resulting in spurious claims of nonconformity.

References

Ausloos, M., Ficcadenti, V., Dhesi, G. and Shakeel, M., (2021). Benford's laws tests on S&P500 daily closing values and the corresponding daily log-returns both point to huge non-conformity. *Physica A: Statistical Mechanics and Its Applications*, 574, 125969. https://doi.org/10.1016/j.physa.2021.125969.

Balashov, V. S., Yan, Y. and Zhu, X., (2021). Using the Newcomb–Benford law to study the association between a country's COVID-19 reporting accuracy and its development. *Scientific Reports*, 11(1), 22914. https://doi.org/10.1038/s41598-021-02367-z.

Benford, F. R., (1938). The Law of Anomalous Numbers. *Proceedings of the American Philosophical Society*, *78*(4), pp. 551–572.

Campanelli, L., (2023). Breaking Benford's law: a statistical analysis of COVID-19 data using the Euclidean distance statistic. *Statistics in Transition new series*, 24(2), pp. 201–215. https://doi.org/10.59170/stattrans-2023-028.

Campolieti, M., (2022). COVID-19 deaths in the USA: Benford's law and under-reporting. *Journal of Public Health*, 44(2), e268–e271. https://doi.org/10.1093/pubmed/fdab161.

Cerqueti, R., Lupi, C., (2023). Severe testing of Benford's law. *Test*, 32(2), pp. 677–694. https://doi.org/10.1007/.

Cerqueti, R., Provenzano, D., (2023). Benford's Law for economic data reliability: The case of tourism flows in Sicily. *Chaos, Solitons & Fractals*, 173, 113635. https://doi.org/10.1016/j.chaos.2023.113635.

Drake, P. D., Nigrini, M. J., (2000). Computer assisted analytical procedures using Benford's Law. *Journal of Accounting Education*, 18(2), pp. 127–146. https://doi.org/doi:10.1016/s0748-5751(00)00008-7.

Eutsler, J., Kathleen Harris, M., Tyler Williams, L. and Cornejo, O. E., (2023). Accounting for partisanship and politicization: Employing Benford's Law to examine misreporting of COVID-19 infection cases and deaths in the United States. *Accounting, Organizations and Society*, 108, 101455. https://doi.org/10.1016/j.aos.2023.101455.

Farhadi, N., (2021). Can we rely on COVID-19 data? An assessment of data from over 200 countries worldwide. *Science Progress*, 104(2), 00368504211021232. https://doi.org/10.1177/00368504211021232.

Farhadi, N., Lahooti, H., (2021). Are COVID-19 Data Reliable? A Quantitative Analysis of Pandemic Data from 182 Countries. *COVID*, 1(1), Article 1. https://doi.org/10.3390/covid1010013.

Farhadi, N., Lahooti, H., (2022a). Forensic Analysis of COVID-19 Data from 198 Countries Two Years after the Pandemic Outbreak. *COVID*, 2(4), Article 4. https://doi.org/10.3390/covid2040034.

Farhadi, N., Lahooti, H., (2022b). Reply to Morillas-Jurado et al. Benford Law to Monitor COVID-19 Registration Data. Comment on "Farhadi, N.; Lahooti,

H. Forensic Analysis of COVID-19 Data from 198 Countries Two Years after the Pandemic Outbreak. COVID 2022, 2, pp. 472–484". *COVID*, 2(7), Article 7. https://doi.org/10.3390/covid2070070.

Fewster, R. M., (2009). A Simple Explanation of Benford's Law. *The American Statistician*, *63*(1), pp. 26–32. https://doi.org/10.1198/tast.2009.0005.

Goodman, W., (2016). The promises and pitfalls of Benford's law. *Significance*, 13(3), pp. 38–41. https://doi.org/10.1111/j.1740-9713.2016.00919.x.

Goodman, W., (2023). Applying and Testing Benford's Law Are Not the Same. *Spanish Journal of Statistics*, 5(1), pp. 43–53. https://doi.org/10.37830/SJS.2023.1.03.

Idrovo, A. J., Fernández-Niño, J. A., Bojórquez-Chapela, I. and Moreno-Montoya, J., (2011). Performance of public health surveillance systems during the influenza A(H1N1) pandemic in the Americas: Testing a new method based on Benford's Law. *Epidemiology and Infection*, 139(12), pp. 1827–1834. https://doi.org/10.1017/S095026881100015X.

Isea, R., (2020). How Valid are the Reported Cases of People Infected with Covid-19 in the World? *International Journal of Coronaviruses*, 1, pp. 53–56. https://doi.org/10.14302/issn.2692-1537.ijcv-20-3376.

Kilani, A., (2021). Authoritarian regimes' propensity to manipulate Covid-19 data: A statistical analysis using Benford's Law. *Commonwealth & Comparative Politics*, 59(3), pp. 319–333. https://doi.org/10.1080/14662043.2021.1916207.

Koch, C., Okamura, K., (2020). Benford's Law and COVID-19 reporting. *Economics Letters*, 196, 109573. https://doi.org/10.1016/j.econlet.2020.109573.

Kolias, P., (2022). Applying Benford's law to COVID-19 data: The case of the European Union. *Journal of Public Health*, 44(2), e221–e226. https://doi.org/10.1093/pubmed/fdac005.

Kossovsky, A. E., (2021). On the Mistaken Use of the Chi-Square Test in Benford's Law. *Stats*, *4*(2), Article 2. https://doi.org/10.3390/stats4020027.

McHugh, M. L. (2013). The Chi-square test of independence. *Biochemia Medica*, pp. 143–149. https://doi.org/10.11613/BM.2013.018.

Neumayer, E., Plümper, T. (2022). Does 'Data fudging' explain the autocratic advantage? Evidence from the gap between Official Covid-19 mortality and excess mortality. *SSM – Population Health*, 19, 101247. https://doi.org/10.1016/j.ssmph.2022.101247.

Nigrini, M. J., (2012). *Benford's Law: Applications for forensic accounting, auditing, and fraud detection* (Vol. 586). John Wiley & Sons.

Rocha Filho, T. M., Mendes, J. F. F., Lucio, M. L. and Moret, M. A., (2023). COVID-19 data, mitigation policies and Newcomb–Benford law. *Chaos, Solitons & Fractals*, 174. http://arxiv.org/abs/2208.11226.

Sambridge, M., Jackson, A., (2020). National COVID numbers—Benford's law looks for errors. *Nature*, 581(7809), pp. 384–385.

World Health Organisation, (2024). *COVID-19 deaths reported (2024 global)*. WHO COVID-19 Dashboard. https://data.who.int/dashboards/covid19/circulation.

STATISTICS IN TRANSITION new series, September 2025
Vol. 26, No. 3, pp. 15–34, https://doi.org/10.59139/stattrans-2025-025
Received – 08.08.2024; accepted – 04.02.2025

Education expansion and income inequality: evidence from Poland (2005–2019)

Leszek Morawski[1]

Abstract

In Poland, between 2005 and 2019, the share of the employed with tertiary education among men increased by about 10 percentage points and among women by about 18 percentage points. We study the impact of this change on relative income poverty and income inequality using a microsimulation decomposition based on a tax-benefit microsimulation model. We show that the educational change reduced poverty and income inequality incidence and depth. Using a microsimulation approach, we estimate that the impact of the above-mentioned educational change on the changing material poverty risk corresponds to 40% of the policy effect associated with changes in tax and benefit regulations. For the Gini index, the educational effect amounted to 91% of the policy effect. The results show that educational changes in Poland between 2005 and 2019 have significantly reduced income inequality and the risk of material poverty.

Key words: relative income poverty, income inequality, tertiary education, tax and benefit regulations, Poland.

1. Introduction

Between 2005 and 2019, Poland experienced an almost uninterrupted period of economic growth, with GDP per capita increasing by 76% and the employment rate rising from 52.8% to 68.2%. According to PHBS (Polish Household Budget Survey), among men, the percentage of employees with a college degree increased from 15.8% in 2005 to 26.0% in 2019. Among women, the change was from 26.7% to 44.7%. At the same time, the shares of employees with at most basic vocational education declined - among men from nearly 50% to 39%, and among women from 28.1% to 17.1%. During this period, income inequality, as measured by the Gini coefficient, decreased from 33.3% in 2006 to 28.5% in 2019 and relative income poverty from 19.1% to 15.4% [2]. Such large changes in the educational structure of the labor force should be reflected in changes in income inequality and the risk of material poverty. The analysis of changes in income inequality in the context of educational changes has a long tradition in economics, dating back to Kuznets, Becker, or Spence. While there is consensus that better education and greater accessibility promote economic development and improve the general standard of living of the population, there is no unanimity on the relationship between change in the educational structure of employment and disposable income disparities.

[1]Faculty of Economic Sciences, University of Warsaw, Poland. E-mail: l.morawski@uw.edu.pl. ORCID: https://orcid.org/000-0000-0003-3464-3963.

[2]Values based on EU-SILC variables ilc_li02 and ilc_di12 from the EUROSTAT database.

In this study, using a microsimulation approach, we examine the consequences of an increase in the share of workers with higher education on changes in income inequality and material poverty risk. The analysis presented looks at changes in Poland between 2005 and 2019, a time of rapid economic growth, significant educational expansion, and favorable changes in inequality and income poverty. This paper aims to contribute to the ongoing debate on the relationship between educational expansion and income inequality.

In what follows, we use the tax-benefit microsimulation approach to analyze the contribution of educational expansion among the employed to changes in income inequality and material poverty. Such an approach has been successfully applied in the previous policy impact studies (Figari et al. 2015). In this paper we use the decomposition method developed in Bargain and Callan (2010). We are extending it by combining it with the reweighting method of DiNardo et al. (2006). We compare the inequalities in the data with those we would have observed if the educational structure of the labor force and the conditional probability of employment had remained unchanged from 2005 to 2019. We analyze the changes in the values of the income inequality indices and material poverty indicators, isolating the direct contribution of changes in tax and benefit (Bargain et al. 2017, Figari et al. 2015).

We describe changes in income inequality using the Gini index and two alternative measures: the Theil index and the Atkinson index. The analysis uses individual and household microdata from the 2005 and 2019 Household Budget Surveys (PHBS). Each survey covers about 100,000 individuals in 30,000 households each year. We show that expanding education has contributed to reducing material poverty and income inequality over the period considered. The magnitude of the educational effect was about 40% of the direct policy effect for relative monetary poverty and was comparable to it when the Gini coefficient measured income inequality.

In the following sections, we present a literature review and background information on changes in the structure of education and their impact on the distribution of gross wages. We also summarize changes in tax and benefit systems. The following section presents the method and the data. Finally, we present the results, interpretations, and conclusions.

2. Literature

The role of education in economic development is deeply rooted in several theoretical frameworks. Whether we refer to growth theory, human capital theory, or signaling theory, we will generally conclude that the expansion of education has a positive impact on individual income and the average income of society. It is much more difficult to predict the impact of changes in the structure of education, for example when the proportion of people with tertiary education increases, on inequality and income poverty.

2.1. Theoretical Perspectives

The work of Kuznets (1955), who suggested that income inequality initially increases as economies grow and eventually decreases, provides the first insight into the relationship between educational changes and changes in income inequality. According to Knight and Sabot (1983), the education expansion has two conflicting effects: the composition effect

and the compression effect. The composition effect prevalent in the first phase of development increases inequalities because there is then a strong demand for skilled workers with a relatively small supply of them. However, after reaching a certain threshold, the increased supply of skilled workers decreases the wage premium for higher-skill workers and thus lowers income inequality; this is the compression effect. The relative strength of these effects explains the directions of change in inequality postulated by Kuznets. Overall, even though initial education expansion is associated with increasing inequality, eventually, it will lead to lower inequality.

This prediction about the long-run relationship is entirely consistent with the postulates of human capital theory, as formulated by Becker (1964). According to this theory in the long term, broader access to education leads to a more equal income distribution by reducing the wage premium for higher education. Schultz (1961), using his modernization theory, was less optimistic. He agreed that education enables structural transformation by reallocating labor from low-productivity to high-productivity sectors. According to him education development promotes economic growth and provides an opportunity to reduce inequality, but only if it provides opportunities to acquire new valuable skills. Inequality can persist or worsen if the supply of educated workers does not match the demand in the labor market.

Signaling theory, introduced by Spence (1973), offers another explanation for why educational expansion does not necessarily cause a decline in income inequality. The theory posits that employers interpret educational attainment as a signal of certain desirable traits, such as intelligence, diligence, and ability to complete tasks. This signaling mechanism influences hiring decisions, wage offers, and career progression. In societies with limited access to education, higher education serves as a distinctive signal for a relatively small group, allowing them to command higher wages and enjoy significant income advantages. However, as education becomes more accessible and a larger share of the population pursues higher qualifications, the signaling value of a degree declines, and the wage premium associated with higher education declines. This phenomenon is known as credential inflation. The predictions of signaling theory are less optimistic than those previously cited. According to it, the expansion of education has a somewhat ambiguous effect on income inequality, and its direction depends largely on the dynamics of the labor market. As higher education becomes more widespread, inequality could persist or worsen if employers tighten qualification requirements, limiting access to good jobs for people from low-income households.

Each of the three main theories provides support for the expectation that an increase in the share of highly educated people will reduce income inequality in the long term. At the same time, each approach recognizes that in the short term, such a change may be associated with an increase in income inequality.

2.2. Empirical Evidence

The ambiguity of the empirical research results is not surprising in the context of diverse theoretical conclusions. However, results showing decreasing inequality when the average level of education increases outweigh those showing the opposite conclusion. Early studies by Knight and Sabot (1983) and Ram (1984) confirmed the Kuznets hypothesis for developing countries. Similar results were obtained by Rehme (2007). Abdullah et al. (2015), using

a meta-regression analysis, found that education reduces top earners' income share and increases bottom earners' share. Hovhannisyan et al. (2019), who analyzed the relationship between education and income inequality in 38 countries (developed and developing) from 1990 to 2014, found that higher educational attainment was consistently associated with lower income inequality. Countries with higher enrollment rates (e.g. Sweden and Finland) exhibited lower Gini coefficients than countries with lower enrollment rates (e.g. Zambia and Guatemala). Lee and Lee (2018), using panel data analysis, studied 95 countries from 1980 to 2015. They confirmed that expanding educational attainment helps reduce educational inequality, which diminishes income inequality. However, Samano-Robles (2018) received ambiguous results for 18 Latin American countries between 2000 and 2010. They found an increase in income inequality in six countries and a decrease in four countries. Makhlouf and Lalley (2023) challenged the wisdom that education expansion reduces income inequality. This study provides insights into the dynamic relationship between these variables using the autoregressive distributed lag (ARDL) model to explore the short- and long-term effects of education inequality on income inequality, incorporating structural transformation as a mediating factor. It examines the relationship between education expansion, income inequality, and structural transformation in 20 OECD countries from 1870 to 2016. It found that expanding education has no immediate effect on reducing income inequality and that in the long run, a negative relationship emerges. The authors justify these results by claiming that education expansion accelerates the shift of workers from low-wage, low-inequality sectors like agriculture to high-wage, high-inequality sectors like services and industry. According to them, the proliferation of education leads to credential inflation, where more individuals compete for the same set of high-paying jobs. This results in lower relative returns for many educated people while preserving high returns for elite disciplines, further widening inequality. Their results align with Kuznet's theory and Schultz's modernization theory. Also, unevenly distributed gains from the education expansion due to credential inflation and the unequal returns to education across sectors are consistent with the signaling theory.

Research on the effects of changing the educational structure is much rarer. Abdullah et al. (2015) found that secondary schooling has a more substantial effect on income inequality than primary schooling, although this finding is not always robust. Hershbein et al. (2020), observed that an increase in tertiary education levels in the United States significantly improved the income situation of lower-income earners, thereby reducing poverty and narrowing income gaps. According to Sianesi and Reenen (2003), the impact of the endowments associated with different levels of education (primary, secondary, and tertiary) on income inequality seems to depend on the country's development level. Tasseva (2021) considered the impact of education expansion on income inequality in Great Britain from 2001 to 2017. She found that the expansion of education increased household incomes. However, the benefits were disproportionately larger at the top of the income distribution, contributing to widening income inequality, particularly in the pre-crisis (2001–2007) and post-crisis (2011–2017) periods. The primary reason was unequal access to quality education and disparities in labor market opportunities for high-skilled jobs. The decrease in inequality due to the compression effect of education was small.

3. Background information

The empirical analysis in this study covers the years 2005-2019, during which the global economic downturn following the 2008 financial crisis occurred. Before the crisis, in 2007, the GDP growth rate was close to 7%. In the following years, the rate gradually declined, reaching almost zero in 2013. After 2013, the GDP growth rate started an upward trend, reaching 6% in 2018. In line with this trend, the unemployment rate declined from 18% in 2005 to below 8% in 2008 and remained around 10% between 2009 and 2013. After that, the unemployment rate gradually decreased, reaching around 5% in 2019. During the economic slowdown, the consolidated gross debt of the general government increased to 55.1% of GDP in 2011, and it dropped to 45.7% of GDP in 2019. The general government deficit increased to 5.0% in 2011, and it was 0.7% in 2019. Although the Polish economy did not fall into recession, the ability to finance new social transfers was significantly reduced between June 2009 and June 2015, when Poland was subject to the EU's Excessive Deficit Procedure (EDP). It made the Ministry of Finance reluctant to accept any changes that lead to higher social spending (Szarfenberg 2023). Despite this, income inequality and the risk of relative income poverty did not increase during this period. The Gini coefficient for disposable income changed from 31.4 in 2006 to 30.2 in 2019, and the risk of income poverty decreased from 17.3% in 2006 to 13.0% in 2019. Between 2009 and 2013, income inequality, as measured by the Gini coefficient, remained virtually unchanged. The value of this coefficient in 2009 was 31.1, and in 2013 it was 30.7. The risk of material poverty decreased slightly then - from 17.1% to 16.2% (Główny Urząd Statystyczny 2022, 2020).

The changes in the macroeconomic situation, income inequality, and material poverty risk were accompanied by a systematic improvement in the educational structure. In 2005, 16.8% of people aged 25-64 completed tertiary education; by 2019, this share risen to 32.4%. For those aged 25-34, the share increased from 25.4% to 44.6%, with the change for women from 30.5% to 54.5% and for men from 20.5% to 37.1% (OECD 2021). According to the PHBS, the proportion of employed males with tertiary education changed from 15.8% in 2005 to 26.0% in 2019, while the corresponding change among women was from 26.7% to 44.7%.

Table 1: Employment according to formal education level (Poland, 2005-2019) (%)

	Males			Females		
	2005	2011	2019	2005	2011	2019
Higher	15.80	20.5	26.0	26.7	35.8	44.7
Secondary	33.40	32.8	35.2	45.2	38.2	35.6
Vocational	40.90	37.9	33.1	20.9	19.8	16.4
Primary	9.90	8.8	5.7	7.2	6.2	3.3
	100.00	100	100	100	100	100

Source: own calculations using PHBS.

The LFS (Labor Force Survey) reveals a very similar pattern. The data indicates that 17.6% of employed men had a tertiary education in 2006 and 28.1% in 2019. Among women, the share increased from 28.1% to 46.6% (Kowalik and Magda 2021).

The tax and benefit system in 2019 was significantly different from that in 2005 (Myck and Najsztub 2016, Morawski and Brzeziński 2023). Table 2 summarizes the fundamental policy changes between 2005 and 2019, while Figure 2 illustrates the differences in budget constraints for selected types of families. The budget constraint shows the relationship between gross income from fixed-term employment and disposable income. The constraints were derived using the SIMPL microsimulation model for Poland (Bargain et al. 2007, Myck and Najsztub 2015, Myck et al. 2015, Haan et al. 2008).

Table 2: Main policy changes in the tax-benefit system in Poland, 2005–2019

Tax-benefit policy	Policy description
SSC	• Reduction of the employee disability social security contribution rate by 2 percentage points from 6.5% to 5.0% (July 2007). • Further reduction of employee disability social security contribution rate to 1.5% and a reduction of the employer rate to 4.5% (January 2008) • Raising the employer disability social security contribution rate by 2.0% to 6.5% (2012) • Introduction of new social security contribution for farmers owning from 50 ha to 300 ha (2012) • Introduction of small health insurance for farmers with farms above 6 ha (2010)
PIT	• Introduction of non-refundable child tax credit; increase in the credit amount from PLN 120 (USD 43) per child to PLN 1145.08 (USD 410) (January–October 2007) • Personal income tax schedule with three tax rates (19%, 30%, 40%) was replaced with a schedule with two rates (18%, 32%) (2009) • Making the child tax credit partially refundable and increasing the amounts for the third and any subsequent child (2014) • Lowering tax rate from 18% to 17%, the introduction of a 0% tax rate for people under 26, a higher universal tax credit for low-income taxpayers and a lower credit for high-income taxpayers
FA	• the amount of family allowance started to depend on the age of children; before that, it depended only on the number of children in a family, resulting in a significant increase in the amount of the allowance (September 2006) • Introduction of a universal one-off childbirth allowance in the amount of PLN 1000 (USD 357) • Introduction of a taper rate ("złotówka za złotówkę) for the family allowance and a longer maternity allowance (2015) • Introduction of parental benefit (PLN 500; USD 178) for second and subsequent children in a family and means-tested for the first child in a family (2016) • Introduction of a parental benefit (PLN 1000; USD 357) for non-working parents of the newly born child (2015) • Introduction of a school starter kid program, PLN 300 yearly per child attending school (2018) • making parental benefit universal (2019)

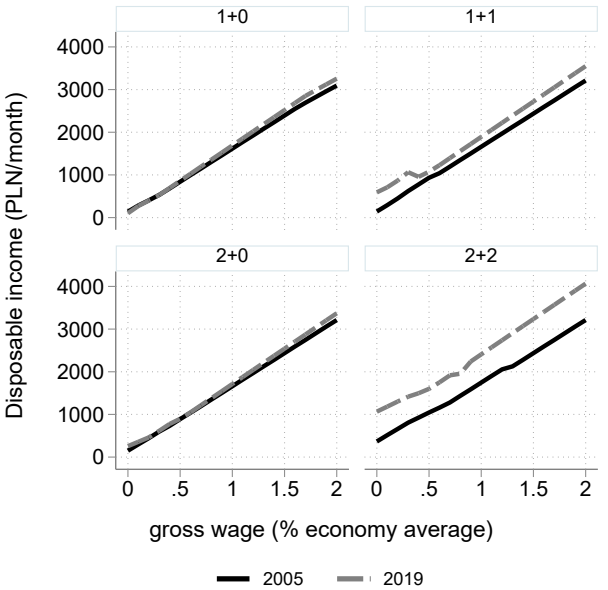


Figure 1: Budget constraints for selected families (Poland 2005 and 2019)
Source: SIMPL model.

The tax and benefit system changes between 2005 and 2019 benefited most households. Childless singles earning less than 10% of the average wage were a particular case where the changes reduced their income. Within this group, the most significant benefit of 5% went to those earning more than the national average wage. The regulatory changes were income-neutral for most single people without children and childless couples. The exception was couples on very low incomes who benefited from an increase in the universal tax credit. The changes were very beneficial for families with children. The most significant gains went to lone parents and parents on low working incomes. However, high- and even very-high-income couples also benefited. For example, the income of a couple with two children and a working partner at the average wage was almost 40% higher in 2019 than in 2005. For such families, but with a salary twice the average, the income was 26% higher.

4. Method and data

Bargain and Callan (2010) used the results from a tax-benefit microsimulation model to assess the impact of tax and benefit law changes on the distribution of household disposable income. Below, we outline that method.

Let I_t denote the value of an inequality or poverty index calculated for the income distribution d_t. In the microsimulation approach to tax-benefit analysis, we assume that income values are generated by a tax-benefit function $d_t(p_t, y_t)$, where p_t is a vector of monetary parameters embedded in tax-benefit regulations, and y_t represents the data required for sim-

ulating taxes and benefits. The data matrix y_t typically contains values for market and replacement incomes (e.g., wages, pensions, disability benefits, survivors' benefits, and social pensions) as well as variables describing the members of households (e.g., age, gender, education, marital status, and health).

A tax-benefit microsimulation model allows us to decompose changes in the value of the index ΔI into two components: the direct effect of regulatory changes and the effect of other changes, such as changes in pre-tax incomes and household characteristics. We present this decomposition as follows:

$$\Delta I_{0,1} = \Delta I^{\pi}_{0,1} + \Delta I^{O}_{0,1} \tag{1}$$

where

- $\Delta I_{0,1}$ is the change in the value of the index between the initial period 0 and the final period 1,
- $\Delta I^{\pi}_{0,1}$ is the direct contribution of regulatory changes introduced during this time,
- $\Delta I^{O}_{0,1}$ accounts for the contribution of other changes.

Bargain and Callan proposed to measure the policy effect as follows:

$$\Delta I^{\pi}_{0,1} = \frac{1}{2}\Big[I[d_1(p_1,y_1)] - I[d_0(\alpha p_0,y_1)]\Big] + \frac{1}{2}\Big[I[d_1(p_1,\alpha y_0)] - I[d_0(\alpha p_0,\alpha y_0)]\Big] \tag{2}$$

where $d_0(.)$ and $d_1(.)$ are the tax-benefit functions for the initial and final periods, respectively. The policy effect is the average marginal effect of implementing legal changes, assuming data distribution from the initial and final periods.

The choice of the index α is significant, as it defines the distribution-neutral scenario for which the policy effect equals zero. We adopt changes in the average wage in the economy so that indexing mechanisms other than wage indexing are treated as policies. For instance, price indexing based on the Consumer Price Index (CPI) applied to family benefits is seen as a policy choice. The applied indexing does not alter the value of ΔI if the tax-benefit system is linearly homogeneous and I is independent of scale. As we demonstrate below, both conditions hold in our case.

The contribution of other factors, i.e., those unrelated to regulatory changes in benefits and taxes accounted for in the microsimulation model is similarly defined:

$$\begin{aligned}\Delta I^{\pi}_{0,1} &= \frac{1}{2}\Big[I[d_0(\alpha p_0,y_1)] - I[d_0(\alpha p_0,\alpha y_0)]\Big] + \frac{1}{2}\Big[I[d_1(p_1,y_1)] - I[d_1(p_1,\alpha y_0)]\Big] \\ &= \frac{1}{2}O(p_0) + \frac{1}{2}O(p_1)\end{aligned} \tag{3}$$

The education level of individuals is one of the variables in the data matrix y_i. Using the microsimulation model, we extract from the $\Delta I^{\pi}_{0,1}$ two effects related to changes in the educational structure, which operate by affecting the distribution of gross wages. The first is the direct composition effect of an increased share of individuals with higher education.

The second is the effect on the change in employment rates by having people with higher education have a lower probability of not working. We reweight datasets to identify effects using the DFL method. For the first effect, we used three distributions of gross wages.

- The observed distribution of gross wages for 2019, reflecting the educational structure edu_1 and other wage determinants X_1. The data set is then $y_1(edu_1, X_1)$.
- A counterfactual distribution of wages in 2019, assuming the educational structure and other wage determinants were as they were in 2005. The data set is then $y_1(edu_0, X_0)$
- A counterfactual distribution of wages in 2019, assuming the educational structure was as in 2019 but other wage determinants were as in 2005. The data set is then $y_1(edu_1, X_0)$

To identify the second effect, we used two counterfactual distribution: a) $y_1(edu_0, X_0; e_1)$, in which employment probabilities conditional on education were as in 2019, b) $y_1(edu_0, X_0; e_0)$ with employment probabilities as in 2005. The extended decomposition using these additional distributions is:

$$\begin{aligned} O(p_1) &= I[d_1(p_1, y_1)] - I[d_1(p_1, \alpha y_0)] \\ &= I[d_1(p_1, y_1(edu_1, X_1))] - I[d_1(p_1, y_1(edu_1, X_0))] \\ &\quad + I[d_1(p_1, y_1(edu_1, X_0))] - I[d_1(p_1, y_1(edu_0, X_0))] \\ &\quad + I[d_1(p_1, y_1(edu_0, X_0; e_1))] - I[d_1(p_1, y_1(edu_0, X_0; e_0))] \\ &\quad + I[d_1(p_1, y_1(edu_0, X_0; e_0))] - I[d_1(p_1, \alpha y_0(edu_0, X_0; e_0))] \\ &= X(p_1) + edu(p_1) + emp(p_1) + res(p_1) \end{aligned} \tag{4}$$

The first difference measures the effect of changes in non-educational determinants of wages acting by changing the distribution of gross wages. The second difference is the changes in the education structure, which changed the distribution of gross wages (the composition effect). The third is the effect of a change in the structure of education on the probability of employment (the employment effect). The last term is the residual term or the new other effect. The decomposition of the second component of the standard other effect is:

$$\begin{aligned} O(p_0) &= I[d_0(\alpha p_0, y_1)] - I[d_0(\alpha p_0, \alpha y_0)] \\ &= I[d_0(\alpha p_0, y_1(edu_1, X_1))] - I[d_0(\alpha p_0, y_1(edu_1, X_0))] \\ &\quad + I[d_0(\alpha p_0, y_1(edu_1, X_0))] - I[d_0(\alpha p_0, y_1(edu_0, X_0))] \\ &\quad + I[d_0(\alpha p_0, y_1(edu_0, X_0; e_1))] - I[d_0(\alpha p_0, y_1(edu_0, X_0; e_0))] \\ &\quad + I[d_0(\alpha p_0, y_1(edu_0, X_0; e_0))] - I[d_0(\alpha p_0, \alpha y_0(edu_0, X_0; e_0))] \\ &= X(p_0) + edu(p_0) + emp(p_0) + res(p_0) \end{aligned} \tag{5}$$

As Figure 2 shows, changes in the educational composition of employed have reduced the share of low wages and increased the importance of wages at the top of the distribution. We expect that changes in the gross wage distribution will affect the distribution of disposable income and the values of the income inequality and poverty indices.

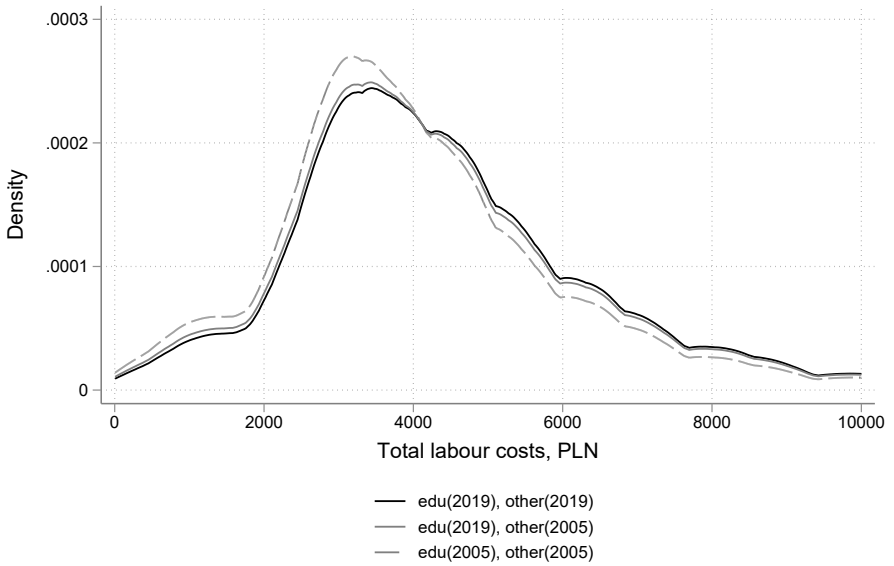
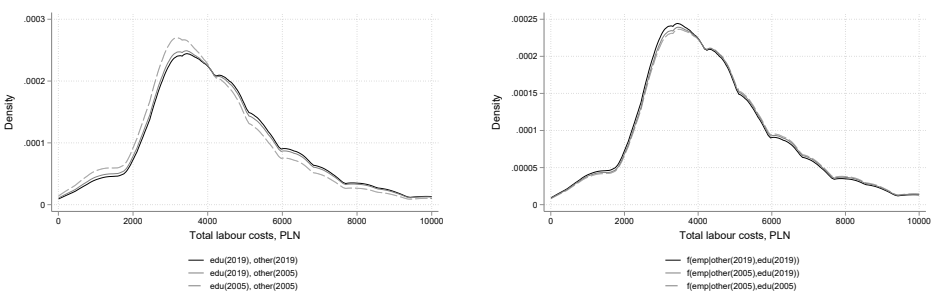


Figure 2: Actual and counterfactual wage distributions for 2019
Source: SIMPL model.



(a) Effect of change in educational structure on gross wage distribution

(b) Employment effect : $\Psi_{e|X0}, \Psi_{e|X1}$

Figure 3: Actual and counterfactual wage distributions for 2019
Source: SIMPL model.

We use the SIMPL tax and benefit microsimulation model to simulate counterfactual income distributions (Bargain et al. 2007, Myck et al. 2015, Haan et al. 2008). The model contains a computer code describing the tax and benefit rules, and the database was created using the PHBS datasets. The PHBS is an annual survey of about 30,000 households and over 100,000 individuals. We used more than 107,000 observations for 2005 and almost

94,000 for 2019. We used simulated equivalised disposable income, so the number of observations with negative or zero income is negligible. In 2005, three people had zero income, and 343 had negative income. The corresponding figures for 2019 are 93 and 373. The survey collects detailed information on individual income from work (permanent employment, temporary employment, and self-employment) and replacement income (pensions, early retirement benefits, disability benefits, unemployment benefits, and maternity pay). Family and household income, such as investment income, property income, survivors' pensions, private transfers, housing benefits, and social assistance, are also included in the data. We use the exact definition of disposable income as in the PHBS. Disposable income is net income from work, family benefits, carer's allowance, housing benefits, and cash social assistance.

Table 3 provides information on the characteristics of the most important variables in the data matrices. Changes in the values of these variables and changes in the value of pre-tax income are the causes of the other effect.

Over several years, the distributions of these characteristics are reasonably stable. However, some changes between 2005 and 2019 could affect the values of the poverty and inequality statistics. Firstly, the proportion of people living in large towns and villages has increased over the period while the proportion of people living in medium-sized towns (200 000 to 500 000 inhabitants) and small towns (20 000 to 100 000 inhabitants) has decreased. Secondly, we see an increase in the proportion of people over 55 and a decrease in the group under 35. Remarkably, there is an increase in the share of the over-65s, accompanied by an increase in the number of households living on pensions. The effect of the aging process is an increase in the proportion of marriages in which both people had retirement incomes and a decline in the importance of families with children. In this study, we are interested in the consequences of the change seen in Table 3 relating to the change in the educational structure. In the data used for the study, the share of people with tertiary education increased from 10% to almost 21% between 2005 and 2019, which, as we have already shown, was accompanied by a significant increase in the importance of people with tertiary education among the working population.

We use a standard at-the-risk poverty rate index with a poverty line of 60 percent of median equivalised disposable income to measure relative monetary poverty. We apply the modified OECD equivalence scale. An analysis of the depth of poverty supplements the incidence analysis. Following the suggestion of Jenkins (2009), we use a portfolio of distribution-sensitive indices since any index of income inequality incorporates value judgments about the relative importance of changes in different parts of the distribution. For example, the Gini index is more sensitive to changes in the middle of the distribution than other inequality measures. It is unreasonable to expect a single numerical index, such as the Gini, to capture the whole distribution in all respects. Therefore, in addition to the Gini index, we use the Theil index, the mean logarithmic deviation (MLD), and the Atkinson index. The Theil index and the MLD belong to the class of generalized inequality indices based on 'entropy,' which measures the deviation from perfect equality. Any index in this class is defined by the parameter describing the sensitivity to changes in a given part of the income distribution. The higher the parameter's value, the more sensitive the index is to changes in the upper tail of the income distribution. We use two values - 0 for the mean log

Table 3: Changes to the input data (%)

Variable	2005	2019
Female	52.32	53.25
Town size (thousand)		
500-	10.44	11.36
200-500	8.96	7.68
100-200	7.03	6.70
20-100	18.42	15.60
<20	11.16	10.28
Village	43.99	48.37
Education		
Higher	10.23	20.98
Secondary	30.89	32.07
Vocational	26.78	26.80
Primary/none	32.10	20.15
Age		
0-19	28.29	22.46
20-24	8.25	4.41
25-34	13.05	11.67
35-44	12.69	13.91
45-54	16.49	12.13
55-64	10.39	14.83
65-74	6.93	13.40
75+	3.91	7.19
Family type		
Single without children	12.62	14.05
Single parent	7.46	6.54
Childless couple	13.30	15.43
Couple with a child	54.61	40.99
Single pensioner	5.90	9.87
Pensioner couple	6.10	13.14

Source: own calculations using PHBS.

deviation and 1 for the Theil index. The value judgment is directly included in the Atkinson index via the inequality aversion parameter in the utility function. A higher value of this parameter gives higher weight to changes in the low-income part of the distribution in social welfare. We consider three values for the aversion parameter: 0.5, 1, and 2.

5. Results

The relative income poverty decreased by 5 percentage points, from 17.05 % in 2005 to 12.02% in 2019, and the poverty gap decreased from 4.59 to 3.72. Both changes were statistically significant at the 5% level.

Table 4: Poverty and inequality, Poland 2005-2019

index	**2005**	95% CI LB	95% CI UB	**2019**	95% CI LB	95% CI UB	
Poverty							
FGT 0	**17.05**	16.94	17.15	**12.02**	11.94	12.11	decrease
FGT 1	**4.59**	4.55	4.63	**3.72**	3.69	3.76	decrease
Gini	**32.09**	31.98	32.21	**30.56**	30.27	30.85	decrease
Theil 0	**17.87**	17.74	18.01	**17.80**	17.43	18.17	no change
Theil 1	**19.73**	19.44	20.02	**26.41**	25.30	27.52	increase
Atkinson (0.5)	**8.83**	8.75	8.90	**9.63**	9.38	9.87	increase
Atkinson (1.0)	**16.37**	16.24	16.49	**16.31**	16.00	16.62	no change
Atkinson (2.0)	**32.67**	32.36	32.97	**33.14**	32.36	33.91	no change
N	93674			107124			
zero income	3			93			
negative income	343			373			

Source: own calculations using SIMPL.

A decrease of 1.5 percentage points in the Gini index indicates a decrease in income inequality. This change is statistically significant. Changes in the Theil index with a parameter of 1 and the Atkinson index with a parameter of 0.5 indicate a statistically significant increase in inequality. The values of the other indices have fallen, suggesting a reduction in inequality. However, at the 5% significance level, the differences are not statistically significant.

A situation in which the value of the Theil 1 index has increased while the value of the Theil 0 index has not changed indicates interesting pattern of changes in income distribution. The increase in the Theil 1 index means that inequality between high-income earners and the rest of the population has increased. This can be explained by a relatively more significant increase in the income of the richest people in the rest of the population or an increase in the concentration of income among the richest. At the same time, the stable value of the Theil 0 index indicates that inequality at the bottom of the income distribution has remained stable. A higher value of A(0.5) means that income inequality among high earners has widened. This is exactly what we have learned from the Theil 1 index. A slight decrease in A(1) shows a small reduction in inequality across the entire income distribution. This is consistent with changes in the Gini index. An increasing value of A(2) indicates that the poorest workers have fallen further behind relative to the rest of the population. Combining these changes with changes in Theil 0, a picture emerges in which inequality at the bottom of the distribution decreases, but at the same time the distance of these people from the wealthy increases. A detailed analysis of the possible reasons for that situation is not the subject of this paper. However, it is worth noting that the results confirm earlier findings about increasing polarisation of income accompanying a decreasing value of the Gini coefficient (Panek 2017).

The results confirm favorable changes in the distribution of disposable income between 2005 and 2019 in Poland - relative poverty risk decreased and income inequality fell, as measured by the change in the value of the Gini index. However, from a welfare perspective, beneficial change in income inequality may be questioned if we attribute more importance to changes between the top and the rest of the income distribution.

Table 5 presents the main results of the extended decomposition analysis. The low homogeneity coefficients confirm that indexing the data and monetary parameters, i.e., expressing 2005 disposable income in 2019 values, did not affect the decomposition results.

Table 5: Decomposition of poverty and inequality indices

index	**change**	**homo**	**policy**	**other:**	**X**	**edu**	**emp**	**residual**
FGT 0	-5.06	0.01	-3.35	-1.71	0.86	-0.92	-0.41	-1.24
FGT 1	-0.87	0.00	-1.23	0.36	0.27	-0.29	-0.15	0.53
Gini	-1.54	0.00	-1.57	0.03	0.75	-0.93	-0.49	0.70
Theil 0	-0.001	0.000	-0.022	0.022	0.009	-0.013	-0.006	0.032
Theil 1	0.067	0.000	-0.024	0.091	0.016	-0.040	-0.013	0.127
Atkinson (0.5)	0.008	0.000	-0.010	0.018	0.005	-0.009	-0.004	0.025
Atkinson (1.0)	-0.001	0.000	-0.019	0.018	0.008	-0.011	-0.005	0.027
Atkinson (2.0)	0.005	0.000	-0.070	0.074	0.016	-0.015	-0.008	0.081

Source: own calculations using SIMPL.

The standard approach results are in line with earlier literature (Myck and Najsztub 2016, Morawski and Brzeziński 2023). Changes in taxes and benefits contributed to a reduction in relative poverty and income inequality, regardless of the measure used in the study. The contributions of other changes were negative, except for poverty incidence. The distributional consequences of economic growth and socio-economic change between 2005 and 2019 were negative. The improvements in inequality indicators observed in the data were due to changes in taxes and benefits. The policy effect was more substantial for its impact on poverty than inequality, which correctly reflects the objectives of social policy.

The new results are the values of the education effects ('edu' and 'emp' columns) and the relationship of these values to the policy effect ('policy' column). In the extended decomposition, the residual part corresponds to the part of the other effect in the standard decomposition that is not related to the educational change. Presumably, the changes that have the most tremendous significance in the residual factor are demographic changes resulting from the aging process. As Table 3 shows, there were visible changes in the structure of family types and the growth of households living on pensions from 2005 to 2019. Changes in income from agriculture and self-employment may also have significantly shaped this effect.

5.1. Relative monetary poverty

Combining the two educational effects, we obtain that they contributed to a 1.3 percentage point reduction in the relative risk of income poverty, which corresponded to 39.7% of the policy effect. The contribution of the composition effect, related to the change in the educational structure, was 27.5%, and the employment effect accounted for 12.2% of the policy effect. We see the positive impact of changes in education on the depth of poverty, too. In this case, the total education effect was 35.9% of the policy effect, of which 23.6% was the contribution of the composition effect, and 12.4% was the contribution of the employment effect.

5.2. Income inequality

The standard two-component decomposition into the policy effect and the other effect revealed the dominant importance of policy changes in explaining the decline in the Gini coefficient. Other changes did not play a significant role according to that approach. An additional decomposition disclosed that the insignificance of the other effect was due to the offsetting effects of changes in the sub-components. We found that the changes in education structure reduced income inequality by 1.42 percentage points, which corresponded to 90.4% of the policy effect. At the same time, changes in other wage determinants contributed to increased income inequality by 0.75 percentage points.

We see a reduction in inequality among high-income earners and among low-income earners while looking at the effect of education from the perspective of the Theil index. Conclusions based on the Atkinson are fully consistent showing a positive effect for the educational change. A larger decrease in the A(2) index than in the A(0.5) index indicates that the effect of education has more strongly reduced inequality at the lower end of the distribution. This suggests that those at the bottom of the income distribution benefited the most from the educational change. It cannot be excluded that changes in the minimum wage may have contributed to this. This conjecture seems worthy of particular interest in future research. It is interesting to compare the magnitude of the education effect for the differentiation measures against the policy effect. For the Theil 1 index, the index has a value of 2.21 and for the Theil 0 index a value of 0.86. This means that for those at the top of the income distribution, the educational change was relatively more important, than for people at the bottom of the distribution. The same conclusion is obtained for the Atkinson index, where for A(0.5) we have 1.3 and for A(2) 0.33.

The results concerning the effect of education are considerably more optimistic than the observations relating to overall changes in income distribution. Looking at changes in the overall distribution, we saw signs of increasing polarisation with generally decreasing inequality. In the case of the educational change, we have a strong basis for claiming that it has contributed to a decrease in inequality in all segments of the distribution.

6. Conclusion

Using the capabilities of a microsimulation approach, we looked at the consequences of the increase in the share of workers with higher education in the workforce in Poland in

2005-2019. We compared the effect of changes in the educational structure of employees with the effect of changes in tax and benefit policy. We found that a 10 percentage point increase in the share of employed men with tertiary education and an 18 percentage point increase in the share of employed women with tertiary education led to:

1. a reduction in the risk of poverty equivalent to 40% of the effect of changes in tax-benefit policy effect,
2. a reduction of income inequality, the size of which, relative to the policy effect, depends on the weight given to changes in different parts of the income distribution. The size of the education effect ranged from 32.9% of the policy effect in the case of the Atkinson(2) effect to 220.8% in the case of the Theil 1 index.
3. We confirmed earlier findings about the generally increasing polarisation of income while showing that changes in the structure of education have contributed to its reduction.

The results underscore the potentially positive externalities of changing the educational structure. The scale of the structural educational effect justifies considering education policy as an important social investment having positive externalities on income distribution.

Our findings imply social policy recommendations at times when the educational structure is improving dynamically. First, it seems reasonable to consider medium- and long-term consequences for changes in income inequality and relative material poverty when assessing the effects of regulations on changes in educational policy. Second, a social policy should address disparities at the lower end of the income distribution by investing in skill development for low-educated workers and improving wage conditions in low-paying sectors. Policies such as vocational training, skills development, or minimum wage increases should be considered. Expanding access to higher education for underprivileged groups can also help reduce the disparities observed in lower-income segments. Without such interventions, the growing gap between the poorest and the rest could undermine social and economic cohesion.

The analysis presented here exploits the advantages of using tax-benefit microsimulation modeling to analyze income distribution changes. In addition to its undoubted advantages, the method also has its limitations. First, accounting for macroeconomic changes is possible but requires a more elaborate model combining the microsimulation part with the macroeconomic model Peichl (2009), Bourguignon and Spadaro (2006). In the case of this analysis, macroeconomic changes (e.g., due to the global financial crisis) were reflected in income before taxes included in the input data. Second, we made a simplifying assumption of no change in the returns to education over the studied period, focusing only on the effect of changes in the education structure. In the future, it will be worthwhile to expand the analysis to include effects related to changes in the value of returns to education.

Acknowledgements

This study was supported by the National Science Centre (Poland) under contract UMO-2018/29/B/HS4/00627.

6.1. Citations

Refer to the literature using Harvard convention, as exemplified by the following articles of Gamrot (2012, 2013), a proceedings paper of Kennickell (1997) or a book of Särndal, Swensson and Wretman (1992).

6.2. Formulae

Expressions should not exceed the text width and may be entered as follows:

$$V(\hat{t}) = \sum_{i,j \in U} \check{y}_i \check{y}_j \Delta_{ij} \tag{6}$$

$$\hat{V}(\hat{t}) = \sum_{i,j \in s} \check{y}_i \check{y}_j \frac{\Delta_{ij}}{\pi_{ij}} \tag{7}$$

6.3. Tables

An example table:

Table 1: Some nicely looking numbers

U	*W*	*X*	*Y*	*Z*
1001	1002	1003	1004	1005
1006	1007	1008	1009	1000
1001	1002	1003	1004	1005
1006	1007	1008	1009	1000

7. The second section

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

8. Conclusions

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas.

Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Acknowledgements

Thanks to anyone for support, funding and such may be included in the non-numbered Acknowledgements section.

References

Abdullah, A., Doucouliagos, H.and Manning, E., (2015). Does education reduce income inequality? A meta-regression analysis. *Journal of Economic Surveys*, 29, pp. 301–316.

Bargain, O., Callan, T., (2010). Analysing the effects of tax-benefit reforms on income distribution: a decomposition approach. *The Journal of Economic Inequality*, 8, pp. 1–21.

Bargain, O., Callan, T., Doorley, K. and Keane, C., (2017). Changes in income distributions and the role of tax-benefit policy during the great recession: An international perspective. *Fiscal Studies*, 38, pp. 559–585.

Bargain, O., Morawski, L., Myck, M. and Socha, M., (2007). As SIMPL as that: introducing a tax-benefit microsimulation model for Poland. *IZA Discussion Papers (2988)* .

Becker, G. S., (1964). Human capital, New York: Columbia University.

Bourguignon, F., Spadaro, A., (2006). Microsimulation as a tool for evaluating redistribution policies. *Journal of Economic Inequality*, 4(1), pp. 77–106.

DiNardo, J., Fortin, N. M. and Thomas, L., (2006). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica*, 64(5), pp. 1001–1044.

Figari, F., Paulus, A. and Sutherland, H., (2015). Microsimulation and policy analysis, in B. Atkinson, A. and F. Bourguignon, eds, Handbook of Income Distribution, Elsevier, pp. 2141–2221.

Główny Urząd Statystyczny, (2020). Zasięg ubóstwa ekonomicznego w Polsce w 2019 r.

Główny Urząd Statystyczny, (2022). Dochody i warunki życia ludności Polski – raport z badania EU-SILC 2021.

Haan, P., Morawski, L. and Myck, M., (2008). Taxes, benefits and financial incentives to work. Britain, Germany and Poland compared. *Bank i Kredyt*, 39(1), pp. 5–33.

Hershbein, B. H., Schettini Kearney, M. and Pardue, L. W., (2020). College attainment, income inequality, and economic security: a simulation exercise, NBER Working Paper Series 26747, pp. 1–42.

Hovhannisyan, A., Castillo-Ponce, R. and Valdez, R. I., (2019). The determinants of income inequality: The role of education. *Scientific Annals of Economics and Business*, 66(4), pp. 451–464.

Jenkins, S. P., (2009). Distributionally-sensitive inequality indices and the gb2 income distribution. *Review of Income and Wealth*, 55(2), pp. 392–398.

Knight, J. B., Sabot, R. H., (1983). Educational expansion and the Kuznets effect. *The American Economic Review*, 73(5), pp. 1132–1136.

Kowalik, Z., Magda, I., (2021). Rynek pracy w Polsce – przemiany i wyzwania. *Ubezpiecenia Społeczne. Teoria i praktyka*, 5, pp. 5–21.

Kuznets, S., (1955). Economic growth and income inequality. *The American Economic Review*, 65(1), pp. 1–28.

Makhlouf, Y., Lalley, C., (2023). Education expansion, income inequality and structural transformation: Evidence from OECD countries. *Social Indicators Research*,(169), pp. 255–281.

Morawski, L., Brzeziński, M., (2023). How do right-wing populist majoritarian governments redistribute? Evidence from Poland, 2005–2019. *Social Policy Administration*, 58(3), pp. 521–539.

Myck, M., Domitrz, A., Morawski, L. and Semeniuk, A., (2015). Financial incentives to work in the context of a complex reform package and growing wages: the Polish experience 2005–2011. *Baltic Journal of Economics*, 15(2), pp. 99–121.

Myck, M., Najsztub, M., (2015). Data and Model Cross-validation to Improve Accuracy of Microsimulation Results: Estimates for the Polish Household Budget Survey. *International Journal of Microsimulation*, 8(1), pp. 33–66.

Myck, M., Najsztub, M., (2016). Distributional Consequences of Tax and Benefit Policies in Poland: 2005-2014, CenEA Working Paper Series (WP02/16), pp. 1–45.

OECD, (2021). Education at a Glance 2021: OECD Indicators, OECD Publishing, Paris.

Panek, T., (2017). Polaryzacja ekonomiczna w Polsce. *Wiadomości Statystyczne (The Polish Statistician)* , 62(1), pp. 5–23.

Peichl, A., (2009). The benefits and problems of linking micro and macro models: Evidence from a flat tax analysis. *Journal of Applied Economics*, 12(2), pp. 301–329.

Ram, R., (1984). Population increase, economic growth, educational inequality, and income distribution: Some recent evidence. *Journal of Development Economics*, 14(3), pp. 419–428.

Rehme, G., (2007). Education, economic growth and measured income inequality. *Economica*, 74(295), pp. 493–514.

Samano-Robles, C., (2018). The impact of education on income inequality in Latin America between 2000 and 2010, in Inequality, Taxation and Intergenerational Transmission (Research on Economic Inequality, Vol. 26), Emerald Publishing Limited, pp. 109–148.

Schultz, T. W., (1961). Investment in human capital. *The American Economic Review*, 51(1), pp. 1–17.

Sianesi, B., Reenen, J. V., (2003). The returns to education: Macroeconomics. *Journal of Economic Surveys*, 17(2), pp. 157–200.

Spence, M., (1973). Job market signaling. *The Quarterly Journal of Economics*, 87(3), pp. 355– 374.

Szarfenberg, R., (2023). Success in agenda setting through failure in policy-making: Exploring a new policy venue in the Polish European Semester 2012–2022, *International Journal of Social Welfare*, 32, pp. 113–125.

Tasseva, I. V., (2021). The changing education distribution and income inequality in Great Britain, *Review of Income and Wealth*, 67(3), pp. 659–683.

STATISTICS IN TRANSITION new series, September 2025
Vol. 26, No. 3, pp. 35–58, https://doi.org/10.59139/stattrans-2025-026
Received – 04.10.2024; accepted – 13.03.2025

Area-biased one-parameter exponential distribution with financial applications

Abdullah Hardan[1], **Loai Alzoubi**[2]

Abstract

Area-biased distributions are special cases of size-biased distributions. We have used the idea of area-biased distributions in this paper to propose a generalisation of a one-parameter linear exponential distribution. The concept is called the area-biased one-parameter linear exponential distribution. Its various characteristics are deduced and thoroughly explored. Some numerical studies are implemented which demonstrate that the distribution is skewed to the right with heavier tail than the normal distribution. The mean waiting and residual life time are also studied. Six methods of estimation are employed to estimate the parameters distribution. A simulation study is conducted which shows that the estimators are approximately unbiased and consistent. Three financial real data sets are applied. They represent the earning per share in the financial, industry and service sectors at the Amman Stock Exchange. The study shows that the suggested distribution has the best fit for these data sets compared to some competence distributions.

Key words: one-parameter linear exponential distribution, area-biased, methods of estimation, earning per share.

1. Introduction

In statistics, modelling lifetime data is an important issue in many fields, including biomedical sciences, economics, finance, engineering, and many others. A lot of continuous distributions have been introduced for modelling such data, because they can tend to be more efficient than the base distributions. Many methods have been used to propose new models such as the combination of two or more distributions.

Weighted distributions, involving a variety of sampling surveys, have been widely applied to model data in nature, offering more insights and adequacy in the modelling. The theory of weighted distributions ensures a collective access to the problems of model specification and data interpretation. It provides a procedure for fitting models to unknown weight functions when samples can be taken from both the original distribution and the developed distribution. They take into account the method of ascertainment by adjusting the probabilities of the actual occurrence of events in order to arrive at a specification of the probabilities of those events as observed and recorded.

[1]Department of Accounting, The world Islamic Sciences and Education University, Amman (11947), Jordan. E-mail: abdullah.hardan@wise.edu.jo. ORCID: https://orcid.org/0009-0007-2545-7719.

[2]Department of Mathematics, Al al-Bayt University, Mafraq (25113), Jordan. E-mail: oai67@aabu.edu.jo. ORCID: https://orcid.org/0000-0002-8795-4100.

The idea of weighted distributions was initially introduced by Fisher (1934) and developed by Rao (1965). Recently, this concept has been employed in a lot of researches related to reliability, ecology, family data analysis, bio-medicine, and some other fields for the improved performance of appropriate statistical models. It is defined by

$$f_w(x) = \frac{w(x)f(x)}{E(w(x))}, \; x > 0, \tag{1}$$

where $w(x)$ is a non-negative weight function.

Let X be a random variable with probability density function (pdf) $f(x)$, then the size-biased distribution can be produced by using the weight function $w(x) = X^m$. It was first studied by Patil and Ord (1976). For $m = 2$, we get the area-biased distribution, which was first employed by Cox (1968) and Zelen (1974). Thus, the resulting pdf takes the form of

$$f_1(x) = \frac{x^2 f(x)}{E(X^2)}. \tag{2}$$

Area-biased distributions, as mentioned above, are special types of weighted distributions. In recent times, many authors have been interested in studying these types of distributions, such as Sharma et al. (2018) who introduced the length- and area-biased Maxwell distribution. Al-Omari et al. (2019a) suggested the power length-biased Suja distribution as a new extension of the length-biased Suja distributions. Saghir et. al. (2017) studied a new class of Maxwell length-biased distributions. Shen et al. (2009) used semi-parametric transformations to model the length-biased data. Al-Omari and Alanzi (2021) suggested and studied the properties of the one-parameter inverse length-biased Maxwell distribution. Das and Roy (2011) suggested the length-biased form of the weighted Weibull distribution. The weighted distributions based on the mixture of two distributions based on weights p_1 and p_2, with $p_1 + p_2 = 1$ are used by many authors, such as: Alzoubi et al. (2022), Benrabia and Alzoubi (2022a), Benrabia and Alzoubi (2022b) and Alzoubi et al. (2022).

In this paper, we propose a new distribution. This distribution is applied to financial data extracted from the Amman Stock Exchange (ASE). We have used Earnings per share (EPS) data to compare the suggested distribution with other distributions. EPS is one way of measuring a company's success. An increase in the EPS indicates higher investor prosperity (Ferniawan et al. 2024). Some researchers concluded that EPS has a significant and positive impact on company value (Kristanti et al. 2024), others indicated a substantial and positive impact of EPS on the stock prices (Taubah et al. 2024 and Dang et al., 2024). The ASE was established in March 1999 as a non-profit independent institution, authorised to function as a regulated market for trading securities in Jordan. ASE aims to operate, manage and develop the operations and activities of security, commodity, and derivatives markets inside and outside of Jordan. It seeks to provide a strong and secure environment to ensure the interaction of supply and demand forces for trading in proper and fair trading practices. It also aims to raise the awareness and knowledge of investing in the financial markets and defining the services provided by the ASE (ASE, 2023).

2. One-parameter linear-exponential distribution

Ghitany et al. (2008) introduced a single parameter distribution called the Lindley distribution. The pdf of this distribution is given by

$$f_l(x) = \frac{\theta^2(1+x)e^{-\theta x}}{1+\theta}; \; x>0, \; \theta>0. \tag{3}$$

Sah (2021) proposed the one-parameter linear-exponential distribution (OPLED). Its pdf and second moment are defined as

$$f_o(x) = \frac{\theta^2(\theta^2+x)e^{-\theta x}}{1+\theta^3}; \; x>0, \; \theta>0 \tag{4}$$

$$E\left(X^2\right) = \frac{2(\theta^3+3)}{\theta^2(1+\theta^3)}. \tag{5}$$

3. Area-biased one-parameter linear-exponential distribution

This section introduces the new proposed distribution, the area-biased one-parameter exponential distribution (ABOPLED). The pdf of this distribution is defined using (2), (4) and (5):

$$f(x) = \frac{\theta^4 x^2\left(x+\theta^2\right)e^{-\theta x}}{2\theta^3+6}; \; x>0, \; \theta>0. \tag{6}$$

The cumulative distribution function (CDF) of ABOPLED is defined as:

$$\begin{aligned} F(x) &= \int_0^x \frac{\theta^4 u^2\left(u+\theta^2\right)e^{-\theta u}}{2\theta^3+6}du \\ &= 1-\frac{\left(\theta^3x^3+\left(\theta^5+3\theta^2\right)x^2+\left(2\theta^4+6\theta\right)x+2\theta^3+6\right)e^{-\theta x}}{2\theta^3+6} \\ &= 1-\left(\frac{\theta^3x^3}{2\theta^3+6}+\frac{\theta^2x^2}{2}+\theta x+1\right)e^{-\theta x}. \end{aligned} \tag{7}$$

$F(x)$ satisfies the conditions of the CDF, since (1) $F(x)$ is right-continuous. (2) $\lim_{x\longrightarrow 0} F(x)=0$, and (3) $\lim_{x\longrightarrow\infty} F(x)=1$

(a) The pdf **(b)** The CDF

Figure 1: The pdf and CDF of ABOPLED for different values of θ.

Figures 1a and 1b show the pdf and CDF plots of ABOPLED for θ values of 1.5-9.5(step(1)). Figure 1a shows that the peak of the distribution gets sharper for smaller values of θ. Figure 1b shows that the CDF reaches 1 faster for larger values of θ.

4. Moments and related measures of the ABOPLED

The moment generating function along with the moments and related measures of the ABOPLED and some tables of the mean, standard deviation, coefficient of skewness, excess kurtosis, and coefficient of variation for certain values of the parameter will be derived in this section.

4.1. Moments

Theorem 1 . *Let X be a random variable following the ABOPLED. The r^{th} moment of X is*

$$E(X^r) = \frac{\left((r+3)!+\theta^3(r+2)!\right)}{2\theta^r(2\theta^3+6)}, \quad r=1,2,.... \tag{8}$$

Proof 1 *The r^{th} moment can be found as*

$$\begin{aligned} E(X^r) &= \int_0^\infty x^r f(x)dx = \int_0^\infty \frac{\theta^4 x^{r+2}\left(x+\theta^2\right)e^{-\theta x}}{2\theta^3+6}dx \\ &= \frac{\theta^4}{2\theta^3+6}\int_0^\infty x^{r+2}\left(x+\theta^2\right)e^{-\theta x}dx = \frac{\left((r+3)!+\theta^3(r+2)!\right)}{2\theta^r(2\theta^3+6)}. \end{aligned}$$

Thus, the first four moments can be calculated by substituting r with 1, 2, 3 and 4 in (8). Hence, we have

$$\mu = E(X) = \frac{\left(12+3\theta^3\right)}{\theta(\theta^3+3)}, \tag{9}$$

$$E(X^2) = \frac{\left(60+12\theta^3\right)}{\theta^2(\theta^3+3)}, \tag{10}$$

$$E(X^3) = \frac{\left(360+60\theta^3\right)}{\theta^3(\theta^3+3)}, \tag{11}$$

$$E(X^4) = \frac{\left(2520+360\theta^3\right)}{\theta^4(\theta^3+3)}. \tag{12}$$

4.2. Related measures

From (9), and (10), the variance (σ^2) and the standard deviation (σ) are as follows:

$$\sigma^2 = E(X^2)-\mu^2 = \frac{\left(60+12\theta^3\right)}{\theta^2(\theta^3+3)} - \left(\frac{\left(12+3\theta^3\right)}{\theta(\theta^3+3)}\right)^2 = \frac{3(\theta^3+2)(\theta^3+6)}{\theta^2(\theta^3+3)^2}$$

$$\sigma = \frac{\sqrt{3(\theta^3+2)(\theta^3+6)}}{\theta(\theta^3+3)}. \tag{13}$$

The coefficient of variation (CV) is defined using (9) and (13) as:

$$CV = \frac{\sqrt{3(\theta^3+2)(\theta^3+6)}}{(12+3\theta^3)}. \tag{14}$$

The skewness (Sk) is defined using (9), (10), (11) and (13) as:

$$Sk(X) = \frac{E(X^3)-3\mu E(X^2)+2\mu^3}{\sigma^3} = \frac{6\left(\theta^9+12\theta^6+36\theta^3+36\right)}{\left(\sqrt{3(\theta^3+2)(\theta^3+6)}\right)^3}. \tag{15}$$

Au et al. (2015) defined the excess kurtosis (*eKur*) as: $eKur(X) = Kur(X)-3$. Thus, for ABOPLED it is defined using (9), (10), (11), (12) and (13) as:

$$\begin{aligned} eKur(X) &= \frac{E(X^4)-4\mu E(X^3)+6\mu^2E(X^2)-3\mu^4}{\sigma^4}-3 \\ &= \frac{\left(5\theta^{12}+80\theta^9+408\theta^6+864\theta^3+648\right)}{((\theta^3+2)(\theta^3+6))^2}-3. \end{aligned} \tag{16}$$

Table 1: Related moments measures for ABOPLED for different values of θ

θ	μ	σ	*Sk*	*eKur*	*CV*	θ	μ	σ	*Sk*	*eKur*	*CV*
1.25	**2.8845**	**1.5686**	**1.045**	**1.614**	**54.3793**	**4.50**	**0.6737**	**0.3889**	**1.154**	**1.9966**	**57.7223**
1.50	**2.3137**	**1.2858**	**1.0738**	**1.7005**	**55.5715**	**4.75**	**0.6373**	**0.3679**	**1.1542**	**1.9975**	**57.7257**
1.75	**1.9194**	**1.0825**	**1.0991**	**1.7846**	**56.4014**	**5.00**	**0.6047**	**0.3491**	**1.1543**	**1.9981**	**57.7281**
2.00	**1.6364**	**0.9315**	**1.1181**	**1.8527**	**56.9275**	**5.25**	**0.5753**	**0.3321**	**1.1544**	**1.9986**	**57.7298**
2.25	**1.426**	**0.8163**	**1.1311**	**1.9019**	**57.2455**	**5.50**	**0.5487**	**0.3168**	**1.1545**	**1.9989**	**57.731**
2.50	**1.2644**	**0.7262**	**1.1395**	**1.9354**	**57.4344**	**5.75**	**0.5244**	**0.3028**	**1.1545**	**1.9992**	**57.732**
2.75	**1.1368**	**0.6542**	**1.1448**	**1.9574**	**57.5469**	**6.00**	**0.5023**	**0.29**	**1.1546**	**1.9994**	**57.7326**
3.00	**1.0333**	**0.5954**	**1.1482**	**1.9716**	**57.6147**	**6.25**	**0.4819**	**0.2782**	**1.1546**	**1.9995**	**57.7332**
3.25	**0.9478**	**0.5465**	**1.1504**	**1.9809**	**57.6564**	**6.50**	**0.4632**	**0.2674**	**1.1546**	**1.9996**	**57.7335**
3.50	**0.8758**	**0.5052**	**1.1518**	**1.9869**	**57.6825**	**6.75**	**0.4459**	**0.2574**	**1.1546**	**1.9997**	**57.7338**
3.75	**0.8144**	**0.4699**	**1.1527**	**1.9909**	**57.6991**	**7.00**	**0.4298**	**0.2481**	**1.1546**	**1.9997**	**57.7341**
4.00	**0.7612**	**0.4393**	**1.1533**	**1.9936**	**57.71**	**7.25**	**0.4149**	**0.2395**	**1.1547**	**1.9998**	**57.7342**
4.25	**0.7147**	**0.4125**	**1.1537**	**1.9954**	**57.7173**	**7.50**	**0.4009**	**0.2315**	**1.1547**	**1.9998**	**57.7344**

Table 1 shows the mean, standard deviation, skewness, excess kurtosis, and coefficient of variation of the ABOPLED distribution for θ values of 0.25-7.5 (step 0.25). The distribution is right-skewed and the table makes this clear regardless of the values of θ as all skewness values are positive. The tails of the proposed distribution are heavier than the tails

of the normal distribution, which is demonstrated by the fact that all excess kurtosis values are positive. Furthermore, Table 1 shows how the mean and standard deviation increases as the value of α increases and how they decrease as θ increases. The coefficient of variation increases as the values of θ increase.

4.3. Moment generating function

Another way of deriving the moments is called the moment generating function (MGF). It is defined by the theorem described below.

Theorem 2 *Assume that random variable X follows the ABOPLED, then the moment generating function of X is given by*

$$M_X(t) = \frac{\theta^4\left(\theta^2(\theta-t)+3\right)}{(\theta-t)^4(\theta^3+3)}, \theta > t. \tag{17}$$

Proof 2

$$M_X(t) = \int_0^\infty e^{tx} f(x)dx = \int_0^\infty e^{tx}\frac{\theta^4 x^2\left(x+\theta^2\right)e^{-\theta x}}{2\theta^3+6}dx = \frac{\theta^4\left(\theta^2(\theta-t)+3\right)}{(\theta-t)^4(\theta^3+3)}$$

The r^{th} derivative of $M_X(t)$ at $t=0$ gives the r^{th} central moment of random variable X, *i.e.* $M^{(r)}(0) = E(X^r)$.

4.4. Mode

The mode is the most frequent value that occurs in data (Manikandan, 2011). When data occur equally, then there is no mode. On the other hand, some data sets can have more than one mode. This happens when the data set has two or more values of an equal frequency which is greater than that of any other value in its neighbourhood. The mode of the ABOPLED is given by equating the derivative of the pdf (6), or equivalently the logarithm of the pdf, with respect to x to zero. Thus, from (6), we have:

$$f'(x) = \frac{\theta^4}{2\theta^3+6}\left(-\theta\left(x^3+\theta^2x^2\right)+\left(3x^2+2\theta^2x\right)\right)e^{-\theta x}. \tag{18}$$

When equating (18) to 0, we receive

$$\begin{aligned} 0 &= -\theta x^3+(3-\theta^3)x^2+2\theta^2 x \\ x &= \frac{\theta^3-3-\sqrt{\theta^6+2\theta^3+9}}{-2\theta}. \end{aligned} \tag{19}$$

Hence, we have one mode only assured by the plot of (19) in Figure 2f. Figure 2f shows the plot of Equation (19). It indicates that the equation has only one solution regardless of the value of θ.

5. Reliability analysis of the ABOPLED

This section introduces the reliability, hazard rate, cumulative hazard rate, reversed hazard rate, and odds rate functions for the ABOPLED, as well as an explanation of their shapes for various values of the distribution parameters.

The reliability function of the ABOPLED can be calculated using (7):

$$R(t) = 1 - F(t) \quad = \quad \frac{\left(\theta^3 t^3 + \left(\theta^5 + 3\theta^2\right) t^2 + \left(2\theta^4 + 6\theta\right) t + 2\theta^3 + 6\right) e^{-\theta t}}{2\theta^3 + 6}. \quad (20)$$

The hazard rate function of ABOPLED, is defined using (6) and (20):

$$h(t) = \frac{f(t)}{1 - F(t)} \quad = \quad \frac{\theta^4 t^2 \left(t + \theta^2\right)}{(\theta^3 t^3 + (\theta^5 + 3\theta^2) t^2 + (2\theta^4 + 6\theta) t + 2\theta^3 + 6)}.$$

Using (20), the cumulative hazard rate function of is defined as:

$$H(t) = -ln\left(R(t)\right) \quad = \quad -ln\left(\frac{\left(\theta^3 t^3 + \left(\theta^5 + 3\theta^2\right) t^2 + \left(2\theta^4 + 6\theta\right) t + 2\theta^3 + 6\right) e^{-\theta t}}{2\theta^3 + 6}\right).$$

The reversed hazard rate function is defined using (6), and (7):

$$RH(t) = \frac{f(t)}{F(t)} \quad = \quad \frac{\theta^4 t^2 \left(t + \theta^2\right) e^{-\theta t}}{2\theta^3 + 6 - ((\theta^3 t^3 + (\theta^5 + 3\theta^2) t^2 + (2\theta^4 + 6\theta) t + 2\theta^3 + 6) e^{-\theta t})}.$$

The odds rate function is defined using (7), and (20):

$$O(t) \quad = \quad \frac{F(t)}{1 - F(t)} = \frac{2\theta^3 + 6 - \left(\left(\theta^3 t^3 + \left(\theta^5 + 3\theta^2\right) t^2 + \left(2\theta^4 + 6\theta\right) t + 2\theta^3 + 6\right) e^{-\theta t}\right)}{((\theta^3 t^3 + (\theta^5 + 3\theta^2) t^2 + (2\theta^4 + 6\theta) t + 2\theta^3 + 6) e^{-\theta t})}.$$

Figure 2 shows the reliability analysis functions of ABOPLED for θ values of 1.5-9.5 (step1). Figure 2a shows the plot of the reliability function. Figure 2b represents the plot of the hazard rate function. The reversed hazard rate function is presented in Figure 2c. Whereas the cumulative hazard rate function is presented in Figure 2d. The odds rate function plot is shown in Figure 2e.

6. Some structural and statistical properties

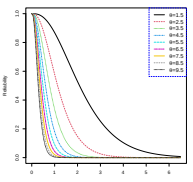
6.1. Order statistics

Let $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ be the order statistics of the random sample $X_1, X_2, \cdots, X_n$ obtained from ABOPLED, then the pdf of the k^{th} order statistics is:

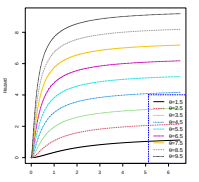
$$f_{(k)}(x) \quad = \quad \frac{n!}{(k-1)!(n-k)!} [F(x)]^{k-1} [1 - F(x)]^{n-k} f(x)$$

$$= \frac{n!}{(k-1)!(n-k)!}\left(1-\left(\frac{\theta^3x^3}{2\theta^3+6}+\frac{\theta^2x^2}{2}+\theta x+1\right)e^{-\theta x}\right)^{k-1} \times \left(\left(\frac{\theta^3x^3}{2\theta^3+6}+\frac{\theta^2x^2}{2}+\theta x+1\right)e^{-\theta x}\right)^{n-k}\left[\frac{\theta^4x^2\left(x+\theta^2\right)e^{-\theta x}}{2\theta^3+6}\right]. \quad (21)$$

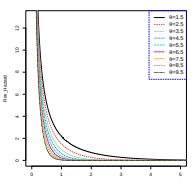
The minimum and maximum order statistics of ABOPLED can be found by replacing $k=1$ and $k=n$, respectively, in (21). As a result, we obtain what follows:



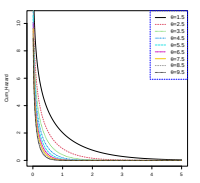
(a) The reliability function.



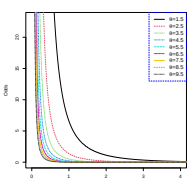
(b) The hazard rate function.



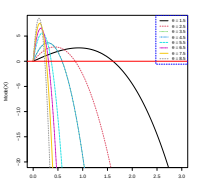
(c) The reversed hazard rate function.



(d) The cumulative hazard rate function.



(e) The odds rate.



(f) The plot of (19)

Figure 2: Reliability analysis functions of ABOPLED for different θ values and the plot of (19).

$$f_{(1)}(x) = n\left(\left(\frac{\theta^3x^3}{2\theta^3+6}+\frac{\theta^2x^2}{2}+\theta x+1\right)e^{-\theta x}\right)^{n-1}\left[\frac{\theta^4x^2\left(x+\theta^2\right)e^{-\theta x}}{2\theta^3+6}\right] \quad (22)$$

$$f_{(n)}(x) = n\left(1-\left(\frac{\theta^3x^3}{2\theta^3+6}+\frac{\theta^2x^2}{2}+\theta x+1\right)e^{-\theta x}\right)^{n-1}\left[\frac{\theta^4x^2\left(x+\theta^2\right)e^{-\theta x}}{2\theta^3+6}\right] \quad (23)$$

6.2. Quartiles

Quartiles are special cases of quantiles. The first quartile (Q1) corresponds to 25% of the values that are below a specific value in the distribution, while the second quartile (Q2) corresponds to 50% of the values that are below a specific value in the distribution. It represents the median of the distribution. The third quartile (Q3) corresponds to 75% of the values that are below a specific value in the distribution. The quantile function can be obtained by finding the inverse of (7), so it can be obtained as:

$$\begin{aligned} q &= F(x_q) = 1 - \frac{\left(\theta^3 x_q^3 + \left(\theta^5 + 3\theta^2\right) x_q^2 + \left(2\theta^4 + 6\theta\right) x_q + 2\theta^3 + 6\right) e^{-\theta x_q}}{2\theta^3 + 6} \\ 1 - q &= \frac{\left(\theta^3 x_q^3 + \left(\theta^5 + 3\theta^2\right) x_q^2 + \left(2\theta^4 + 6\theta\right) x_q + 2\theta^3 + 6\right) e^{-\theta x_q}}{2\theta^3 + 6}, \end{aligned} \tag{24}$$

where q is a random variable following the uniform distribution, *i.e.* $q \in (0,1)$. For $q = 0.5$, we obtain the median of the distribution. Equation (24) does not have an explicit solution. Figure 3 shows the plot of this equation for θ = 1.5 and q values of 0.05, 0.1, 0.25, 0.4, 0.5, 0.6, 0.75, 0.8 and 0.95. It shows that (3b) it has exactly one solution.

(a) The order statistics plot **(b)** The quantile function

Figure 3: The pdf of order statistics for θ =2.5 and n = 10 and the quantile function when θ =2.5.

6.3. Mean waiting time

Significantly, the mean waiting time is the measure used to verify the effectiveness of the service. The form, which is most effective in decreasing waiting time, can be determined by comparing many different services patterns. Pollaczek (1957) introduced a formula for the mean waiting time in a G/G/1 queue. Rosenberg (1968) introduced the mean waiting time to measure the effectiveness of the service because it is the easiest property of the waiting time distribution to calculate. Otsuka et al. (2010) proposed a theoretical analysis of the mean waiting time for message delivery in lattice ad hoc networks. Romero-Silva and Hurtado (2017) studied the difference in mean waiting times between two classes of customers in a single-server FIFO queue. For the ABOPLED, the mean waiting time can be written as:

$$
\begin{aligned}
mw(t) &= \frac{1}{F(t)}\int_0^t F(x)dx \\
&= \frac{\left(\begin{array}{c}\left(\left(2\theta^4+6\theta\right)t-6\theta^3-24\right)e^{\theta t}+\theta^3t^3+\left(\theta^5+6\theta^2\right)t^2\\ +\left(4\theta^4+18\theta\right)t+6\theta^3+24\end{array}\right)}{\theta\left(2\theta^3+6-\left(\theta^3t^3+\left(\theta^5+3\theta^2\right)t^2+\left(2\theta^4+6\theta\right)t+2\theta^3+6\right)\right)}. \qquad (25)
\end{aligned}
$$

6.4. Mean residual life time

Several studies have attempted to find substitutes that do not depend on the entire right tail of the pdf, such as mean residual lifetime and median residual life, and the corresponding residual life quantiles. Several studies have addressed this issue. First, Schmittlein and Morrison (1981) introduced the median residual lifetime and characterisation theorem along with its application. Joe and Proschan (1984a) and Joe and Proschan (1984b) examined the comparison of two life distributions based on their percentile residual life functions. Lillo (2005) studied the median residual lifetime and its properties. Jeong et al. (2008) investigated the nonparametric inference on median residual life function. Recently, Zamanzade et al. (2024) analysed the estimation of the mean residual life based on ranked set sampling. For the ABOPLED, the mean residual lifetime can be written as"

$$
\begin{aligned}
MR(t) &= \frac{1}{\overline{F}(t)}\int_t^\infty \overline{F}(x)dx = \frac{2\theta^3+6}{\left(\theta^3t^3+\left(\theta^5+3\theta^2\right)t^2+\left(2\theta^4+6\theta\right)t+2\theta^3+6\right)e^{-\theta t}} \\
&\quad\times\int_t^\infty \frac{\left(\theta^3x^3+\left(\theta^5+3\theta^2\right)x^2+\left(2\theta^4+6\theta\right)x+2\theta^3+6\right)e^{-\theta x}}{2\theta^3+6}dx \\
&= \frac{2\theta^3+6}{\left(\theta^3t^3+\left(\theta^5+3\theta^2\right)t^2+\left(2\theta^4+6\theta\right)t+2\theta^3+6\right)e^{-\theta t}} \\
&\quad\times\frac{\left(\theta^2t^3+\left(\theta(\theta^3+6)t^2+(4\theta^3+18)t+6\theta^2+\frac{24}{\theta}\right)e^{-\theta t}\right.}{2\theta^3+6} \\
&= \frac{\left(\theta^2t^3+\left(\theta(\theta^3+6)t^2+(4\theta^3+18)t+6\theta^2+\frac{24}{\theta}\right)\right.}{\left(\theta^3t^3+\left(\theta^5+3\theta^2\right)t^2+\left(2\theta^4+6\theta\right)t+2\theta^3+6\right)}. \qquad (26)
\end{aligned}
$$

6.5. Entropy

Olbryś and Ostrowski (2021) introduced introduced a new procedure for the measurement of stock market depth and market liquidity. An algorithm inferring the initiator of a trade supports the proposed Shannon entropy-based market depth indicator. The findings of the empirical experiments for real high-frequency data specify that this new entropy-based method can be considered as a good market depth and liquidity proxy with an intuitive base for both theoretical and the empirical analyses in financial markets. In this section, we have derived the theoretical entropies with some numerical results.

The Shannon (Shannon, 1948) (S_ρ), R*é*nyi (R_ρ) (R*é*nyi, 1961) and Tsallis (T_ρ) (Tsallis, 1988) entropies of the ABOPLED are defined as:

$$\begin{aligned}
S_\rho &= -\int_0^\infty f(x)\log(f(x))dx \\
&= -\int_0^\infty \left(\frac{(\theta^4)(\theta^2+x)x^2}{(2\theta^3+6)}e^{-\theta x}\right)\log\left(\frac{(\theta^4)(\theta^2+x)x^2}{(2\theta^3+6)}e^{-\theta x}\right)dx \qquad (27)\\
R_\rho &= \frac{1}{1-\rho}\log\left(\int_0^\infty (f(x))^\rho dx\right),\ \rho>0,\ \rho\neq 0 \\
&= \frac{1}{1-\rho}\log\left(\int_0^\infty \left(\frac{(\theta^4)(\theta^2+x)x^2}{(2\theta^3+6)}e^{-\theta x}\right)^\rho dx\right) \\
&= \frac{1}{1-\rho}\log\left(\left(\frac{(\theta^4)}{(2\theta^3+6)}\right)^\rho \int_0^\infty \sum_{k=0}^{\rho}\binom{\rho}{k}\theta^{2k}x^{3\rho-k}e^{-\rho\theta x}dx\right) \\
&= \frac{1}{1-\rho}\log\left(\left(\frac{(\theta^4)}{(2\theta^3+6)}\right)^\rho \sum_{k=0}^{\rho}\binom{\rho}{k}\frac{(3\rho-k)!}{\rho^{3\rho-k+1}\theta^{3\rho+k+1}}\right) \qquad (28)\\
T_\rho &= \frac{1}{\rho-1}\left(1-\int_0^\infty (f(x))^\rho dx\right),\ \rho>0,\ \rho\neq 0 \\
&= \frac{1}{\rho-1}\left(1-\left(\left(\frac{(\theta^4)}{(2\theta^3+6)}\right)^\rho \sum_{k=0}^{\rho}\binom{\rho}{k}\frac{(3\rho-k)!}{\rho^{3\rho-k+1}\theta^{3\rho+k+1}}\right)\right). \qquad (29)
\end{aligned}$$

Table 2 shows the numerical results for Shannon, R*é*nyi and Tsallis entropies for ABOPLED using different values of θ of 0.05-4.25 (step(0.1)) when ρ = 5. It also shows the mean waiting time and the mean residual life time. The table clarifies that as the values of θ increase, all entropy values decrease. It shows the values of the mean waiting time and the mean residual life time for the same values of θ. The mean waiting time values rise as the values of θ grow. On the other hand, the values of the mean residual life time decrease as the values of θ increase.

Table 2: Numerical results for Shannon, R*é*nyi and Tsallis entropies, the mean waiting time and mean residual life time for ABOPLED using different θ values with ρ=5.

θ	S_ρ	R_ρ	T_ρ	mw	MR	θ	S_ρ	R_ρ	T_ρ	mw	MR
0.05	**3.477**	**4.704**	**0.250**	**0.010**	**71.213**	**2.25**	**1.099**	**0.771**	**0.239**	**1.081**	**0.653**
0.25	**3.410**	**3.089**	**0.250**	**0.054**	**15.729**	**2.45**	**1.003**	**0.674**	**0.233**	**1.300**	**0.569**
0.45	**2.822**	**2.501**	**0.250**	**0.101**	**8.374**	**2.65**	**0.916**	**0.586**	**0.226**	**1.540**	**0.503**
0.65	**2.453**	**2.132**	**0.250**	**0.154**	**5.384**	**2.85**	**0.836**	**0.506**	**0.217**	**1.792**	**0.451**
0.85	**2.180**	**1.860**	**0.250**	**0.214**	**3.701**	**3.05**	**0.763**	**0.432**	**0.206**	**2.050**	**0.409**
1.05	**1.958**	**1.639**	**0.250**	**0.283**	**2.630**	**3.25**	**0.695**	**0.364**	**0.192**	**2.308**	**0.374**
1.25	**1.769**	**1.449**	**0.249**	**0.365**	**1.924**	**3.45**	**0.631**	**0.300**	**0.175**	**2.563**	**0.345**
1.45	**1.603**	**1.281**	**0.249**	**0.462**	**1.454**	**3.65**	**0.572**	**0.241**	**0.155**	**2.813**	**0.320**
1.65	**1.455**	**1.132**	**0.247**	**0.579**	**1.138**	**3.85**	**0.516**	**0.185**	**0.131**	**3.058**	**0.299**
1.85	**1.323**	**0.998**	**0.245**	**0.719**	**0.920**	**4.05**	**0.463**	**0.132**	**0.102**	**3.299**	**0.281**
2.05	**1.205**	**0.878**	**0.243**	**0.886**	**0.766**	**4.25**	**0.413**	**0.082**	**0.070**	**3.535**	**0.264**

7. Parameters estimation

7.1. Maximum likelihood method

Let $X_1, X_2, \cdots, X_n$ be a random sample from ABOPLED, where $x_1, x_2, \cdots, x_n$ are the observed values of the random sample. The likelihood function is:

$$L = \prod_{i=1}^{n}\left[\frac{\theta^4 x_i^2\left(x_i+\theta^2\right)e^{-\theta x_i}}{2\theta^3+6}\right] = \left[\frac{\theta^4}{2\theta^3+6}\right]^n\left(\prod_{i=1}^{n}x_i^2\right)\left(\prod_{i=1}^{n}(x_i+\theta^2)\right)\left(e^{-\theta\sum_{i=0}^{n}x_i}\right)$$

Thus, the log-likelihood function is:

$$\ell = ln(L) = 4nln(\theta) - nln(2\theta^3+6) + \sum_{i=0}^{n} ln\left(x_i^2\right) + \sum_{i=0}^{n} ln\left((x_i+\theta^2)\right) - \theta\sum_{i=0}^{n}x_i \quad (30)$$

With respect to θ, we receive

$$\frac{\partial\ell}{\partial\theta} = \frac{4n}{\theta} - \frac{6n\theta^2}{2\theta^3+6} + \sum_{i=1}^{n}\frac{2\theta}{x_i+\theta^2} - \sum_{i=0}^{n}x_i.$$

Nonlinear equation $\frac{\partial\ell}{\partial\theta}=0$ can be solved numerically as there is no explicit solution, and the maximum likelihood estimate (MLE) of θ is this solution.

7.2. Least square methods

This subsection describes the least squares methods for estimating the ABOPLED parameters. These methods were summarised by Swain et al. (1988) as follows: Let $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ be the order statistics of random sample $X_1, X_2, \cdots, X_n$ obtained from ABOPLED with the observed ordered values of $x_{(1)}, x_{(2)}, \cdots, x_{(n)}$. The ordinary least squares (OLS) method is defined as:

$$R_{OLS} = \sum_{i=1}^{n}\left[F(x_{(i)}) - \frac{i}{n+1}\right]^2$$

$$= \sum_{i=1}^{n}\left[\frac{n+1-i}{n+1} - \left(\frac{\theta^3 x_{(i)}^3}{2\theta^3+6} + \frac{\theta^2 x_{(i)}^2}{2} + \theta x_{(i)} + 1\right)e^{-\theta x_{(i)}}\right]^2. \quad (31)$$

Thus, the OLS estimator of θ is the solution of $\frac{\partial R_{OLS}}{\partial\theta}=0$.

The weighted least squares (WLS) estimate can be determined as:

$$W_{WLS} = \sum_{i=1}^{n}\frac{(n+1)^2(n+2)}{i(n+1-i)}\left[1 - \left(\frac{\theta^3 x_{(i)}^3}{2\theta^3+6} + \frac{\theta^2 x_{(i)}^2}{2} + \theta x_{(i)} + 1\right)e^{-\theta x_{(i)}} - \frac{i}{n+1}\right]^2. \quad (32)$$

Again, the WLS estimators of θ is the solution of $\frac{\partial W_{WLS}}{\partial\theta}=0$.

7.3. Maximum product spacings method

The maximum product of spacings (MPS) (Cheng and Amin, 1983) estimator, $\hat{\theta}_{MPS}$, of θ can be obtained using

$$\begin{aligned} NL(\alpha,\beta|x) &= \frac{1}{n+1}\sum_{i=1}^{n+1} ln\left[F(x_{(i)};\alpha,\beta)-F(x_{(i-1)};\alpha,\beta)\right] \\ &= \frac{1}{n+1}\sum_{i=1}^{n+1} ln\left[\left[1-\left(\frac{\theta^3x_{(i)}^3}{2\theta^3+6}+\frac{\theta^2x_{(i)}^2}{2}+\theta x_{(i)}+1\right)e^{-\theta x_{(i)}}\right]\right. \\ &\quad \left.-\left[1-\left(\frac{\theta^3x_{(i-1)}^3}{2\theta^3+6}+\frac{\theta^2x_{(i-1)}^2}{2}+\theta x_{(i-1)}+1\right)e^{-\theta x_{(i-1)}}\right]\right]. \end{aligned} \quad (33)$$

$\hat{\theta}_{MPS}$ can be obtained by solving nonlinear equation $\frac{\partial NL(\theta|x)}{\partial\theta}=0$ with respect to θ parameter.

7.4. Cramer-Von Mises and Anderson-Darling methods

The Cramer-Von Mises (CVM) method (Cramér, 1928 and Von Mises, 1928) for estimating ABOPLED parameters is defined as:

$$\begin{aligned} CVM &= \frac{1}{12n}+\sum_{i=1}^{n}\left[F(x_{(i)},\theta)-\frac{2i-1}{2n}\right]^2 \\ &= \frac{1}{12n}+\sum_{i=1}^{n}\left[1-\left(\frac{\theta^3x_{(i)}^3}{2\theta^3+6}+\frac{\theta^2x_{(i)}^2}{2}+\theta x_{(i)}+1\right)e^{-\theta x_{(i)}}-\frac{2i-1}{2n}\right]^2. \end{aligned} \quad (34)$$

The estimator is the solution of the following system of nonlinear equation $\frac{\partial CVM}{\partial\theta}=0$.

The Anderson-Darling (AD) estimators (Anderson, 1962) of θ can be obtained as:

$$\begin{aligned} AD &= -n-\frac{1}{n}\sum_{i=1}^{n}(2i-1)\left\{ln[F(x_{(i)};\alpha,\beta\theta)]+ln\overline{F}(x_{(n+1-i)};\alpha,\beta)\right\} \\ &= -n-\sum_{i=1}^{n}\frac{(2i-1)}{n}\left[\begin{array}{c} ln\left(1-\left(\frac{\theta^3x_{(i)}^3}{2\theta^3+6}+\frac{\theta^2x_{(i)}^2}{2}+\theta x_{(i)}+1\right)e^{-\theta x_{(i)}}\right) \\ +ln\left(\left(\frac{\theta^3x_{(n+1-i)}^3}{2\theta^3+6}+\frac{\theta^2x_{(n+1-i)}^2}{2}+\theta x_{(n+1-i)}+1\right)e^{-\theta x_{(n+1-i)}}\right) \end{array}\right]. \end{aligned} \quad (35)$$

The AD estimator of θ is the solution of the non-linear equation $\frac{\partial AD}{\partial\theta}=0$.

The estimator presented in this section will be estimated using a simulation study in the next part of the article.

8. Simulation study

A simulation study is conducted to examine the efficiency of the estimators used and the precision of the methods applied to estimate the ABOPLED parameters is discussed in Section 7.

The following algorithm is used to estimate the distribution parameters:

- A Monte Carlo simulation study is carried out using different sample sizes: n = 30, 50, 100, 200, 300 and 500 to assess the performance of the ABOPLED via the *R* package (R Core Team, 2021)
- 1,000 samples are simulated using the true parameters values θ =2.
- The MLE, OLS, WLS, MPS, CVM and AD estimators are obtained through the non-linear equations by maximising or minimising Equations (30), (31), (32), (33), (34), (35), respectively, as required by the method with respect to θ.
- The AB and MSEs of all estimates are calculated.
- For each sample, the estimates of the parameter θ, MSE and the bias are obtained. Then, we calculate the AB and the MSE as follows: $AB(\hat{\theta}) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}-\theta)$, $MSE = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}-\theta)^2$. The results of this simulation are summarised in Table 3.

Table 3 shows the estimate of θ and its AB and MSE. It shows that these values decrease as the sample size increases, indicating that the estimate behaves consistently for $\hat{\theta}$, therefore it is unbiased and consistent. Based on AB and MSE, we recommend using the MLE method to estimate the parameter.

Table 3: AB and MSE for estimated ABOPLED parameters

Method	n	$\hat{\theta}$	AB	MSE	Method	n	$\hat{\theta}$	AB	MSE
MLE	30	2.0047	0.0047	0.0035	MLE	200	2.0017	0.0017	0.0006
OLS		2.1161	0.1161	0.2267	OLS		2.0621	0.0621	0.0313
WLS		2.0868	0.0868	0.2594	WLS		2.0186	0.0186	0.0305
CVM		2.2910	0.2910	0.3560	CVM		2.0310	0.0310	0.0860
MPS		2.4773	0.4773	0.7012	MPS		2.1164	0.1164	0.3551
AD		2.1014	0.1014	0.3807	AD		2.0301	0.0301	0.0860
MLE	50	2.0025	0.0025	0.0021	MLE	300	2.0016	0.0016	0.0003
OLS		2.0852	0.0852	0.1282	OLS		2.0577	0.0577	0.0222
WLS		2.0489	0.0489	0.1364	WLS		2.0145	0.0145	0.0207
CVM		2.0516	0.0516	0.2017	CVM		2.0305	0.0305	0.0746
MPS		2.2317	0.2317	0.5507	MPS		2.1148	0.1148	0.3138
AD		2.0516	0.0516	0.2017	AD		2.0250	0.0250	0.0746
MLE	100	2.0022	0.0022	0.0010	MLE	500	2.0008	0.0008	0.0002
OLS		2.0750	0.0750	0.0612	OLS		2.0568	0.0568	0.0137
WLS		2.0363	0.0363	0.0641	WLS		2.0134	0.0134	0.0113
CVM		2.0442	0.0442	0.1297	CVM		2.0299	0.0299	0.0660
MPS		2.1517	0.1517	0.5069	MPS		2.1066	0.1066	0.3011
AD		2.0452	0.0452	0.2017	AD		2.0130	0.0130	0.0660

9. Real data applications

This section compares the proposed distribution's goodness of fit to a few other existing distributions in order to demonstrate its flexibility based on three financial real data sets.

The data were collected from companies listed at the ASE. The sample selection criteria depended on data availability for three years: 2021, 2022, 2023. The population comprised of all companies listed at the ASE, i.e. a total of 203 companies, the final sample consisted of 92 companies, excluding those suspended from trading, with incomplete data and those that incurred losses during the study periods. The total number of observations of EPS and ROE was 273 each.

Earnings per share is the amount of income earned on a share of common stock during an accounting period. It is calculated by dividing the profit or the loss attributable to ordinary equity holders of the parent entity (the numerator) by the weighted average number of ordinary shares outstanding (the denominator) during the period (IAS, 2023). The ratio indicated the company's ability to produce a profit for common shareholders. It is widely used by analysts and other external users of financial statements, as well as by management.

The following distributions are used for this comparison:

- Gharaibeh distribution (Gharaibeh, 2021):

$$f(x) = \frac{\theta^6}{120(\theta^6+\theta^4+\theta^2+1)}\left(x^5+20x^3+120x+120\theta\right)e^{-\theta x}; x>0,\ \theta>0;$$

- Exponential distribution (Exp) (Kingman, 1982):

$$f(x) = \alpha e^{-\theta x},\ x>0,\ \theta>0;$$

- Lindley distribution (Ghitany et al., 2008):

$$f(x) = \frac{\theta^2(1+x)e^{-\theta x}}{1+\theta}; x>0, \theta>0;$$

- Length bias Benrabia distribution (LBBD) (Almakhareez and Alzoubi, 2024a):

$$f_l(x;\alpha,\beta) = \frac{\beta(\alpha\Gamma(\alpha-1)\beta x+\beta^{\alpha}x^{\alpha-1})e^{-\beta x}}{(\alpha-\beta+\alpha\beta)\Gamma(\alpha-1)}, x>0,\ \alpha>1, \beta>0;$$

- Area bias Benrabia distribution (ABBD) (Almakhareez and Alzoubi, 2024b):

$$f_a(x;\alpha,\beta) = \frac{(\alpha\Gamma(\alpha-1)\beta^3x^2+\beta^{\alpha+2}x^{\alpha})e^{-\beta x}}{(\alpha^2\beta-\alpha\beta+2\alpha)\Gamma(\alpha-1)};\ x,\beta>0,\ \alpha>1;$$

- Karam distribution (Gharaibeh and Sahtout, 2022):

$$f(x) = \frac{\theta^6\left(x^5+x^4+x^2+1\right)}{\theta^5+2\theta^3+24\theta+120}e^{-\theta x};\ x>0;\ \theta>0;$$

- OPLED (See (4));

- size-biased Ishita distribution (SBID) (Al-Omari et al., 2019b):

$$f_s(x) = \frac{\theta^4}{\theta^3+6}x(\theta+x^2)e^{-\theta x},\ x>0,\ \theta>0.$$

Tables 4-6 show the data sets used in this section. They show the financial ratios per year for financial, industry and service sectors during the 2021–2023 period.

Table 4: Dataset I: EPS for the financial sector during the 2021–2023 period

Company Name	Financial Ratios/Year		
	2023	2022	2021
JORDAN ISLAMIC BANK	0.31151	0.30555	0.29529
SAFWA ISLAMIC BANK	0.1751	0.15112	0.1406
ISLAMIC INTERNATIONAL ARAB BANK	0.35326	0.35497	0.4
JORDAN KUWAIT BANK	0.39405	0.12455	0.05159
JORDAN COMMERCIAL BANK	0.09571	0.0945	0.05837
THE HOUSING BANK FOR TRADE AND FINANCE	0.43406	0.41111	0.33531
ARAB JORDAN INVESTMENT BANK	0.12384	0.12039	0.11394
BANK AL ETIHAD	0.23569	0.21455	0.20312
ARAB BANKING CORPORATION /(JORDAN)	0.04546	0.06063	0.08729
INVEST BANK	0.24629	0.19826	0.17812
CAPITAL BANK OF JORDAN	0.27317	0.33007	0.39407
CAIRO AMMAN BANK	0.18571	0.18218	0.17263
BANK OF JORDAN	0.22012	0.2007	0.18004
JORDAN AHLI BANK	0.09266	0.08422	0.07092
ARAB BANK	0.58648	0.51113	0.2436
MIDDLE EAST INSURANCE	0.03929	0.10864	0.07403
AL-NISR AL-ARABI INSURANCE	0.63802	0.231	0.29263
JORDAN INSURANCE	0.06033	0.00036	0.02921
DELTA INSURANCE	0.10801	0.0777	0.01752
JERUSALEM INSURANCE	0.18775	0.20698	0.16436
THE UNITED INSURANCE	0.21804	0.1675	0.16527
GULF INSURANCE GROUP/ JORDAN	0.36495	0.28091	0.25762
NATIONAL INSURANCE	0.16729	0.13781	0.11505
EURO ARAB INSURANCE GROUP	0.19688	0.19041	0.11778
THE MEDITERRANEAN & GULF INSURANCE COMPANY-JORDAN P.L.C	0.0208	0.03486	0.00737
FIRST INSURANCE	0.10379	0.07202	0.07
THE ISLAMIC INSURANCE	0.11616	0.11066	0.13
AL-AMAL FINANCIAL INVESTMENTS	0.01476	0.02963	0.09023
BABELON INVESTMENTS	0.07179	0.0246	0.00509
DARAT JORDAN HOLDINGS	0.04654	0.01134	0.10751

FIRST FINANCE	**0.03256**	**0.04269**	**0.02772**
INMA INVESTMENT AND FINANCIAL ADVANCES	**0.03367**	**0.04155**	**0.02566**
JORDAN LOAN GUARANTEE CORPORATION	**0.06282**	**0.03956**	**0.03886**
JORDAN MORTGAGE REFINANCE	**0.3677**	**0.35538**	**0.43181**
JORDANIAN MANAGEMENT AND CONSULTING COMPANY	**0.29049**	**0.17916**	**0.19246**
AD-DULAYL INDUSTRIAL PARK & REAL ESTATE COMPANY P.L.C	**0.06236**	**0.05714**	**0.04944**
AL-TAJAMOUAT FOR TOURISTIC PROJECTS CO PLC	**0.02698**	**0.02374**	**0.00992**
AMAD INVESTMENT & REAL ESTATE DEVELOPMENT	**0.00437**	**0.07656**	**0.01058**
CONTEMPRO FOR HOUSING PROJECTS	**0.00308**	**0.01992**	**0.0209**
JORDAN MASAKEN FOR LAND & INDUSTRIAL DEVELOPMENT PROJECTS	**0.0148**	**0.00919**	**0.00289**
NOOR CAPITAL MARKTS FOR DIVERSIFIED INVESTMENTS	**0.12639**	**0.19163**	**0.317**
THE PROFESSIONAL COMPANY FOR REAL ESTATE INVESTMENT AND HOUSING	**0.02423**	**0.0337**	**0.02976**
THE REAL ESTATE & INVESTMENT PORTFOLIO CO.	**0.00286**	**0.05566**	**0.02754**
NOOR ASSETS MANAGEMENT AND LEASING CO.	**0.2158**	**0.13914**	**0.12803**

Table 5: Dataset II: EPS for the industry sector during the 2021–2023 period

Company Name	**Financial Ratios/Year**		
	2023	**2022**	**2021**
THE INDUSTRIAL COMMERCIAL & AGRICULTURAL	**0.05789**	**0.03362**	**0.06274**
THE ARAB PESTICIDES & VETERINARY DRUGS MFG. CO.	**0.27107**	**0.26055**	**0.23649**
JORDAN CHEMICAL INDUSTRIES	**0.02470**	**0.14547**	**0.05689**
UNITED CABLE INDUSTRIES	**0.02930**	**0.01857**	**0.01208**
READY MIX CONCRTE AND CONSTRUCTION SUPPLIES	**0.17116**	**0.07234**	**0.02440**
ARABIAN STEEL PIPES MANUFACTURING	**0.13898**	**0.09263**	**0.06792**
AL-QUDS READY MIX	**0.15671**	**0.01634**	**0.02434**
ASSAS FOR CONCRETE PRODUCTS CO. LTD	**0.05061**	**0.07530**	**0.04958**
JORDAN DAIRY	**0.16668**	**0.05058**	**0.08475**
GENERAL INVESTMENT	**0.21088**	**0.24084**	**0.20878**
UNIVERSAL MODERN INDUSTRIES	**0.07028**	**0.11759**	**0.13885**
NUTRI DAR	**0.00842**	**0.04065**	**0.02242**
JORDAN VEGETABLE OIL INDUSTRIES	**0.35863**	**0.25515**	**0.25914**
SINIORA FOOD INDUSTRIES PLC	**0.15817**	**0.19213**	**0.27924**
ARAB ALUMINIUM INDUSTRY /ARAL	**0.00100**	**0.06988**	**0.12114**
JORDAN PHOSPHATE MINES	**1.80000**	**8.67159**	**4.05965**
NORTHERN CEMENT CO.	**0.07000**	**0.11382**	**0.13499**
THE ARAB POTASH	**3.51000**	**7.21611**	**2.60108**
INVESTMENTS & INTEGRATED INDUSTRIES CO. PLC (HOLDING CO)	**0.04000**	**0.03953**	**0.01875**
DAR AL DAWA DEVELOPMENT & INVESTMENT	**0.09033**	**0.06637**	**0.03005**
HAYAT PHARMACEUTICAL INDUSTRIES CO.	**0.24502**	**0.37144**	**0.38339**
PHILADELPHIA PHARMACEEUTICALS	**0.10763**	**0.09348**	**0.05470**
THE JORDAN WORSTED MILLS	**0.12776**	**0.14262**	**0.09987**

Table 6: Dataset III: EPS for the service sector during the 2021–2023 period

Company Name	Financial Ratios/Year		
	2023	2022	2021
BINDAR TRADING & INVESTMENT	0.262608	0.40806	0.170304
COMPREHENSIVE LEASING COMPANY	0.243656	0.22314	0.388221
JORDAN INTERNATIONAL TRADING CENTER	0.07173	0.093165	0.057167
JORDAN TRADE FACILITIES	0.374112	0.242446	0.238253
JORDANIAN DUTY FREE SHOPS	0.336582	0.375173	0.00718
AL-ISRA FOR EDUCATION AND INVESTMENT	0.205352	0.332692	0.263241
AL-ZARQA EDUCATIONAL & INVESTMENT	0.046396	0.039709	0.033073
PETRA EDUCATION COMPANY	0.140622	0.167452	0.230803
PHILADELPHIA INTERNATIONAL EDUCATIONAL INVESTMENT	0.11588	0.005117	0.019367
THE ARAB INTERNATIONL FOR EDUCATION & INVESTMENT	0.118594	0.056378	0.108322
JORDAN PETROLEUM REFINERY	1.226189	0.961974	0.520464
AFAQ FOR ENERGY CO. P.L.C	4.914741	3.491291	0.211068
NATIONAL PETROULEUM	2.495661	5.240506	0.378521
THE CONSULTANT & INVESTMENT GROUP	0.041827	0.044151	0.041361
ARAB INTERNATIONAL HOTELS	0.026987	0.008047	0.004793
AL-FARIS NATIONAL COMPANY FOR INVESTMENT & EXPORT	0.009391	0.0315	0.044468
JORDAN TELECOM	0.243991	0.234832	0.13933
JORDAN NATIONAL SHIPPING LINES	0.078482	0.257447	0.193284
SALAM INTERNATIONL TRANSPORT & TRADING	0.013732	0.117071	0.076555
TRUST INTERNATIONAL TRANSPORT	0.080481	0.007918	0.013924
IRBID DISTRICT ELECTRICITY	0.657753	2.070165	0.690224
JORDAN ELECTRIC POWER	0.187703	0.170691	0.114224
ELECTRICITY DISTRIBUTION	0.507459	1.578454	0.618777
CENTRAL ELECTRICITY GENERATING	4.424406	0.822969	0.398052

Table 7 shows the summary of the three data sets used in this study.

Table 7: Summary of the datasets

Dataset	Min.	1st Qu	Median	Mean	3rd Qu	Max.
I	0.00036	0.03950	0.11449	0.14826	0.20887	0.63802
II	0.00100	0.05058	0.09987	0.51149	0.21088	8.67159
III	0.004793	0.053883	0.190493	0.538412	0.380946	5.240506

Tables 8 - 10 show that the suggested distribution has the lowest values of $-ln(L)$, AIC, CAIC, BIC, HQIC and KS with the highest p-value. Therefore, the suggested distribution is preferred over the competence distributions. The 95% CIs of the parameter θ are calculated in these tables.

Table 8: Application I

Distribution	$-ln(L)$	*AIC*	*CAIC*	*BIC*	*HQIC*	*KS*	*p*-value	parameter	Estimate	*SE*	95%CI	
Gharaibeh	**268.9182**	**498.5167**	**498.6809**	**504.66**	**501**	**0.79057**	**2.23E-05**	θ	**1.40952**	**0.07806**	**1.256519**	**1.562515**
ABOPLED	**214.7046**	**247.6625**	**247.8083**	**249.848**	**248.5**	**0.108**	**0.5671**	θ	**2.45937**	**0.148823**	**2.16768**	**2.751065**
Karam	**220.2601**	**254.0115**	**254.1574**	**256.197**	**254.8**	**0.1224**	**0.409**	θ	**0.926**	**0.162**	**0.60848**	**1.24352**
OPLED	**240.4058**	**246.3666**	**248.4974**	**250.326**	**249.1**	**0.19233**	**0.043**	θ	**0.0924**	**0.009725**	**0.07334**	**0.111461**
Exponential	**254.1606**	**412.3661**	**412.4869**	**415.313**	**413.5**	**0.523845**	**0.000126**	θ	**0.20058**	**0.027815**	**0.146062**	**0.255096**
Lindley	**283.1807**	**493.5926**	**493.9845**	**499.837**	**496**	**0.835538**	**2.99E-07**	θ	**0.89039**	**0.055254**	**0.782089**	**0.998684**
ABBD	**339.3911**	**546.5599**	**546.8783**	**551.633**	**548.5**	**0.952397**	**2.42E-12**	α	**1.32772**	**0.131967**	**1.069067**	**1.586376**
								β	**0.59652**	**0.047479**	**0.503466**	**0.689583**
LBBD	**292.5332**	**519.4088**	**519.8325**	**526.16**	**522**	**0.42493**	**4.91E-09**	α	**13.4936**	**1.916729**	**9.736853**	**17.25043**
								β	**0.79805**	**0.093192**	**0.615392**	**0.980704**
SBID	**342.8647**	**555.0833**	**555.1819**	**557.489**	**556**	**0.910601**	**1.46E-12**	θ	**0.8271**	**0.215**	**0.405702**	**1.248502**

Table 9: Application II

Distribution	$-ln(L)$	*AIC*	*CAIC*	*BIC*	*HQIC*	*KS*	*p*-value	parameter	Estimate	*SE*	95%CI	
Gharaibeh	**131.4238**	**264.8477**	**264.9074**	**267.0818**	**265.734**	**0.40153**	**4.35E-10**	θ	**1.100993**	**0.062543**	**0.978409**	**1.223577**
ABOPLED	**629.0371**	**1262.074**	**1262.198**	**1267.284**	**1264.183**	**0.095631**	**0.3198**	θ	**0.129883**	**0.011658**	**0.107032**	**0.152733**
Karam	**91.09243**	**184.1849**	**184.2446**	**186.419**	**185.0712**	**0.26513**	**0.000123**	θ	**1.97627**	**0.106528**	**1.767475**	**2.185065**
OPLED	**89.82951**	**181.659**	**181.7187**	**183.8931**	**182.5454**	**0.28699**	**2.32E-05**	θ	**1.408618**	**0.078005**	**1.255727**	**1.561509**
Exponential	**130.9789**	**263.9578**	**264.0175**	**266.1919**	**264.8441**	**0.4477**	**1.94E-12**	θ	**0.407271**	**0.049029**	**0.311173**	**0.503368**
Lindley	**119.311**	**240.622**	**240.6817**	**242.8561**	**241.5083**	**0.40044**	**4.90E-10**	θ	**0.653587**	**0.049124**	**0.557304**	**0.74987**
ABBD	**396.0404**	**796.0808**	**796.2547**	**800.6341**	**797.8935**	**0.15922**	**0.05196**	α	**4.546753**	**1.956473**	**0.712066**	**8.38144**
								β	**0.032501**	**0.001715**	**0.02914**	**0.035863**
LBBD	**116.3911**	**234.7822**	**234.9201**	**236.2161**	**235.2496**	**0.14041**	**0.1951**	α	**2.444609**	**1.21023**	**0.072558**	**4.816659**
								β	**0.008821**	**0.000642**	**0.007563**	**0.010078**
SBID	**96.31876**	**194.6375**	**194.6972**	**196.8716**	**195.5239**	**0.29669**	**1.06E-05**	θ	**1.353036**	**0.072632**	**1.210677**	**1.495395**

Table 10: Application III

Distribution	$-ln(L)$	*AIC*	*CAIC*	*BIC*	*HQIC*	*KS*	*p*-value	parameter	Estimate	*SE*	95%CI	
Gharaibeh	**190.437**	**382.016**	**382.04**	**383.505**	**382.619**	**0.15386**	**1.76E-02**	θ	**0.53498**	**0.13245**	**0.27537**	**0.79458**
ABOPLED	**63.473**	**127.746**	**127.771**	**128.788**	**128.168**	**0.04221**	**0.6136**	θ	**0.40425**	**0.05325**	**0.29988**	**0.50862**
Karam	**268.019**	**537.53**	**537.56**	**539.474**	**538.317**	**0.2158**	**0.000123**	θ	**0.59548**	**0.16864**	**0.26495**	**0.92602**
OPLED	**333.624**	**669.248**	**669.289**	**671.854**	**670.303**	**0.92175**	$< 2.2E-16$	θ	**0.63064**	**0.15134**	**0.33401**	**0.92727**
Exponential	**243.719**	**488.92**	**488.95**	**490.85**	**489.701**	**0.17301**	**5.02E-03**	θ	**0.10125**	**0.04523**	**0.0126**	**0.1899**
Lindley	**74.541**	**149.55**	**149.56**	**150.159**	**149.797**	**0.06768**	**0.5881**	θ	**0.18657**	**0.06413**	**0.06089**	**0.31226**
ABBD	**321.539**	**647.077**	**647.201**	**652.287**	**649.186**	**0.44059**	$< 2.2E-16$	α	**4.87299**	**0.75101**	**3.401**	**6.34498**
								β	**0.41435**	**0.05011**	**0.31614**	**0.51256**
LBBD	**144.212**	**290.243**	**290.299**	**292.611**	**291.201**	**0.12309**	**9.66E-02**	α	**3.95669**	**1.67379**	**0.67606**	**7.23733**
								β	**0.23114**	**0.03585**	**0.16087**	**0.3014**
SBID	**188.221**	**377.572**	**377.595**	**379.046**	**378.169**	**0.15845**	**1.32E-02**	θ	**0.40413**	**0.13112**	**0.14713**	**0.66112**

10. Conclusions

In this paper, we propose the area-biased one-parameter linear exponential distribution. The main properties of this distribution are derived such as the moments and the related measures, the harmonic mean and the mode. The reliability analysis functions are derived along with the pdfs of the minimum, maximum and the k^{th} order statistics; additionally, the quantile function; additionally, the mean absolute deviations from the mean and the median jointly with the mean waiting and residual lifetime. A simulation study using the MLE, OLS, WLS, MPS, CVM and AD methods of estimating parameters is conducted showing that the estimators are unbiased and consistent. Three real financial data applications prove the goodness of fit for this distribution. They show that the suggested distribution fits the real data better than the competence distributions.

Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable comments.

References

Al-Omari, A., Al-Nasser, A., and Ciavolino, E., (2019). A Size-Biased Ishita Distribution and Application to Real Data. *Quality & Quantity: International Journal of Methodology*, 53(1), pp. 493–512.

Al-Omari, A., Alanzi, A., (2021). Inverse Length Biased Maxwell Distribution: Statistical Inference with an application . *Computer Systems Science and Engineering*, 39(1), pp. 147–164.

Al-Omari, A., Alhyasat, K., Ibrahim, K. and Abu-Bakar, M., (2019). Power length biased Suja distribution: properties and application . Electronic Journal of App.ied Statistical Analysis, 12(2), pp. 429–452.

Almakhareez, S., Alzoubi, L., (2024). Extension to Benrabia Distribution with applications and Parameter Estimation. *RESEARCH IN STATISTICS*, 2(1).

Almakhareez, S. and Alzoubi, L., (2024). Length-Biased Benrabia Distribution with applications. *Electronic Journal of App.ied Statistical Analysis (EJASA)*, 17(2), pp. 255–277.

Alzoubi, L., Gharaibeh, M., Alkhazaalh, A. and Benrabia, M., (2022). Estimation of Sameera Distribution Parameters with applications to Real Data. *App.ied Mathematics Information Sciences*, 16, pp. 1057–1071.

Alzoubi, L., Gharaibeh, M., Alkhazaalh, A. and Berabia, M., (2022). Loai distribution: Properties, parameters estimation and application to covid-19 real data. *Mathematical Statistician and Engineering applications*, 7(4), pp. 1231–1255.

Anderson, T., (1962). Distribution of the two-sample cramer–von mises criterion. *Annals of Mathematical Statistics*, 33(3), pp. 1148–1159.

ASE, (2021-2023). Amman stock exchange. Online.

Au, M., Agba, B. and Gagnon, F., (2015). Fast identification of partial discharge sources using blind source separation and kurtosis. *Electronics Letters*, 51(25), pp. 2132–2134.

Benrabia, M., Alzoubi, L., (2022). Alzoubi Distribution: Properties and applications. *Journal of Statistics applications & Probability, An International Journal*, 11(2), pp. 625–640.

Benrabia, M., Alzoubi, L., (2022). Benrabia Distribution: Properties and applications. *Electronic Journal of App.ied Statistical Analysis*, 15(2), pp. 300–317.

Cheng, R. and Amin, A., (1983). Estimating Parameters in Continuous Univariate Distributions with a Shifted Origin. *Journal of the Royal Statistical Society. Series B (Methodological)*, 45(3), pp. 394–403.

Cox, D., (1968). *Some Sampling Problems in Technology*. In: Johnson, N.L. and Smith, H., Eds., New Developments in Survey Sampling.

Cramér, H., (1928). On the Composition of Elementary Errors. *Scandinavian Actuarial Journal*, 1, pp. 13–74.

Dang, N. , Dam, H., Dinh, A. and Le, T., (2024). The factors influencing the stock prices of construction industry enterprises: An empirical study in Vietnam. Revista de Gestão Social e Ambiental, 18(3), pp. 1–17.

Das, K., Roy, T., (2011). On Some Length-Biased Weighted Weibull Distribution. *Advances in App.ied Science Research*, 2, pp. 465–475.

Ferniawan, M., Kusumawati, A. and Afdal, (2024). The influence of earnings per share (eps), price earnings ratio (per), price to book value (pbv), and debt equity ratio(der) on the stock return. *Akrual: Jurnal Bisnis dan Akuntansi Kontemporer*, 17(1), pp. 114–130.

Fisher, R., (1934). The Effect of Methods of Ascertainment upon the Estimation of Frequencies. *The Annals of Human Genetics*, 6, pp. 13–25.

Gharaibeh, M., Sahtout, M., (2022). Karam distribution: A new lifetime distribution with real data application. *Journal of App.ied Probability & Statistics*, 17(1), pp. 71–85.

Gharaibeh, M., (2021). Gharaibeh Distribution and Its applications. *Journal of Statistics applications & Probability*, 10(2), pp. 441–452.

Ghitany, M., Atieh, B. and Nadarajah, S., (2008). Lindley Distribution and its applications. *Mathematics and Computers in Simulation*, 78, pp. 493–506.

IAS, (2021–2023). International accounting standards (ias,33).

Jeong, J., Jung, S., and Costantino, J., (2008). Nonparametric inference on median residual life function. *Biometrics*, 64(1), pp. 157–163.

Joe, H., Proschan, F. (1984a). Comparison of Two Life Distributions on the Basis of Their Percentile Residual Life Functions. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 12(2), pp. 91–97.

Joe, H., Proschan, F., (1984). Percentile Residual Life Functions. *Operations Research*, 32(3), pp. 668–678.

Kingman, J., (1982). The coalescent. *Stochastic Processes and their applications*, 13(3), pp. 235–248.

Kristanti, F. , Nurhayati, I. and Fariska, P., (2024). Financial performance's impact on firm value: How managerial ownership mediates the relationship. *Journal of Ecohumanism*, 3(6), pp. 204–216.

Lillo, R., (2005). On the Median Residual Lifetime and its Aging Properties: A Characterization Theorem and applications. *Naval Research Logistics (NRL)*, 52(4), pp. 370–380.

Manikandan, S., (2011). Measures of central tendency: Median and mode. *Journal of Pharmacology and Pharmacotherapeutics*, 2(3), pp. 214–215.

Olbryś, J., Ostrowski, K., (2021). An Entropy-Based App.oach to Measurement of Stock Market Depth. *Entropy*, 23(5).

Otsuka, A., Nakano, K. and Miyakita, K. (2010). Theoretical Analysis of Mean Waiting Time for Message Delivery in Lattice Ad Hoc Networks. *Journal of Circuits, Systems and Computers*, 19(08), pp. 1711–1741.

Patil, G., Ord, J., (1976). On size-biased sampling and related form-invariant weighted distributions. Sankhya: The Indian Journal of Statistics, 38(1), pp. 48–61.

Pollaczek, F., (1957). *Problémes stochastiques posés par le phénoméne de formation d´une queue d´attente á un guichet et par des phénoménes app.rentés*. Gauthier-Villars.

R Core Team, (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rao, C., (1965). On Discrete Distributions Arising out of Methods of Ascertainment. *Sankhya: The Indian Journal of Statistics, Series A*, (1961-2002), 27(2/4), pp. 311–324.

Rényi, A., (1961). On Measures of Entropy and Information. Proceedings of the 4th Berkeley *Symposium on Mathematical Statistics and Probability*, 1, pp. 547–561.

Romero-Silva, R., Hurtado, M., (2017). The difference of mean waiting times between two classes of customers in a single-server fifo queue: An experimental study. *Cogent Engineering*, 4(1), pp. 500–517.

Rosenberg, L. (1968). Mean Waiting Time as a Measure of Effectiveness. *The American Statistician*, 22(4), pp. 31–34.

Saghir, A., Khadim, A. and Lin, Z., (2017). The Maxwell-length-biased distribution: Properties and Estimation. *Journal of Statistical Theory and Practice*, 11(1), pp. 26–40.

Sah, B., (2021). One-Parameter Linear-Exponential Distribution. *International Journal of Statistics and App.ied Mathematics*, 6(6), pp. 6–15.

Schmittlein, D., Morrison, D., (1981). The Median Residual Lifetime: A Characterization Theorem and an application. *Operations Research*, 29(2), pp. 392–399.

Shannon, C., (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27:379–423, pp. 623–656.

Sharma, V., Dey, S., Singh, S. and Manzoor, U., (2018). On length and area biased maxwell distributions . *Communications in Statistics - Simulation and Computation*, 47(5):1506– 1528.

Shen, Y., Ning, J. and Qin, J. (2009). Analyzing length-biased data with semiparametric transformation and accelerated failure time models. *Journal of the American Statistical Association*, 104(487), pp. 1192–1202.

Swain, J., Venkatraman, S. and Wilson, J., (1988). Least-squares estimation of distribution functions in johnson's translation system. Journal of Statistical Computation and Simulation, 29(4), pp. 271–297.

Taubah, I., Roslina, N. Y., Damayanti, I., and Tansar, I., (2024). The influence of earning per share (eps) and price earnings ratio (per) on stock prices in the food and beverage sector on the indonesian stock exchange (bei) for the 2015-2019 period. *Journal of Economics, Management, and Entrepreneurship*, 2(1), pp. 49–63.

Tsallis, C., (1988). Possible generalization of Boltzmann-Gibbs statistics. Journal of Statistical Physics, 52, pp. 479–487.

Von Mises, R., (1928). *Wahrscheinlichkeit, Statistik und Wahrheit*. Julius Springer.

Zamanzade, E., Zamanzade, E. and Parvardeh, A., (2024). Estimation of a decreasing mean residual life based on ranked set sampling with an application to survival analysis. *The international journal of biostatistics*, 20(2), pp. 571–583. https://doi.org/10.1515/ijb-2023-0051

Zelen, M., (1974). Problems in cell kinetics and the early detection of disease. *Reliability and Biometry, SIAM*, 56(3), pp. 701–706.

STATISTICS IN TRANSITION new series, September 2025
Vol. 26, No. 3, pp. 59–79, https://doi.org/10.59139/stattrans-2025-027
Received – 30.12.2022; accepted – 30.12.2024

Bayesian estimation of two-parameter power Rayleigh distribution and its application

Mohd Irfan[1], **Anup Kumar Sharma**[2]

Abstract

This paper explores classical and Bayesian approaches to the estimation of unknown parameters and reliability functions for the power Rayleigh distribution. The maximum likelihood estimator (MLE) method is considered in classical estimation. The Bayesian estimation, on the other hand uses several loss functions under informative and non-informative prior distributions, utilizing the Lindley technique and Markov chain Monte Carlo (MCMC) methods for Bayesian computations. Approximate confidence intervals are established based on the MLEs using the delta technique, while Bayes credible intervals are determined using the MCMC method. A simulation study is conducted to compare the performance of these methods in terms of biases and mean square errors, revealing that Bayesian estimators outperform their classical counterparts. Additionally, two real datasets are presented for illustrative purposes.

Key words: Power Rayleigh distribution, delta method, Lindley approximation, Metropolis-Hasting algorithm, highest posterior density credible intervals, Monte Carlo simulation, coverage probability, goodness of fit.

1. Introduction

Parameter estimation is a fundamental aspect of statistics, playing a crucial role in various statistical analyses and decision-making processes. Estimating parameters in statistical distributions involves two primary methodologies: frequentist and Bayesian. In the frequentist paradigm, estimates are derived from observed data, treating parameters as fixed but unknown values. A common technique within this framework is maximum likelihood estimation (MLE), where parameter values are selected to maximize the likelihood function. In contrast, the Bayesian approach treats parameters as random variables with associated probability distributions, acknowledging the inherent uncertainty. Bayesian parameter estimation combines prior beliefs about parameters with observed data using Bayes' theorem, resulting in a posterior distribution that reflects updated knowledge. While the Bayesian approach provides a systematic means to incorporate prior information and adapt beliefs as more data becomes available, it necessitates specifying a prior distribution, introducing subjectivity that may impact results. Several authors have used different lifetime models to

[1]National Institute of Technology Raipur, India. E-mail: irfan.maya786@gmail.com. ORCID: https://orcid.org/0009-0000-2791-1005.

[2]National Institute of Technology Raipur, India., E-mail: aksharma.ism@gmail.com. ORCID: https://orcid.org/0000-0002-3958-1901.

study parametric inference under frequentist and Bayesian approaches. For example, Soliman (2000), Shuo-Jye-Wu *et al.* (2006), Soliman and Al-Aboud (2008), Dey (2009), Khan *et al.* (2010), Ahmad *et al.* (2013), Asgharzadeh and Azizpour (2016), Ghazal and Hasaballah (2017), Talukdar (2019), Yilmaz and Kara (2022), Irfan and Sharma (2023), and Irfan and Sharma (2024) have considered different distributions for parameter estimation under classical and Bayesian approaches. Some of them also discussed life testing and reliability estimates under various loss functions.

The Rayleigh distribution, a statistical model, stands out for its unique application in representing the magnitudes of vector components featuring random amplitudes. Named after the British scientist Lord Rayleigh, this distribution finds widespread use in diverse domains like wireless communication, radar systems, and image processing. The Rayleigh distribution is a crucial model used in reliability theory, survival analysis, physical sciences, medical imaging, and various branches of engineering. This study focuses on the power Rayleigh distribution (PRD). PRD is the extension of the Rayleigh distribution and was introduced by Bhatt and Ahmad (2020). They also studied some exciting properties like moment-generating function, hazard rate, mean residual life, order statistics, quantiles, etc. PRD offers flexibility for modelling the real-life dataset with a long-tailed, right-skewed curve. Due to the wide practical utility of the PRD, various scholars have studied it for different purposes. For example, Mahmoud *et al.* (2020) discussed the lifetime performance index of PRD under progressive first failure censored data. Kilany *et al.* (2023) obtained the classical estimates for PRD under a complete sample with COVID-19 application. Migdadi *et al.* (2023) derived the Bayes estimates of the parameters of PRD under adaptive type II progressive censored sample. Migdadi *et al.* (2023) discussed the optimal design for the PRD under censored sample for the k-level step-stress accelerated life test. Further, Migdadi *et al.* (2023) obtained the Bayesian and classical estimates of PRD under joint progressive censoring scheme.

This study estimates the parameters and reliability characteristics of PRD using maximum likelihood and derives approximate confidence intervals (ACIs) via the Fisher information matrix. In the Bayesian framework, Bayes estimators under SELF, GELF, and LLF are computed using informative and non-informative priors. Since closed-form solutions are unavailable, we use MCMC and Lindley approximation methods, with Bayesian credible intervals (BCIs) obtained via the Metropolis-Hastings algorithm. A Monte Carlo simulation evaluates the methods, and two real datasets illustrate their application.

The novelty of this paper comes from the fact that no previous study has been found on the reliability and parameter estimation for PRD in the Bayesian context.

The rest of the paper is structured as follows: Section 2 provides an overview of the power Rayleigh distribution. Section 3 discusses the frequentist method for estimating unknown parameters. Section 4 focuses on obtaining Bayes estimators for the unknown model parameters, employing Lindley's method, and reliability estimation through different loss functions. Additionally, Section 4 incorporates MCMC methods. Section 5 presents a simulation study to assess the performance of the estimators in terms of biases and mean square errors (MSE). Section 6 applies two real datasets for practical demonstration and application purposes. Section 7 concludes the work with some suggestions for future research in this field.

2. Power Rayleigh Distribution

Let Y be a random variable that follows Rayleigh distribution with parameter ν whose cumulative distribution function (CDF) and probability distribution function (PDF) are, respectively, given by

$$F(y) = 1 - exp\left(-\frac{y^2}{2\nu^2}\right); y > 0, \nu > 0, \tag{1}$$

and

$$f(y) = \frac{y}{\nu^2} exp\left(-\frac{y^2}{2\nu^2}\right); y > 0, \nu > 0. \tag{2}$$

Now, the transformation $X = Y^{\frac{1}{\tau}}$ will follow power Rayleigh distribution (PRD) with parameter ν and τ be obtained as

$$\begin{aligned} F(x) &= p(X \leq x) \\ &= p(x^{\tau}). \end{aligned} \tag{3}$$

Using equations (1), (2), and (3), the cumulative distribution function (CDF) and probability distribution function (PDF) of PRD are, respectively, given as

$$F(x) = 1 - exp\left(-\frac{x^{2\tau}}{2\nu^2}\right), x > 0, \nu, \tau > 0, \tag{4}$$

and

$$f(x) = \frac{\tau}{\nu^2} x^{2\tau-1} exp\left(-\frac{x^{2\tau}}{2\nu^2}\right), x > 0, \nu, \tau > 0, \tag{5}$$

where ν and τ are scale and shape parameters of PRD, respectively.

Remark: If $\tau = 1$ then the equation (4) is reduced to the cumulative distribution function of the Rayleigh distribution.

The survival and hazard rate function of PRD are, respectively, given by

$$R(t) = exp\left(-\frac{t^{2\tau}}{2\nu^2}\right); t > 0, \tag{6}$$

and

$$H(t) = \frac{\tau}{\nu^2} t^{2\tau-1}; t > 0. \tag{7}$$

The plots of the hazard rate function for some values of the parameter are depicted in Figure 1.

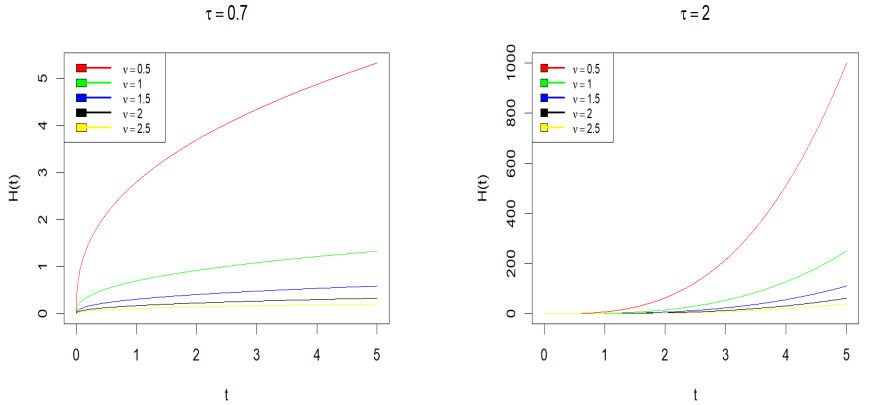


Figure 1: The Hazard rate function fo PRD.

3. Frequentist Approach

In this section, we will obtain PRD's unknown model parameters and survival functions using the maximum likelihood estimation (MLE) method. We will also establish the approximate confidence intervals for the parameters and any functions of them.

3.1. Maximum likelihood estimation

Let $x_1, x_2, ..., x_n$ be a random sample drawn from the PRD with PDF (5). Therefore, the likelihood function, $L(\nu, \tau|x)$, is the joint density of the random sample, $x_1, x_2, ..., x_n$, and is given by

$$L(\tau, \nu|x) = \prod_{i=1}^{n} \left[\frac{\tau}{\nu^2} x_i^{2\tau-1} exp\left(-\frac{x^{2\tau}}{2\nu^2} \right) \right]. \tag{8}$$

On simplifying, we get

$$L(\tau, \nu|x) = \left(\frac{\tau}{\nu^2} \right)^n \prod_{i=1}^{n} x_i^{2\tau-1} exp\left(-\frac{\sum_{i=1}^{n} x_i^{2\tau}}{2\nu^2} \right). \tag{9}$$

On taking the logarithm of equation (9), the log-likelihood function can be expressed as follows:

$$l(\tau, \nu|x) = n\log(\tau) - 2n\log(\nu) + (2\tau - 1)\sum_{i=1}^{n} \log(x_i) - \frac{\sum_{i=1}^{n} x_i^{2\tau}}{2\nu^2}. \tag{10}$$

Now, differentiating equation (10) with respect to τ and ν and equating the resulting

terms to zero, the normalizing equations are given as

$$\frac{\partial l}{\partial \tau} = \frac{n}{\tau} + 2\sum_{i=1}^{n} \log(x_i) - \frac{\sum_{i=1}^{n} x_i^{2\tau} \log(x_i)}{\nu^2} = 0, \tag{11}$$

and

$$\frac{\partial l}{\partial \nu} = -\frac{2n}{\nu} + \frac{\sum_{i=1}^{n} x_i^{2\tau}}{\nu^3} = 0. \tag{12}$$

Let $\hat{\tau}$ and $\hat{\nu}$ represent MLEs of τ and ν, respectively. Then $\hat{\tau}$ is obtained by solving the nonlinear equation (11) using Newton-Rapson iterative procedures. Thereafter, $\hat{\nu}$ can be obtained by substituting $\hat{\tau}$ into equation (12) and solving the resulting expression as follows:

$$\hat{\nu} = \left(\frac{\sum_{i=1}^{n} x_i^{2\hat{\tau}}}{2n} \right)^{\frac{1}{2}}. \tag{13}$$

Using invariance property of the MLE the $R(t)$, say $\hat{R}(t)$, can be written as

$$\hat{R}(t) = exp\left(-\frac{t^{2\hat{\tau}}}{2\hat{\nu}^2} \right). \tag{14}$$

3.2. Approximate confidence intervals

Since all the second-order partial derivatives exist in their domain. Therefore, the interval estimates for the parameters τ and ν are derived using the Fisher information matrix as follows:

$$I(\tau, \nu) = -E \begin{bmatrix} I_{\tau\tau} & I_{\tau\nu} \\ I_{\nu\tau} & I_{\nu\nu} \end{bmatrix}, \tag{15}$$

where $I_{\tau\tau} = \frac{\partial^2 l}{\partial \tau^2} = -\frac{n}{\tau^2} - \frac{2\sum_{i=1}^{n} x_i^{2\tau} (\log(x_i))^2}{\nu^2}$,
$I_{\tau\nu} = I_{\nu\tau} = \frac{\partial^2 l}{\partial \tau \partial \nu} = \frac{\partial^2 l}{\partial \nu \partial \tau} = \frac{2}{\nu^3} \sum_{i=1}^{n} x_i^{2\tau} \log(x_i)$, and $I_{\nu\nu} = \frac{2n}{\nu^2} - \frac{3}{\nu^4} \sum_{i=1}^{n} x_i^{2\tau}$.

From (15), it is observed that the exact solution of expectation is tedious to evaluate. Therefore, approximate variance and covariance matrix $I^{-1}(\hat{\tau}, \hat{\nu})$ is given by

$$I^{-1}(\hat{\tau}, \hat{\nu}) = \begin{bmatrix} -I_{\tau\tau} & -I_{\tau\nu} \\ -I_{\nu\tau} & -I_{\nu\nu} \end{bmatrix}^{-1}_{(\hat{\tau}, \hat{\nu})}. \tag{16}$$

According to Gomez-Deniz (2010), the MLE vector $(\hat{\tau}, \hat{\nu})^T$ is consistent and asymptotically normally distributed as follows:

$$\sqrt{n}\left[(\hat{\tau}, \hat{\nu})^T - (\tau, \nu)^T \right] \to N(0, I^{-1}(\tau, \nu)).$$

The $100(1-\xi)\%$ confidence intervals τ and ν can be approximated by $\hat{\tau} \pm Z_{\frac{\xi}{2}} \sqrt{var(\hat{\tau})}$ and $\hat{\nu} \pm Z_{\frac{\xi}{2}} \sqrt{var(\hat{\nu})}$, respectively, where the diagonal elements of $I^{-1}(\tau,\nu)$ represent the variance of parameters of τ and ν while the off-diagonal elements represent the covariance between τ and ν, and $Z_{\frac{\xi}{2}}$ is the $(1-\frac{\xi}{2})$ quantile of the standard normal distribution.

To derive the confidence intervals for $R(t)$, we must know their variance, and for this the delta method is adapted to obtain the variances of $R(t)$. For more details about the delta method, one may refer to W. H. Greene (2003). Utilising the delta technique, the approximate variance of $\hat{R}(t)$ (at their MLEs $\hat{\tau}$ and $\hat{\nu}$) can be expressed as follows:

$$var(\hat{R}(t)) = \left[\nabla \hat{R}(t)\right]^T I^{-1}(\hat{\tau},\hat{\nu})[\nabla \hat{R}(t)], \tag{17}$$

where T is stands for transpose operator and $\nabla \hat{R}(t)$ is the gradient of reliability function, respectively, with respect to τ and ν, i.e.,

$$[\nabla R(t)]^T = \left[\frac{\partial R(t)}{\partial \tau}, \frac{\partial R(t)}{\partial \nu}\right]_{(\hat{\tau},\hat{\nu})}. \tag{18}$$

From equation (6) the derivative of $R(t)$ with respect to τ and ν is obtained as follows:

$$\frac{\partial R(t)}{\partial \tau} = -\frac{t^{2\tau}}{\nu^2}\, \log(t)\, exp\left(-\frac{t^{2\tau}}{2\nu^2}\right), \quad \frac{\partial R(t)}{\partial \nu} = \frac{t^{2\tau}}{\nu^3}\, exp(-\frac{t^{2\tau}}{2\nu^2}).$$

Therefore, the $100(1-\xi)\%$ confidence intervals for $R(t)$ can be derived as follows:

$$\hat{R}(t) \pm Z_{\frac{\xi}{2}} \sqrt{var(\hat{R}(t))}.$$

4. Bayesian inference

In this section, we will consider the Bayesian estimation method to estimate the parameters and any parametric function of PRD based on the complete sample.

4.1. Loss functions

In Bayesian estimation, the loss function quantifies the cost of estimation errors, guiding decision-making to minimize expected loss by balancing precision and bias. Common loss functions include SELF (symmetric), GELF, and LINEX (asymmetric). Let $\tilde{\phi}$ be an estimator of ϕ. The SELF, GELF, and LLF are defined as follows:

$$L_{SE}(\phi,\tilde{\phi}) = (\phi-\tilde{\phi})^2, \tag{19}$$

$$L_{GE}(\phi,\tilde{\phi}) = \left(\frac{\tilde{\phi}}{\phi}\right)^q - q\, \log\left(\frac{\tilde{\phi}}{\phi}\right) - 1, q \neq 0, \tag{20}$$

and

$$L_{LI}(\phi,\tilde{\phi}) = e^{-q(\phi-\tilde{\phi})} - p(\phi-\tilde{\phi}) - 1, q \neq 0. \tag{21}$$

4.2. Prior distribution

In Bayesian analysis, the prior represents initial beliefs about parameters before data integration. Selecting priors is critical but lacks a fixed rule. Here, a non-informative prior is assumed for τ, and a conjugate prior for ν, defined as:

$$\pi_1(\tau) = \frac{1}{\tau};\ \ \tau > 0,$$

and

$$\pi_2(\nu) = \frac{1}{\nu^{a+1}} exp\left(-\frac{b}{2\nu^2}\right);\ \ a,b > 0,$$

where $a > 0$ and $b > 0$ are hyperparameters. The joint prior distribution for τ and ν is given as:

$$\pi(\tau,\nu) = \frac{1}{\tau\nu^{a+1}} exp\left(-\frac{b}{2\nu^2}\right);\ \tau > 0, \nu, a, b > 0. \tag{22}$$

4.3. Posterior analysis

The posterior density function of τ and ν is obtained by combining equation (9) and (22), which is written as follows:

$$\pi(\tau,\nu|x) = K^{-1} \frac{\tau^{n-1}}{\nu^{2n+a+1}} \prod_{i=1}^{n} x_i^{2\tau-1} exp\left(-\frac{b+\sum_{i=1}^{n} x_i^{2\tau}}{2\nu^2}\right), \tag{23}$$

where $K = \int_0^\infty \int_0^\infty \frac{\tau^{n-1}}{\nu^{2n+a+1}} \prod_{i=1}^{n} x^{2\tau-1} exp\left(-\frac{b+\sum_{i=1}^{n} x_i^{2\tau}}{2\nu^2}\right) d\tau d\nu$, is called the normalising constant. Equation (23) cannot be converted analytically due to the complex form of the likelihood function. In order to derive the Bayes estimates, the Lindley approximation technique is used for further analysis.

4.4. Lindley's method

Let $u(\tau,\nu)$ be the function of τ and ν, then by using the equation (23), the expected value of $u(\tau,\nu)$ is given by:

$$E(u(\tau,\nu|x)) = \frac{\int_\tau \int_\nu u(\tau,\nu) e^{l(\tau,\nu|\underline{\mathbf{x}})+\eta(\tau,\nu)} d\tau d\nu}{\int_\tau \int_\nu e^{l(\tau,\nu|\underline{\mathbf{x}})+\eta(\tau,\nu)} d\tau d\nu}.$$

The Bayes estimator $u(\tau,\nu)$ is the solution of the above equation. Unfortunately, it is very hard to obtain the Bayes estimator analytically due to its dependence on the ratio of two integrals. To overcome this difficulty, the Lindley's technique introduced by Lindley's 1980, is used. Let the I be the approximate value of $E(u(\tau,\nu|x))$ then

$$
\begin{aligned}
I(x) \quad = \quad & \hat{u}(\tau,\nu)+0.5(\hat{u}_{\tau\tau}\hat{\sigma}_{\tau\tau}+\hat{u}_{\nu\nu}\hat{\sigma}_{\nu\nu})+\hat{u}_{\tau\nu}\hat{\sigma}_{\tau\nu}+\hat{u}_{\tau}(\hat{\sigma}_{\tau\tau}\hat{\eta}_{\tau}+\hat{\sigma}_{\nu\tau}\hat{\eta}_{\nu})+\\
& \hat{u}_{\nu}(\hat{\sigma}_{\tau\nu}\hat{\eta}_{\tau}+\hat{\sigma}_{\nu\nu}\hat{\eta}_{\nu})+0.5\hat{l}_{\tau\tau\tau}(\hat{u}_{\tau}\hat{\sigma}^{2}_{\tau\tau}+\hat{u}_{\nu}\hat{\sigma}_{\tau\tau}\hat{\sigma}_{\tau\nu})+0.5\hat{l}_{\tau\tau\nu}\\
& \left(3\hat{u}_{\tau}\hat{\sigma}_{\tau\tau}\hat{\sigma}_{\tau\nu}+\hat{u}_{\nu}(\hat{\sigma}_{\tau\tau}\hat{\sigma}_{\nu\nu}+2\hat{\sigma}^{2}_{\tau\nu})\right)+0.5\hat{l}_{\tau\nu\nu}\left(\hat{u}_{\tau}(\hat{\sigma}_{\tau\tau}\hat{\sigma}_{\nu\nu}+2\hat{\sigma}^{2}_{\tau\nu})\right.\\
& \left.+3\hat{u}_{\nu}\hat{\sigma}_{\tau\nu}\hat{\sigma}_{\nu\nu}\right)+0.5\hat{l}_{\nu\nu\nu}(\hat{u}_{\tau}\hat{\sigma}_{\tau\nu}\hat{\sigma}_{\tau\tau}+\hat{u}_{\nu}\hat{\sigma}^{2}_{\nu\nu}), \qquad (24)
\end{aligned}
$$

where $\hat{u}(\tau,\nu)$ is the function of τ and ν evaluated at $\hat{\tau}$ and $\hat{\nu}$ and σ_{ij} is the $(i,j)^{th}$ element of matrix $[-\hat{l}_{ij}]^{-1}; i,j=1,2$.

The other notation is interpreted with the following definition, such that:
$\hat{u}_{\tau}=\frac{\partial u}{\partial\tau}$, $\hat{u}_{\nu}=\frac{\partial u}{\partial\nu}$, $\hat{u}_{\nu\tau}=\hat{u}_{\tau\nu}=\frac{\partial^2 u}{\partial\nu\partial\tau}$, $\hat{\eta}_{\nu}=\frac{\partial\log\pi(\tau,\nu)}{\partial\nu}$, $\hat{\eta}_{\tau}=\frac{\partial\log\pi(\tau,\nu)}{\partial\tau}$, $\hat{l}_{\nu}=\frac{\partial l}{\partial\nu}$, $\hat{l}_{\tau}=\frac{\partial l}{\partial\tau}$

$\hat{l}_{\nu\tau}=\hat{l}_{\tau\nu}=\frac{\partial^2 l}{\partial\nu\tau}$, $\hat{l}_{\nu\nu}=\frac{\partial^2 l}{\partial\nu^2}$, $\hat{l}_{\tau\tau}=\frac{\partial^2 l}{\partial\tau^2}$, $\hat{l}_{\nu\nu\nu}=\frac{\partial^3 l}{\partial\nu^3}$, $\hat{l}_{\tau\tau\tau}=\frac{\partial^3 l}{\partial\tau^3}$, $\hat{l}_{\nu\nu\tau}=\frac{\partial^3 l}{\partial\nu^2\partial\tau}$, and $\hat{l}_{\nu\tau\tau}=\frac{\partial^3 l}{\partial\nu\tau^2}$.

The Bayes estimates of τ, ν, and $S(t)$ under the SELF, GELF, and LLF are obtained in the following subsections.

4.4.1 Bayesian estimation under SELF

One of the very popular symmetric loss function is SELF, which was first addressed by Legendre (1805), which endows equal weight for overestimation and underestimation.

The Bayes estimate (BE) of parameter τ under SELF is found as follows:

$$
\hat{\tau}_{BS}=E(\tau|x)=\frac{\int_0^\infty\int_0^\infty \tau^n \frac{1}{\nu^{2n+a+1}}\prod_i^n x_i^{2\tau-1} exp\left(\frac{b+\sum_i^n x_i^{2\tau}}{2\nu^2}\right)\, d\tau d\nu}{\int_0^\infty\int_0^\infty \tau^{n-1} \frac{1}{\nu^{2n+a+1}}\prod_i^n x_i^{2\tau-1} exp\left(\frac{b+\sum_i^n x_i^{2\tau}}{2\nu^2}\right)\, d\tau d\nu}. \qquad (25)
$$

If $u(\tau,\nu)=\tau$, then $u_{\tau}=1$, $u_{\nu\nu}=u_{\nu}=u_{\tau\tau}=u_{\nu\tau}=u_{\tau\nu}=0$. Then the BE of τ is written as follows:

$$
\begin{aligned}
\hat{\tau}_{BS} =& \hat{\tau}+(\hat{\eta}_{\tau}\hat{\sigma}_{\tau\tau}+\hat{\eta}_{\nu}\hat{\sigma}_{\nu\tau})+0.5\left[\hat{L}_{\tau\tau\tau}\hat{\sigma}^{2}_{\tau\tau}+3\hat{L}_{\tau\tau\nu}\hat{\sigma}_{\tau\nu}\hat{\sigma}_{\tau\tau}+\hat{L}_{\tau\nu\nu}(\hat{\sigma}_{\nu\nu}\hat{\sigma}_{\tau\tau}+2\hat{\sigma}^{2}_{\tau\nu})\right.\\
& \left.+\hat{L}_{\nu\nu\nu}\hat{\sigma}_{\tau\nu}\hat{\sigma}_{\nu\nu}\right]. \qquad (26)
\end{aligned}
$$

The Bayes estimate (BE) of parameter ν under SELF is found as follows:

$$
\hat{\nu}_{BS}=E(\nu|x)=\frac{\int_0^\infty\int_0^\infty \tau^{n-1} \frac{1}{\nu^{2n+a}}\prod_i^n x_i^{2\tau-1} exp\left(\frac{b+\sum_i^n x_i^{2\tau}}{2\nu^2}\right)\, d\tau d\nu}{\int_0^\infty\int_0^\infty \tau^{n-1} \frac{1}{\nu^{2n+a+1}}\prod_i^n x_i^{2\tau-1} exp\left(\frac{b+\sum_i^n x_i^{2\tau}}{2\nu^2}\right)\, d\tau d\nu}.
$$

If $u(\tau,\nu)=\nu$ then $u_{\nu}=1$, $u_{\nu\nu}=u_{\tau}=u_{\tau\tau}=u_{\nu\tau}=u_{\tau\nu}=0$.

Thus, the BE of ν under SELF is obtained as follows:

$$\hat{\nu}_{BS} = \hat{\nu} + (\hat{\eta}_{\nu}\hat{\sigma}_{\nu\nu} + \hat{\eta}_{\tau}\hat{\sigma}_{\tau\nu}) + 0.5\left[\hat{L}_{\tau\tau\tau}\hat{\sigma}_{\tau\tau}\hat{\sigma}_{\tau\nu} + 3\hat{L}_{\tau\nu\nu}\hat{\sigma}_{\nu\nu}\hat{\sigma}_{\tau\nu} + \hat{L}_{\tau\tau\nu}(\hat{\sigma}_{\nu\nu}\hat{\sigma}_{\tau\tau} + 2\hat{\sigma}^2_{\tau\nu}) + \hat{L}_{\nu\nu\nu}\hat{\sigma}^2_{\nu\nu}\right]. \quad (27)$$

If $u(\tau,\nu) = exp\left(-\frac{t^{2\tau}}{2\nu^2}\right)$ then $u_{\nu} = \frac{t^{2\tau}}{\nu^3}exp\left(-\frac{t^{2\tau}}{2\nu^2}\right)$,
$u_{\nu\nu} = -\frac{3t^{2\tau}}{\nu^4}exp\left(-\frac{t^{2\tau}}{2\nu^2}\right) + \left(\frac{t^{2\tau}}{\nu^3}\right)^2 exp\left(-\frac{t^{2\tau}}{2\nu^2}\right)$,
$u_{\tau} = -\frac{t^{2\tau}\log(t)}{2\nu^2}exp\left(-\frac{t^{2\tau}}{2\nu^2}\right)$, $u_{\tau\tau} = -\frac{2t^{2\tau}(\log(t))^2}{\nu^2}exp\left(-\frac{t^{2\tau}}{2\nu^2}\right) + \left(\frac{2t^{2\tau}\log(t)}{\nu^2}\right)^2 exp\left(-\frac{t^{2\tau}}{2\nu^2}\right)$,
and $u_{\nu\tau} = u_{\tau\nu} = \frac{t^{2\tau}\log(t)}{\nu^3}exp\left(-\frac{t^{2\tau}}{2\nu^2}\right) - \frac{t^{4\tau}\log(t)}{\nu^5}exp\left(-\frac{t^{2\tau}}{2\nu^2}\right)$.

Then BE of the survival function under SELF is given by:

$$\begin{aligned}\hat{R}_{BS}(t) &= \hat{R} + 0.5(\hat{u}_{\tau\tau}\hat{\sigma}_{\tau\tau} + \hat{u}_{\nu\nu}\hat{\sigma}_{\nu\nu}) + \hat{u}_{\tau\nu}\hat{\sigma}_{\tau\nu} + \hat{u}_{\tau}(\hat{\sigma}_{\tau\tau}\hat{\eta}_{\tau} + \hat{\sigma}_{\nu\tau}\hat{\eta}_{\nu}) + \\ &\quad \hat{u}_{\nu}(\hat{\sigma}_{\tau\nu}\hat{\eta}_{\tau} + \hat{\sigma}_{\nu\nu}\hat{\eta}_{\nu}) + 0.5\hat{L}_{\tau\tau\tau}(\hat{u}_{\tau}\hat{\sigma}^2_{\tau\tau} + \hat{u}_{\nu}\hat{\sigma}_{\tau\tau}\hat{\sigma}_{\tau\nu}) + 0.5\hat{L}_{\tau\tau\nu} \\ &\quad \left(3\hat{u}_{\tau}\hat{\sigma}_{\tau\tau}\hat{\sigma}_{\tau\nu} + \hat{u}_{\nu}(\hat{\sigma}_{\tau\tau}\hat{\sigma}_{\nu\nu} + 2\hat{\sigma}^2_{\tau\nu})\right) + 0.5\hat{L}_{\tau\nu\nu}\left(\hat{u}_{\tau}(\hat{\sigma}_{\tau\tau}\hat{\sigma}_{\nu\nu} + 2\hat{\sigma}^2_{\tau\nu})\right. \\ &\quad \left. + 3\hat{u}_{\nu}\hat{\sigma}_{\tau\nu}\hat{\sigma}_{\nu\nu}\right) + 0.5\hat{L}_{\tau\tau\tau}(\hat{u}_{\tau}\hat{\sigma}_{\tau\nu}\hat{\sigma}_{\tau\tau} + \hat{u}_{\nu}\hat{\sigma}^2_{\nu\nu}). \end{aligned} \quad (28)$$

4.4.2 Bayesian estimation under GELF

GELF is an asymmetric loss function and it was proposed by Calabria and Pulcini (1994). BE of τ under this loss function is derived as follows:

$$\hat{\tau}_{BG} = [E(\tau^{-q}|x)]^{-\frac{1}{q}};\ q \neq 0. \quad (29)$$

If $u(\tau,\nu) = \tau^{-q}$ then $u_{\tau} = -q\tau^{-(q+1)}$, $u_{\tau\tau} = q(q+1)\tau^{-(q+2)}$, $u_{\tau} = u_{\nu\nu} = u_{\tau\nu} = u_{\nu\tau} = 0$.

Then using equation (24) we have

$$\begin{aligned}E(\tau^{-q}|x) &= \hat{\tau}^{-q} + 0.5\hat{u}_{\tau\tau}\hat{\sigma}_{\tau\tau} + \hat{u}_{\tau}(\hat{\sigma}_{\tau\tau}\hat{\eta}_{\tau} + \hat{\sigma}_{\nu\tau}\hat{\eta}_{\nu}) + 0.5\hat{L}_{\tau\tau\tau}\hat{u}_{\tau}\hat{\sigma}^2_{\tau\tau} \\ &\quad + 1.5\hat{L}_{\tau\tau\nu}\hat{u}_{\tau}\hat{\sigma}_{\tau\tau}\hat{\sigma}_{\tau\nu} + 0.5\hat{L}_{\tau\nu\nu}\left(\hat{u}_{\tau}(\hat{\sigma}_{\tau\tau}\hat{\sigma}_{\nu\nu} + 2\hat{\sigma}^2_{\tau\nu})\right) + 0.5\hat{L}_{\nu\nu\nu}\hat{u}_{\tau}\hat{\sigma}_{\tau\nu}\hat{\sigma}_{\nu\nu}.\end{aligned}$$

If $u(\tau,\nu) = \nu^{-q}$, $u_{\nu} = -q\nu^{-(q+1)}$, $u_{\nu\nu} = q(q+1)\nu^{-(q+2)}$, $u_{\tau} = u_{\tau\tau} = u_{\tau\nu} = u_{\nu\tau} = 0$. Then the BE of ν under GELF is obtained as:

$$\hat{\nu}_{BG} = [E(\nu^{-q}|x)]^{-\frac{1}{q}};\ q \neq 0, \quad (30)$$

where

$$\begin{aligned}E(\nu^{-q}|x) &= \hat{\nu}^{-q} + 0.5\hat{u}_{\nu\nu}\hat{\sigma}_{\nu\nu} + \hat{u}_{\nu}(\hat{\sigma}_{\nu\nu}\hat{\eta}_{\nu} + \hat{\sigma}_{\tau\nu}\hat{\eta}_{\tau}) + 0.5\hat{L}_{\tau\tau\tau}\hat{u}_{\nu}\hat{\sigma}_{\tau\nu}\hat{\sigma}_{\tau\tau} \\ &\quad + 1.5\hat{L}_{\tau\nu\nu}\hat{u}_{\nu}\hat{\sigma}_{\nu\nu}\hat{\sigma}_{\tau\nu} + 0.5\hat{L}_{\tau\tau\nu}\left(\hat{u}_{\nu}(\hat{\sigma}_{\tau\tau}\hat{\sigma}_{\nu\nu} + 2\hat{\sigma}^2_{\tau\nu})\right) + 0.5\hat{L}_{\nu\nu\nu}\hat{u}_{\nu}\hat{\sigma}^2_{\nu\nu}.\end{aligned}$$

If $u(\tau,\nu)=exp\left(\frac{qt^{2\tau}}{2\nu^2}\right)$, $u_\nu=-\frac{qt^{2\tau}}{\nu^3}exp\left(\frac{qt^{2\tau}}{2\nu^2}\right)$,

$u_{\nu\nu}=\frac{3qt^{2\tau}}{\nu^4}exp\left(\frac{qt^{2\tau}}{2\nu^2}\right)+\left(\frac{qt^{2\tau}}{\nu^3}\right)^2exp\left(\frac{qt^{2\tau}}{2\nu^2}\right), u_\tau=\frac{qt^{2\tau}log(t)}{2\nu^2}exp\left(\frac{qt^{2\tau}}{\nu^2}\right)$,
$u_{\tau\tau}=\frac{2qt^{2\tau}(log(t))^2}{\nu^2}exp\left(\frac{qt^{2\tau}}{2\nu^2}\right)+\left(\frac{qt^{2\tau}log(t)}{\nu^2}\right)^2exp\left(\frac{qt^{2\tau}}{2\nu^2}\right)$,
$u_{\nu\tau}=u_{\tau\nu}=-\frac{2qt^{2\tau}log(t)}{\nu^3}exp\left(\frac{qt^{2\tau}}{2\nu^2}\right)-\frac{q^2t^{4\tau}log(t)}{\nu^5}exp\left(\frac{qt^{2\tau}}{2\nu^2}\right)$.

Then, BE of $R(t)$ under GELF is obtained as

$$\hat{R}(t)_{BG}=[E(R(t)^{-q}|x)]^{-\frac{1}{q}};\ q\neq 0, \tag{31}$$

where

$$\begin{aligned}E(R(t)^{-q}|x) &= \hat{R}(t)^{-q}+0.5\hat{u}_{\tau\tau}\hat{\sigma}_{\tau\tau}+\hat{u}_\tau(\hat{\sigma}_{\tau\tau}\hat{\eta}_\tau+\hat{\sigma}_{\nu\tau}\hat{\eta}_\nu)+\\ &\quad 0.5\hat{L}_{\tau\tau\tau}\hat{u}_\tau\hat{\sigma}_{\tau\tau}^2+1.5\hat{L}_{\tau\tau\nu}\hat{u}_\tau\hat{\sigma}_{\tau\tau}\hat{\sigma}_{\tau\nu}+0.5\hat{L}_{\tau\nu\nu}\left(\hat{u}_\tau(\hat{\sigma}_{\tau\tau}\hat{\sigma}_{\nu\nu}+2\hat{\sigma}_{\tau\nu}^2)\right)\\ &\quad +0.5\hat{L}_{\nu\nu\nu}\hat{u}_\tau\hat{\sigma}_{\tau\nu}\hat{\sigma}_{\nu\nu}.\end{aligned}$$

4.4.3 Bayesian estimation under LLF

LLF is an asymmetric loss function that proposed by Klebanov (1972) and later used by Varian (1975).

If $u(\tau,\nu)=exp(-q\tau),\ u_\tau=-q\ exp(-q\tau),\ u_{\tau\tau}=q^2\ exp(-q\tau),\ u_\nu=u_{\nu\nu}=u_{\tau\nu}=u_{\nu\tau}=0$. Then, BE of τ under LLF is obtained as:

$$\hat{\tau}_{BL}=-\frac{1}{q}\ ln\ E(exp(-q\tau)|x));\ q\neq 0, \tag{32}$$

where

$$\begin{aligned}E(exp(-q\tau)|x) &= exp(-q\hat{\tau})+0.5\hat{u}_{\tau\tau}\hat{\sigma}_{\tau\tau}+\hat{u}_\tau(\hat{\sigma}_{\tau\tau}\hat{\eta}_\tau+\hat{\sigma}_{\nu\tau}\hat{\eta}_\nu)+0.5\hat{L}_{\tau\tau\tau}\hat{u}_\tau\hat{\sigma}_{\tau\tau}^2\\ &\quad +1.5\hat{L}_{\tau\tau\nu}\hat{u}_\tau\hat{\sigma}_{\tau\tau}\hat{\sigma}_{\tau\nu}+0.5\hat{L}_{\tau\nu\nu}\left(\hat{u}_\tau(\hat{\sigma}_{\tau\tau}\hat{\sigma}_{\nu\nu}+2\hat{\sigma}_{\tau\nu}^2)\right)\\ &\quad +0.5\hat{L}_{\nu\nu\nu}\hat{u}_\tau\hat{\sigma}_{\tau\nu}\hat{\sigma}_{\tau\tau}.\end{aligned}$$

If $u(\tau,\nu)=exp(-q\nu),\ u_\nu=-q\ exp(-q\nu),\ u_{\nu\nu}=q^2\ exp(-q\nu),\ u_\tau=u_{\tau\tau}=u_{\tau\nu}=u_{\nu\tau}=0$. Then

$$\hat{\nu}_{BL}=-\frac{1}{q}\ ln\ E(exp(-q\nu)|x);\ q\neq 0, \tag{33}$$

where

$$\begin{aligned}E(exp(-q\nu)|x) &= exp(-q\hat{\nu})+0.5\hat{u}_{\nu\nu}\hat{\sigma}_{\nu\nu}+\hat{u}_\nu(\hat{\sigma}_{\nu\nu}\hat{\eta}_\nu+\hat{\sigma}_{\tau\nu}\hat{\eta}_\tau)+0.5\hat{L}_{\tau\tau\tau}\hat{u}_\nu\hat{\sigma}_{\tau\nu}\hat{\sigma}_{\tau\tau}\\ &\quad +1.5\hat{L}_{\tau\nu\nu}\hat{u}_\nu\hat{\sigma}_{\tau\tau}\hat{\sigma}_{\tau\nu}+0.5\hat{L}_{\tau\tau\nu}\left(\hat{u}_\nu(\hat{\sigma}_{\tau\tau}\hat{\sigma}_{\nu\nu}+2\hat{\sigma}_{\tau\nu}^2)\right)+0.5\hat{L}_{\nu\nu\nu}\hat{u}_\nu\hat{\sigma}_{\nu\nu}^2.\end{aligned}$$

$$\text{If } u(\tau,\nu)=exp\left(-q\,exp\left(-\tfrac{t^{2\tau}}{2\nu^2}\right)\right),\ u_\nu=-\tfrac{qt^{2\tau}}{\nu^3}exp\left[-\left\{\left(q\,exp\left(-\tfrac{t^{2\tau}}{2\nu^2}\right)\right)+\tfrac{t^{2\tau}}{2\nu^2}\right\}\right],$$

$$u_{\nu\nu}=\frac{qt^{2\tau}}{\nu^3}exp\left[-\left\{\left(q\,exp\left(-\frac{t^{2\tau}}{2\nu^2}\right)\right)+\frac{t^{2\tau}}{2\nu^2}\right\}\right]$$
$$\left\{\frac{3}{\nu}+q\left(exp\left(-\frac{t^{2\tau}}{2\nu^2}\right)+\frac{t^{2\tau}}{2\nu^2}\right)\left(qexp(-\frac{t^{2\tau}}{2\nu^2})\frac{t^{2\tau}}{\nu^3}-\frac{t^{2\tau}}{\nu^3}\right)\right\}, \tag{34}$$

$$u_\tau=\tfrac{qt^{2\tau}log(t)}{2\nu^2}exp\left[-\left\{\left(q\,exp\left(-\tfrac{t^{2\tau}}{2\nu^2}\right)\right)+\tfrac{t^{2\tau}}{2\nu^2}\right\}\right],$$

$$u_{\tau\tau}=\tfrac{qt^{2\tau}(log(t))^2}{\nu^2}exp\left[-\left\{\left(q\,exp\left(-\tfrac{t^{2\tau}}{2\nu^2}\right)\right)+\tfrac{t^{2\tau}}{2\nu^2}\right\}\right]\left\{2+\tfrac{qt^{2\tau}}{\nu^2}\left(exp(\left(-\tfrac{t^{2\tau}}{\nu^2}\right)-1\right)\right\},$$

and

$$u_{\nu\tau}=u_{\tau\nu}=-exp\left[-\left\{\left(q\,exp\left(-\frac{t^{2\tau}}{2\nu^2}\right)\right)+\frac{t^{2\tau}}{2\nu^2}\right\}\right]$$
$$\left\{\frac{2qt^{2\tau}log(t)}{\nu^3}-\frac{qt^{2\tau}log(t)}{\nu^2}\left(\frac{qt^{2\tau}}{\nu^3}exp\left(-\frac{t^{2\tau}}{2\nu^2}\right)-\frac{t^{2\tau}}{\nu^3}\right)\right\}. \tag{35}$$

BE of $R(t)$ under LLF is given by

$$\hat{R}(t)_{BL}=-\frac{1}{q}\,ln\,\left(E(exp(-q\,R(t))|x)\right);\ \ q\neq 0, \tag{36}$$

where

$$\begin{aligned}E(exp(-qR(t))|x) &= \hat{R}(t)^{-q}+0.5\hat{u}_{\tau\tau}\hat{\sigma}_{\tau\tau}+\hat{u}_\tau(\hat{\sigma}_{\tau\tau}\hat{\eta}_\tau+\hat{\sigma}_{\nu\tau}\hat{\eta}_\nu)+0.5\hat{L}_{\tau\tau\tau}\hat{u}_\tau\hat{\sigma}^2_{\tau\tau}\\ &\quad +1.5\hat{L}_{\tau\tau\nu}\hat{u}_\tau\hat{\sigma}_{\tau\tau}\hat{\sigma}_{\tau\nu}+0.5\hat{L}_{\tau\nu\nu}\left(\hat{u}_\tau(\hat{\sigma}_{\tau\tau}\hat{\sigma}_{\nu\nu}+2\hat{\sigma}^2_{\tau\nu})\right)\\ &\quad +0.5\hat{L}_{\nu\nu\nu}\hat{u}_\tau\hat{\sigma}_{\tau\nu}\hat{\sigma}_{\nu\nu}.\end{aligned}$$

Using Lindley's approximation, the Bayes estimates of the parameters τ, ν, and $R(t)$ are derived, but the highest posterior density (HPD) credible interval is not possible to be constructed. Therefore, the Metropolis-Hasting (M-H) algorithm is introduced to compute the Bayes estimates as well as HPD credible intervals.

4.4.4 Metropolis-Hastings (M-H) Algorithm

The Metropolis-Hastings algorithm is a Markov Chain Monte Carlo (MCMC) technique used for sampling from complex probability distributions. It operates by iteratively proposing candidate states, evaluating their acceptance probability based on a defined proposal distribution and target distribution, and accepting or rejecting them accordingly. The algorithm addresses challenges in sampling from distributions that are difficult to directly

sample from by introducing a transition probability function. For more about the M-H algorithm one may refers to Metropolis *et al.* (1953) and Hastings (1970). Using the M-H algorithm, the conditional posterior density function of parameters τ and ν are given by

$$\pi_1(\tau|\nu,x) \propto \tau^{n-1}\prod_{i=1}^{n} x_i^{2\tau-1}\ exp\left(-\frac{\sum_{i=1}^{n} x_i^{2\tau}}{2\nu^2}\right), \tag{37}$$

and

$$\pi_2(\nu|\tau,x) \propto \frac{1}{\nu^{2n+a+1}}\ exp\left(-\frac{b+\sum_{i=1}^{n} x_i^{2\tau}}{2\nu^2}\right). \tag{38}$$

respectively. Since the conditional posterior distributions of the parameters τ and ν in the equations (35) and (36) are unknown. Therefore, we can use M-H algorithm with normal proposal distribution to generate the posterior sample from (35) and (36) respectively. The M-H algorithm consists of the following steps:

Step 1: Set $j = 1$ and start with initial values $\tau^{(0)} = \hat{\tau}$ and $\nu^{(0)} = \hat{\nu}$.
Step 2: Use the M-H algorithm steps to generate posterior samples for $\tau^{(j)}$ and $\nu^{(j)}$ from the conditional distributions $\pi_1(\tau^{(j-1)}|\nu^{(j-1)},x)$ and $\pi_2(\nu^{(j-1)}|\tau^{(j-1)},x)$ using normal proposal distributions $N(\tau^{(j-1)}, var(\tau))$ and $N(\nu^{(j-1)}, var(\nu))$ respectively.
Step 3: Set j = j+1.
Step 4: Replicate steps 2-3 N times to extract samples $\phi^{(j)} = (\tau^{(j)}, \nu^{(j)}, R^{(j)}(t))$ for $j = 1,2,...,N$.
Step 5: The Bayes estimates of the parameters τ, ν, and $R(t)$ under SELF, GELF, and LLF can be obtained from the following expressions:

$$\hat{\phi}_{BS} = \frac{1}{N-M}\sum_{j=M+1}^{N}\phi^{(j)}, \tag{39}$$

$$\hat{\phi}_{BG} = \left(\frac{1}{N-M}\sum_{j=M+1}^{N}(\phi^{(j)})^{-q}\right)^{-\frac{1}{q}}, \tag{40}$$

$$\hat{\phi}_{BL} = -\frac{1}{q}\ \ln\left(\frac{1}{N-M}\sum_{j=M+1}^{N} exp(-q\phi^{(j)})\right), \tag{41}$$

where M is the burn in period of the Markov chain.

Step 6: To construct the HPD credible interval of $\phi = (\tau, \nu, R(t))$ order the MCMC sample of ϕ, where $\phi^{(j)} = (\tau^{(j)}, \nu^{(j)}, R^{(j)}(t))$, $j = 1,2,...,N$, for sufficiently large N, then for arbitrary, $0 < \xi < 1$ the $100(1-\xi)\%$ credible interval ϕ can be obtained as $(\phi^{[k]}, \phi^{[k+N-(\xi N+1)]})$, where $k = 1,2,...,[N\xi]$. Therefore, the $100(1-\xi)\%$ credible interval can be constructed based on the condition given below:

$$(\phi^{[k^*+N-(\xi N+1)} - \hat{\tau}^{[k^*]}) = min_{k=1}^{N\xi}\ (\phi^{[k+N-(\xi N+1)} - \phi^{[k]}). \tag{42}$$

where [z] denotes the greatest integer less or equal to z.

5. Simulation Study

We conduct an extensive Monte Carlo simulation study to assess the relative precision of the proposed estimates. We generate 10^4 samples from PRD with parameters τ and ν for each combination of different parameter values and sample size. Here, we choose the parameters as $\tau = 0.5, 1, 2$, $\beta = 1, 1.5, 1.5$ and the corresponding values of $R(t)$ at $t = 0.5, 1.50, 1.25$ are 0.7788, 0.6065, and 0.5813, and sample sizes as $n = 20, 40, 60, 80, 100$. The MLEs and Bayes estimates using Lindley and MCMC techniques are applied to simulate the data. In the Bayes estimates, we consider the non-informative prior for scale parameter τ and informative gamma prior for shape parameter τ. In addition, the values of hyper-parameters $a = 2, 3$ and $b = 2, 2$ are chosen such that the prior mean is equal to the true value of the parameter. Moreover, the Bayes estimates are obtained using SELF, GELF, and LINEX loss functions. For GELF and LLF, the constant q is taken to be -0.5 and 0.5, respectively. The evaluation of the estimates has been conducted with consideration given to the following standpoint:

- **Average absolute bias (AAB):** Let ψ and $\hat{\psi}$ denote the actual and predicted value of the parameters, and N represent the total number of replications. Then the average absolute value is defined as follows:

$$AAB = \frac{1}{N}\sum_{i=1}^{N} |\psi_i - \hat{\psi}_i|.$$

  A smaller AAB value suggests the experimental data exhibits higher accuracy with the predictive model.

- **Mean squared error (MSE):** The MSE is defined as follows:

$$MSE = \frac{1}{N}\sum_{i=1}^{N} (\psi_i - \hat{\psi})^2.$$

  The smaller value signifies superior performance of the estimates.

- **Average length (AL):** AL of the interval estimates at a significance level of ξ has been assessed. A shorter length indicates superior performance in the estimation of intervals.

- **Coverage probability (CP):** The probability of containing the actual parameter values within the estimated interval ranges.

The average absolute bias (AAB) and mean square error (MSE) of τ, ν and $R(t)$ are presented in Tables 1, 3, and 5, respectively. Moreover, the associated 95% approximate confidence and HPD credible intervals are also obtained and listed in Tables 2, 4, and 6, respectively. The following interpretations can be obtained from these Tables:
(i) Tables 1, 3, and 5 show that as the sample size n increases, the average absolute bias (AAB) and mean square error (MSE) of all estimates decrease as expected. This suggests

that all the estimators are asymptotically efficient.
(ii) The performance of the Bayes estimates is better than the classical approach in terms of AAB and MSE.
(iii) In Bayesian estimation, the MCMC approach provides superior outcomes to the Lindley approximation techniques regarding AAB and MSE. Comparing all the estimators, the following mathematical relation can be obtained:

$$AAB_{MLE} > AAB_{Lindley} > AAB_{MCMC},$$

and

$$MSE_{MLE} > MSE_{Lindley} > MSE_{MCMC}.$$

(iv) Concerning the loss functions, GELF ($q = 0.5$) under the MCMC approach is a better choice for the parameters τ and ν and LLF ($q = -0.5$) is better for the reliability function in terms of AAB and MSE.
(v) In Tables 2, 4, and 6, the highest posterior density (HPD) credible interval is shorter than the approximate confidence intervals (ACIs) in terms of the average length of confidence/credible intervals.
(vi) The coverage probability (CP) of HPD credible interval is higher than the approximate confidence intervals.

Table 1: Simulation results for classical and Bayes estimator of parameters $\tau = 0.5$, $\nu = 1$ and $R(0.5) = 0.7788$ with biases (first row) and MSEs (second row)

			Lindley					MCMC				
	n	MLE	SELF	GELF		LLF		SELF	GELF		LLF	
				q = -0.5	q = 0.5	q =-0.5	q = 0.5		q = -0.5	q =0.5	q = -0.5	q = 0.5
τ	20	0.0769	0.0740	0.0730	0.0722	0.0750	0.0792	0.0737	0.0728	0.0713	0.0745	0.0729
		0.0114	0.0110	0.0102	0.0106	0.0108	0.0109	0.0103	0.0100	0.0095	0.0105	0.0100
ν		0.1371	0.1281	0.1288	0.1258	0.1312	0.1295	0.1266	0.1244	0.1207	0.1302	0.1234
		0.0355	0.0326	0.0312	0.0288	0.0330	0.0300	0.0317	0.0303	0.0277	0.0344	0.0295
$R(t)$		0.0610	0.0542	0.0550	0.0560	0.0541	0.0552	0.0538	0.0542	0.0553	0.0536	0.0541
		0.0058	0.0046	0.0048	0.0049	0.0046	0.0048	0.0046	0.0046	0.0048	0.0045	0.0046
τ	40	0.0551	0.0544	0.0538	0.0534	0.0540	0.0539	0.0534	0.0529	0.0522	0.0537	0.0530
		0.0052	0.0047	0.0048	0.0047	0.0050	0.0049	0.0049	0.0048	0.0046	0.0049	0.0048
ν		0.0870	0.0851	0.0846	0.0822	0.0859	0.0849	0.0844	0.0835	0.0819	0.0856	0.0832
		0.0127	0.0120	0.0118	0.0118	0.0127	0.0120	0.0120	0.0117	0.0112	0.0123	0.0116
$R(t)$		0.0434	0.0417	0.0416	0.0410	0.0414	0.0412	0.0403	0.0404	0.0407	0.0402	0.0403
		0.0029	0.0027	0.0027	0.0028	0.0027	0.0027	0.0026	0.0026	0.0026	0.0025	0.0026
τ	60	0.0422	0.0419	0.0418	0.0410	0.0417	0.0416	0.0413	0.0411	0.0408	0.0415	0.0411
		0.0030	0.0029	0.0029	0.0028	0.0029	0.0029	0.0029	0.0028	0.0028	0.0029	0.0029
ν		0.0721	0.0711	0.0710	0.0712	0.0718	0.0704	0.0710	0.0705	0.0696	0.0717	0.0704
		0.0088	0.0086	0.0085	0.0084	0.0088	0.0085	0.0086	0.0084	0.0082	0.0088	0.0084
$R(t)$		0.0350	0.0339	0.0338	0.0336	0.0339	0.0337	0.0335	0.0335	0.0338	0.0334	0.0335
		0.0019	0.0019	0.0019	0.0019	0.0018	0.0018	0.0018	0.0018	0.0018	0.0017	0.0018
τ	80	0.0360	0.0358	0.0356	0.0359	0.0360	0.0357	0.0355	0.0354	0.0352	0.0356	0.0354
		0.0021	0.0021	0.0021	0.0021	0.0021	0.0021	0.0021	0.0020	0.0020	0.0021	0.0021
ν		0.0057	0.0057	0.0055	0.0055	0.0057	0.0055	0.0056	0.0055	0.0054	0.0056	0.0055
		0.0298	0.0288	0.0285	0.0290	0.0289	0.0290	0.0287	0.0287	0.0289	0.0286	0.0287
$R(t)$		0.0298	0.0290	0.0290	0.0289	0.0289	0.0288	0.0287	0.0287	0.0289	0.0286	0.0287
		0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
τ	100	0.0331	0.0329	0.0326	0.0326	0.0329	0.0328	0.0326	0.0326	0.0324	0.0327	0.0326
		0.0018	0.0018	0.0018	0.0018	0.0018	0.0018	0.0018	0.0017	0.0017	0.0018	0.0018
ν		0.0554	0.0551	0.0550	0.0544	0.0549	0.0549	0.0543	0.0541	0.0536	0.0546	0.0540
		0.0050	0.0050	0.0048	0.0048	0.0049	0.0048	0.0048	0.0048	0.0047	0.0049	0.0048
$R(t)$		0.0277	0.0271	0.0266	0.0269	0.0268	0.0268	0.0266	0.0266	0.0267	0.0266	0.0266
		0.0012	0.0012	0.0012	0.0011	0.0012	0.0011	0.0011	0.0011	0.0011	0.0011	0.0011

Table 2: The 95% confidence interval and HPD credible interval for parameters $\tau = 0.5$, $\nu = 1$ and $R(0.5) = 0.7788$

Parameters	n	ACI	CPs	HPD	CPs	
τ	20	0.3679	0.9410	0.3350	0.9450	
ν		0.6281	0.9390	0.5652	0.9460	
$R(0.5)$		0.2849	0.9010	0.2460	0.9220	
τ	40	0.2527	0.9430	0.2368	0.9550	
ν		0.4243	0.9370	0.3957	0.9450	
$R(0.5)$		0.2051	0.9280	0.1864	0.9340	
τ	60	0.2030	0.9510	0.1916	0.9410	
ν		0.3466	0.9510	0.3277	0.9350	
$R(0.5)$		0.1677	0.9220	0.1557	0.9250	
τ	60	0.1742	0.9470	0.1652	0.9560	
ν		0.2950	0.9540	0.2790	0.9670	
$R(0.5)$		0.1464	0.9500	0.1365	0.9520	
τ	100	0.1560	0.9370	0.1479	0.9380	
ν		0.2638	0.9440	0.2497	0.9580	
$R(0.5)$		0.1309	0.9430	0.1226	0.9530	

Table 3: Simulation results for classical and Bayes estimator of parameters $\tau = 1$, $\nu = 1.5$ and $R(1.5) = 0.6065$ with biases (first row) and MSEs (second row)

			Lindley					MCMC				
	n	MLE	SELF	GELF		LLF		SELF	GELF		LLF	
				q = -0.5	q = 0.5	q = -0.5	q = 0.5		q = -0.5	q = 0.5	q = -0.5	q = 0.5
τ	20	0.1616	0.1430	0.1428	0.1442	0.1456	0.1425	0.1362	0.1362	0.1370	0.1376	0.1352
		0.0477	0.0322	0.0312	0.0311	0.0319	0.0305	0.0303	0.0300	0.0297	0.0312	0.0295
ν		0.2942	0.2433	0.2372	0.2352	0.2525	0.2388	0.2329	0.2291	0.2240	0.2462	0.2232
		0.1850	0.1083	0.1032	0.0874	0.1163	0.0879	0.0989	0.0936	0.0852	0.1177	0.0864
$R(1.5)$		0.0748	0.0686	0.0690	0.0722	0.0674	0.0681	0.0671	0.0684	0.0713	0.0666	0.0676
		0.0087	0.0073	0.0074	0.0082	0.0073	0.0074	0.0073	0.0075	0.0081	0.0072	0.0074
τ	40	0.1054	0.0982	0.0982	0.0986	0.0990	0.0971	0.0968	0.0968	0.0970	0.0973	0.0964
		0.0185	0.0156	0.0154	0.0157	0.0152	0.0151	0.0145	0.0145	0.0144	0.0148	0.0144
ν		0.1879	0.1754	0.1747	0.1745	0.1785	0.1737	0.1673	0.1659	0.1638	0.1715	0.1639
		0.0618	0.0485	0.0472	0.0451	0.0502	0.0448	0.0463	0.0452	0.0434	0.0494	0.0439
$R(1.5)$		0.0528	0.0504	0.0512	0.0519	0.0507	0.0506	0.0498	0.0503	0.0513	0.0496	0.0500
		0.0044	0.0041	0.0041	0.0043	0.0040	0.0042	0.0040	0.0041	0.0043	0.0040	0.0040
τ	60	0.0850	0.0811	0.0811	0.0813	0.0818	0.0812	0.0808	0.0808	0.0808	0.0811	0.0805
		0.0123	0.0111	0.0114	0.0116	0.0112	0.0113	0.0106	0.0105	0.0104	0.0107	0.0104
ν		0.1478	0.1409	0.1407	0.1405	0.1484	0.1481	0.1388	0.1380	0.1366	0.1413	0.1368
		0.0384	0.0348	0.0332	0.0322	0.0350	0.0332	0.0326	0.0319	0.0308	0.0342	0.0312
$R(1.5)$		0.0416	0.0412	0.0414	0.0420	0.0419	0.0420	0.0407	0.0410	0.0416	0.0406	0.0408
		0.0027	0.0026	0.0027	0.0027	0.0026	0.0026	0.0026	0.0026	0.0027	0.0026	0.0026
τ	80	0.0722	0.0698	0.0694	0.0694	0.0698	0.0698	0.0688	0.0687	0.0688	0.0690	0.0686
		0.0083	0.0076	0.0076	0.0076	0.0076	0.0076	0.0075	0.0074	0.0074	0.0075	0.0074
ν		0.1244	0.1252	0.1236	0.1197	0.1188	0.1199	0.1187	0.1182	0.1174	0.1202	0.1175
		0.0259	0.0238	0.0235	0.0229	0.0240	0.0232	0.0228	0.0225	0.0220	0.0236	0.0222
$R(1.5)$		0.0357	0.0354	0.0354	0.0356	0.0351	0.0352	0.0350	0.0352	0.0356	0.0349	0.0351
		0.0020	0.0020	0.0020	0.0020	0.0020	0.0020	0.0019	0.0020	0.0020	0.0019	0.0019
τ	100	0.0622	0.0615	0.0614	0.0616	0.0619	0.0620	0.0608	0.0608	0.0609	0.0609	0.0607
		0.0063	0.0059	0.0059	0.0059	0.0060	0.0059	0.0059	0.0059	0.0059	0.0059	0.0058
ν		0.1111	0.1096	0.1089	0.1090	0.1098	0.1097	0.1081	0.1079	0.1077	0.1089	0.1075
		0.0208	0.0191	0.0194	0.0193	0.0194	0.0193	0.0190	0.0188	0.0185	0.0195	0.0186
$R(1.5)$		0.0325	0.0325	0.0324	0.0325	0.0323	0.0324	0.0324	0.0325	0.0329	0.0323	0.0325
		0.0017	0.0016	0.0016	0.0017	0.0017	0.0016	0.0016	0.0016	0.0017	0.0016	0.0016

Table 4: The average lengths of 95% confidence intervals and coverage probability for parameters $\tau = 1$, $\nu = 1.5$ and $R(1.5) = 0.6065$

Parameters	n	ACI	CPs	HPD	CPs	
τ	20	0.7401	0.9400	0.6385	0.9460	
ν		1.403	0.949	1.109	0.9150	
$R(0.5)$		0.3441	0.9150	0.3094	0.9270	
τ	40	0.5027	0.9560	0.4550	0.9580	
ν		0.9016	0.9390	0.7859	0.9470	
$R(0.5)$		0.2464	0.9260	0.2277	0.9330	
τ	60	0.4067	0.9380	0.3759	0.9130	
ν		0.7220	0.9560	0.6534	0.9220	
$R(0.5)$		0.2020	0.9480	0.1897	0.9420	
τ	60	0.3483	0.9590	0.3238	0.9460	
ν		0.6164	0.9530	0.5636	0.9540	
$R(0.5)$		0.1752	0.9420	0.1647	0.9310	
τ	100	0.3107	0.9520	0.2909	0.9570	
ν		0.5452	0.9560	0.5026	0.9580	
$R(0.5)$		0.1570	0.9460	0.1482	0.9530	

Table 5: Simulation results for classical and Bayes estimator of parameters $\tau = 2$, $\nu = 1.5$ and $R(1.25) = 0.5813$ with biases (first row) and MSEs (second row)

			Lindley					MCMC				
	n	MLE	SELF	GELF		LLF		SELF	GELF		LLF	
				q = -0.5	q = 0.5	q = -0.5	q = 0.5		q = -0.5	q = 0.5	q = -0.5	q = 0.5
τ	20	0.3327	0.2815	0.2822	0.2826	0.2889	0.2807	0.2679	0.2672	0.2674	0.2756	0.2628
		0.2065	0.1430	0.1419	0.1329	0.1494	0.1436	0.1239	0.1217	0.1187	0.1348	0.1158
ν		0.2989	0.2417	0.2403	0.2377	0.2499	0.2381	0.2275	0.2224	0.2143	0.2442	0.2153
		0.1896	0.1093	0.1023	0.0911	0.1384	0.0923	0.0992	0.0922	0.0813	0.1288	0.0824
$R(1.25)$		0.0748	0.0667	0.0671	0.0670	0.0665	0.0672	0.0648	0.0658	0.0688	0.0645	0.0651
		0.0087	0.0070	0.0073	0.0080	0.0070	0.0071	0.0068	0.0071	0.0080	0.0068	0.0069
τ	40	0.2126	0.2078	0.2060	0.2066	0.2080	0.2023	0.1935	0.1933	0.1934	0.1959	0.1918
		0.0752	0.0611	0.0606	0.0606	0.0622	0.0599	0.0590	0.0586	0.0581	0.0612	0.0574
ν		0.1917	0.1888	0.1756	0.1758	0.1871	0.1781	0.1709	0.1695	0.1674	0.1754	0.1674
		0.0665	0.0514	0.0495	0.0480	0.0542	0.0480	0.0494	0.0479	0.0456	0.0533	0.0462
$R(1.25)$		0.0537	0.0521	0.0525	0.0536	0.0517	0.0521	0.0510	0.0515	0.0526	0.0509	0.0512
		0.0045	0.0041	0.0043	0.0043	0.0042	0.0042	0.0040	0.0041	0.0043	0.0040	0.0041
τ	60	0.1648	0.1671	0.1671	0.1689	0.1683	0.1695	0.1690	0.1575	0.1581	0.1582	0.1570
		0.0443	0.0411	0.0409	0.0411	0.0406	0.0396	0.0390	0.0390	0.0390	0.0396	0.0386
ν		0.1444	0.1390	0.1389	0.1393	0.1399	0.1381	0.1366	0.1361	0.1358	0.1382	0.1353
		0.0358	0.0324	0.0321	0.0302	0.0324	0.0304	0.0314	0.0301	0.0294	0.0316	0.0296
$R(1.25)$		0.0427	0.0424	0.0426	0.0422	0.0420	0.0423	0.0421	0.0405	0.0412	0.0409	0.0412
		0.0029	0.0028	0.0028	0.0029	0.0028	0.0028	0.0027	0.0028	0.0028	0.0027	0.0027
τ	80	0.1509	0.1430	0.1427	0.1425	0.1442	0.1420	0.1334	0.1326	0.1324	0.1348	0.1321
		0.0366	0.0330	0.0328	0.0326	0.0336	0.0325	0.0319	0.0317	0.0316	0.0316	0.0314
ν		0.1309	0.1236	0.1228	0.1213	0.1257	0.1219	0.1040	0.1130	0.1152	0.1159	0.1128
		0.0281	0.0249	0.0245	0.0238	0.0258	0.0241	0.0229	0.0238	0.0228	0.0238	0.0220
$R(1.25)$		0.0372	0.0360	0.0361	0.0363	0.0359	0.0360	0.0350	0.0350	0.0351	0.0340	0.0349
		0.0021	0.0020	0.0020	0.0021	0.0020	0.0020	0.0019	0.0019	0.0019	0.0019	0.0019
τ	100	0.1302	0.1259	0.1259	0.1260	0.1265	0.1255	0.1157	0.1144	0.1170	0.1163	0.1098
		0.0269	0.0248	0.0247	0.0247	0.0251	0.0245	0.0235	0.0237	0.0238	0.0243	0.0235
ν		0.1117	0.1074	0.1069	0.1060	0.1087	0.1063	0.0974	0.0968	0.0961	0.0984	0.0961
		0.0204	0.0185	0.0183	0.0179	0.0190	0.0180	0.0174	0.0173	0.0180	0.0179	0.0171
$R(1.25)$		0.0319	0.0312	0.0313	0.0315	0.0312	0.0312	0.0305	0.0304	0.0304	0.0299	0.0305
		0.0016	0.0016	0.0016	0.0016	0.0016	0.0016	0.0015	0.0015	0.0015	0.0015	0.0016

Table 6: The average lengths of 95% confidence intervals and coverage probability for the parameters $\tau = 2$, $\nu = 1.5$ and $R(1.25) = 0.5813$

Parameters	n	ACI	CPs	HPD	CPs	
τ	20	1.498	0.939	1.337	0.948	
ν		1.431	0.956	1.165	0.9560	
$R(1.25)$		0.3471	0.9210	0.3195	0.9410	
τ	40	1.0089	0.9540	0.9824	0.9590	
ν		0.9134	0.9470	0.8443	0.9450	
$R(1.25)$		0.2483	0.9290	0.2414	0.9440	
τ	60	0.8065	0.9540	0.7753	0.9580	
ν		0.7117	0.9460	0.6901	0.9400	
$R(1.25)$		0.2038	0.9430	0.2035	0.9450	
τ	80	0.6999	0.9480	0.6439	0.9480	
ν		0.6212	0.9510	0.5988	0.9550	
$R(1.25)$		0.1766	0.9420	0.1835	0.9460	
τ	100	0.6219	0.9470	0.5914	0.9540	
ν		0.5483	0.9540	0.5846	0.9530	
$R(1.25)$		0.1582	0.9490	0.1494	0.9550	

6. Application

In this section, we employ two real datasets to demonstrate the process of calculating estimators for the unknown model parameters.

Dataset I

The dataset considered by Bing Long (2023) represents the failure time of the mechanical components. To check whether dataset I better fits the considered model or not, we utilized different goodness of fit criteria, namely the Akaike information criterion (AIC), Bayesian information criterion (BIC), Hannan-Quinn information criterion (HQIC), corrected Akaike information criterion (AICC). In addition, K-S distance and p-value are obtained and listed in Table 7. Afterwards, MLEs of the parameters τ and ν are completed and presented in Table 7. For comparison purposes, we have considered different lifetime models such as the Rayleigh distribution (RD), Exponentiated Rayleigh distribution (ERD), Weibull Rayleigh distribution (WRD), and Transmuted Rayleigh distribution (TRD), and fitted them to the empirical dataset depicted in Figure 2. It has been observed that the PRD is a good fit for dataset I in comparison with other lifetime models. The estimated values of τ, ν and $R(t)$ under frequentist and Bayesian approaches for the real dataset I are listed in Table 8. The interval estimates are also derived and presented in Table 9.

Table 7: Goodness of fit measures and MLEs for the dataset I

Model	$\hat{\tau}$	$\hat{\nu}$	-2 log(l)	AIC	BIC	HQIC	AICC	K-S	p-value
PRD	1.2089	1.1055	48.55	52.55	55.35	53.45	52.99	0.07901	0.985
RD	-	0.9897	50.23	52.23	53.63	52.67	52.37	0.09540	0.924
ERD	1.4863	0.6511	48.02	52.02	54.82	52.91	52.46	0.07908	0.984
WRD	0.8641	0.6840	48.09	52.09	54.89	52.99	52.54	0.38940	0.00013
TRD	0.85215	-0.68747	48.85	52.85	55.65	53.74	53.29	0.08019	0.982

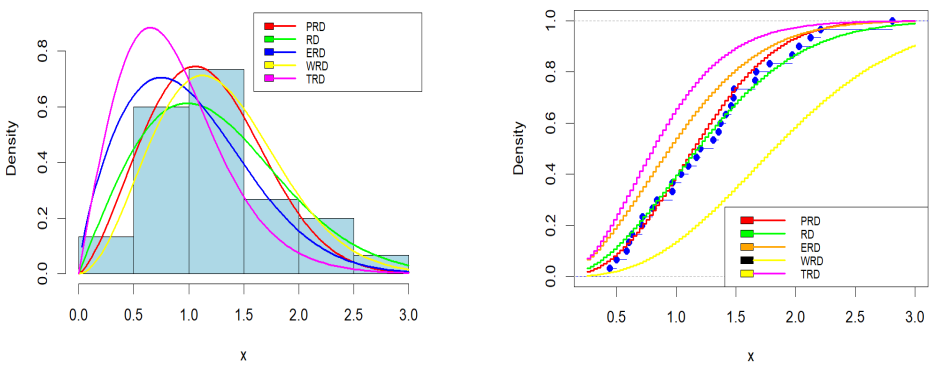


Figure 2: The estimated density and fitted plot of PRD and different lifetime models for dataset I.

Table 8: Estimated values of τ, ν and $R(0.5)$ based on the real dataset I

		Lindley					MCMC				
	MLE	SELF	GELF		LLF		SELF	GELF		LLF	
			q = -0.5	q = 0.5	q =-0.5	q = 0.5		q = -0.5	q =0.5	q = -0.5	q = 0.5
τ	1.2089	1.0817	1.0437	1.1832	1.0807	1.0709	1.2052	1.1988	1.1861	1.2131	1.1975
ν	1.1055	1.2123	1.2083	1.2176	1.1125	1.1074	1.1321	1.1269	1.1169	1.1384	1.12616
$R(0.5)$	0.9263	0.9110	0.9106	0.9102	0.9126	0.9106	0.9211	0.9207	0.9201	0.9214	0.9208

Table 9: The 95% ACI and HPD credible intervals of the parameters τ, ν and $R(t)$ based on the real dataset I

Parameters	ACI	HPD
τ	(0.8785,1.5394)	(0.8824,1.5346)
ν	(0.8220,1.3889)	(0.8718,1.3112)
$R(0.5)$	(0.8624,0.9901)	(0.8603,0.9787)

Dataset II

Dataset II represents the breaking stress of carbon fibers of 50 mm length (GPa) that has been analyzed by Al-Aqtash et al. (2014). Later on, Bhat and Ahmad (2020) considered the same dataset and demonstrated that PRD best fit to dataset II based on the different goodness of fit tests and various graphs. The MLEs and BEs of the parameters τ, ν, and $R(t)$ at $t = 2$ are presented in Table 10. For BEs, 5000 MCMC samples are generated, and the first 500 samples are discarded to avoid the initial guess. Note that non-informative prior information is considered because no prior information is available in this experiment. It has been observed that the MLEs and BEs estimates are pretty close to each other. The 95% ACI/HPD credible intervals are constructed and listed in Table 11. It can be seen that HPD credible intervals are more faithful than ACI.

Table 10: Estimated values of τ, ν and $R(2)$ based on the real dataset II

		Lindley					MCMC				
	MLE	**SELF**	**GELF**		**LLF**		**SELF**	**GELF**		**LLF**	
			q = -0.5	q = 0.5	q =-0.5	q = 0.5		q = -0.5	q =0.5	q = -0.5	q = 0.5
τ	1.721	1.694	1.684	1.751	1.726	1.725	1.749	1.743	1.733	1.757	1.740
ν	4.850	5.043	4.956	5.369	4.864	4.802	5.258	5.179	5.027	5.736	4.887
$R(2)$	0.7939	0.7976	0.7939	0.7967	7698	0.7667	0.8010	0.8004	0.7993	0.8015	0.8006

Table 11: The 95% ACI and HPD credible intervals for the parameters τ, ν and $R(2)$ based on the dataset II

Parameters	**ACI**	**HPD**
τ	(1.3965, 2.0445)	(1.5514, 1.8801)
ν	(2.8181, 6.8833)	(3.6821, 5.9394)
$R(2)$	(0.6813, 0.8872)	(0.7115, 0.8611)

7. Conclusions

This paper explores frequentist and Bayesian inference for the parameters and reliability estimation of the power Rayleigh distribution using a complete sample. The maximum likelihood estimates and approximate confidence intervals of the parameters have been computed using iterative procedures via the “nleqslv" package. Further, two approximation techniques have been considered: the Lindley and M-H algorithm for Bayesian computation under various loss functions. The performance of all estimators has been assessed through a Monte Carlo simulation study. The simulation experiment demonstrates that the Bayes approach dominates the frequentist approach based on absolute bias and mean square error. When comparing the Lindley and MCMC methods, the MCMC approach is more ef-

ficient than the Lindley technique. Moreover, the performance of the GELF and LINEX loss functions is better for the parameters and survival function in all cases, respectively. Finally, the Bayesian MCMC approach is recommended. In future studies, the scope of this study could be extended to hierarchical and empirical Bayesian estimation methods, especially for censored data and record values.

References

Ahamad, A., Ahmad, S. P. and Reshi, J. A., (2013). Bayesian analysis of Rayleigh distribution. *International Journal of Scientific and Research Publications*, 3(10), pp. 1–9.

Ahmed, A. A. J., Batah, F. S. M., (2023). The Estimation reliability for parallel system of stress–Strength model of power Rayleigh distribution. *AIP Conference Proceedings*, Vol. 2839(1). AIP Publishing.

Al-Aqtash, R., Lee, C. and Famoye, F., (2014). Gumbel-Weibull distribution: properties and applications. *Journal of Modern Applied Statistical Methods*, 13, pp. 201-225.

Soliman, A. A., Al-Aboud, F. M., (2008). Bayesian inference using record values from Rayleigh model with application. *European Journal of Operational Research*, 185(2), pp. 659–672.

Bhat, A. A., Ahmad, S. P., (2020). A New Generalization of Rayleigh Distribution: Properties and Applications. *Pakistan journal of statistics*, 36(3), pp. 225–250.

Calabria, R., Pulcini, G., (1994). An Engineering Approach to Bayes Estimation for the Weibull Distribution. *Micro-electron Reliability* 34(5), pp. 789–802.

Dey, S., (2009). Comparison of Bayes Estimators of the Parameter and Reliability Function for Rayleigh Distribution under Different Loss Functions. *Malaysian Journal of Mathematical Sciences*, 3(2), pp. 249–266.

Ghazal, M. G. M., Hasaballah, H. M., (2017). Exponentiated Rayleigh distribution: A Bayes study using MCMC approach based on unified hybrid censored data. *Journal of Advance in Mathematics*, 12(12), pp. 6863–6880.

Greene, W. H., (2003). Econometric analysis. 7th ed. Pearson Prentice-Hall, Upper Saddle River.

Hastings, W. K., (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57, pp. 97–109.

Irfan, M., Sharma, A. K., (2024). Bayesian Estimation and Prediction for Inverse Power Maxwell Distribution with Applications to Tax Revenue and Health Care Data. Journal of Modern Applied Statistical Methods, 23.

Irfan, M., Sharma, A. K., (2023, February). Bayesian Modeling for a Shape Parameter of Weibull-Lomax Distribution with an Application to Health Data. In International

Conference on Biomedical Engineering Science and Technology (pp. 284-299). Cham: Springer Nature Switzerland.

Kilany, N. M., Mahmoud, M. A. and El-Refai, L. H., (2023). Power Rayleigh Distribution for Fitting Total Deaths of Covid-19 in Egypt. *Journal of Statistics Applications & Probability*, Vol. 12(3), pp. 1073–1085.

Long, B., (2023). Estimation and prediction for the Rayleigh distribution based on double Type-I hybrid censored data. *Communications in Statistics-Simulation and Computation*, 52(8), pp. 3553–3567.

Lindley, D. V., (1980). Approximate Bayesian methods. *Trabajos de estadística y de investigaciónoperativa*, 31(1), pp. 223–245.

Mahanta, J., Talukdar, M. B. A., (2019). A Bayesian approach for estimating parameter of Rayleigh distribution. *Journal of Scientific Research*, 11(1), pp. 23–39.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6), pp. 1087–1092.

Migdadi, H. S., Al-Olaimat, N. M. and Meqdadi, O., (2023). Inference and optimal design for the k-level step-stress accelerated life test based on progressive Type-I interval censored power Rayleigh data. *Mathematical Biosciences and Engineering*, 20(12), pp. 21407–21431.

Migdadi, H. S., Al-Olaimat, N. M., Mohiuddin, M. and Meqdadi, O., (2023). Statistical inference for the Power Rayleigh distribution based on adaptive progressive Type-II censored data. *AIMS Mathematics*, 8(10), pp. 22553–22576.

Mahmoud, M. A. W., Kilany, N. M. and El-Refai, L. H., (2020). Inference of the lifetime performance index with power Rayleigh distribution based on progressive first-failure–censored data. *Quality and Reliability Engineering International*, 36(5), pp. 1528–1536.

Soliman, A. A., (2000). Comparison of linex and quadratic Bayes estimators for the Rayleigh distribution. *Communication in statistics - Theory and Method*, 291(1), pp. 95–107.

Varian, H. R., (1975). A Bayesian Approach to Real Estate Assessment. Econometrics and Statistics in Honor of Leonard J. Savage, Stephen, E. Fienberg; Arnold Zellner. *Amsterdam, North Holland*, pp. 195–208.

Yilmaz, A.; Kara, M., (2022). Reliability estimation and parameter estimation for inverse Weibull distribution under different loss functions. *Kuwait Journal of Science*, 49, pp. 1–24.

STATISTICS IN TRANSITION new series, September 2025
Vol. 26, No. 3, pp. 81–98, https://doi.org/10.59139/stattrans-2025-028
Received – 30.08.2024; accepted – 14.02.2025

The impact of the COVID-19 pandemic on forecast uncertainty of macroeconomic data releases

Krzysztof Brania[1], **Henryk Gurgul**[2]

Abstract

Macroeconomic data releases are very important benchmarks of the economy. Therefore, the vast majority of financial market analysts and traders closely monitor both the projected estimates and, the intuitively more impactful actual values. In this research, we focus on the uncertainty associated with macroeconomic data forecasts measured by the surprise indicator (SI). Moreover, we examine whether the distribution of SI depends on the economy, category of indicator or time, considering pre-pandemic, pandemic and post-pandemic periods in the context of the COVID-19 crisis. We also propose the construction of a sentiment indicator that is intended to aggregate all information that is jointly released through macroeconomic indicators.

Key words: macroeconomic, data, release, uncertainty, surprise, sentiment, indicator, forecast, actual

1. Introduction

In contemporary economics the performance of stock exchanges or exchange rates has an essential impact on the real economy, while data on the real economy impacts stock exchanges and exchange rates. The ultimate direction of interaction depends on the size of the economy that is the source of the data. Therefore, we can expect that the most influential macroeconomic data comes from the US. It is clear that due to globalization not only the US stock exchanges, but also other stock exchanges can be impacted by both domestic and the US macroeconomic data. Macroeconomic news announcements are the most important risk factors for financial markets. This is the case because the state of the economy reflected in these announcements is one of the main sources of risk. Moreover, this source of risk cannot be accounted for as a diversifiable risk.

Our extended study comprises macroeconomic data of selected economies and takes into account a number of announced macroeconomic indicators and their relations to the median of the forecasted values of these indicators and risk given by the standard deviation of forecasts. The study uses intraday data. It is clear that if the median of forecasts is not in line with the actual value of an announced macroeconomic indicator (i.e. its expected value

[1]Department of Applications of Mathematics in Economics, Faculty of Management, AGH University of Science and Technology, 30-067 Krakow, Poland. E-mail: kbrania@agh.edu.pl. ORCID: https://orcid.org/0000-0002-5622-7605.

[2]Department of Applications of Mathematics in Economics, Faculty of Management, AGH University of Science and Technology, 30-067 Krakow, Poland. E-mail: gurgul@agh.edu.pl. ORCID: https://orcid.org/0000-0002-6192-2995.

significantly differs from the actual value) the reaction of various markets to this announcement may be stronger.

This paper is organized as follows:

- In Section 2 we present an overview of the selected literature concerning the impact of macroeconomic announcements on the financial markets. Most of the results overviewed in this section are based on event study methodology. The second possible approach refers to the simple notion, which is not widespread in the financial literature, namely the macroeconomic surprise indicator. We reviewed two recent studies which used the definition of this indicator.

- Section 3 explains the underlying macroeconomic data examined in this research.

- Section 4 contains the results of the empirical studies.

- In Section 5 we propose an aggregate index that could be used as a proxy of domestic market sentiment.

- Section 6 contains the final remarks, conclusions and potential fields for further research.

The macroeconomic data used in the study are from the United States of America, the United Kingdom, China, the Eurozone, Germany, Japan and Poland.

The next section provides a short (selected) literature overview of the importance of macroeconomic news announcements.

2. Literature overview

Macroeconomic data announcements have a great impact on the treasury bond market, stock markets, exchange rates, interest rates and other economic variables.

News on macroeconomic data from different countries plays a particularly important role for stock markets. In this short overview we will concentrate on the literature concerning this issue, especially the importance of the US macroeconomic data announcements. The empirical results presented in the contributions listed below were mostly calculated by means of event study methods. There is a much lower number of contributions which referred to the surprise variable. If the data released correspond precisely to the expected value, this variable defined below will be equal to zero. At the end of this section we review two recent papers using surprise variables in order to quantify the impact of macroeconomic surprises on different economic variables such as equity prices or the exchange rates of different currencies.

Taking into account the role of the US economy throughout the world, information concerning its condition is of great interest in the economic literature. The effects of main US macroeconomic news announcements on stock prices have been checked for markets at different stages of development: developed stock markets (Andersen et al. (2007), Boyd

et al. (2005), Harju & Hussain (2011), Li & Hu (1998), Pearce & Roley (1983), Nikkinen & Sahlström (2004), Schwert (1981)), as well as emerging markets, e.g. in Central and Eastern Europe (see Gurgul et al. (2012), Gurgul et al. (2013), Gurgul & Wójtowicz (2014), Gurgul & Wójtowicz (2015), Hanousek et al. (2009)). In their well-known and frequently cited contribution Nikkinen & Sahlström (2004) examine the impact of the US and domestic macroeconomic news on the German and Finnish markets. They discovered the dominant role of macroeconomic information from the US. According to the authors, the impact of announcements from the US on the volatility of both markets under consideration was stronger than the effect of domestic information. This investigation was continued by Nikkinen et al. (2006). The contributors examined the impact of the US macroeconomic news announcements on a broad group of 35 stock markets around the world. Among the stock markets under study there were some developed and emerging markets in Europe. The study concludes that unexpected macroeconomic news from the US economy affects volatility on developed stock markets in Europe and Asia. However, volatility on emerging European markets (including Poland) did not react to unexpected announcements of American macroeconomic data. This demonstrated the main difference between developed and emerging European markets with respect to the impact of information coming from the US. Quite different results were presented by Gurgul et al. (2012). They detected a significant reaction of the daily returns of the main Warsaw stock index WIG20 to unexpected news about inflation and industrial production in the United States. The basic difference between these results may follow from an analysis of data from different periods. Nikkinen et al. (2006) take into consideration returns from July 1995 to March 2002, i.e., from the initial period of the development of WSE. The study by Gurgul et al. (2012) is based on more recent data.

More precise results on the impact of the US macroeconomic news on European markets were obtained by applying intraday data. For example, Harju & Hussain (2011) use 5-minute returns to investigate the impact of US macroeconomic news announcements on the British, French, German, and Swiss stock markets. They prove the significant and immediate impact of such news on volatility and 5-minute returns on these markets. Similar results are presented by Dimpfl (2011), who analyzes 1-minute DAX returns, and by Gurgul & Wójtowicz (2015), who analyze 5-minute ATX returns. Additionally, the study of the changes in the strength of the reaction of ATX returns to US data announcements over subsequent years presented by Gurgul & Wójtowicz (2015) leads to the conclusion that investors on the Vienna Stock Exchange reacted most strongly during the global financial crisis of 2007-2009.

An empirical analysis of the effect of macroeconomic news on intraday data has been conducted for European emerging markets by Hanousek et al. (2009). The authors conclude that the Czech and the Hungarian stock markets reacted to macroeconomic news from both the US and EU. However WSE is affected only by announcements from the Eurozone. This stream of research was continued by Hanousek & Kočenda (2011). They show that emerging stock markets in the CEE region are mainly sensitive to macroeconomic information from the EU. Quite different results are presented in Gurgul & Wójtowicz (2014). They detect that unexpected news from the US economy implies a significant and very strong reaction of the WIG20 index as early as the first minute after the announcements. This

extended analysis proves that the US macroeconomic announcements also significantly influence medium and small stock indices of the stock exchange in Warsaw. However, the reaction is slower than in the case of the largest companies.

In our computations, we use the surprise index (SI). An analogous measure of uncertainty is used in the contributions by Heinlein & Lepori (2022) and Jäggi et al. (2019). An important difference is that our study (unlike the study below) is based on high frequency data. Heinlein & Lepori (2022) investigated the response of the UK asset prices to a large set of domestic scheduled, macroeconomic announcements using data at a daily frequency from 1998 to 2017. Surprises about retail sales, claimant count rate, GDP, and industrial production have the most prevalent effects across the four asset classes in their dataset. A large number of macroeconomic announcements increase trading activity on the stock market, whereas there is barely any (only minor) evidence that announcements (surprises) affect the volatility of asset prices. Heinlein & Lepori (2022) also proved that the effects of macroeconomic surprises depend not only upon the state of the economy but also on the state of the stock market (bull vs bear).

The impact of macroeconomic surprises can be detected not only on the capital market, since it has also important monetary implications. Jäggi et al. (2019) studied the reaction of CHF and JPY to macroeconomic surprises and changes in a broader market environment before and during the crisis using high-frequency data. The results published by the contributors demonstrate that CHF and JPY are traditionally more sensitive to macroeconomic surprises than other currencies. This reflects the fact that macroeconomic surprises have an effect on uncertainty and risk aversion. Jäggi et al. (2019) stressed that this link was further increased during the crisis. The authors underlined that it could not be broken by the specific measures adopted by monetary authorities to limit the appreciation trend. In addition the contributors found some evidence that, during the crisis, CHF and JPY responded more strongly to surprises generating an appreciation than to surprises leading to a depreciation. Moreover, both currencies also systematically reacted to changes in the general market environment. According to the authors, this result is resistant to the applications of two measures of the market environment: VIX and a novel index based on Bloomberg wires.

In the following section the datasets, including extensive intraday macroeconomic data announcements for selected economic areas are presented.

3. Datasets

In the first part of this section we describe data from time period 2018 – 2023 given in three subintervals, each of two years in length. In the second part we describe macroeconomic data in detail.

We have decided to divide our macroeconomic data sample into three intervals:

- Pre-pandemic period (2018-2019);
- COVID-19 pandemic (2020-2021);
- Post-pandemic period (2022-2023).

There are two main reasons that support this division. Firstly, we designed equally long periods in order to have comparable number of observations in all three time intervals. Secondly, we deem this partitioning as justified in the context of COVID-19 data.

The source of the macroeconomic data is Bloomberg Terminal, which allows professionals and researchers across multiple industries to access both real-time and historical feeds across a wide range of categories.

Examples of fields that are available in the feed: Date, Time, Event, Period, Surv(M), Actual, Prior, Revised, S, Ticker, Day, Surv(A), Surv(H), Surv(L), Freq., Ests., Std Dev, Surprise, Category, Subcategory, R, First Rev., Last Rev., Country/Region.
For our research the quantitative metrics that help us to define the surprise index are the most crucial ones:

Definition 1 *For the given quantitative macroeconomic data release (at the time t_0) we define surprise index (SI) as follows:*

$$SI_{t_0} = \frac{Surv(M)_{t_0} - Actual_{t_0}}{StdDev_{t_0}}, \tag{1}$$

where Surv(M) is the median of survey forecasts referring to this particular macroeconomic data release at the time t_0, Actual is an immediate announcement of the indicator (not including further updates) and StdDev is the standard deviation of forecasts.

Therefore, the surprise index is a simple metric that takes into account two dimensions that we wished to capture:

- How far away from the market consensus the actual macroeconomic data release was;
- Uncertainty of market analysts forecasts.

Based on the construction of the surprise index as a ratio of the above factors, we deem that it is indeed a good proxy of the importance of new information appearing on the market. We assume that the efficient markets hypothesis is true, that new information should not be included in the asset prices at the moment of release, but it is likely to impact them in the very short run.

Another important metric that is leveraged in this study is the relevance factor. This is an indicator ranging from 0 to 100 that aims to measure the importance of a particular macroeconomic indicator. This is approximately the ratio between the number of Bloomberg Terminal's users that have set up the alerts on the particular indicator to the total number of users that have created at least one notification on their accounts.

The composition of the economies was due to a few reasons. Firstly, we have picked the biggest economies in the world: the United States of America, China, Japan, Germany and the United Kingdom. Moreover, Germany is the biggest European economy and the largest trading partner of Poland. The Eurozone's macroeconomic data is considered to be important data as it is aggregated data across the countries that use euro as the official currency. It should be noted that euro is the most important foreign exchange reserve currency, obviously after USD. We included Poland as a minor emerging market.

All available macroeconomic data announcements were downloaded from the Bloomberg Terminal. Then, we applied a few filters to facilitate later quantitative procedures:

- Filtering out events for which "Actual" value was empty or not available;
- Leaving out cases where "Surprise" value cannot be calculated (missing median or standard deviation of forecasts);
- Omitting events for which the time was only approximate, i.e. the whole day instead of exact time point during the day.

For instance, the US data was the richest dataset available amongst the economies considered. From the initial number of 10827 announcements (from 117 categories), 7273 (from 99 categories) remained after filtering.

In all economies under study there is no visible relationship between surprise and relevance. This statement is supported by Pearson's linear correlations and rank correlations (Spearman's and Kendall's), which are very close to zero in all cases, as denoted in Table 1.

Table 1: Relationship between relevance and surprise of macroeconomic data releases across economies – measured by correlation metrics

	China	Eurozone	Germany	Japan	Poland	UK	USA
Pearson	-0.0393	0.0177	-0.0336	0.0057	0.0231	0.0073	-0.0177
Spearman	-0.0468	0.0269	-0.0329	0.0072	0.0158	0.0089	-0.0085
Kendall	-0.0304	0.0179	-0.0219	0.005	0.0109	0.0061	-0.0058

Therefore, it seems as if the quality of analysts' forecasts does not depend on the importance of the metric. It might be counter-intuitive as one might suspect that most important releases receive much more attention from market participants than minor announcements.

It is also important to acknowledge that in order to measure the impact of macroeconomic news announcements on any financial instruments, we have to consider only those indicators that are released when the particular asset is tradeable. As one might expect, the major factor that influences the time of day of the majority of releases is the economy's time zone – the scatterplot can clearly divide the data into visible clusters.

After the preliminary characteristics of the relationships between surprise and relevance indicators for the selected macroeconomic releases, in the next section we will present the empirical results of our computations and testing procedures.

4. Empirical results – single indicator level

We would like to begin with some simple exploratory data analysis. Standard descriptive statistics are available from the authors upon request. These are, namely, minimum, maximum, mean, skewness and kurtosis. They are calculated for surprise indicator values, separately for each type of macroeconomic event, per the 2-year periods defined previously.

Then, we calculated the p-values of the Kruskal-Wallis and ANOVA tests for all the economies examined. The order is as follows: Table 2 – Germany, Table 3 – China, Table 4

– the Eurozone, Table 5 – Japan, Table 6 – Poland, Table 7 – the United Kingdom and Tables 8, 9 – the United States. The significance level is set to 5% – all categories of macroeconomic announcements that have at least one p-value smaller or equal than 5% are marked in bold font.

Table 2: P-values of Kruskal-Wallis, ANOVA, Kolmogorov-Smirnov and Anderson-Darling tests for Germany

Indicator	K.W	ANOVA	K.S_1	K.S_2	K.S_3	A.D_1	A.D_2	A.D_3
Budget Maastricht % of GDP	0.7408	0.952	0.6667	1	1	0.1675	1	0.7852
Capital Investment QoQ	0.1133	0.1432	0.6601	0.2827	0.087	0.2697	0.1713	0.056
CPI EU Harmonized MoM	0.1148	0.448	**0.041**	0.2055	0.7972	**0.0136**	0.0731	0.8507
CPI EU Harmonized YoY	0.2635	0.9008	**0.0381**	0.0744	0.7762	**0.0154**	**0.03**	0.9227
CPI MoM	0.6595	0.9492	0.3651	0.5885	0.9921	0.2924	0.3982	0.77
CPI YoY	0.606	0.9414	0.3074	0.4458	0.7204	0.2208	0.2737	0.5636
Current Account Balance	0.3833	0.4134	0.4726	0.14	0.3261	0.3612	0.1816	0.4626
Exports SA MoM	0.8164	0.9485	0.6794	0.686	0.9935	0.6141	0.8176	0.9913
Factory Orders MoM	0.3911	0.3689	0.2628	0.4426	0.6798	0.2589	0.4137	0.9129
Factory Orders WDA YoY	0.3775	0.2586	0.686	0.686	0.8982	0.2536	0.3372	0.6643
GDP NSA YoY	0.4667	0.7482	0.5055	0.5425	0.5425	0.2155	0.3469	0.3022
GDP SA QoQ	0.8432	0.4858	0.4056	0.6483	0.7673	0.1981	0.3762	0.8715
GDP WDA YoY	0.1785	0.1935	0.0873	0.4257	0.1553	**0.022**	0.1306	0.1683
GfK Consumer Confidence	0.6585	0.5697	0.8982	0.1307	0.6794	0.6998	0.2469	0.5011
Government Spending QoQ	0.2189	0.5031	0.5594	0.0948	0.1103	0.2778	0.2823	0.2112
IFO Business Climate	0.7614	0.7163	0.7067	0.4425	0.8728	0.5933	0.4683	0.868
IFO Current Assessment	0.9419	0.9531	0.9348	0.8983	0.5961	0.9184	0.9926	0.7517
IFO Expectations	0.4837	0.8513	0.7503	0.2628	0.4542	0.8069	0.249	0.7243
Import Price Index MoM	**0.05**	0.0826	**0.0189**	0.6985	0.2578	0.0161	0.5344	0.1593
Import Price Index YoY	0.0676	0.0884	**0.0238**	0.601	0.4224	**0.0235**	0.3967	0.2426
Imports SA MoM	0.7712	0.8501	0.9942	0.8938	0.449	0.8892	0.9041	0.491
Industrial Production SA MoM	0.4547	0.309	0.8822	0.4358	0.8938	0.8262	0.2756	0.5704
Industrial Production WDA YoY	0.4826	0.2985	0.6723	0.6725	0.4286	0.6435	0.3329	0.5971
PPI MoM	0.244	0.1273	0.2395	0.8982	0.4955	0.1177	0.9288	0.25
PPI YoY	0.1508	0.0957	0.4325	0.6044	0.117	0.1014	0.8236	0.1486
Private Consumption QoQ	0.6855	0.921	0.6601	0.9801	0.6224	0.391	0.6339	0.5674
Retail Sales MoM	0.2867	0.247	0.4294	0.686	0.2578	0.3847	0.531	0.1147
Retail Sales NSA YoY	0.8509	0.9403	1	0.8939	0.6804	0.9701	0.5545	0.3321
Trade Balance	0.5013	0.427	0.686	0.182	0.6539	0.7046	0.2291	0.7201
Unemployment Change (000's)	0.0967	0.285	0.686	0.2537	0.137	0.6718	0.1703	0.0553
Unemployment Claims Rate SA	0.1953	0.2583	**0.0484**	0.4229	0.2183	0.6124	0.1053	0.2288
ZEW Survey Current Situation	0.7293	0.8761	0.6666	0.6727	1	0.771	0.5936	0.9956
ZEW Survey Expectations	0.7958	0.5245	0.686	0.8982	0.6804	0.6708	0.6834	0.4678

In the case of China we have only one significant indicator, for Eurozone – two, for Germany – one, for Japan – four, for Poland – two, for the United Kingdom – eight and for United States – eleven.

In the above-mentioned tables, we also report the p-values of the Kolmogorov-Smirnov and Anderson-Darling tests for all the economies examined. The significance level is unchanged (5%) – all categories of macroeconomic announcements that have at least one p-value smaller or equal than 5% are marked in bold font.

Note that we have chosen the following notation:

- Tests referring to the comparison between pre-crisis and crisis periods are marked as **1**;
- For periods before crisis and after crisis – **2**;
- Crisis and post-crisis tests are denoted as **3**.

Table 3: P-values of Kruskal-Wallis, ANOVA, Kolmogorov-Smirnov and Anderson-Darling tests for China

Indicator	K.W	ANOVA	K.S_1	K.S_2	K.S_3	A.D_1	A.D_2	A.D_3
1-Year Loan Prime Rate	0.5878	0.6106	0.0873	0.0666	0.9254	**0.0427**	**0.0396**	1
5-Year Loan Prime Rate	0.8926	0.8041	0.1538	0.287	0.272	**0.031**	0.2148	0.613
Aggregate Financing CNY	0.9	0.7601	0.6397	0.9239	0.4425	0.6253	0.9186	0.6507
CPI YoY	0.1365	0.1365	0.0968	0.0889	0.8719	0.3024	0.1779	0.792
Exports YoY	0.0856	0.2048	0.3008	0.7666	0.0778	0.1689	0.865	**0.0324**
Exports YoY CNY	0.1263	0.0903	0.2506	0.8531	0.2191	0.1114	0.8809	0.1502
Fixed Assets Ex Rural YTD YoY	0.8749	0.6077	0.6127	0.8604	0.3866	0.3711	0.6433	0.4125
Foreign Reserves	0.9834	0.9907	0.888	0.5607	0.8184	0.8334	0.6387	0.5699
GDP SA QoQ	0.4788	0.3813	0.9702	0.2124	0.6601	0.4656	0.1137	0.3893
GDP YoY	0.1621	0.2992	0.5594	0.2528	**0.0145**	0.5199	0.0911	**0.0247**
GDP YTD YoY	0.1996	0.2198	**0.0482**	0.6713	0.2827	0.0931	0.3892	0.3728
Imports YoY	0.4508	0.4035	0.5901	0.6274	0.5713	0.5152	0.619	0.1664
Imports YoY CNY	0.0724	0.1089	0.9692	0.1133	**0.0273**	0.9598	0.0743	**0.0219**
Industrial Production YoY	0.7127	0.9823	0.5623	0.1745	0.9813	0.3457	0.227	0.9522
Industrial Production YTD YoY	0.8952	0.5659	0.8316	1	0.8574	0.6018	0.9955	0.6266
Manufacturing PMI	0.8595	0.6377	0.686	0.9942	0.6726	0.339	0.7705	0.6659
Money Supply M0 YoY	0.2927	0.5252	0.2642	0.2181	0.72	0.3593	0.2091	0.7086
Money Supply M1 YoY	0.1802	0.0525	0.1364	0.6794	0.4402	0.0533	0.4108	0.2089
Money Supply M2 YoY	0.146	0.1251	0.4167	**0.0477**	0.6558	0.1661	**0.0323**	0.289
New Yuan Loans CNY	0.9633	0.8712	0.9935	0.658	0.4295	0.97	0.4766	0.2608
Non-manufacturing PMI	0.6472	0.7859	0.8882	0.6726	0.2628	0.7882	0.6756	0.2378
PPI YoY	**0.0092**	**0.0068**	0.0539	0.0529	**0.0298**	0.059	**0.0262**	**0.0058**
Retail Sales YoY	0.6713	0.8002	0.9813	0.3283	0.832	0.8232	0.3366	0.8429
Retail Sales YTD YoY	0.6002	0.4347	0.8236	0.3253	0.6179	0.908	0.2897	0.2842
Trade Balance	0.7331	0.79	0.56	0.9551	0.8248	0.5568	0.8711	0.7766
Trade Balance CNY	0.7304	0.6855	0.5304	0.7374	0.6558	0.6391	0.531	0.7091

Table 4: P-values of Kruskal-Wallis, ANOVA, Kolmogorov-Smirnov and Anderson-Darling tests for the Eurozone

Indicator	K.W	ANOVA	K.S_1	K.S_2	K.S_3	A.D_1	A.D_2	A.D_3
Consumer Confidence	0.356	0.2166	**0.0412**	0.1364	0.6234	**0.0326**	0.1096	0.6587
CPI Core YoY	0.3232	0.4001	0.5148	0.273	0.7023	0.4285	0.1151	0.5524
CPI Estimate YoY	0.4883	0.4536	0.4845	0.0416	0.3923	0.1392	**0.017**	0.2995
CPI MoM	0.5966	0.325	0.2608	0.173	0.1884	0.2468	0.1322	0.386
CPI YoY	0.4897	0.2484	0.1019	0.2556	0.7387	0.3076	0.9699	0.4786
Economic Confidence	**0.0018**	**0.002**	0.0656	0.8893	**0.0111**	**0.0069**	0.7846	**0.0009**
GDP SA QoQ	0.5419	0.6523	0.1363	0.4346	0.9642	**0.0461**	0.3436	0.6967
GDP SA YoY	0.1522	0.5229	**0.0271**	0.8871	**0.0372**	**0.0143**	0.9997	**0.0246**
Govt Expend QoQ	0.6448	0.9218	0.2124	0.9879	0.8364	0.5281	0.7212	0.7675
Gross Fix Cap QoQ	0.767	0.4308	0.9801	0.5859	0.5859	0.7778	0.4691	0.5962
Household Cons QoQ	0.4605	0.9482	0.205	0.5998	0.9751	0.1572	0.3344	0.9836
Industrial Confidence	**0.0035**	**0.0071**	**0.0037**	0.6691	**0.0042**	**0.0022**	0.6116	**0.001**
Industrial Production SA MoM	0.519	0.4803	1	0.2466	0.2537	0.9955	0.3459	0.5772
Industrial Production WDA YoY	0.5595	0.4635	0.8815	0.429	0.2576	0.6258	0.4247	0.1352
M3 Money Supply YoY	0.1613	0.2062	0.6446	0.2271	0.1996	0.5606	0.1682	0.1329
PPI MoM	0.6791	0.5654	0.6255	0.3921	0.4041	0.6179	0.4767	0.4445
PPI YoY	0.5709	0.7642	0.8828	0.4295	0.4195	0.6146	0.632	0.5074
Retail Sales MoM	0.1917	0.4482	0.4358	0.124	0.2531	0.532	0.0628	0.1248
Retail Sales YoY	0.5641	0.6399	0.4296	0.6794	0.1362	0.6528	0.8087	0.2143
Sentix Investor Confidence	0.555	0.5496	0.2527	0.8982	0.6795	0.2302	0.7289	0.5174
Services Confidence	0.4949	0.475	0.4425	0.6794	0.6798	0.4758	0.7335	0.5508
Trade Balance SA	0.743	0.8618	0.9159	0.3625	0.8596	0.9762	0.5694	0.7622
Unemployment Rate	0.3828	0.4488	0.4279	0.9053	0.2303	0.8876	0.9747	0.7178

To summarize the results of testing the differences in the distributions of macroeconomic surprises in the three periods under consideration, the fraction of indicators of significant differences (at least one test rejecting the null hypothesis) oscillates between 9.3% for Japan to 30.8% for China (for the Eurozone – 26.1%, for Germany – 18.2%, for Poland – 11.5%,

Table 5: P-values of Kruskal-Wallis, ANOVA, Kolmogorov-Smirnov and Anderson-Darling tests for Japan

Indicator	K.W	ANOVA	K.S_1	K.S_2	K.S_3	A.D_1	A.D_2	A.D_3
Annualized Housing Starts	0.3706	0.7978	0.6558	0.2575	0.2342	0.5767	0.2477	0.1753
BoP Current Account Adjusted	0.7209	0.232	0.4361	0.6723	0.6726	0.4681	0.8262	0.2723
BoP Current Account Balance	0.6336	0.6705	0.4425	0.686	0.252	0.2503	0.4497	0.2546
Capital Spending Ex Software YoY	0.5234	0.6999	0.2827	0.6601	0.6601	0.2738	0.3332	0.338
Capital Spending YoY	0.3027	0.8994	0.2827	0.9801	0.2827	0.182	0.83	0.2884
Coincident Index	0.1633	0.4681	0.2528	0.2843	0.052	0.1881	0.336	**0.0443**
Consumer Confidence Index	**0.0138**	0.0647	**0.0054**	0.0659	0.342	**0.0042**	**0.0287**	0.2734
Core Machine Orders MoM	0.85	0.7333	0.9942	0.9024	0.8939	0.8968	0.9188	0.8744
Core Machine Orders YoY	0.9265	0.9248	0.9024	0.9942	0.686	0.7574	0.9667	0.8715
Eco Watchers Survey Current SA	0.1283	**0.0362**	0.2578	0.2575	**0.0299**	**0.0447**	0.1352	**0.0468**
Eco Watchers Survey Outlook SA	0.5566	0.3626	0.2628	0.2613	0.8982	0.3117	0.3245	0.5998
Exports YoY	0.2887	0.4837	0.4445	0.4364	0.6736	0.2155	0.1354	0.4004
GDP Annualized SA QoQ	0.4754	0.4199	0.9522	0.0879	0.3033	0.8778	0.1266	0.3885
GDP Business Spending QoQ	0.4122	0.5469	0.6036	0.3959	0.0928	0.6533	0.3094	0.0781
GDP Deflator YoY	0.3629	0.07	0.3139	0.2308	0.3787	0.5749	0.2469	0.2087
GDP Nominal SA QoQ	0.7669	0.4487	0.7229	0.6594	0.6225	0.9155	0.6595	0.631
GDP Private Consumption QoQ	0.5953	0.3638	0.3749	0.0927	0.517	0.279	0.2448	0.6062
GDP SA QoQ	0.4824	0.5109	0.5743	0.206	0.2916	0.6548	0.12	0.2857
Household Spending YoY	0.3009	0.3655	0.9115	0.4493	0.2628	0.8205	0.2806	0.2014
Housing Starts YoY	0.2886	0.5132	0.6736	0.6794	0.0678	0.4561	0.5307	0.1297
Imports YoY	0.5667	0.5722	0.4425	0.8843	0.6731	0.3823	0.6441	0.5757
Industrial Production MoM	0.7919	0.7639	0.9024	0.686	0.686	0.7316	0.7387	0.7256
Industrial Production YoY	0.6782	0.6641	0.4426	0.6798	0.6752	0.1253	0.4024	0.6261
Job-To-Applicant Ratio	0.7114	0.9219	0.1825	0.9696	0.5581	0.2803	0.9533	0.6459
Jobless Rate	0.3474	0.219	0.3875	0.6208	0.1078	0.2237	0.3216	0.1218
Labor Cash Earnings YoY	0.2202	0.332	0.3281	0.2653	0.6794	0.3406	0.1183	0.5474
Leading Index CI	0.2205	0.1548	0.5845	0.061	0.4044	0.4193	0.0713	0.4948
Monetary Base End of period	0.3801	0.4375	0.3333	0.6667	1	0.2975	0.6267	1
Monetary Base YoY	0.3189	0.6064	0.5	0.6	1	0.1247	0.3491	0.9762
Money Stock M2 YoY	0.214	0.1116	0.2138	0.5958	0.1039	0.061	0.3675	0.1464
Money Stock M3 YoY	0.1094	**0.0214**	0.0861	0.7311	0.2764	**0.0118**	0.3669	0.1015
Natl CPI Ex Fresh Food YoY	0.2421	0.2623	0.3177	0.3639	0.6821	0.0657	0.0746	0.9804
Natl CPI YoY	0.3988	0.4425	0.7193	0.6387	0.999	0.2019	0.6741	0.8336
PPI MoM	0.2148	0.1186	0.2273	0.9909	0.4195	0.1749	0.7889	0.2065
PPI Services YoY	0.764	0.7845	0.9828	0.4457	0.947	0.8227	0.5287	0.8737
PPI YoY	0.4237	0.3774	0.6349	0.4324	0.9916	0.3116	0.1867	0.8628
Real Cash Earnings YoY	0.7447	0.7997	0.4205	0.6526	0.9942	0.3327	0.6624	0.9312
Retail Sales MoM	0.6351	0.4622	0.4425	0.8938	0.6794	0.2491	0.6778	0.4816
Retail Sales YoY	0.8635	0.8926	0.686	0.449	0.8982	0.5033	0.2586	0.7907
Tankan Large All Industry Capex	0.9489	0.9813	0.6601	0.6601	0.9801	0.637	0.5407	0.6688
Tankan Large Mfg Index	0.2787	0.3366	0.2827	0.2827	0.6601	0.0856	0.3955	0.5445
Tankan Large Mfg Outlook	0.8325	0.663	0.9702	0.6601	0.9702	0.9141	0.6183	0.9479
Tankan Large Non-Mfg Index	0.717	0.7724	0.9801	0.6601	0.9801	0.83	0.5308	0.6472
Tankan Large Non-Mfg Outlook	0.2144	0.1509	0.9702	0.2528	0.2827	0.8873	0.161	0.1595
Tankan Small Mfg Index	0.4439	0.5356	0.2572	0.6601	0.6601	0.1288	0.4257	0.637
Tankan Small Mfg Outlook	0.077	**0.0465**	0.6601	0.087	0.1702	0.3664	**0.0289**	0.1945
Tankan Small Non-Mfg Index	0.3589	0.539	0.2827	0.2827	0.9702	0.2265	0.1661	0.9203
Tankan Small Non-Mfg Outlook	0.6914	0.656	0.6224	0.6224	0.9801	0.8558	0.7994	0.7641
Tertiary Industry Index MoM	0.3706	0.1836	0.2576	0.3075	0.177	0.3024	0.178	0.2179
Tokyo CPI Ex-Fresh Food YoY	0.4978	0.7377	0.1791	0.228	0.293	0.3219	0.2115	0.2504
Tokyo CPI YoY	0.4997	0.5494	0.7664	0.8777	0.6285	0.5838	0.5442	0.4852
Trade Balance	0.6945	0.6137	0.9928	0.8982	0.8939	0.9481	0.6486	0.7499
Trade Balance Adjusted	0.9066	0.8031	0.8982	0.6794	0.8791	0.5868	0.709	0.6929
Trade Balance BoP Basis	0.1884	0.1631	0.6634	0.8928	0.1322	0.5484	0.5262	0.0647

for the United Kingdom – 23.7%, for United States – 26.3%). In general, we observe significant differences in the distributions of surprise indicators related to inflation and unemployment across most of the economies under consideration in these periods. Another conclusion is that, in general, the bigger the economy, the higher the fraction of indicators of significant differences (with the exception of Japan). Across examined economies, the

Table 6: P-values of Kruskal-Wallis, ANOVA, Kolmogorov-Smirnov and Anderson-Darling tests for Poland

Indicator	K.W	ANOVA	K.S_1	K.S_2	K.S_3	A.D_1	A.D_2	A.D_3
Average Gross Wages MoM	0.3927	0.1263	0.1366	0.686	0.8891	0.19	0.2035	0.5778
Average Gross Wages YoY	0.4008	0.2329	0.2628	0.6794	0.4435	0.1638	0.2808	0.354
Construction Output YoY	0.9081	0.6002	0.8982	0.2575	0.2628	0.7662	0.1387	0.3768
CPI Core MoM	0.5374	0.4214	0.7427	0.9533	0.9406	0.7062	0.9632	0.9673
CPI Core YoY	0.6387	0.9479	0.934	0.6839	0.9533	0.7072	0.9497	0.6878
CPI MoM	0.1532	0.1186	0.0539	0.7858	0.0979	**0.0438**	0.8467	0.1185
CPI YoY	0.0533	**0.0236**	**0.0165**	0.4197	0.0656	**0.0109**	0.3313	0.1215
Current Account Balance	0.732	0.9319	0.4425	0.2628	0.4425	0.5631	0.3064	0.585
Employment MoM	0.6443	0.3327	0.1775	0.6441	0.3687	0.2355	0.6488	0.1528
Employment YoY	0.6369	0.7597	0.8106	0.8043	0.8605	0.6072	0.8081	0.5227
Exports	0.8011	0.8606	0.9928	0.686	0.8982	0.9182	0.7198	0.8388
GDP Annual YoY	0.1017	0.0548	0.3333	0.3333	0.3333	0.0835	0.0835	0.0835
GDP SA QoQ	0.6171	0.7552	0.2827	0.6224	0.2827	0.1135	0.4352	0.1486
GDP YoY	0.4371	0.2993	0.6601	0.6601	0.087	0.2727	0.487	0.1013
Imports	0.447	0.3042	0.6794	0.6726	0.2628	0.6076	0.402	0.1321
Money Supply M3 MoM	0.4917	0.6216	0.7012	0.3649	0.6794	0.9286	0.4069	0.6261
Money Supply M3 YoY	0.3251	0.6887	0.436	0.1328	0.1322	0.4549	0.12	0.347
PPI MoM	0.0585	**0.0268**	0.228	0.2384	**0.0238**	0.1067	0.142	**0.0231**
PPI YoY	0.1712	0.1867	0.25	0.686	0.1316	0.0701	0.5461	0.1717
Retail Sales Real YoY	0.5711	0.2449	0.686	0.992	0.8982	0.5022	0.9665	0.4132
Retail Sales YoY	0.1662	0.0674	0.2526	0.8982	0.6794	0.081	0.6567	0.3498
Sold Industrial Output MoM	0.4511	0.2994	0.4425	0.2509	0.1287	0.2763	0.5976	0.264
Sold Industrial Output YoY	0.3009	0.2399	0.9024	0.1398	0.1361	0.6009	0.2947	0.122
Trade Balance	0.626	0.3254	0.8982	0.6794	0.9024	0.8624	0.5034	0.796
Unemployment Rate	0.9327	0.7941	0.8312	0.9819	0.8404	0.7847	0.9998	0.8188
Unemployment Rate Quarterly	0.9398	0.9368	0.9851	0.6601	0.3998	0.9932	0.5282	0.4681

ANOVA test seems to be more sensitive in detecting significant differences in distribution than the Kruskal-Wallis test. Results in three sub-sample periods indicate that the Anderson Darling test is superior to Kolmogorov-Smirnov test in the purpose of detecting differences in distributions. This is in line with the statistical intuition as the Anderson Darling test puts more weight into the comparison of the tail distribution.

In the next section, we first define the aggregated sentiment (AS). We also include the p-values of the Kruskal-Wallis, ANOVA, Kolmogorov-Smirnov and Anderson-Darling tests for these economies in the tables.

5. Empirical results – sentiment level

In order to quantitatively aggregate the impact of new market information from macroeconomic data announcements, we have suggested using the aggregated sentiment. For multiple events occurring at the same time, this indicator is also normalized by the sum of Relevance Factors of all simultaneous macroeconomic data releases. This means that the sentiment is calculated as the weighted average where the weights are based on the values of the relevance factor per each indicator.

Definition 2 *Aggregated sentiment (AS) at the time t is the product sum of the surprise indicator (SI) and relevance factor (RF), summed through all economic releases taking place simultaneously – let us assume that they are annotated by natural numbers* $(1,\ldots,n)$*:*

$$AS_t = \frac{\sum_{i=1}^{n} SI_{i,t} \cdot RF_{i,t}}{\sum_{i=1}^{n} RF_{i,t}} \tag{2}$$

Table 7: P-values of Kruskal-Wallis, ANOVA, Kolmogorov-Smirnov and Anderson-Darling tests for the United Kingdom

Indicator	K.W	ANOVA	K.S_1	K.S_2	K.S_3	A.D_1	A.D_2	A.D_3
Average Weekly Earnings 3M/YoY	0.6103	0.3073	0.24	0.5995	0.9924	0.2353	0.3435	0.8719
Bank of England Bank Rate	0.2405	0.3185	1	1	0.4678	0.7461	1	**0.0337**
BRC Sales Like-For-Like YoY	0.0643	0.1164	0.1703	0.5455	0.1538	0.0575	0.5716	**0.0454**
CBI Business Optimism	0.3493	0.4249	0.4156	0.5357	0.6786	0.2748	0.5186	0.2906
CBI Retailing Reported Sales	0.9677	0.7446	0.9024	0.8514	0.9686	0.887	0.9786	0.9276
CBI Total Dist. Reported Sales	0.5722	0.97	0.6667	1	0.5714	0.5681	1	0.6483
CBI Trends Selling Prices	0.5624	0.2396	0.6336	0.4802	0.5966	0.5261	0.4433	0.6487
CBI Trends Total Orders	0.2703	0.344	0.1362	0.449	0.449	0.1056	0.3097	0.4377
Construction Output MoM	0.2712	0.297	0.686	0.6794	0.0678	0.6846	0.6636	0.1067
Construction Output YoY	0.3912	0.4427	0.686	0.686	0.1398	0.5517	0.7122	0.1883
CPI Core YoY	0.1122	0.6117	0.0958	**0.0193**	0.9909	**0.0229**	**0.0119**	0.9874
CPI MoM	0.1638	0.5642	0.0468	**0.0187**	0.9924	**0.0207**	**0.0162**	0.8693
CPI YoY	0.1343	0.4897	0.1154	0.0529	0.9907	**0.038**	**0.0439**	0.978
CPIH YoY	0.0885	0.3294	**0.0466**	0.0532	0.9889	**0.0392**	**0.0319**	0.9981
Current Account Balance	0.9981	0.9619	0.9801	0.9702	0.9801	0.773	0.7412	0.9184
Employment Change 3M/3M	0.8368	0.9336	0.9942	0.9938	0.7679	0.9552	0.982	0.8079
Exports QoQ	0.141	0.4233	0.2557	0.2942	0.9534	0.1311	0.0751	0.9659
GDP QoQ	0.4975	0.4552	0.2961	0.4593	0.9112	0.2185	0.2596	0.8996
GDP YoY	**0.0311**	0.1026	0.1173	**0.031**	0.2102	**0.0367**	**0.0096**	0.1817
GfK Consumer Confidence	0.4859	0.2556	0.8797	0.2393	0.4604	0.7948	0.2835	0.371
Government Spending QoQ	0.0763	0.1293	0.0557	0.0815	0.7387	**0.0397**	0.0573	0.8355
Gross Fixed Capital Formation QoQ	0.5089	**0.0428**	0.3607	0.5953	0.4131	0.4583	0.3823	0.2253
House Price Index YoY	0.2081	0.6762	0.3435	**0.0481**	0.1011	0.2918	0.0639	0.0944
ILO Unemployment Rate 3Mths	0.7351	0.3993	0.1416	0.7478	0.3165	0.2467	0.4811	0.2421
Imports QoQ	0.7023	0.3155	0.2653	0.9038	0.7315	0.5285	0.6205	0.4653
Index of Services 3M/3M	0.6214	0.9634	0.3196	0.4497	0.4195	0.4716	0.3862	0.7146
Index of Services MoM	0.9474	0.7642	0.5168	0.9321	0.8982	0.355	0.7257	0.6031
Industrial Production MoM	0.6693	0.4079	0.9935	0.6794	0.8916	0.9936	0.51	0.6169
Industrial Production YoY	0.392	0.303	0.2576	0.4425	0.4425	0.1927	0.3492	0.2705
Manufacturing Production MoM	0.465	0.2407	0.4364	0.6794	0.9024	0.2775	0.3526	0.5283
Manufacturing Production YoY	0.608	0.4662	0.4425	0.686	0.6794	0.1672	0.4925	0.2713
Monthly GDP (3M/3M)	0.0827	0.1153	**0.0168**	0.7086	0.3956	**0.0387**	0.5078	0.2779
Monthly GDP (MoM)	0.6648	0.5717	0.6056	0.7305	0.6794	0.4144	0.6306	0.4939
Mortgage Approvals	0.8066	0.9949	0.8891	0.8982	0.8946	0.5233	0.7986	0.6571
Nationwide House PX MoM	0.3041	0.189	0.2854	0.8728	0.2032	0.2181	0.9464	0.4242
Nationwide House Px NSA YoY	0.1203	0.0845	0.0733	0.9549	0.0856	0.0751	0.8581	0.1382
Net Consumer Credit	0.2129	0.2242	0.9024	0.6691	0.1398	0.6399	0.593	0.0781
Net Lending Sec. on Dwellings	0.4213	0.3822	0.9024	0.4033	0.4163	0.7698	0.3158	0.3332
PPI Input NSA MoM	0.1385	**0.0493**	0.256	0.4425	0.2628	0.1125	0.5957	0.1017
PPI Input NSA YoY	0.1442	**0.0407**	0.1398	0.693	0.5644	0.0548	0.5783	0.2578
PPI Output NSA MoM	0.5306	0.5397	0.4324	0.8828	0.8916	0.2567	0.9014	0.592
PPI Output NSA YoY	0.5912	0.8369	0.413	0.3744	0.1899	0.466	0.3696	0.3862
Private Consumption QoQ	0.1966	0.4396	0.1009	0.3447	0.8559	0.0876	0.2948	0.8097
PSNB ex Banking Groups	0.8869	0.7864	0.4357	0.449	0.9024	0.2796	0.4615	0.4898
Public Sector Net Borrowing	0.6792	0.5422	0.9024	0.4089	0.6462	0.5387	0.6351	0.5494
Retail Price Index	**0.0416**	0.133	0.129	**0.011**	0.8982	0.105	**0.0139**	0.9058
Retail Sales Ex Auto Fuel MoM	0.3642	0.3023	0.8982	0.686	0.686	0.6024	0.2237	0.5422
Retail Sales Ex Auto Fuel YoY	0.4678	0.3525	0.449	0.2628	0.449	0.4279	0.3151	0.3832
Retail Sales Inc Auto Fuel MoM	0.3914	0.409	0.449	0.1338	0.8982	0.4167	0.1094	0.8732
Retail Sales Inc Auto Fuel YoY	0.2431	0.1088	0.449	0.4425	0.2628	0.3311	0.1212	0.2688
RICS House Price Balance	**0.0257**	0.2031	**0.0299**	1	**0.0299**	**0.0085**	0.9998	**0.0054**
RPI Ex Mort Int.Payments (YoY)	0.1383	0.4749	0.2546	0.1135	0.3679	0.0946	0.151	0.3788
RPI MoM	0.079	0.1965	0.1063	0.1067	0.9877	0.0521	0.0595	0.9857
RPI YoY	**0.0254**	0.0639	0.1212	0.0215	0.9909	**0.0366**	**0.0072**	0.872
Total Business Investment QoQ	0.8197	0.991	0.5882	0.4269	0.499	0.325	0.7206	0.2983
Total Business Investment YoY	0.0771	**0.023**	**0.006**	0.7794	0.1429	**0.0114**	0.6977	0.1233
Trade Balance GBP/Mn	0.5853	0.2875	0.9942	0.9024	0.4379	0.6923	0.6858	0.4068
Visible Trade Balance GBP/Mn	0.862	0.3787	0.9024	0.9024	0.9024	0.9255	0.8022	0.8752
Weekly Earnings ex Bonus 3M/YoY	0.5193	0.4773	0.5757	0.976	0.6133	0.7922	0.6293	0.4317

Table 10 presents the p-values of the Kruskal-Wallis and ANOVA tests applied to the sentiment data calculated for all economies under consideration. The division of the sample

Table 8: P-values of Kruskal-Wallis, ANOVA, Kolmogorov-Smirnov and Anderson-Darling tests for United States (1/2)

Indicator	K.W	ANOVA	K.S_1	K.S_2	K.S_3	A.D_1	A.D_2	A.D_3
ADP Employment Change	0.6999	0.8979	0.6795	0.6249	0.2694	0.6082	0.5618	0.2161
Advance Goods Trade Balance	0.8299	0.8633	0.9315	0.8823	0.8982	0.9454	0.754	0.7363
Average Hourly Earnings MoM	0.1764	0.2271	0.4041	0.2214	0.9847	0.1214	0.174	0.9696
Average Hourly Earnings YoY	0.5612	0.6671	0.575	0.8253	0.9847	0.5	0.3599	0.949
Average Weekly Hours All Employees	**0.0166**	**0.0494**	**0.0003**	0.1971	0.1004	**0**	0.0546	**0.0414**
Building Permits	0.994	0.9797	0.9942	0.6794	0.9942	0.8492	0.735	0.967
Building Permits MoM	1	0.9766	0.9928	0.6726	0.9942	0.8218	0.6979	0.9873
Business Inventories	0.2025	0.2095	0.5315	0.5712	0.0547	0.6281	0.8064	0.155
Cap Goods Orders Nondef Ex Air	0.2428	0.203	0.6455	0.0844	0.5099	0.4886	0.1625	0.2547
Cap Goods Ship Nondef Ex Air	0.559	0.2604	0.2762	0.6326	0.3328	0.2672	0.6854	0.5332
Capacity Utilization	0.2794	0.5081	0.0593	0.8982	0.2413	0.1194	0.8988	0.3687
Change in Manufact. Payrolls	0.5187	0.5237	0.9024	1	0.6794	0.55	0.9896	0.6145
Change in Nonfarm Payrolls	0.1017	0.0876	0.686	0.1398	0.2584	0.4908	0.1207	**0.0348**
Change in Private Payrolls	0.2388	0.2019	0.9942	0.2576	0.4354	0.8638	0.2341	0.1771
Chicago Fed Nat Activity Index	0.809	0.5306	0.6023	0.5792	0.5479	0.7402	0.7349	0.7093
Conf. Board Consumer Confidence	0.8655	0.9109	0.9024	0.8982	0.9942	0.8673	0.9228	0.9648
Construction Spending MoM	0.5246	0.1743	0.9024	0.4426	0.9024	0.6774	0.3267	0.8067
Consumer Credit	0.936	0.5845	0.9024	0.8982	0.9024	0.6675	0.6967	0.8012
Continuing Claims	0.6419	0.6926	0.3647	0.3027	0.1404	0.5156	0.0811	0.1365
Core PCE Price Index QoQ	0.7627	0.8378	0.7145	0.7976	0.8385	0.4524	0.521	0.8464
CPI Core Index SA	0.6156	0.6799	0.8676	0.4272	0.9024	0.8303	0.3938	0.5571
CPI Ex Food and Energy MoM	0.4512	0.394	0.8191	0.5177	0.8434	0.3869	0.1178	0.771
CPI Ex Food and Energy YoY	0.6345	0.2384	0.8253	0.5114	0.575	0.3303	0.5174	0.385
CPI Index NSA	0.3665	0.2212	0.0656	0.6727	0.9024	0.123	0.4559	0.6542
CPI MoM	0.238	0.145	0.0756	0.5177	0.8253	**0.0453**	0.313	0.4755
CPI YoY	0.3076	0.1449	0.1836	0.8253	0.6397	0.0678	0.432	0.385
Current Account Balance	0.5054	0.6265	0.9801	0.6601	0.9801	0.8487	0.3394	0.8638
Dallas Fed Manf. Activity	0.3606	0.6519	0.9942	0.2628	0.2628	0.7131	0.3266	0.2071
Durable Goods Orders	0.3953	0.2335	0.6536	0.1981	0.4279	0.8683	0.2367	0.4933
Durables Ex Transportation	**0.0282**	0.2389	**0.0113**	**0.012**	0.83	0.1085	**0.0113**	0.4914
Empire Manufacturing	0.5301	0.317	0.6795	0.1366	0.1398	0.6978	0.1439	0.1805
Employment Cost Index	0.3164	0.2179	0.205	0.5594	0.6224	0.1719	0.408	0.3789
Existing Home Sales	**0.018**	**0.0195**	**0.004**	0.4435	0.2576	**0.0067**	0.3638	0.0884
Existing Home Sales MoM	**0.0176**	**0.0213**	**0.0043**	0.2628	0.2628	**0.0066**	0.2926	0.0924
Export Price Index MoM	0.2687	0.1254	0.2497	0.6794	0.6691	0.2375	0.5433	0.2318
Export Price Index YoY	0.1845	0.3941	0.3636	0.9474	**0.0398**	0.2397	0.8803	0.0548
Factory Orders	0.2795	0.3202	0.2497	0.6475	0.2577	0.0979	0.3735	0.1026
FHFA House Price Index MoM	0.1382	0.2466	0.4361	**0.0113**	0.449	0.2486	**0.0382**	0.6466
FOMC Rate Decision (Lower Bound)	0.0695	0.068	0.2143	1	0.35	0.9944	0.4566	0.8011
FOMC Rate Decision (Upper Bound)	0.0695	0.068	0.2143	1	0.35	0.9944	0.4566	0.8011
GDP Annualized QoQ	0.5571	0.3942	0.6444	0.1792	0.057	0.6381	0.3629	0.1085
GDP Price Index	0.3611	0.6551	0.2832	**0.031**	0.836	0.2078	**0.0454**	0.563
Housing Starts	0.6112	0.3325	0.686	0.686	0.449	0.3605	0.5112	0.4213
Housing Starts MoM	0.2334	0.1699	0.686	0.0651	0.686	0.4059	0.0954	0.6279
Import Price Index ex Petroleum MoM	0.0877	0.0594	0.0657	0.2628	0.2628	**0.0435**	0.4252	0.2444
Import Price Index MoM	0.3341	0.4266	0.2522	0.9935	0.1398	0.2314	0.8664	0.1755
Import Price Index YoY	0.3178	0.8949	0.4289	0.4309	**0.0256**	0.1585	0.7307	**0.0342**
Industrial Production MoM	0.9205	0.831	0.443	0.686	0.6794	0.4743	0.774	0.3609
Initial Jobless Claims	0.2653	0.1653	0.2143	0.918	0.3492	0.157	0.7839	0.1978

data is exactly the same as in the previous examinations for single macroeconomic indicators. The economies for which both p-values are smaller than or equal to 5% are (in order of significance) the United States, the Eurozone, the United Kingdom and Japan, which means that the null hypothesis about the same distribution in the three periods with respect to the COVID-19 pandemic was rejected. In the case of China no significant differences in these periods were found. Somewhat "intermediate" cases are the German and Polish economies - in these examples the ANOVA test indicates definitely insignificant differences, while the Kruskall-Wallis test is significant at the level of 10%.

Table 9: P-values of Kruskal-Wallis, ANOVA, Kolmogorov-Smirnov and Anderson-Darling tests for United States (2/2)

Indicator	K.W	ANOVA	K.S_1	K.S_2	K.S_3	A.D_1	A.D_2	A.D_3
Interest on Reserve Balances Rate	0.0824	0.0749	0.1818	1	0.1818	**0.0457**	1	**0.0354**
ISM Employment	0.6242	0.6886	0.6667	0.5333	0.5817	0.5681	0.5275	0.6638
ISM Manufacturing	0.1287	0.1295	0.2581	0.4428	**0.0298**	0.2989	0.3414	**0.0229**
ISM New Orders	**0.0434**	0.2658	**0.0351**	**0.049**	0.1637	**0.0129**	0.0803	0.2332
ISM Prices Paid	**0.0419**	0.057	0.0678	0.686	0.0678	0.0767	0.5974	**0.0254**
ISM Services Index	0.1067	0.1221	0.1398	0.6794	0.1364	0.1382	0.6982	0.0708
JOLTS Job Openings	0.1918	0.239	0.0678	0.4426	0.6804	0.1317	0.664	0.2581
Kansas City Fed Manf. Activity	0.0991	0.1593	**0.0177**	0.2695	0.6601	**0.0186**	0.1405	0.5798
Labor Force Participation Rate	0.0521	**0.0411**	0.093	0.3505	0.1548	**0.0321**	0.3298	0.2098
Leading Index	**0.007**	**0.0072**	0.3568	0.3175	**0.0057**	0.2102	0.0886	**0.0005**
Manufacturing (SIC) Production	0.7954	0.9839	0.9888	0.6546	0.8916	0.737	0.5727	0.6412
MNI Chicago PMI	0.23	0.3279	0.4435	0.9024	0.2628	0.2532	0.5425	0.1613
Monthly Budget Statement	0.0628	0.346	0.2227	0.4216	0.061	0.0568	0.4559	**0.0346**
NAHB Housing Market Index	0.1201	0.1735	0.9909	0.2074	0.1236	0.9731	0.0715	**0.0377**
New Home Sales	0.9156	0.9851	0.6794	0.686	0.9942	0.8126	0.7561	0.9777
New Home Sales MoM	0.3053	0.4056	0.4362	0.2628	0.6794	0.473	0.1449	0.6292
NFIB Small Business Optimism	0.754	0.742	0.6794	0.6794	0.5311	0.8121	0.6172	0.4853
Nonfarm Productivity	0.609	0.6193	0.8663	0.2677	0.4188	0.7812	0.4751	0.2879
PCE Core Deflator MoM	0.1804	0.1583	0.2378	0.0683	0.7076	**0.0186**	0.1985	0.3135
PCE Core Deflator YoY	0.5963	0.9192	0.1275	0.6964	0.9784	0.0907	0.5426	0.7773
PCE Deflator MoM	0.7685	0.79	0.9053	0.5889	0.7193	0.2599	0.1356	0.8617
PCE Deflator YoY	0.2844	0.5361	0.1256	0.3634	0.7518	0.086	0.1695	0.8456
Pending Home Sales MoM	0.8158	0.6056	0.9024	0.9935	0.8982	0.6324	0.888	0.7242
Pending Home Sales NSA YoY	0.673	0.3487	0.4398	0.9087	0.781	0.3799	0.861	0.7062
Personal Consumption	0.1163	0.3203	0.9495	0.2294	0.1305	0.868	0.1062	0.0546
Personal Income	0.2843	0.3714	0.1129	0.3127	0.2101	0.0523	0.2707	0.1503
Personal Spending	0.678	0.8254	0.8253	0.3447	0.2258	0.2539	0.1914	0.1629
Philadelphia Fed Business Outlook	0.0868	**0.0434**	0.449	0.2581	0.1398	0.3676	0.2143	**0.0289**
PPI Ex Food and Energy MoM	0.206	0.0851	0.1236	0.6208	0.2273	0.1185	0.4236	0.084
PPI Ex Food and Energy YoY	0.4723	0.1908	0.8916	0.9909	0.6691	0.3704	0.6944	0.4358
PPI Final Demand MoM	0.0851	0.0898	0.24	0.9899	0.2497	0.0665	0.9196	0.0526
PPI Final Demand YoY	0.1328	0.235	0.1249	0.8778	**0.0272**	0.0775	0.6319	0.0814
Real Personal Spending	0.1757	0.6699	0.2503	0.0976	0.1154	0.1401	0.0755	0.0514
Retail Inventories MoM	0.4232	0.7639	0.1877	0.9525	0.1287	0.2583	0.9526	0.1979
Retail Sales Advance MoM	0.7557	0.4813	0.4425	0.1315	0.6657	0.6313	0.1568	0.3665
Retail Sales Control Group	0.3802	0.3369	0.9924	0.4324	0.449	0.9828	0.1665	0.305
Retail Sales Ex Auto and Gas	0.0791	**0.0464**	0.6794	0.0619	0.2532	0.284	**0.0202**	0.2773
Retail Sales Ex Auto MoM	0.2886	0.2016	0.8946	0.1305	0.4301	0.7714	0.0974	0.4145
Richmond Fed Manufact. Index	0.3498	0.3933	0.2524	0.6794	0.4425	0.1311	0.4946	0.3196
S&P CoreLogic CS 20-City MoM SA	**0.0402**	**0.0216**	0.2456	**0.0105**	0.2524	0.1768	**0.0058**	0.2194
S&P CoreLogic CS 20-City YoY NSA	0.1227	0.1018	0.2628	0.4425	0.686	0.0598	0.2211	0.5328
S&P CoreLogic CS US HPI YoY NSA	0.2159	0.2603	0.1429	0.5714	0.8571	0.0751	0.448	0.871
Trade Balance	0.3588	0.2939	0.4294	0.686	0.1354	0.3291	0.6974	0.1359
U. of Mich. Current Conditions	0.8699	0.4701	0.879	0.9904	0.6491	0.6107	0.9854	0.4734
U. of Mich. Expectations	0.7509	0.6247	0.7299	0.6478	0.4458	0.7103	0.7167	0.34
U. of Mich. Sentiment	0.9769	0.8463	0.8456	0.8457	0.6905	0.8196	0.7684	0.5082
Unemployment Rate	0.1639	0.1696	**0.0463**	0.8588	0.4199	0.171	0.7705	0.162
Unit Labor Costs	0.532	0.5395	0.4892	0.8469	0.2057	0.4438	0.7722	0.2793
Wholesale Inventories MoM	0.6487	0.4644	0.8787	0.4998	0.8265	0.6047	0.5891	0.5095
Wholesale Trade Sales MoM	0.5499	0.5503	0.7863	0.2987	0.6074	0.7363	0.6027	0.4816

Figure 1 presents the results of the Tukey HSD test for China, the Eurozone, Germany and Japan. Analogously, Figure 2 refers to Poland, the United Kingdom and the United States. The groups have been assigned as follows: I – pre-crisis, II – crisis, III – post-crisis.

The Tukey HSD test provides more insightful information about differences in distributions than the ANOVA or Kruskall-Wallis tests. The Tukey HSD test does not only indicate whether there is a statistically significant difference between groups, but also helps to understand which pair of groups causes that result. For example, when we look at the chart for the Eurozone, we can prove that there is a statistically significant difference between

Table 10: P-values of Kruskal-Wallis, ANOVA, Kolmogorov-Smirnov and Anderson-Darling tests for Sentiment Indicator, per examined economies

Country	K.W	ANOVA	K.S_1	K.S_2	K.S_3	A.D_1	A.D_2	A.D_3
China	0.5209	0.5656	0.6741	0.5824	0.2989	0.5828	0.5068	0.3246
Eurozone	**0.0013**	**0.0164**	**0.0393**	0.7086	**0.0018**	**0.006**	0.3777	**0.0005**
Germany	0.064	0.6128	**0.0235**	0.1268	0.4016	**0.0117**	**0.0248**	0.1996
Japan	**0.042**	**0.0322**	**0.0131**	0.1422	0.2792	**0.0083**	0.2886	0.1829
Poland	0.0746	0.249	0.0613	0.3991	0.0597	0.0642	0.2792	0.0527
United Kingdom	**0.0098**	0.0636	**0.0252**	0.1407	0.1595	**0.0025**	0.0502	0.0891
United States	**0**	**0**	**0**	0.3678	**0.0018**	**0**	0.1575	**0.0003**

periods I/II and between periods II/III, while there is no evidence of a difference for pair I/III (this means that the difference in distributions of the sentiment indicator in periods I and III is statistically not significant).

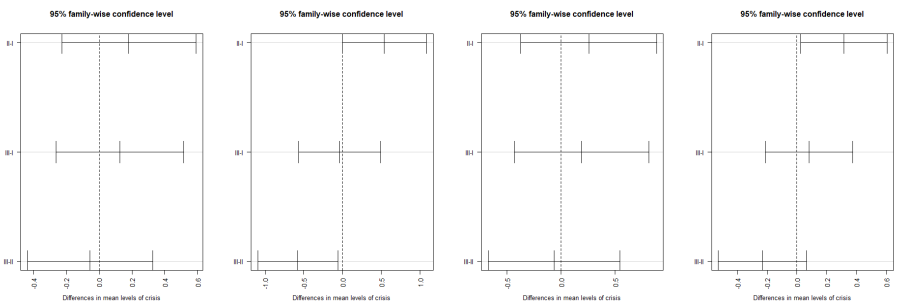


Figure 1: Visualization of the results of the Tukey HSD test for China, Eurozone, Germany and Japan

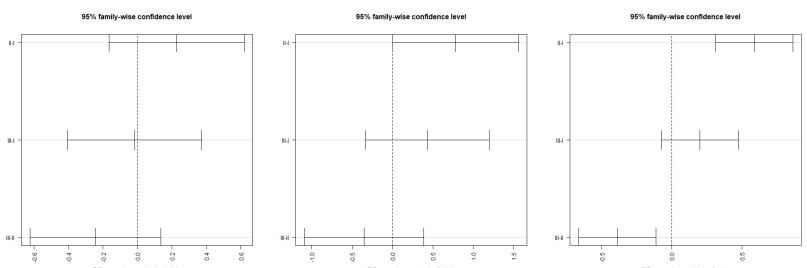


Figure 2: Visualization of the results of the Tukey HSD test for Poland, the United Kingdom and United States

The interpretation of the above visualizations of the Tukey HSD test for economies under study is left to the reader.

For all economies, excluding China and Poland, we obtained significant differences between the distributions in the periods under consideration (as per Table 10). In the case

of China, the breach of 5% significance is quite large, although for Poland the differences are only slightly greater than 5%.

As can be seen from columns $K-S\_1$ and $A-D\_1$, we should reject the null hypothesis for the same sentiment's indicator distributions before and during the crisis for the five economies considered. From the p-values in columns $K-S\_2$ and $A-D\_2$, it follows that the sentiment's indicator distributions before and after the crisis are not significantly different (except for Germany in the Anderson-Darling test). Lastly, columns $K-S\_3$ and $A-D\_3$ indicate that the differences in distribution of the sentiment's indicator during the crisis and post-crisis time also appear, but not as significantly as in the pre-crisis versus crisis period.

Therefore, the sentiment indicator seems to be a good proxy of the state of the economy. This explanation might follow from the fact that the increased uncertainty of economic forecasts, which is an attribute of the crisis period, makes the results of analysts' projections less reliable. In other words, surprises in macroeconomic data releases follow different distributions in crisis and non-crisis periods. This conclusion is strongly supported by the p-values of the Kolmogorov-Smirnov and Anderson-Darling tests reported above for the three periods under consideration.

As observed in the previous section, the fraction of macroeconomic indicators exhibiting significant distribution differences in the periods under consideration was moderate across economies concerned. However, aggregation of individual surprises into country level sentiment makes most of the distribution differences strongly pronounced.

The next, final section summarizes the results, concludes and indicates future directions for research.

6. Conclusions

A striking observation based on the empirical results is the relatively small percentage of macroeconomic indicators exhibiting statistically significant differences in the distributions of the surprise indicator across all economic areas examined, taking into account the proposed division of the data sample into three sub-intervals (pre-pandemic, pandemic and post-pandemic). Strictly speaking, only 1 out of 26 indicators for China, 2 out of 26 for the Eurozone, 1 out of 33 for Germany, 4 out of 57 for Japan, 2 out of 26 for Poland, 8 out 59 for the United Kingdom and 11 out of 101 for the United States.

However, when we consider the sentiment indicator defined in Section 5, we obtain a different picture. The economic areas for which we found a statistically significant difference with respect to the sentiment indicator in the three sub-periods are the United States, the Eurozone, the United Kingdom and Japan. Only China exhibits an insignificant difference in the sentiment indicator across these periods. In the case of Germany and Poland, the Kruskall-Wallis and ANOVA tests do not coincide, while the Kruskal-Wallis test is almost significant in both cases and ANOVA is definitely not significant. At the first glance, it might seem counter-intuitive that the aggregated approach leads to detecting more significant differences in distributions. However, in the case of the vast majority of the statical tests, the power of the test as the function of the number of observations is strictly increasing. Furthermore, using the relevance factor as the weighting coefficient guarantees that

more important economic announcements are more impactful on the sentiment indicator's value than minor releases.

To the best of our knowledge, the methodology of the aggregated sentiment, where the relevance factors are used as the weighting coefficients has been applied for the first time. The second novelty of the paper is using the intraday data to calculate the sentiment.

The above-mentioned conclusions are also supported by the results of the Kolmogorov-Smirnov and Anderson-Darling tests, which were performed separately for individual macroeconomic indicators and aggregated sentiment. Therefore, the sentiment indicator seems to be more sensitive to the economic crisis, in comparison with the values of the surprise indicator per single category of macroeconomic data announcements.

The next observation based on the Tukey HSD test is that the pre-pandemic and post-pandemic periods with respect to the sentiment indicator are mostly convergent, while in most cases both significantly differ from the pandemic period.

We found no evidence that the importance of the macroeconomic data release, measured by the relevance indicator, impacts the distribution of the surprise indicator. Therefore, it is quite improbable that financial analysts' forecasts for the most important categories of macroeconomic indicators are of better quality than for the minor ones - neither the mean nor variance of the surprise indicator seems to be sensitive to the relevance factor.

As far as the direction of future research is concerned, we plan to examine the reaction of various asset classes (for example stocks, equity indices, bonds, foreign exchange rates, commodity prices, etc.) to macroeconomic data releases with a strong emphasis on the explanatory power of the sentiment indicator in selected periods to determine trends in prices and the volatility of securities. Intuitively, the higher the absolute value of the sentiment, the more pronounced the price, volatility and trading volume movements should be. Furthermore, the sign (positive or negative) of the sentiment indicator should explain the direction of the metric movements of the securities, mentioned above. We expect a sharper impact for negative surprises (in other words, sentiment indicator values below 0).

Acknowledgement

The authors are grateful to two anonymous reviewers for their valuable comments and suggestions made on the previous draft of this manuscript.

Funding

Krzysztof Brania and Henryk Gurgul were financed by AGH University of Krakow (institutional subsidy for maintaining Research Capacity Grant 16.16.200/396.).

References

Andersen, T., Bollerslev, T., Diebold, F. and Vega, C., (2007). Real-time price discovery in global stock, bond and foreign exchange markets. *Journal of International Economics*, 73, pp. 251–277.

Boyd, J. H., Hu, J. and Jagannathan, R., (2005). The stock market's reaction to unemployment news: why bad news is usually good for stocks. *Journal of Finance*, 60, pp. 649–672.

Dimpfl, T., (2011). The impact of US news on the German stock market — An event study analysis. *The Quarterly Review of Economics and Finance*, 51(4), pp. 389–398.

Gurgul, H., Suliga, M. and Wójtowicz, T., (2012). Responses of the Warsaw Stock Exchange to the U.S. macroeconomic data announcements. *Managerial Economics*, 12, pp. 41–60.

Gurgul, H., Suliga, M. and Wójtowicz, T., (2013). The reaction of intraday WIG returns to the U.S. macroeconomic news announcements. *Quantitative Methods in Economics*, 14(1), pp. 150–159.

Gurgul, H., Wójtowicz, T., (2014). The impact of US macroeconomic news on the Polish stock market. The importance of company size to information flow. *Central European Journal of Operations Research*, 22, pp. 795–817.

Gurgul, H., Wójtowicz, T., (2015). The response of intraday ATX returns to U.S. macroeconomic news. *Finance a úvěr - Czech Journal of Economics and Finance*, 65(3), pp. 230–253.

Gurgul, H., Wójtowicz, T., (2020). Wpływ informacji makroekonomicznych na transakcje na rynkach akcji. CH Beck. Warszawa.

Hanousek, J., Kočenda, E. and Kutan, A. M., (2009). The Reaction of Asset Prices to Macroeconomic Announcements in New EU Markets: Evidence from Intraday Data. *Journal of Financial Stability*, 5(2), pp. 199–219.

Hanousek, J., Kočenda, E., (2011). Foreign News and Spillovers in Emerging European Stock Markets. *Review of International Economics*, 19(1), pp. 170–188.

Harju, K., Hussain, S. M., (2011). Intraday seasonalities and macroeconomic news announcements. *European Financial Management*, 17, pp. 367–390.

Heinlein, R., Lepori, G. M., (2022). Do financial markets respond to macroeconomic surprises? Evidence from the UK. *Empirical Economics*, 62, pp. 2329–2371.

Jäggi, A., Schlegel, M., and Zanetti, A., (2019). Macroeconomic surprises, market environment, and safe-haven currencies. *Swiss Journal of Economics and Statistics*, 155(5), pp. 1–21.

Li, L., Hu, Z. F., (1998). Responses of stock markets to macroeconomic announcements across economic states. *IMF Working Paper*, No 79.

Nikkinen, J., Sahlström, P., (2004). Scheduled Domestic and US Macroeconomic News and Stock Valuation in Europe. *Journal of Multinational Financial Management*, 14, pp. 201–245.

Nikkinen, J., Omran, M., Sahlström, M. and Äijö, A., (2006). Global stock market reactions to scheduled U.S. macroeconomic news announcements. *Global Finance Journal*, 17(1), pp. 92–104.

Pearce, D. K., Roley, V. V., (1983). The reaction of stock prices to unanticipated changes in money: a note. *Journal of Finance*, 38, pp. 1323–1333.

Schwert, G. W., (1981). Measuring the effects of regulation: evidence from the capital markets. *Journal of Law and Economics*, 24, pp. 121–145.

STATISTICS IN TRANSITION new series, September 2025
Vol. 26, No. 3, pp. 99–115, https://doi.org/10.59139/stattrans-2025-029
Received – 02.03.2021; accepted – 11.12.2024

Bayesian and frequentist modelling of West African economic growth: a dynamic panel approach

Nureni Olawale Adeboye[1], Olumide Sunday Adesina[2]

Abstract

The empirical outcomes of previous studies examining the relationship between economic growth and socio-economic indicators have been inconclusive and conflicting. To further probe into the study area, the current research employed a dynamic panel model estimated via three robust dynamic panel data estimators of the generalized method of moment (GMM), frequentist instrumental variable (IV) and the Bayesian IV on real and simulated data. Various model performance criteria such as Wald statistics, leave-out-one cross-validation and the Pareto *k* checks were used for validity verification. The results of the robust diagnostics checks and a model strength metric showed that the family of IV models outperformed the GMM. Thus, the estimation provided by the Bayesian IV is upheld and recommended in modelling dynamic panel data as it provides robust estimates of the parameters of interest.

Key words: dynamic panel data, economic growth, generalized method of moment, instrumental variable, socio-economic indicators.

1. Introduction

National economic development alludes to an expansion in the total efficiency of a nation or landmass. It is the amount more the economy produces than it did in the earlier period. To be exact, the correlation should eliminate the impacts of expansion (Becsi and Wang, 2002). Financial development is the advancement of Total national output (Gross domestic product) in the short, medium and long haul. It is the aftereffect of an expansion in esteem added delivered by every one of the organizations working inside a country. The increment in the worth added during a given period implies that the worldwide abundance of a country is rising and this shows itself in the development of per capita income and in a more significant level of prosperity.

[1] Corresponding Author. Department of Statistics, Faculty of Basic and Applied Sciences, Osun State University, Osogbo, Osun State, Nigeria. E-mail: nureni.adeboye@uniosun.edu.ng. ORCID: https://orcid.org/0000-0002-8023-221X.

[2] School of Computing and Sciences, Nigerian University of Technology and Management, Apapa, Lagos, Nigeria & Johannesburg Business School, University of Johannesburg, South Africa. E-mail: adesinao@uj.ac.za. ORCID: https://orcid.org/0000-0003-1476-8080.

A wide scope of studies has explored the variables fundamental to economic development utilizing varying calculated and strategic perspectives, these investigations have set accentuation on an alternate arrangement of informative boundaries and offered different bits of knowledge to the wellsprings of economic development (Lensink and Morrissey, 2006). Venture is the most major determinant of economic development distinguished in the literature. The significance appended to speculation has prompted a colossal measure of experimental investigations analyzing the connection among venture and economic development (Artelaris *et al.,* 2007). It is additionally conceivable to accomplish total economic development without an expanded normal negligible efficiency yet through additional immigrants or higher rates of birth (Obadan, 2006).

Basu *et al.* (2005) noticed that Africa is the world's least fortunate continent. Various nations have as of late arose out of common conflicts that have seriously interfered with their formative endeavors while in different pieces of the continent, new outfitted struggles have erupted. These contentions and other antagonistic factors, outstandingly helpless climate conditions and crumbling as far as exchange, have prompted misfortune in monetary energy in the district in the course of the most recent twenty years. The authors recommend that what is required is a maintained and a considerable expansion in genuine per capita Gross domestic product development rates in these nations, combined with huge enhancements in friendly conditions. Endeavor to appraise the African mainland development is dependent on its Gross domestic product advancement and resident's buying equality. The monetary and social circumstance in sub-Saharan Africa accordingly stays delicate and defenseless against homegrown and outside shocks. Speculation stays curbed, restricting endeavors to broaden financial designs and lift development (Nkurunziza and Bates, 2004). This is in sharp difference to the happenings in the OECD nations where expansion in reserve funds and venture rate lead to economic development (Becsi and Wang, 2002).

Hu *et al.* (2014) suggest a generalized method of moment with individual specific fixed and threshold effects simultaneously. The issue of endogeneity in GMM was resolved by confirming that the symmetry conditions proposed by Arellano and Bond (1991) are legitimate. The proposed GMM estimator shows that the edge and incline boundary can be assessed precisely with consistency, and furthermore the finite sample dissemination of slant boundaries is well approximated by the asymptotic distribution (Blundell and Bond, 2000; Al-Sadoon et al., 2019).

Bardi et al. (2016) established empirically a positive and critical connection between structural policy and economic development utilizing a generalized moment method developed within dynamic panel structure. Sharma (2018) equally employs generalized method of moment estimator to re-examine wellbeing development relationship utilizing an unequal panel of 17 developed economies. The estimator takes care of endogeneity issues and through alternate model specifications it was established

that population apply a positive and critical impact on both genuine income per capita just as development.

A few other authors who have written extensively on the estimation of economic growth both in Africa and globally as well as the practical application of GMM technique in modelling dynamic panel data are Lichtenberg (1992); Kiviet (1995); Blundell and Bond (1998); Agiomirgianakis *et al.* (2002); Ajayi (2003); Bengoa and Sanchez-Robles (2003); Agbeyegbe (2006); Obadan (2006); Lensink and Morrissey (2006); Dreher (2006); Levina (2011); Meraj (2013) and Adeboye *et al.* (2023). While GMM estimators depends strongly on the ratio of variance of the individual-specific effect and the variance of the general error term (see, e.g. Bun and Carree 2005), the IV largely depends on their individual specific effects that are uncorrelated with the explanatory variables x_{it}. A recent technique with limited approach in the literature is the Bayesian inference, which provides robust estimates of parameter of interest given because it involves updating the information based on prior statistics (Adesina and Obokoh, 2024). Limited studies have employed Bayesian statistics especially in recent times to estimate the parameters of interest in panel data. Some of the studies include Cho and Zheng (2021).

Dynamic panel estimation techniques were employed to establish the econometric bond between the selected macro-economic indicators of economic growth and purchasing power parity (PPP) across West African countries so that we can examine some desirable implications. Panel data has been established in the literature as all encompassing, in the areas of economic analysis [see the work of Adeboye and Agunbiade, 2019a and Adeboye and Agunbiade, 2019b]. Dynamic panel data estimation includes the work of Li et al. (2021) and Jin et al. (2021) who provided GMM estimation for dynamic panel models. The aim is to estimate the economic panel data with classical and Bayesian models using two-stage Least Square (2SLS) instrumental variable technique and compare with the GMM estimator proposed by Hu *et al.* (2014) to determine the approach that will provide the best estimates for dynamic panel. The adopted variables of measurement to validate the position of Basu *et al.* (2005) on the estimation of African continent growth are based on its GDP evolution and citizen's purchasing parity. The remaining part of the paper comprises Section 2, the material and methods, the results are presented in Section 3, and finally, Section 4 the provides conclusion.

2. Materials and Methods

The life data utilized were obtained mainly from UNESCO data site, which covers a period of 10 years ranging from 2008-2017 for selected West African countries as retrieved in the year 2018 while Monte Carlo simulation scheme was carried out using a data-generating procedure specified within a dynamic panel data model.

2.1. Model Specification

The relational model for this study is specified as

$$lGDP_{it} = \beta_{0it} + \beta_1(lPPP)_{1,it} + \beta_2(lGNI)_{2,it} + e_{it} \quad (1)$$

where i and t indicate the cross-sectional units (countries) and years under consideration respectively. GDP is the gross domestic product, PPP is the purchasing power parity and GNI is the gross national income of the West African countries while $\boldsymbol{e_{it}}$ is the unestimated residual. Considering the fact that the countries are diverse, a panel unit root test was carried out on the variables through the adoption of IPS (2003) test for individual unit root process given as

$$\Delta y_{it} = \rho_{it} y_{i,t-1} + \sum_{L=1}^{p_i} \emptyset_{iL} \Delta y_{i,t-L} + z_{it}^{/} y + u_{it} \quad (2)$$

The LPS test is based on the assumption that the unit root can differ across the cross-sectional units in the model.

2.2. Estimation Methods

Two dynamic panel data estimation methods of the generalized method of moment and instrumental variable were employed. GMM was estimated according to the Arellano and Bond approach having fully taken care of endogeneity phenomenon and IV estimated via a 2sls technique.

2.2.1 Generalized Method of Moment (GMM)

Considering the first order model

$$y_{it} = X_{it}\beta + \delta y_{i,t-1} + \alpha_i + \varepsilon_{it} \quad (3)$$

and adopting the principle established by Hu *et al.* (2014), the first difference equation of (3) was observed to get rid of constant time of individual effects as

$$\Delta\, y_{it} = \Delta X_{it}(\beta)'\, \alpha_i + \Delta y_{i,t-1}(\delta)' + \Delta\varepsilon_{it} \quad (4)$$

$$y_{it} - y_{i,t-1} = \propto_i (x_{it} - x_{i,t-1})\beta' + (y_{i,t-1} - y_{i,t-2})\, \delta' + (\varepsilon_{it} - \varepsilon_{i,t-1}) \quad (5)$$

Considering that $\alpha_1 \neq \alpha_2$, it was established that the orthogonality conditions that exist between lagged values of y_{it} and the residual term ε_{it} are also valid in model (5). And without loss of generality, for any given t

$$X_{it}(\beta) = (x_{it}, 0)' or\ X_{it}(\beta) = (0. x_{it})' \quad (6)$$

And the first difference yields

$$\Delta X_{it}(\beta) = X_{i,t} - X_{i,t-1} \quad (7)$$

It should be noted that $\mathbf{X_{i,t-1}}$ satisfies the conditions of endogeneity and $\boldsymbol{\varepsilon_{it}}'\boldsymbol{s}$ are serially uncorrelated. Thus, the orthogonality conditions are given by

$$E\left(X_{i,t-s}\, \Delta\varepsilon_{it}\right) = 0, \quad for\ s = 1, \cdots, t-1; t = 2, \cdots, T$$

Thus, within T observations in group i, Arellano and Bond (1995) suggested the fact that

$$\mathbf{E}\left[\begin{pmatrix} x_{1it} \\ x_{2it} \\ z_{1i} \\ \bar{X}_{1i} \end{pmatrix} (\eta_{it} - \bar{\eta}_{it})\right] = \mathbf{0} \quad \text{for some s} \neq \text{t.} \tag{8}$$

In principle, each valid instrument is extrinsic with respect to η_{it} subject to current, lagged, and future periods. Thus, there are a total of $[T(C1 + C2) + D1 + C1)]$ moment conditions for every observation.

$$\textbf{Let } W_i = \begin{pmatrix} w'_{i1} \\ w'_{i2} \\ \vdots \\ w'_{iT} \end{pmatrix} \quad \text{and} \quad y_i = \begin{pmatrix} y'_{i1} \\ y'_{i2} \\ \vdots \\ y'_{iT} \end{pmatrix} \tag{9}$$

W_i is assumed to be a $T\times$ $(1+ C1 + C2 + D1 + D2)$ matrix and T +1 observations available on y_i. Considering a matrix V_i consisting of $Ti - 1$ rows with instrument $\mathbf{v}'_{it}$ given as

$$V_i = \begin{bmatrix} v'_{i1} & 0' & \cdots & 0' \\ \vdots & v'_{i1} & \ddots & \vdots \\ 0' & \cdots & & a'_i \end{bmatrix} \tag{10}$$

Considering the transformation matrix, H, constructed as

$$\mathbf{H} = \begin{pmatrix} M^{01} \\ T^{-1}\iota' T \end{pmatrix} \tag{11}$$

where M^{01} denotes the first T −1 rows of the matrix (M^0) that creates deviations from group means. Thus, H replaces the last row of M^0 with a row of T^{-1}.

Let the $T \times 1$ column vector of disturbances be represented as

$$\eta_i = [\eta_{i1}, \eta_{i2}, \ldots, \eta_{iT}] = [(\varepsilon_{i1} + u_i), (\varepsilon_{i2} + u_i), \ldots, (\varepsilon_{iT} + u_i)]', \tag{12}$$

then

$$H_\eta = \begin{pmatrix} \eta_{i1} - \bar{\eta}_i \\ \vdots \\ \eta_{i,T-1}\bar{\eta}_i \\ \bar{\eta}_i \end{pmatrix}$$

$$E[V' H_{\eta i}] = E[g_i] = 0. \tag{13}$$

The moment condition that follows from (13) is given as

$$plim\, n^{-1} \sum_{i=1}^{n} V_i' H_{\eta_i} \tag{14}$$

Explicitly given as

$$plim\ n^{-1} \sum_{i=1}^{n} V_i' H \begin{bmatrix} y_{i1} - \delta y_{i0} - x'_{1i1}\beta_1 - x'_{2i1}\beta_2 - z'_{1i}\alpha_1 - z'_{2i}\alpha_2 \\ y_{i2} - \delta y_{i1} - x'_{1i2}\beta_1 - x'_{2i2}\beta_2 - z'_{1i}\alpha_1 - z'_{2i}\alpha_2 \\ \vdots \\ y_{iT} - \delta y_{i,T-1} - x'_{1iT}\beta_1 - x'_{2iT}\beta_2 - z'_{1i}\alpha_1 - z'_{2i}\alpha_2 \end{bmatrix} \tag{15}$$

$$= plim\ n^{-1} \sum_{i=1}^{n} m_i = plim \bar{m} \tag{16}$$

Then the GMM estimator $\hat{\theta}$ is obtained by minimizing

$$q_{it} = \bar{m}^{/} A \bar{m} \tag{17}$$

The best weighting result of matrix A is derived as the inverse of the asymptotic covariance matrix of $\sqrt{n}\bar{m}$ and the solution to the minimizing problem of q_{it} with respect to the parameter vector $\boldsymbol{\theta}$ is the GMM estimator given as

$$\hat{\theta}_{GMM} = [(\sum_{i=1}^{n} W_i^{/} HV_i)(\sum_{i=1}^{n} V_i^{/} H\eta_i\eta_i^{/} H^{/} V_i)^{-1}(V_i^{/} H^{/} W_i)]^{-1} \\ [(\sum_{i=1}^{n} W_i^{/} HV_i)(\sum_{i=1}^{n} V_i^{/} H\eta_i\eta_i^{/} H^{/} V_i)^{-1}(V_i^{/} H^{/} W_i)] \tag{18}$$

2.2.2. Instrumental Variable (IV)

This is used to estimate causal relationships when controlled experiments are not feasible. Going by equation (3), the first order model becomes

$$y_{it} = x_{it}^{/}\boldsymbol{\beta} + z_i^{/}\boldsymbol{\alpha} + \varepsilon_{it}. \tag{19}$$

The underlying assumption of equation (19) clearly specifies that individual specific effects z_i are uncorrelated with the explanatory variables x_{it}. Thus, the model becomes

$$y_{it} = \mathbf{x}^{/}_{1i\,it}\boldsymbol{\beta}_1 + \mathbf{x}^{/}_{2i\,it}\boldsymbol{\beta}_2 + \mathbf{z}^{/}_{1i}\boldsymbol{\alpha}1 + \mathbf{z}^{/}_{2i}\boldsymbol{\alpha}_2 + \varepsilon_{it} + u_i \tag{20}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ and $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1\ \boldsymbol{\alpha}_2)$.

The strategy for estimation involved deviations of group means to have

$$y_{it} - \bar{y}_i = (\mathrm{x}_{1it} - \bar{x}_{1i})^{/}\beta_1 + (\mathrm{x}_{2it} - \bar{x}_{2i})^{/}\beta_2 + \varepsilon_{it} - \bar{\varepsilon}_i. \tag{21}$$

Representing the model variables as a weighted instrument given as

$$w_{it}^{/} = (x_{1it}^{/}, x_{2it}^{/}, z_{1i}^{/}, z_{2i}^{\prime}). \tag{22}$$

The transformed variables of equation (22) becomes

$$w_{it}^{*/} = w_{it}^{/} - \hat{\theta}\bar{w}_i^{/} \text{ and } y_{it}^{*} = y_{it}^{/} - \hat{\theta}\bar{y}_i^{/} \tag{23}$$

where $\hat{\theta}$ is a BLUE of θ. Thus, instrumental variables are given as

$$v_{it}^{/} = [(\mathrm{x}_{1it} - \bar{x}_{1i})^{/} + (\mathrm{x}_{2it} - \bar{x}_{2i})^{/} + z_{1t} - \bar{x}_{1t}\ ] \tag{24}$$

And these are pile up in the rows of matrix $nTx(C1 + C2) + D1 + C1)$ denoted as **V**. The time-invariant variables and group means are repeated for the 3rd and 4th set of instruments, and the instrumental variable estimator becomes

$$(\beta\alpha)_{iv} = [(\mathrm{W}^{*/}\mathrm{V})(\mathrm{V}^{/}\mathrm{V})^{-1}(\mathrm{V}^{/}\mathrm{W}^{*})]^{-1}[(\mathrm{W}^{*/}\mathrm{V})(\mathrm{V}^{/}\mathrm{V})^{-1}(\mathrm{V}^{/}\mathrm{y}^{*})]. \tag{25}$$

The Bayesian alternative to the frequentist IV is presented in Section 2.3.

2.3. The Bayesian Implementation for the Instrumental Variable (IV)

Beyond the frequentist IV, the Bayesian IV was conducted based on multilevel approach. The Bayesian statistics involves the combination of likelihood and the prior distribution to obtain another distribution known as posterior distribution.

2.3.1. Prior Distributions and Sampling procedure

Prior distribution at group level assumed that parameters of interest come from a multivariate normal distribution having zero mean and unknown covariance matrix $\boldsymbol{\Sigma}$.

$$\epsilon \sim N(0, \boldsymbol{\Sigma}) \tag{26}$$

Covariances between group-level parameters are generally of different groupings factors and assumed to be zero. The model can be simplified to

$$\epsilon_i \sim N(0, \boldsymbol{\Sigma_i}) \tag{27}$$

where i indexes grouping factors. In cases where there are different levels with additional level indexed by j and the grouping factors are not dependent, Eq. (27) leads to:

$$\epsilon_{ij} \sim N\left(0, \mathbf{M_j}\right) \tag{28}$$

The model parameters will result from the covariance matrices $\mathbf{M_j}$, and No-U-Turn Sampler (NUTS) to sample $\mathbf{M_j}$ as recommended by Hoffman and Gelman (2014). The parameters of $\boldsymbol{M}_j$ are selected in terms of correlation matrix $\boldsymbol{\Omega_j}$ and a vector of standard deviations σ_j through

$$\boldsymbol{M}_j = \boldsymbol{D}\left(\sigma_j\right)\boldsymbol{\Omega_j}\boldsymbol{D}\left(\sigma_j\right) \tag{29}$$

2.3.2. The Sampling and Diagnostics Checks

The sampling method is the NUTS Sampler. NUTS is an extended Hamiltonian Monte-Carlo (HMC) which allows setting parameters and eliminates the need for hand-tuning Hoffman and Gelman (2014). Software package by R core team (2024) was used to fit the model with brms package by Bürkner (2017), which uses stan processor. Diagnostic plots for acceptance of NUTS plots were conducted, Adesina (2021) has the details of the procedure.

The study adopted the Leave-one-out cross-validation (LOO-CV) for the diagnostic tests. In Bayesian analysis, the data are repeatedly subdivided into a training set y_{train} and a holdout set $y_{holdout}$ with the objective of fitting y_{train} yielding a posterior distribution

$$p_{train}(\theta) = p_{train}(\theta | y_{train}) \tag{30}$$

The Bayesian LOO-CV estimate of out-of-sample predictive fit is

$$pd_{loocv} = \sum_{i=1}^{n} log\, p_{post(-i)}\,(y_i) \tag{31}$$

and estimated as

$$\sum_{i=1}^{n} log\left(\frac{1}{S}\sum_{s=1}^{S} log\, p\,(y_i|\theta^{is})\right) \tag{32}$$

To compare between two or more models the lowest LOO suggests better model fit.

The k Pareto also assesses the reliability and approximate convergence rate of the Pareto smoothed importance sampling (PSIS). It follows that if $k < 0.5$ ('good') then the central limit theorem holds. If $0.5 \leq k < 1$, ('ok') then the variance of the raw importance ratios is infinite, but the mean exists. On the other hand, if $k > 0.7$ ('bad'), unreasonable convergence rates are observed and unreliable Monte Carlo error estimates, and finally, if $k \geq 1$ ('very bad'), the variance and the mean of the raw importance ratios does not exist.

2.4 Monte Carlo Simulation Scheme

Monte Carlo simulations method was used to generate alternative data necessary for fitting and validation of the suitability of the proposed economic growth model. According to Hu *et al.* (2014), the data generating procedure (DGP) is given by

$$y_{it} = \delta y_{i,t-1} + \mathrm{x}'_{1it}\beta_1 + \mathrm{x}'_{2it}\beta_2 + \mathrm{z}'_{1i}\alpha_1 + \mathrm{z}'_{2i}\alpha_2 + \varepsilon_{it} + u_{it} \quad (33)$$

$for\ i = 1, \cdots, N\ and\ t = 1, \cdots, T\ where\ \varepsilon_{it} \sim i.i.d.N(0,1), \quad u_{it} \sim i.i.d.N(0,1),$
$\alpha_1 = \alpha_2 = 0. z_{it,}\ u_{it}, \varepsilon_{it}$ are mutually independent random variables.

The design of Monte Carlo simulations was carried out to further examine both the effectiveness and finite sample properties of different estimators of parameter α. The cross-sectional units are as small as 20 while $T = 10$ is the largest time dimension used in the study. A balanced panel data was first simulated and the data was made dynamic by the deletion of 2nd time period (time 4) for all individuals. It was assumed rho and alpha are 0, while the parameters used are uniformly distributed.

3. Results and Discussions

The results of both real life and simulated data are presented in the following tables:

Table 1: Results of Generalized method of Moments (GMM)

	Real Life Data		Simulated Data	
Economic Growth Indicators	**GMM (One step) Parameter Estimate**		**GMM (One step) Parameter Estimate**	
l_GDP(-1)	0.936024	(0.0001)	0.0727352	(0.9891)
Constant	–0.00519	(0.9951)	-0.0088614	(0.8455)
l_PPP	0.32395	(0.1575)	-0.0237805	(0.7934)
l_GNI	0.04064	(0.4341)	0.0016380	(0.9269)

Table 2: GMM Model Diagnostic

	Real Life Data		Simulated Data	
Test	**Estimates**		**Estimates**	
Test for AR(1) errors	-1.33091	(0.1832)	-0.7451126	(0.4562)
Test for AR(2) errors	-0.47378	(0.6357)	0.1865387	(0.85202)
Sargan over-identification test	87.3077	(0.0001)	14.40015	(1.0000)
Wald (joint) test	6298.11	(0.0000)	0.1324148	(0.9979)

Table 3: Results of Instrumental Variable

Real Life Data			Simulated Data		
Indicators	**2SLS Estimates**		**Indicators**	**2SLS Estimates**	
Constant	13.1335	(4.42e-112)***	Constant	12.47	(2e-16)***
PPP	-3.63842	(0.0238)**	(gdp,1)	0.000001408	(2e-16)***
GNI	5.6721	(7.36e-06)*	PPP	-0.00007405	(0.767)
			GNI	0.00003354	(0.550)

Table 4: IV Model Diagnostic

Real Life Data			Simulated Data	
Test	**Estimates**		**Estimates**	
F- statistic	219.195	(0.0000)	823.615	(2e-16)***
Wald (joint) test	18.5464	(0.0001)	226900	(2.2e-16)***
R^2	0.69433		0.9997	

Note that the P-values are in parenthesis.

GMM and IV models specified from Tables 1 and 3 are given as

$$l_{GDP_{it}} = -0.00519 + \left(0.93602_{GDP(it-1)}\right) + 0.32395(l_{PPP})_{1,it} + 0.04064(l_{GNI})_{2,it} \quad (34)$$

$$l_{GDP_{it}} = -0.00886 + \left(0.0727_{GDP(it-1)}\right) - 0.0237(l_{PPP})_{1,it} + 0.0406423(l\_GNI)_{2,it} \quad (35)$$

$$GDP_{it} = 13.1335 - 3{,}63842(PPP)_{1,it} + 5.6721(GNI)_{2,it} \quad (36)$$

$$GDP_{lt} = 12.47 + 0.000001408(GDP_{-1}) - 0.00007405(PPP)_{1,it} + 0.00003354(GNI)_{2,it} \quad (37)$$

Models (34) – (37) represent the empirical growth models estimated from both real life and simulated data. It is pertinent to note that models from the GMM technique give negative projections of African economic growth at constant values of the predictors, despite the absence of exogeneity while that of IV give positive projections with a more superior significant values as presented in Tables 2 and 4. Thus, the model in which its explanatory variables are more significant with improved validity checks is that of the instrumental variable.

The validity checks further revealed the absence of serial correlation among the variables due to the results of AR(1) and AR(2) while the Sargan test validates the instrumental variables. Similarly, the results reported in Table 3 shows that this instrument can be considered as exogenous given that the null hypothesis is not rejected at both 1% and 5% percent level, as posited by Bascle (2008). The other two macroeconomic instruments were individually and simultaneously tested for exogeneity to increase our confidence that both instruments can be considered as exogenous in this setting. Table 5 contains the estimates based on Bayesian Multilevel IV model.

Table 5: Bayesian Multilevel IV model

Specification	Estimate	Est.Error	l-95%CI	u-95%CI	$\widehat{R}$	Bulk ESS	Tail ESS
sd(GDP_Intercept)	181914.25	109221.3	5516.60	409371.7	1.01	126	185
sd(GDP_GNI)	0.00	0.00	0.00	0.00	1.00	231	440
sd(GDP_PPP)	316677.86	214973.8	7218.73	66169.88	1.01	109	154
sd(logGDP_Int.)	2.60	0.48	1.83	3.71	1.00	282	343
sd(logGDP_PPP)	1.19	0.55	0.21	2.41	1.00	387	412
sd(logGDP_GNI)	0.00	0.00	0.00	0.00	1.00	262	306
cor(GDP_Int, GDP_GNI)	0.02	0.47	-0.83	0.87	1.02	85	221
cor(GDP_Int, GDP_PPP)	0.05	0.49	-0.83	0.88	1.00	239	457
cor(GDP_GNI, GDP_PPP)	-0.02	0.48	-0.88	0.80	1.00	446	473
cor(logGDP_Int, logGDP_PPP)	-0.57	0.28	-0.95	0.08	1.00	534	383
cor(logGDP_Int, logGDP_GNI)	0.31	0.51	-0.78	0.95	1.00	175	332
cor(logGDP_PPP, logGDP_GNI)	-0.04	0.49	-0.86	0.86	1.00	253	425

NB: Int-Intercept.

There are two models in Table 5, the standard deviation estimate model, and the correlation model. The estimates are provided in the second column. The estimation error in the third column, the upper and lower 96% confidence interval in the fourth and fifth column. The Rhat ($\hat{R}$) in the sixth column which serves as potential scale reduction factor on split chains. The Bulk ESS, and Tail ESS in the seventh and eight column. The Bulk ESS is a diagnostic test to determine sampling efficiency while Tail ESS is used to determine the sampling efficiency in the tails of the posterior respectively.

The Bayesian multilevel IV model based on Table 5 can be expressed in terms of random intercepts and random slope, correlations between predictors and Bayesian priors as given in equations (38) and (39) below, which represent the models with standard deviation and correlation respectively:

$$GDP_{it} = 181914.25 + 0.000(GPD\_GNI)_{(it)} + 316677.86(GDP\_PPP)_{it} + 2.60(GDP\_Intrecept)_{it} + 1.19(logGDP\_PPP)_{it} + 0.000(logGPD\_GNI)_{(it)} \quad (38)$$

$$GDP_{it} = 0.02(GPD\_Int)(GDP\_GNI)_{(it)} + 0.05(GPD\_Int)(GDP\_PPP)_{(it)} - 0.02(GPD\_GNI)(GDP\_PPP)_{(it)} - 0.57(logGDP\_Int)(logGDP\_PPP)_{it} + 0.31(logGPD\_Int)(logGDP\_GNI)_{it} - 0.04(logGDP\_PPP)(logGDP\_GNI)_{it} \quad (39)$$

The models predicted GDP using several predictors ($GDP\_GNI, GDP\_PPP,$ log$GPD_GNI, logGPD_GNI$ *and* $logGPD\_Int$). The random intercept estimated in equation (38) accounts for the variation across units and it is assumed to follow a normal distribution with a standard deviation of 181914.25, which indicates substantial variation in GDP across countries. The zero variation between GDP and GNI implies that the countries' GNI has impacted favorably on the GDP without any variation while that of PPP suggests a large variation in the GDP of countries as occasioned by the countries' PPP. According to the correlation estimates provided in equation (39), the predictors are correlated with each other at different degrees, with correlations close to zero, suggesting little association between variables.

The estimated standard errors of the model as contained in the column 3 of Table 5 are negligible except for that of the interaction between GDP and PPP. This implies that the estimates of all other predictors are reliable with negligible uncertainty in their estimation. This was supported with credible intervals provided for all the predictors, with intervals which include zero except for that of PPP mentioned earlier. All the $\hat{R}$ are greater than 1.00 in all the cases indicating a good convergence for the Markov Chain Monte Carlo (MCMC) chains, with high values of ESS (ESS > 100), which suggests that the estimates are reliable and that the posterior distribution has been adequately sampled. This opinion is in tune with the work of Bürkner (2017) and Jiménez et al. (2022).

Table 6 contains the estimates of the response variables based on the intercept.

Table 6: Comparison of Intercept Estimates

Specification	Estimate	Est. Error	l-95% CI	u-95% CI	$\hat{R}$	Bulk ESS	Tail ESS
GDP_Intercept	4702.15	3421.77	-2638.15	10147.91	1.06	21	59
logGDP_Intercept	11.86	0.64	10.53	13.10	1.01	94	226

From Table 6, the regression estimate of the $(GDP)_{int}$ = 4702.15, zero included, while $(logGDP)_{int}$ = 11.86 is significant. The Rhat is close to 1 in both cases, which shows that the chain converged. The Bulk ESS and Tail ESS of $(GDP)_{int}$ are less than 100, whereas Tail ESS of $(log\ GDP)_{int}$ is greater than 100 showing that there is efficient sampling in the tails of the posterior distribution.

Draws were sampled using sampling (NUTS). For each parameter, Bulk ESS and Tail ESS are effective sample size measures, and Rhat is the potential scale reduction factor on split chains (at convergence, Rhat = 1). Table 7 shows the posetrior summary.

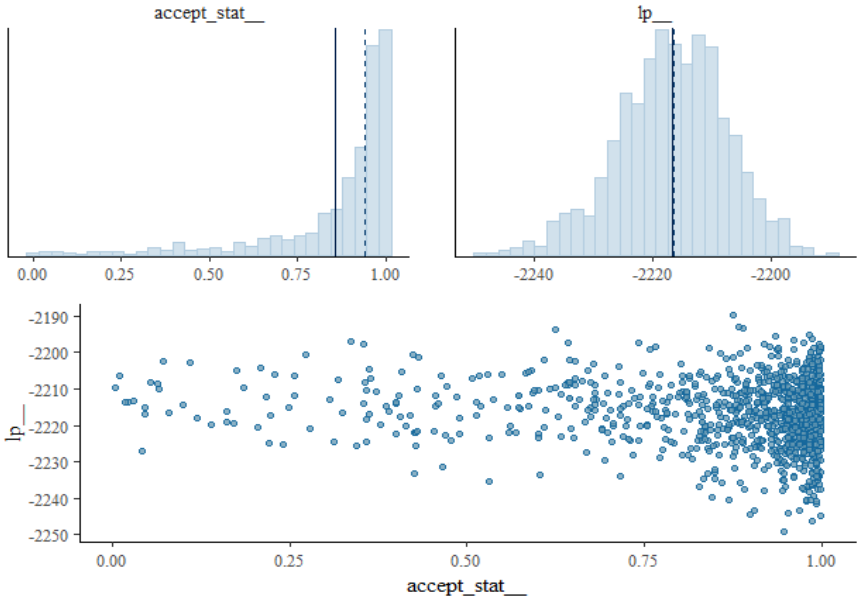


Figure 1: Acceptance diagnostic checks for NUT Sampler

Figure 1 shows that the acceptance probability of the sampler is nearly 100%, which shows the efficiency of the sampler for the model. The density on the upper right in Figure 1 shows that the data are well distributed. The accept_stat_ in Figure 1 shows that the sample cluster around 1.00 and majority close to 1.00 showing a high acceptance rate.

Table 7: Posterior Distribution Estimates

Specification	Estimate	Est.Error	Q2.5	Q97.5
b_GDP_Intercept	4.7021e+03	3.4217e+03	-2.6381e+03	1.0147e+04
b_logGDP_Intercept	1.1860e+01	6.4158e-01	1.0531e+01	1.3098e+01
sd_ID__GDP_Intercept	1.8191e+05	1.0922e+05	5.5165e+03	4.0937e+05
sd_ID__GDP_GNI	1.7202e-07	5.0712e-08	9.3678e-08	2.8880e-07
sd_ID__GDP_PPP	3.1667e+05	2.1497e+05	7.2187e+03	7.6616e+05
sd_ID__logGDP_Intercept	2.6008e+00	4.7777e-01	1.8265e+00	3.7144e+00

Table 7: Posterior Distribution Estimates (cont.)

Specification	Estimate	Est.Error	Q2.5	Q97.5
sd_ID__logGDP_PPP	1.1864e+00	5.5401e-01	2.1320e-01	2.4061e+00
sd_ID__logGDP_GNI	1.0017e-13	6.7751e-14	3.2062e-14	2.8442e-13
cor_ID__GDP_Intercept __GDP_GNI	2.1594e-02	4.7442e-01	-8.2537e-01	8.7270e-01
cor_ID__GDP_Intercept __GDP_PPP	4.5414e-02	4.9268e-01	-8.2671e-01	8.8462e-01
cor_ID__GDP_GNI __GDP_PPP	-2.2093e-02	4.7607e-01	-8.8198e-01	8.0275e-01
cor_ID__logGDP_Intercept __logGDP_PPP	-5.6823e-01	2.8381e-01	-9.4641e-01	7.9702e-02
cor_ID__logGDP_Intercept __logGDP_GNI	3.0527e-01	5.0828e-01	-7.8195e-01	9.4885e-01
cor_ID__logGDP_PPP __logGDP_GNI	-4.3449e-02	4.8786e-01	-8.5531e-01	8.6349e-01
sigma_GDP	2.7149e+05	1.6926e+04	2.4059e+05	3.0696e+05
sigma_logGDP	3.3815e-01	2.2685e-02	2.9746e-01	3.8724e-01
Intercept_GDP	4.7021e+03	3.4217e+03	-2.6381e+03	1.0147e+04
Intercept_logGDP	1.1860e+01	6.4158e-01	1.0531e+01	1.3098e+01
rescor__GDP__logGDP	7.0127e-01	4.7306e-02	5.9782e-01	7.8346e-01

Table 7 shows the posterior summary, which shows the model is similar to that of the estimates in Table 5. It was computed from 1000 by 143 log-likelihood matrix. The models for the standard deviation and the correlation estimates can as well be specified in the neighborhood of model (38) and (39). Table 8 and Table 9 contain the leave-out-one cross validation estimates and Pareto k diagnostic tests.

Table 8: LOO

Specification	Estimate	SE
elpd_loo	-2031.1	76.6
p_loo	69.5	34.4
Looic	4062.3	153.3

The elpd_loo (-2031.1) is the Bayesian leave-one-out (LOO) estimate of the expected log pointwise predictive density (ELPD), it can either be positive or negative. Large ELPD values indicate good estimated predictive performance, when comparing models, a larger ELPD suggests a better predictive performance. The p_loo is the difference between elpd_loo and the non-cross-validated log posterior predictive density. If p_loo < the number of parameters p, then the model is likely to be misspecified. The p_loo is 69.5 greater than the number of parameters.

Table 9: Pareto k diagnostic values

Specification	Count	Pct.	Min. ESS
(-Inf, 0.67] (good)	139	97.2%	71
(0.67, 1] (bad)	1	0.7%	-
(1, Inf) (very bad)	3	2.1%	-

Table 9 shows that out of 143 data points, 139 (97.2%) fall under good samples, 1 (0.7%) falls under bad sample, while 3 (2.1%) fall under very bad samples. The model proved to be a very good one.

Table 10 shows the R^2 statistics for both Bayes and (leave-out-one) LOO.

Table 10: Measure of Determination for Bayes and LOO Estimates

Specification	Estimate	Est.Error	Q2.5	Q97.5
	Baye_R2			
R2GDP	0.9174490	0.005425372	0.9055567	0.9261627
R2logGDP	0.9796951	0.001253107	0.9770527	0.9819441
	LOO_R2			
R2GDP	0.9150547	0.05972399	0.7529397	0.9818688
R2logGDP	0.9758178	0.01030771	0.9491834	0.9894607

The R^2 for both Bayes and LOO are very high with the R^2 values of 0.9174490 and 0.9150547 (91.75% and 91.51%) respectively, higher than that of frequentist IV model (0.69433). The two tail 95% confidence interval Q2.5 and Q97.5 shows that both R^2 are significant since the interval does not include zero.

4. Conclusions

Instances of African economic development have become the concern of many international agencies and governments at various levels, hence the needs for its continuous evaluation. Moreover, the opinions posited by previous studies examining the relationship between economic growth and socio-economic indicators have been indecisive and conflicting due to different sample periods, variables used, countries studied and econometric techniques employed. Thus, dynamic panel data estimation techniques of the generalized method of moment and instrumental variables (both the classical and the Bayesian) were employed to revisit the estimation. GMM was estimated according to the Arellano and Bond approach having fully taken care of endogeneity phenomenon as established by Hu *et al.* (2014) and IV estimated via a two-stage least square (2SLS) technique; it was discovered that the instrumental variable technique outperformed GMM based on robustness of the estimated models and the adopted model selection criteria. The preferred technique works well for both life and simulated data and Monte Carlo simulations reveal that the two methods have very good finite sample performance and give a positive projection of African economic growth compared to GMM, which gives a negative projection with weak validity criteria.

It is pertinent to emphasize the robustness of the adopted Bayesian IV in providing more reliable policy insights in terms of its consistency in handling endogeneity issues in data-driven approaches, which can improve the accuracy of policy recommendations. As established with the fitted IV models, policy makers should prioritize the

growth of the countries national income and create more leverage in the purchasing power parity of citizenries to enhance sustainable economic development.

It is pertinent to note that the greater focus of this current research is in the area of opinionating a robust estimation technique for a dynamic panel model through the modeling of African economic growth, and this has been vigorously established in the Bayesian IV. This technique, however, is recommended for the expedition of current economic data with more diverse econometric variables for a more robust contribution to the field of econometrics, as it concerns the modeling of economic growth.

Acknowledgement

The authors wish to acknowledge the online data producers through which the data for this research were sourced.

References

Adeboye, N. O., Abimbola, V. O., Folorunso, S. O. and Akinbo, R. Y., (2023). Modelling the volatility of African capital markets in the presence of covid-19 pandemic: evidence from the five emerging economies in Africa. *Statistics in Transition new series*, 24(2), pp. 17–36. https://doi.org/10.59170/stattrans-2023-018.

Adeboye, N. O., Agunbiade, D. A., (2019a). Monte Carlo estimation of heteroscedasticity and periodicity effects in a panel data regression model. *International Journal of Mathematical and Computational Sciences*, 13(6), pp. 157–166.

Adeboye, N. O., Agunbiade, D. A., (2019b). Monte Carlo estimation of heterogeneity effects in a panel data regression model. *International Journal of Mathematics and Computation*, 30(1), pp. 62–77

Adesina, O. S, Obokoh, L. O., (2024). Bayesian modelling of tail risk using extreme value theory with application to currency exchange. *International Journal of Safety and Security Engineering*, 14 (5), pp. 1447–1454

Adesina, O. S., (2021). Bayesian multilevel models for count data. *Journal of the Nigerian Society of Physical Sciences*, 3, pp. 224–253.

Agbeyebge, T. D., Stotsky, J. and Woldemariam, A., (2006). Trade liberalization, exchange rate changes, and tax revenue in sub-Saharan Africa. *Journal of Asian Economics*, 17(2), pp. 261-284. https://doi.org/10.1016/j.asieco.2005.09.003.

Agiomirgianakis, G., Asteriou, D. and Monastiriotis, V., (2002). Human capital and economic growth revisited: a dynamic panel data study. *International Advances in Economic Research*, 8(3), pp. 177–187. https://doi.org/10.1007/BF02297955.

Ajayi, S. I., (2003). Globalization and Africa. *Journal of African Economies*, 12(1), pp. 120–150. https://doi.org/10.1093/jae/12.suppl_1.120.

Al-Sadom, M., Jimenez-Martin, S. and Labeaga, J., (2019). *Simple methods for consistent estimation of dynamic panel data sample selection models*. Barcelona GSE Working Paper Series. Working Paper n^0 1069.

Artelaris P., Arvanitidis P. and Petrakos G., (2007). *Theoretical and methodological study on dynamic growth regions and factors explaining their growth performance*. Paper Presented at the 2nd Workshop of DYNREG in Athens, 9–10 March.

Arellano, M., Bond, S., (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58 (2), pp. 277–97. JSTOR. https://doi.org/10.2307/2297968.

Arellano, M., Bover, O., (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68(1), pp. 29–51. https://doi.org/10.1016/0304-4076(94)01642-DGet rights and content

Bardi, W., Sadraoui, T. and Bardi, A., (2016). A dynamic panel data analysis for relationship between structural policy and economic growth. *International Journal of Econometrics and Financial Management*, 4(1), pp. 1–10. http://pubs.sciepub.com/ijefm/4/1/1

Basu, A., Calamitsis E. A. and Ghura, D., (2005). Adjustment and growth in Sub-Saharan Africa. Economic Issue is based on IMF Working Paper 99/51.

Becsi, Z., Wang, P., (1997). Financial development and growth. Economic Review, 82(4), 46–62. https://EconPapers.repec.org/RePEc:fip:fedaer, vol. 82, no. 4.

Bengoa, M., Sanchez-Robles, B., (2003). Foreign direct investment, economic freedom and growth: new evidence from Latin America, *European Journal of Political Economy*, 19(3), pp. 529–545. https://doi.org/10.1016/S0176-2680(03) 00011-9.

Blundell, R., Bond, S., (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1), pp. 115–143. https://doi.org/10.1016/S0304-4076(98)00009-8.

Blundell, R., Bond, S., (2000). GMM estimation with persistent panel data: an application to production functions. *Econometric Reviews*, 19(3), pp. 321–340. https://doi.org/10.1080/07474930008800475.

Bürkner, P., (2017). Brms: an *R* package for Bayesian multilevel models using stan. *Journal of Statistical Software*, 80(1), pp. 1–28. https://doi.org/10.18637/jss.v080.i01.

Dreher, A., (2006). Does globalization affect growth? Evidence from a new index of globalization. *Applied Economics*, 38(10), pp. 1091–1110. https://doi.org/10.1080/00036840500392078.

Bun, M. J. G., Carree, M. A., (2005). Bias-corrected estimation in dynamic panel data models. *Journal of Business & Economic Statistics*, 23(2), pp. 200–210. http://www.jstor.org/stable/27638812.

Bascle, G., (2008). Controlling for endogeneity with instrumental variables in strategic management research. *Strategic Organization*, 6(3), pp. 285–387. https://doi.org/10.1177/1476127008094339.

Cho, M., Zheng, X., (2021). Bayesian estimation of dynamic panel data gravity model. *Econometric Reviews*, pp. 607–634 https://doi.org/10.1080/07474938.2021.1889203.

Jiménez, J., Godinho, R., Pinto, D., Lopes, S., Castro, D., Cubero, Dm., Osorio, A., Piqué, J., Moreno-Opo, R., Quiros, P., González-Nuevo, D., Hernandez-Palacios, O. and Kéry, M., (2022). *The cantabrian capercaillie: a population on the edge. Sci Total Environ.* 821, 153523. https://doi.org/10.1016/j.scitotenv.2022.153523. Epub 2022 Jan 29. PMID: 35104529.

Hoffman, M. D., Gelman, A., (2014). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *The Journal of Machine Learning Research*, 15 (2014), 1593.

Jin, F., Lee, L. and Yu, J., (2021). Sequential and efficient GMM estimation of dynamic short panel data models. *Econometric Reviews,* pp. 1007–1037. https://doi.org/10.1080/ 07474938.2021.1889178.

Kiviet, J. F. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics*, 68(1), pp. 53–78. https://doi.org/10.1016/0304-4076(94)01643-E.

Lm, K. S., Pesaran, M. H. and Shin, Y., (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics*, 115(1), pp. 53–74. https://doi.org/10.1016/S0304-4076(03)00092-7.

Lensink, R., Morrissey, O., (2006). Foreign direct investment: flows, volatility and the impact on growth. *Review of International Economics*, 14(3), pp. 478–493. https://doi.org/10.1111/j.1467-9396.2006.00632.x.

Levina, O., (2011). *FDI, economic freedom and growth. are they related?* Being an unpublished MA, Department of Economics, Central European University.

Lichtenberg, F., (1992). R&D Investment and international productivity Differences. *NBER Working Paper*, no. 4161.

Li, T, Maasoumi, E, Xiao, Z., (2021). Econometric reviews honors Cheng Hsiao. *Econometric Reviews*, 40, 6, pp. 535–539.

Meraj, M., (2013). Impact of globalization and trade openness on economic growth in Bangladesh. *Ritsumeikan Journal of Asia Pacific Studies*, 31, pp. 40–50. https://doi.org/10.34409/rjaps.40.1_1.

Nkurunziza, J. D., Bates, R. H., (2003). Political institutions and economic growth in Africa. *Center for International Development Working Paper*, no. 98. Harvard University.

Obadan, M., (2006). Prospects for diversification in Nigeria's export trade. Annual Conference of the Nigerian Economic Society, *Heinemann press*, *Ibadan*, pp. 33–53.

Sharma, R., (2018). Health and economic growth: evidence from dynamic panel data of 143 years. *PLoS ONE*, 13(10), e0204940. https://doi.org/10.1371/journal.pone.0204940.

R Core Team, (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/>.

Yi, Hu, Dongmei, G., Ying, D. and Shouyang, W., (2014). *Estimation of nonlinear dynamic panel data models with individual effects*. Mathematical Problems in Engineering, Article ID 672610. http://dx.doi.org/10.1155/2014/ 672610.

STATISTICS IN TRANSITION new series, September 2025
Vol. 26, No. 3, pp. 117–133, https://doi.org/10.59139/stattrans-2025-030
Received – 04.09.2024; accepted – 02.04.2025

Exploring the stochastic production frontier in the presence of outliers: a simulation study

Anik Djuraidah[1], **Ismail Pranata**[2]

Abstract

Technical efficiency measures the performance of an observation unit in generating outputs effectively. The stochastic production frontier (SPF), which is a commonly used method for this purpose, determines how close a unit is to achieving maximum output based on its inputs. However, the outliers in the data can distort the accuracy of SPF models. To address this issue, various error distribution modifications like gamma, Student's-t, Weibull and Rayleigh distributions have been proposed. However, there is limited research comparing these distributions in handling outliers. This study describes a simulation conducted to compare five SPF models: Normal-half Normal, Normal-Gamma, Normal-Weibull, Normal-Rayleigh, and Student's-t-half Normal. Applying simulated data across nine scenarios with varying data amounts and outlier percentages, the findings demonstrate that the SPF Student's t-half Normal model provides the most accurate prediction of technical efficiency. Using a heavy-tailed distribution, such as the Student's t distribution, for the disturbance component is more effective in handling outliers in the response variable than modifying the inefficiency of the component distribution.

Key words: robust, outlier, simulation, stochastic production frontier, technical efficiency.

1. Introduction

Stochastic Frontier Analysis (SFA) is one of the analytical methods used to measure the performance of individual units by estimating the technical efficiency level in transforming input into output. SFA modelling in performance measurement assumes that the outputs generated by individual units are not always efficient or do not reach the maximum output limit (Coelli et al. 2005). Additionally, the SFA model assumes that inefficiency effects do not solely cause the deviations occurring in SFA modelling but can also be attributed to factors of statistical errors/disturbances (noise) (Mariyono

[1] School of Data Science, Mathematics, and Informatics, IPB University, Indonesia. E-mail: anikdjuraidah@apps.ipb.ac.id. ORCID: https://orcid.org/0000-0002-3163-4343.

[2] BPS-Statistics Indonesia, Indonesia. E-mail: pranata@bps.go.id ORCID: https://orcid.org/0009-0000-7973-7133.

2006). The error component of SFA is a combination of inefficiency and disturbance components. Both components are assumed to have specific parametric distributions. The disturbance component has a symmetric distribution that is assumed to be independently and identically distributed, while the inefficiency component is assumed to have a one-sided distribution independent of the disturbance component (Aigner et al. 2023).

SFA modelling is supported by only one assumption: the production approach that can utilize distance, income, cost, or profit functions (O'Donnell 2018). In general, the SFA model uses a distance approach to calculate the efficiency level of individual units, where the model measures the distance between the observed production and the estimated maximum production (frontier) at a given input. The SFA model using the distance function approach is called the Stochastic Production Frontier (SPF). The SPF model will estimate the maximum production by incorporating disturbance components into its estimation process, thus yielding a stochastic estimation of the maximum production. The estimation of the performance level of an individual unit in the SPF model begins with estimating the inefficiency. Estimating inefficiency in the SPF model can be measured through two approaches, namely the output-oriented and the input-oriented approach (Kumbhakar et al. 2015). The SPF model in measuring inefficiency estimation using an output-oriented approach is the difference between the actual production and the maximum possible production by a certain amount of input. The production function is theoretically common in various literature, namely the Cobb-Douglas (CB) and the Transcendental Logarithmic (Translog) (Zulkarnain et al. 2021).

The distribution of the standard SPF model assumes that the disturbance component follows a Normal distribution ($N[0, \sigma_v^2]$) and the inefficiency component follows a half-normal distribution ($N^+[0, \sigma_u^2]$) (Meeusen et al. 1977; Aigner et al. 2023). The standard SPF model, in its application, has limitations in estimating technical efficiency values when there are outliers present in a dataset (Campos et al. 2022). The presence of outliers in the SPF model can cause parameter estimation to become inaccurate and lead to an excessive spread in the estimated values of technical efficiency (Wheat et al. 2019). One approach to deal with outliers in a dataset when forming an SPF model is modifying the distribution assumptions with a more flexible parameterization to describe the data distribution (Greene 1990).

SPF models that use a more flexible distribution have been widely implemented to improve robustness against outliers. The SPF Normal-Gamma model replaces the standard half-normal inefficiency distribution with a Gamma distribution, enhancing the accuracy and distribution of technical efficiency estimates (Greene 1990). The SPF Normal-Gamma model can also provide smaller standard deviations and ranges of estimated technical efficiency compared to the standard SPF model. Similarly, the

Normal-Weibull SPF model modifies the inefficiency distribution to follow a Weibull distribution, which has a longer tail and higher kurtosis than the Gamma distribution, helping to mitigate outlier effects (Tsionas 2007). The SPF Normal-Rayleigh model assumes a two-sided normal distribution for disturbances and a one-sided Rayleigh distribution for inefficiency, allowing for non-zero inefficiency estimates and reducing the risk of underestimation (Hajargasht 2015). Meanwhile, the SPF Student's t-half Normal model modifies the disturbance component to follow a Student's t-distribution while keeping the half-normal inefficiency distribution. With heavier tails in the disturbance component, this model improves inefficiency estimation, resulting in smaller standard deviations, a narrower range, and a more favorable distribution pattern (Wheat et al. 2019).

This research compares various SPF models with different parameter distributions for disturbance and inefficiency components to identify the most accurate model for estimating technical efficiency in the presence of outliers. The optimal SPF model is one that effectively handles outliers, assessed through its application to both real and simulated data. Previous studies on SPF model development using real data have examined the SPF Normal-half Normal, SPF Normal-Gamma, and SPF Student's t-half Normal models (Pranata et al. 2023). This study expands the analysis by incorporating simulation data to evaluate additional models, including SPF Normal-Weibull and SPF Normal-Rayleigh. The simulation data consists of varying sample sizes with different outlier proportions.

2. Stochastic Production Frontier

The Stochastic Production Frontier (SPF) is a Stochastic Frontier Analysis (SFA) model that measures the performance of individual units by measuring the distance between actual production and the estimated production frontier, given the same inputs. The production frontier function in the model is a production function that incorporates disturbance effects in its estimates, making it inherently stochastic. The disturbance component describes the possibility of random shocks that affect the production process and cannot be controlled, such as weather, economy, etc. (Coelli et al. 2005). In the SPF model, there is also an inefficiency component (Aigner et al. 1968). The measurement of estimation inefficiency in the SPF model can be categorized into output-oriented inefficiency and input-oriented inefficiency (Kumbhakar et al. 2015). Output-oriented inefficiency calculates the distance between the produced output and the estimated output frontier (achievable maximum production) for a given input. On the other hand, input-oriented inefficiency calculates the distance between the actual input and the estimated input frontier (minimum required) to achieve a certain output. In this study, the approach for measuring inefficiency in the model uses the output-

oriented inefficiency measurement. The form of the SPF model using the Cobb-Douglas production function is as follows:

$$\ln y_i = \beta_0 + \sum_{j=1}^{n} \beta_j \ln x_{ij} + \varepsilon_i, \quad \varepsilon_i = v_i - u_i \tag{1}$$

where ε_i represents the error model for observation i. The error model consists of two components, v_i is the disturbance component following symmetric distribution for observation i and u_i is the inefficiency component following one-sided distribution that is always positive for observation i.

One of the parameter estimation methods in the SPF model is maximizing the log-likelihood function based on the marginal density function of the error. When the marginal density function of the error is not in a closed form, the estimates of the parameter cannot directly maximize the log-likelihood function, but use the simulated maximum likelihood method (Train 2009). The equation for the log-likelihood function is as follows:

$$\ln \mathcal{L} = \ln \prod_{i=1}^{n} f(\varepsilon_i) \tag{2}$$

where $\mathcal{L}$ represents the likelihood function and $f(\varepsilon)$ represents the marginal density function of the error SPF model. The probability density function of the error in the SPF model is formed by composing the probability density functions of the disturbance and the inefficiency component. The probability density function for the error in the SPF model begins with the formation of a joint probability density function between the disturbance and the inefficiency component. Since both components are independent, their joint probability density function is the product of their density functions. The joint probability density function of the two components is as follows:

$$f(v, u) = f(v) \times f(u) \tag{3}$$

where $f(v)$ represents the probability density function of the disturbance component and $f(u)$ represents the probability density function of the inefficiency component. The joint probability density function $f(\varepsilon, u)$ is constructed by transforming the disturbance component using $v = \varepsilon + u$. The error's probability density function is then obtained by integrating the joint density over the inefficiency component. The equation is given as follows:

$$f(\varepsilon) = \int_0^{\infty} f(\varepsilon, u)\, du \tag{4}$$

Estimating inefficiency in the SPF model begins by forming the probability density function of inefficiency values where the error values are known ($f(u|\varepsilon)$). This function can be determined as follows:

$$f(u|\varepsilon) = \frac{f(u, \varepsilon)}{f(\varepsilon)} \tag{5}$$

In the SPF model, inefficiency is estimated by calculating the expected value of inefficiency given the error $E(u|\varepsilon)$. The form of this expected value is as follows:

$$E(u|\varepsilon) = \int_0^{\infty} u\, f(u|\varepsilon)\, du \tag{6}$$

The estimation of technical efficiency ($\widehat{TE}$) in the SPF model follows the equation introduced by Jondrow et al. (1982), where technical efficiency is computed as the exponent of negative estimated inefficiency. The equation as follows:

$$\widehat{TE} = \exp(-E(u|\varepsilon)) \tag{7}$$

Several types of SPF models include the following:

1. **The SPF Normal-half Normal Model:** The SPF Normal-half Normal model is a standard SPF model in which the disturbance component follows a symmetric Normal distribution ($N(0, \sigma_v^2)$), and the inefficiency component follows a half Normal distribution ($N^+(0, \sigma_u^2)$) Meeusen & Van Den Broeck (1977) and (Aigner et al., 2023). The marginal density function of the error in the SPF Normal-half Normal model is closed-form so that the estimation of the parameter maximizes the log-likelihood function.
2. **The SPF Normal-Gamma Model**: The SPF Normal-Gamma model is a SPF model where the disturbance component follows a symmetric Normal distribution ($N(0, \sigma_v^2)$), and the inefficiency component follows a Gamma distribution ($G(\alpha, \theta)$) as introduced by (Greene, 1990). Since the marginal density function of the error lacks a closed-form solution, its log-likelihood function cannot be directly maximized. Instead, parameter estimation can be performed using the maximum simulated likelihood method, as suggested (Greene 2003).
3. **The SPF Normal-Weibull Model**: The SPF Normal-Weibull model assumes that the disturbance component follows a symmetric Normal distribution ($N(0, \sigma_v^2)$), and the inefficiency component follows a Weibull distribution. This model was first introduced by (Tsionas 2007). The Weibull distribution has a long-tailed curve shape, classified as heavy-tailed. The characteristic of a long-tailed curve shape allows the Weibull distribution to handle outliers or extreme observations in data distributions. The formed marginal density function of the error in the SPF Normal-Weibull model does not have a closed-form expression, so parameter estimation using simulated maximum likelihood (Nurwulan et al. 2023).
4. **The SPF Normal-Rayleigh Model**: The SPF Normal-Rayleigh model assumes that the disturbance component follows a symmetric Normal distribution ($N(0, \sigma_v^2)$), and the inefficiency component follows a Rayleigh distribution $Ra(0, \sigma_u^2)$, as introduced by Hajargasht (2015). The characteristics of the Rayleigh distribution in the SPF model are different from the half-normal and exponential distributions.

The half-normal and exponential distributions tend to produce inefficiency estimates close to zero, raising concerns about potential underestimation. In contrast, the Rayleigh distribution exhibits a broader spread of data, reducing concentration near zero and potentially providing more accurate inefficiency estimates. The marginal density function of the error in the SPF Normal-Rayleigh model has a closed-form so that the estimation of the parameter maximizes the log-likelihood function.

5. **The SPF Student's t-half Normal Model**: The SPF Student's t-half Normal model assumes that the disturbance component follows a symmetric Student's t distribution ($T(a)$) with degrees of freedom (a), while the inefficiency component follows a half-normal distribution ($N^{+}(0, \sigma_u^2)$) (Wheat et al., 2019). The Student's t distribution is a two-tailed distribution characterized by a long-tailed curve on both sides of the distribution. Like the Weibull distribution, the Student's t distribution is also categorized as heavy-tailed. The marginal density function of the error in the SPF Student's t-half Normal model is not a closed form. As a result, the parameter estimation process for this SPF model utilizes simulated maximum likelihood.

3. Method

This study compares the estimated coefficients of model parameters, inefficiency values, and technical efficiency among five Stochastic Production Frontier (SPF) models. The five SPF models include the SPF Normal-half Normal model, SPF Normal-Gamma model, SPF Normal-Weibull model, SPF Normal-Rayleigh model, and SPF Student's t-half Normal model. Additionally, this study will also identify the best-performing SPF model or the SPF model robust to outliers in the response variable observations. The data used in this research is generated through simulation, where simulated data is created for nine dataset scenarios. These scenarios are formed based on combinations of predefined data quantities and outlier observation percentages. The process of modelling SPF in this study utilizes the "sfaR" package (Dakpo et al. 2022) and the "rfrontier" package (Wheat et al. 2019).

3.1. Generating Simulated Data

This study uses data simulation with Monte Carlo simulations. Monte Carlo simulation data repeatedly generates random numbers or values of a variable with a specific distribution (Sartono 2005). The number of generated data in the simulation consists of three different data quantities: 200 data points as the small data category, 500 points data as the medium data category, and 1000 points data as the large data category. The varying sample data sizes aim to observe the influence of data quantity on enhancing estimation accuracy. The simulation outlier percentages are divided into three levels: 1%, 3%, and 5% of the response variable. It aims to assess the model's consistency in handling data containing outlier observations.

The production function used in the SPF model is the Cobb-Douglas production function, where the model's parameter coefficients are determined as follows: $\beta_0 = 0{,}1$, $\beta_1 = 2$, and $\beta_2 = 2$. Thus, the form of the simulated data SPF model is as follows:

$$\ln y_i = 0{,}1 + 2 \ln x_{1i} + 2 \ln x_{2i} + v_i - u_i \tag{8}$$

Generating simulated data starts by creating independent variable data, disturbance components, and inefficiency components. The model uses two independent variables, where both are mutually independent and are generated using a Uniform distribution (10, 20). In generating independent variable data, using a variance-covariance matrix is an identity matrix to ensure that the generated data for both variables are mutually independent. The disturbance components are assumed to follow a Normal distribution. The disturbance components will be generated with a mean of 0 and a variance of 0.25 $(N(0, 0{,}25))$. Meanwhile, the inefficiency components are assumed to follow a half-normal distribution. The inefficiency components will be generated with a mean of 0 and a variance of 1 $(N^+(0{,}1))$. The results of generated data from independent variables, disturbance component, and inefficiency component will be input into the model to obtain the generated data for the response variable. The process will be repeated one hundred times for each set of generated data.

The outlier data in the simulated data is a set of data significantly different from the generated data. The outlier will replace a randomly selected portion of the response variable data, and the predetermined outlier percentage will determine the amount of these replaced data.

3.2. Modelling

The SPF model has a characteristic that the residual distribution is skewed to the left (Kumbhakar et al., 2015). Therefore, an error distribution test is performed on the generated data before conducting the SPF modelling to ensure the error distribution is skewed to the left. The testing can be done by calculating the statistically significant Coelli Test (M3T) (Mutz et al. 2017). The M3T test counts the standardized third moment of the error from the least squares model and has an asymptotic distribution of a standard normal random variable. The equation for the M3T test is as follows: (Coelli 1995)

$$M3T = {m_3}\Big/{\sqrt{6m_2^3/N}} \tag{11}$$

where m_2 and m_3 are the second and third moments of the least square error, forming a distribution $N(0, 6m_2^3/N)$. The Hypothesis is as follows:

H_0: The error has a symmetric distribution.

H_1: The error has a skewed distribution.

The generated data can be used for SPF modelling if the M3T test results reject the null hypothesis (the M3T value falls outside the critical range of ±1.96) at a significance level of α=5% and the M3T value is negative. This condition indicates that the distribution of least squares errors from the generated data has a left-skewed curve.

Once it is confirmed that the generated data for SPF modelling has a left-skewed distribution, the next step is to proceed with SPF modelling using the five different SPF model forms to obtain parameter estimation values, inefficiency values, and technical efficiency values from the models.

3.3. Estimation Performance

The criterion for selecting the best SPF model is its ability to provide estimated technical efficiency close to the true values. It can be determined by the mean square error (MSE) of the estimated inefficiency and technical efficiency compared to the true values. The SPF model that yields the smallest MSE can be considered the best SPF model. The MSE for estimating inefficiency is the average of the squared differences between the estimated inefficiency and the true inefficiency across all observations and repetitions. The calculation of the MSE for inefficiency is as follows:

$$\mathrm{MSE(u)} = \frac{1}{\mathrm{NR}} \sum_{\mathrm{r=1}}^{\mathrm{R}} \sum_{\mathrm{i=1}}^{\mathrm{N}} (\hat{\mathrm{u}}_{\mathrm{i,r}} - \mathrm{u}_{\mathrm{i,r}})^2 \tag{9}$$

where N is the number of generated dataset scenarios, R is the number of repetitions, $\hat{\mathrm{u}}_{\mathrm{i,r}}$ is the estimated inefficiency for observation i and repetition r, $\mathrm{u}_{\mathrm{i,r}}$ is the true inefficiency (generated) for observation i and repetition r. Meanwhile, the MSE for technical efficiency is the average of the squared differences between the estimated technical efficiency and the true technical efficiency across all observations and repetitions. The MSE for technical efficiency is as follows (Sakouvogui et al. 2021):

$$\mathrm{MSE(TE)} = \frac{1}{\mathrm{nR}} \sum_{\mathrm{r=1}}^{\mathrm{R}} \sum_{\mathrm{i=1}}^{\mathrm{N}} (\widehat{\mathrm{TE}}_{\mathrm{i,r}} - \mathrm{TE}_{\mathrm{i,r}})^2 \tag{10}$$

where $\widehat{\mathrm{TE}}_{\mathrm{i,r}}$ is the estimated technical efficiency for observation i and repetition r and $\mathrm{TE}_{\mathrm{i,r}}$ is the true value (generated) technical efficiency for observation i and repetition r.

4. Results and Discussion

The simulation of data of the SPF model begins by generating data for nine scenarios. These nine scenarios consist of three sets of generated data, each with a sample size of 200, 500, and 1000. Within each set there are three variations of outlier percentages: 1%, 3%, and 5%. The distribution of generated data in nine scenarios can be observed in Figure 1. The data distribution plots in Figure 1 represent a sample from

one dataset of the independent variable ($\ln x_1$) and response variable ($\ln y$) generated from a hundred repetitions for each scenario.

The generated data from the nine scenarios is tested using the M3T test to ensure that the generated data has the error distribution criteria of the SPF model. The results of the minimum and maximum M3T values across a hundred repetitions are presented in Table 1. The M3T test uses a significance level of $\alpha = 5\%$, and the critical values for rejecting H_0 a are when the M3T value is less than -1.96 or greater than 1.96. The range of the M3T values from the nine scenarios with a hundred repetitions for each simulation yields values smaller than -1.96. Hence, all the generated datasets reject the null hypothesis with the negative M3T values. In other words, all the generated data distributions have a left-skewed error distribution with a 95% confidence level.

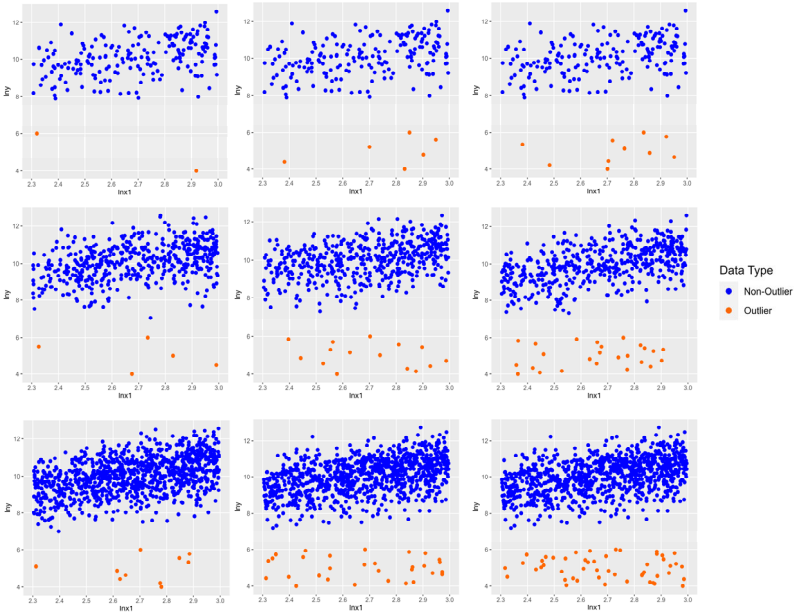


Figure 1: Plot the distribution of generated data between the response variable ($\ln y$) and the independent variable ($\ln x_1$) with outlier observation percentages 1%; 3%; and 5%, (a) a sample size of 200, (b) a sample size of 500, (c) a sample size of 1000

Source: own study.

Table 1. The range of M3T values across the nine simulation scenarios with 100 repetitions

Simulation Scenarios	M3T Score	
	Minimum	**Maximum**
Sample size=200		
Outlier 1%	-17.99	-5.63
Outlier 3%	-16.31	-10.15
Outlier 5%	-14.89	-10.01

Table 1: The range of M3T values across the nine simulation scenarios with 100 repetitions (cont.)

Simulation Scenarios	M3T Score	
	Minimum	Maximum
Sample size=500		
Outlier 1%	-22.07	-13.99
Outlier 3%	-24.67	-19.05
Outlier 5%	-23.07	-18.88
Sample size=1000		
Outlier 1%	-28.63	-20.44
Outlier 3%	-34.28	-27.39
Outlier 5%	-33.01	-28.12

Source: own study.

4.1. SPF Model Parameter Estimates

The distribution of estimated intercept (b_0) from the SPF Normal-half Normal model, SPF Normal-Gamma model, SPF Normal-Weibull model, SPF Normal-Rayleigh model, and SPF Student's t-half Normal model is presented in Figure 2. From the generated simulation data, the SPF Student's t-half Normal model is the model that can estimate the parameter value β_0 close to the true value. This can be seen from the distribution of estimated b_0, which are around the true parameter, and the median of b_0 has the smallest bias to the true parameter compared to the other four SPF models across the nine simulation scenarios.

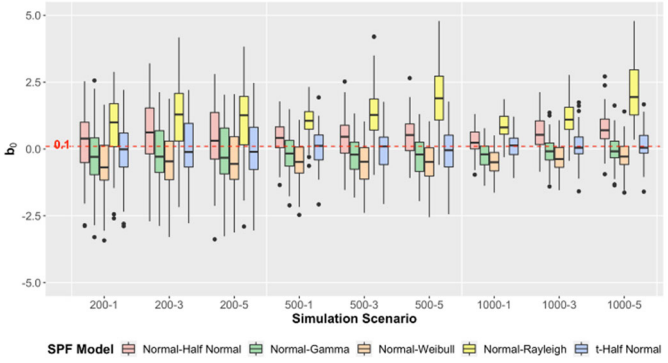


Figure 2: The estimation distribution of b_0 across nine simulation data scenarios using the SPF Normal-half Normal, SPF Normal-Gamma, SPF Normal-Weibull, SPF Normal-Rayleigh, and SPF Student's t-half Normal models

Source: own study.

Meanwhile, the SPF Normal-Rayleigh model exhibits significant estimation bias, with the distribution of b_0 estimates deviating considerably from the true parameter

value and the estimated median of b_0 being far from the actual value. The median of the b_0 in the Normal-half Normal SPF model is greater than that in the SPF Student's t-half Normal model. Furthermore, in the Normal-Half Normal SPF model, the median of the b_0 increases with higher percentages of outliers. This indicates that the accuracy of the standard SPF model in estimating intercept coefficients is greatly affected when outlier observations are present in the data. The SPF Normal-Gamma model and SPF Normal-Weibull model have similar characteristics in estimating intercept parameter coefficients. Both SPF models have an estimated distribution of b_0 that tends to be lower than the true intercept. Despite having a heavier tail distribution than the SPF Normal-Gamma model, the SPF Normal-Weibull model does not yield better intercept estimations than the SPF Normal-Gamma model. A box plot in Figure 2 shows that as the sample size used in the SPF model increases, the variance of b_0 estimation becomes smaller. It is indicated by the decreasing length of the box plot as the simulation data size grows.

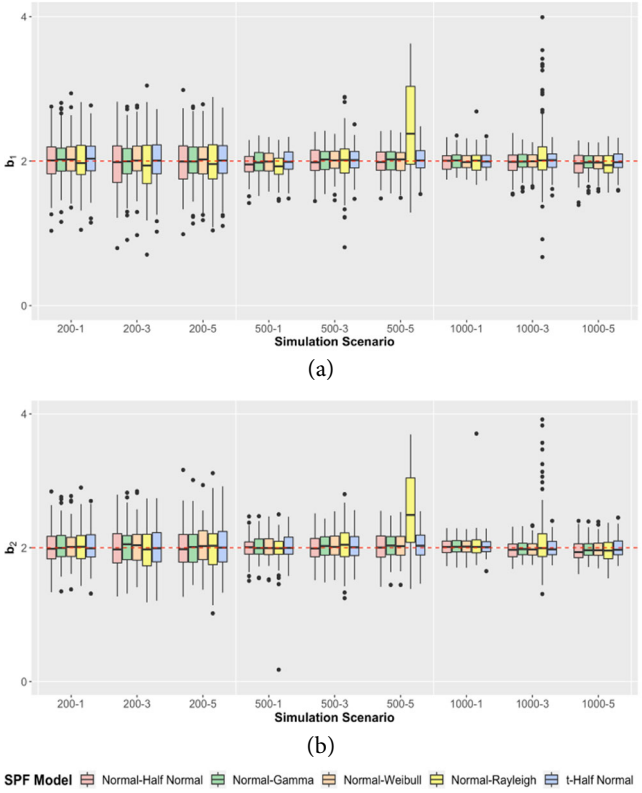


Figure 3: Distribution of estimated slope parameter across nine simulation data scenarios using the SPF Normal-half Normal, SPF Normal-Gamma, SPF Normal-Weibull, SPF Normal-Rayleigh, and SPF Student's t-half Normal models, (a) estimated of b_1 (b) estimated of b_2.

Source: own study.

The distribution of the estimated slope from nine simulation scenarios, composed of two coefficient parameters, namely the estimated b_1 and b_2, is depicted in the box plot diagram presented in Figure 3. The five SPF models were used to estimate the coefficients b_1 and b_2, which generally had a median parameter estimate close to the true values. However, in simulations with 500 observations and a 5% outlier percentage, the SPF Normal-Rayleigh model yields median slope estimates that deviate from the true values.

4.2. Inefficiency and Technical Efficiency Prediction

Figure 4 compares the estimated inefficiency distributions of the SPF Normal-Half Normal, SPF Normal-Gamma, SPF Normal-Weibull, SPF Normal-Rayleigh, and SPF Student's t-Half Normal models across nine simulation scenarios. The SPF Normal-Half Normal and SPF Normal-Rayleigh models tend to overestimate inefficiency, while the SPF Normal-Gamma and SPF Normal-Weibull models generally underestimate it. Among these models, the SPF Student's t-Half Normal model provides the most accurate inefficiency estimates, with its distribution closely aligning with the true inefficiency values. This is evident from the median of the estimated inefficiency distribution being close to the true inefficiency median, with the overall estimated distribution falling within the range of the true inefficiency values. In the SPF Student's t-Half Normal model, the number of estimated inefficiency outliers closely matches the true inefficiency outliers. In contrast, the other four SPF models tend to produce significantly more outliers than the true inefficiency distribution.

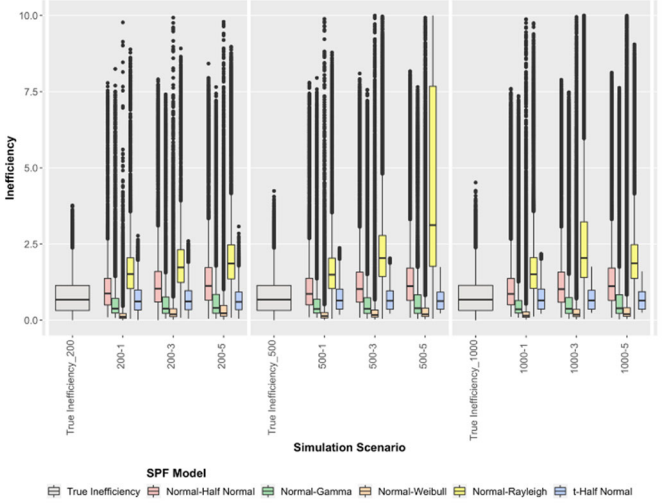


Figure 4: The distribution of true inefficiency and the estimated distributions of inefficiency using the SPF Normal-half Normal, SPF Normal-Gamma, SPF Normal-Weibull, SPF Normal-Rayleigh, and SPF Student's t-half Normal models across nine simulation scenarios

Source: own study.

In addition to comparing the estimated inefficiency scores of the five SPF models using box-and-whisker plots, the mean squared error (MSE) was calculated to assess how accurately each model predicts the true inefficiency. The MSE results, presented in Table 2, compare the estimated inefficiency with the true values across nine simulation scenarios using the five SPF models. The SPF Student's t-Half Normal model consistently produces lower MSE values than the other four SPF models across all nine simulation scenarios. In contrast, the SPF Normal-Weibull model performs the worst in predicting inefficiency, as indicated by its significantly higher MSE compared to the other models. Increasing the outlier percentage in the response variable generally leads to higher MSE values. However, this trend does not apply to the SPF Normal-Weibull model.

Table 2: Mean square error of the estimated inefficiency of five SPF models across nine simulation scenarios

Simulation scenario	MSE Scores of SPF				
	Normal-half Normal	Normal-Gamma	Normal-Weibull	Normal-Rayleigh	Student's t-half Normal
Sample size=200					
Outlier 1%	0.4619	0.4890	85.5771	1.1897	0.1941
Outlier 3%	1.1106	0.8847	3.1932e+05	2.2624	0.2100
Outlier 5%	1.7314	1.2923	6.4129	3.2150	0.2313
Sample size=500					
Outlier 1%	0.4524	0.4668	1.5906e+10	12.6826	0.1666
Outlier 3%	1.0874	0.8626	1.1455e+05	4.6814	0.1897
Outlier 5%	1.7247	1.2943	3.6382	39.2181	0.2081
Sample size=1000					
Outlier 1%	0.4326	0.4251	8.7324e+08	1.5060	0.1635
Outlier 3%	1.0787	0.8547	21.1244e+05	7.2978	0.1806
Outlier 5%	1.7409	1.3054	34.9856	290.9171	0.1982

Source: own study.

The distribution of estimated technical efficiency from the five SPF models is presented in Figure 5, with each model yielding a distinct distribution. Among the nine simulation scenarios, the SPF Student's t-half Normal model produces estimated technical efficiency distributions that closely align with the true distribution. This is evident from the box-and-whisker plot, which closely resembles the true technical efficiency plot, with the estimated median exhibiting minimal bias relative to the true median. The estimated technical efficiency of the Student's t-half Normal and SPF Normal-half Normal models does not contain outliers, whereas the other three SPF models do.

The accuracy of the model in predicting technical efficiency values is measured using the Mean Squared Error (MSE) between the estimated and true technical efficiency values. The MSE results for nine simulation scenarios are presented in Table 3. The SPF Student's t-half Normal model has the smallest MSE among all SPF

models across all nine simulated data scenarios. In contrast, the SPF Normal-Weibull and SPF Normal-Rayleigh models have the highest MSE among the five SPF models.

These results indicate that handling outliers in the response variable by replacing the distribution of inefficiency components with a heavy-tailed distribution does not impact the SPF model's ability to improve the accuracy of predicting true technical efficiency. Conversely, modifying the distribution of disturbance components to a heavier-tailed distribution can enhance the model's accuracy in predicting true technical efficiency.

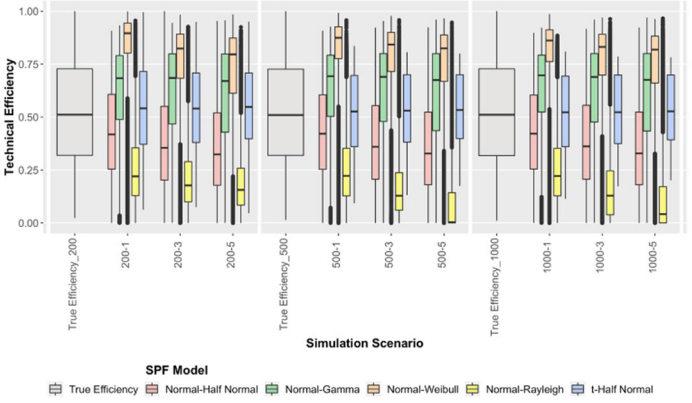


Figure 5: The distribution of true technical efficiency and the estimated technical efficiency using the SPF Normal-half Normal, SPF Normal-Gamma, SPF Normal-Weibull, SPF Normal-Rayleigh, and SPF Student's t-half Normal models across nine simulation scenarios

Source: own study.

Table 3: Mean square error of the estimated technical efficiency of five SPF models across nine simulation scenarios

Simulation scenario	MSE Scores				
	SPF Normal-half Normal	SPF Normal-Gamma	SPF Normal-Weibull	SPF Normal-Rayleigh	SPF Student's t-half Normal
Sample size=200					
Outlier 1%	0.0429	0.0520	0.1467	0.1058	0.0376
Outlier 3%	0.0578	0.0531	0.1044	0.1331	0.0371
Outlier 5%	0.0682	0.0564	0.0946	0.1491	0.0391
Sample size=500					
Outlier 1%	0.0419	0.0497	0.1312	0.1069	0.0314
Outlier 3%	0.0573	0.0523	0.1118	0.1715	0.0329
Outlier 5%	0.0683	0.0563	0.1039	0.2501	0.0344
Sample size=1000					
Outlier 1%	0.0419	0.0470	0.1205	0.1072	0.0308
Outlier 3%	0.0564	0.0517	0.1044	0.1768	0.0318
Outlier 5%	0.0676	0.0559	0.1005	0.2336	0.0330

Source: own study.

5. Conclusions

The findings suggest that the SPF Student's t-half Normal model produces more accurate intercept and slope estimates than the SPF Normal-half Normal, SPF Normal-Gamma, SPF Normal-Weibull, and SPF Normal-Rayleigh models. Additionally, it provides the most accurate inefficiency and technical efficiency predictions among the SPF models. Changing the assumption of the distribution in the disturbance component to a distribution with heavy tails, such as the Student's t distribution, is more effective in handling outliers in the response variable than altering the assumption of the distribution in the inefficiency component. The SPF Student's t-half Normal model is robust in dealing with outliers in the response variable when predicting technical efficiency.

Acknowledgement

The authors express gratitude to IPB University and the Statistics of Indonesia for their support, which has enabled the completion of this research.

References

Aigner, D. J., Chu, S. F., (1968). On estimating the industry production function. *American Economic Review*, 58(4), pp. 826–839.

Aigner, D., Lovell, C. A. K. and Schmidt, P., (2023). Reprint of: Formulation and estimation of stochastic frontier production function models. *J. Econometrics*, 234(S), pp. 15–24. doi: 10.1016/j.jeconom.2023.01.

Behr, A., (2015). Production and efficiency analysis with R. *Springer,* Cham.

Campos, M. S., Costa, M. A., Gontijo, T. S., Lopes-Ahn, A. L., (2022). Robust stochastic frontier analysis applied to the Brazilian electricity distribution benchmarking method. *Decis. Anal.*, *3*, 100051. doi:/10.1016/j.dajour.2022.100051.

Coelli, T., (1995). Estimators and hypothesis tests for a stochastic frontier function: A Monte Carlo analysis. *Journal of Productivity Analysis*, 6(3), pp. 247–268. doi: 10.1007/BF01076978.

Coelli, T. J., Rao, D. S. P., O'Donnell, C. J. and Battese, G.E., (2005). An introduction to efficiency and productivity analysis. *Springer*, Boston, MA, USA.

Dakpo, K. H., Desjeux, Y. and Latruffe, L., (2022). {sfaR}: Stochastic frontier analysis using R.

Greene, W. H., (1990). A gamma-distributed stochastic frontier model. *J. Econometrics*, 46, pp. 141–163.

Greene, W. H., (2003). Simulated likelihood estimation of the normal-gamma stochastic frontier function. *Journal of Productivity Analysis*, 19, pp. 179–190.

Hajargasht, G., (2015). Stochastic frontiers with a Rayleigh distribution. *Journal of Productivity Analysis,* 44(2), pp 199-208. doi: 10.1007/s11123-014-0417-8.

Jondrow, J., Knox Lovell, C., Materov, I. S., Schmidt, P., (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *J. Econometrics*., 19(2-3), 233-238. Doi: 10.1016/0304-4076(82)90004-5.

Kumbhakar, C. S., Wang, J. H. and Horncastle, P. A., (2015). A practitioner's guide to stochastic frontier analysis using Stata. *Cambridge University Press*, New York.

Mariyono, J., (2006). Geographical distribution of technical efficiency in Indonesian rice production during the period of 1979-1994. *Ekon. Pembang*., 11, pp. 33–48.

Meeusen, W., Van Den Broeck, J., (1977). Efficiency estimation from Cobb-Douglas production function with composed error. *International Economic Review (Philadelphia)*, 18.

Mutz, R., Bornmann, L. and Daniel, H.-D., (2017). Are there any frontiers of research performance? Efficiency measurement of funded research projects with the Bayesian stochastic frontier analysis for count data. *JOI*, 11(3), pp. 613–628. doi: 10.1016/j.joi.2017.04.009.

Nurwulan, R., Djuraidah, A. and Fitrianto, A., (2022). Handling outliers in the stochastic frontier model using Cauchy and Rayleigh distributions to measure technical efficiency of rice farming business. *Indonesian Journal of Artificial Intelligence and Data Mining*, 5(2), pp. 98–110.

O'Donnell, C. J., (2018). Productivity and efficiency analysis. *Springer*, Singapore.

Pranata, I., Djuraidah, A. and Aidi, M., (2023). Robust stochastic production frontier to estimate technical efficiency of rice farming in Sulawesi Selatan. *Barekeng Journal of Mathematics and Applications*, 17(1). doi: 10.30598/barekengvol17iss1pp0391-0400.

Sakouvogui, K., Shaik, S., Doetkott, C. and Magel, R., (2021). Sensitivity analysis of stochastic frontier analysis models. *Monte Carlo Methods and Applications*, 27 (1), pp. 71-90. doi: 10.1515/mcma-2021-2083.

Sartono, B. , (2005). Pembangkitan bilangan acak untuk simulasi Monte Carlo non-parametrik. *Forum Statistika dan Komputasi*, 10, pp. 8–11.

Train, K., (2009). *Discrete choice methods with simulation*. 2nd ed. *Cambridge University Press*, Cambridge.

Tsionas, G. E., (2007). Efficiency measurement with the Weibull stochastic frontier. *Oxford Bulletin of Economics and Statistics*, 69(5), pp. 693–706. doi: 10.1111/j.1468-0084.2007.

Wheat, P., Stead, A. D. and Greene, W. H., (2019). Robust stochastic frontier analysis: A Student's t-half normal model with application to highway maintenance costs in England. *Journal of Productivity Analysis*, 51(1), pp. 21–38. doi: 10.1007/s11123-018-0541-y.

Zulkarnain, R., Djuraidah, A., Sumertajaya, I. M. and Indahwati, (2021). Utilization of Student's-T distribution to handle outliers in technical efficiency measurement. *Media Statistika*, 14(1), pp. 56–67.

STATISTICS IN TRANSITION new series, September 2025
Vol. 26, No. 3, pp. 135–154, https://doi.org/10.59139/stattrans-2025-031
Received – 17.12.2024; accepted – 11.03.2025

Regional differentiation of income distributions in Poland

Alina Jędrzejczak[1], Małgorzata Misztal[2], Dorota Pekasiewicz[3]

Abstract

Income inequality is observed to have recently increased both at the country and regional level. Consequently, inequality, poverty and social stratification have become important issues of debate among economists, sociologists and social policy-makers. Thus, an in-depth statistical analysis of the income situation of households is particularly important when counteracting the social effects of the discussed phenomenon. In the literature, the situation of regions in Poland is compared primarily in terms of the differences in GDP or average wages observed for households. The aim of the article is to analyze the regional differences in the entire income distribution in Poland, taking into account not only average income levels, but also income inequality and poverty parameters. The study, based on individual data from the Household Budget Survey, used parametric and non-parametric methods for estimating inequality and poverty measures, as well as cluster analysis methods. In the parametric approach, the empirical income distributions in Poland were approximated using the theoretical Dagum distribution. This enabled the segmentation of voivodships in terms of the estimated characteristics of the equivalent household income distribution. The results of the calculations confirmed the assumption that income distributions in Poland differ significantly across regions, and the obtained clusters allowed the detection of groups of regions that may require a separate social policy aimed at improving the material situation of households.

Key words: income distribution; inequality; poverty; Dagum model; cluster analysis.

1. Introduction

Reducing regional inequality was one of the key means of promoting the "harmonious development" within Europe envisioned in the EEC Treaty of 1957.

[1] Department of Statistical Methods, University of Lodz; Lodz, Poland & Statistical Office in Lodz, Lodz, Poland. E-mail: alina.jedrzejczak@uni.lodz.pl, ORCID: https://orcid.org/0000-0002-5478-9284

[2] Department of Statistical Methods, University of Lodz, Lodz, Poland. E-mail: malgorzata.misztal@uni.lodz.pl, ORCID: https://orcid.org/0000-0002-8719-2097

[3] Department of Statistical Methods, University of Lodz, Lodz, Poland. E-mail: dorota.pekasiewicz@uni.lodz.pl, ORCID: https://orcid.org/0000-0001-8275-3345

The pursuit of "economic, social and territorial cohesion" through ever closer regional and national harmonization was also proclaimed in the 2007 Lisbon Treaty, but deepening European integration has not always been matched with convergence in living standards between sub-national regions. The gap between poorer and richer areas increased during the last economic crisis even in some developed economies, and the income discrepancy between richer and poorer regions is likely to widen further as government-spending cuts disproportionately hurt less prosperous regions. Over the last few decades, income inequality measured by the Gini index has been growing not only in Europe (see: *Growing Unequal*, OECD 2008; *Divided We Stand. Why Inequality Keeps Rising*, OECD 2011; *In It Together: Why Less Inequality Benefits All*, OECD 2015), and in 2014 most OECD countries recorded the highest level of inequality in 30 years. An even more pessimistic picture of income and wealth inequalities can be seen in a recent publication of OECD (2024), where it has been revealed that more than one-third of people in the OECD are at risk of poverty. The total income inequality can be decomposed into between- and within- regions inequality. Recent trends on income inequality in the EU confirm that, by decomposing overall inequality, as much as 85% is explained by within countries inequality (Bonesmo Fredriksen, 2012), while the findings of Barca (2009) reveal how prosperous regions in EU countries show at the same time strong internal, i.e. subnational, disparities. It has also been found that nearly 40% of total income goes on average to people in the highest income quintile, and less than 10% to people in the first quintile (Eurostat, 2014). In a working paper of Savoia (2020), devoted to the question whether the EU Cohesion Policy may have contributed to reduce the inequalities within countries or regions, it has been found that income inequality observed within NUTS 2 is converging, but to a higher level, so the regions "have tended to become equally more unequal". As a result of the increasing inequalities within and across countries and regions, the reduction in the number of persons at risk of poverty or social exclusion was one of the five key targets of the Europe 2020 strategy.

Income inequality in Poland increased significantly during the process of transformation from a centrally-planned to a market economy - the Gini index went up by approximately 10 percentage points. After the period of rapid economic changes the rate of growth of the Gini index slowed down and for a long time we could observe only slight fluctuations, at about the level of 0.34-0.35, according to the Household Budget Survey (HBS) data, and 0.31 according to EU-SILC. After the decline between 2016 and 2019 the Gini ratio in Poland began increasing again, and in 2023 it reached a very high level of 0.414 (GUS, 2024, p. 212), with high discrepancy between income distributions for socio-economic groups.

Regional inequalities can be measured and analyzed in many different ways - the extent of inequality may be mapped in terms of demography, income and wealth, labor markets, and education and skills. The main focus of this presentation is on the analysis of regional inequalities in terms of household income distribution. Income distributions can differ in many aspects and it seems definitely not enough to compare them only from the point of view of average income levels. Economic inequalities are particularly important today and can be measured by income inequality between and within regions, among other things. Regional income inequality can be strongly linked to the poverty and social exclusion observed across regions and socioeconomic groups. Having said that, though, it is worth noting that the dependencies between average income, inequality and poverty are not straightforward and they depend on many other socio-economic factors.

It is well known that high income inequality can have several undesirable political and social consequences, such as poverty and the polarization of particular economic groups. Although they are usually perceived as similar and are in fact highly related concepts, inequality and poverty may not always come together. One can imagine a strictly egalitarian distribution of incomes, where all the income receivers are poor, or a highly dispersed population without poverty. Setting aside these not very likely situations, there is a strong empirical evidence based on income data from many countries that confirms a strong positive correlation between inequality and poverty. As a consequence, the countries with a more dispersed income distribution tend to have a higher relative level of income poverty, with only a few exceptions. The situation can be different, however, when analyzing these variables across regions. The relationships between economic development (measured by GDP per capita or average household income) and the corresponding income inequality (measured by, e.g. the Gini index) can also be different across countries and regions, what have been discussed in Jędrzejczak (2015) for the Polish and Italian regions. The observed regularities seem to partially confirm the well-known Kuznets' "inverted-U" relationship between the level of economic development and income inequality.

It is worth noting that the discussion on the possible relationship between GDP and the inequality level, which has been present in the economic literature since the mid-1950s, has produced very inconclusive results. We can find many countries (e.g. the Czech Republic) where the process of transformation was connected with no substantial inequality growth, while for many developed countries the inequality first declined, then increased again after a tipping point had been reached (e.g.: Italy in the 1990s). Deininger and Squire (1998, pp. 259-287), using their famous panel data on income inequality, did not find any significant relationship between income inequality and the level of development, even when country-effects were included into the

analysis. Li, Squire and Zou (1998, pp. 26-43) found out that the Kuznets relationship seems to work better in cross-sectional than time-series analyses.

Polish voivodeships differ in the level of economic development. There are many reasons for this differentiation and they concern various areas of socio-economic policy. The most important determinants of economic development include: the levels of regional GDP per capita, average incomes of households and individuals, characteristics of the labor market, demographic indicators, infrastructure and many others. Various works have recently been published on this issue, including: Malina (2020), Dańska-Borsiak (2024), where the authors take into account many indicators, the basic ones being GDP and household income, and then use multivariate statistical methods, which leads to interesting results. The data on GDP per capita and the estimates of household income suggest that there are substantial differences in regional income levels across the country, but little can be deduced from this about differences within regions and the relative number of people in different regions with income below the poverty line, as defined at the national level. In our study, we focus on income distribution characteristics which constitute an important dimension of regional disparities.

In the paper we would like to compare and contrast the regions of Poland (voivodeships) from the point of view of differences in their income distributions. The main aim is to analyze regional differences in the entire income distribution in Poland, taking into account not only average income levels, but also the parameters of income inequality and poverty. The study used non-parametric and parametric methods for estimating inequality and poverty measures, as well as cluster analysis methods for grouping the regions. This enabled the segmentation of voivodeships in terms of the estimated characteristics of the equivalent household income distribution. The applied parametric approach fills the gap in the existing grouping of regions, allowing for the estimation of income distribution parameters based on the Dagum distribution model, which has a strong stochastic basis.

The empirical evidence was micro-data coming from the Household Budget Survey conducted by Statistics Poland (GUS) for the year 2021. The detailed objectives of the paper are the following:

- analysis of income distributions in Polish voivodeships and their approximation based on the Dagum distribution,
- segmentation of Polish voivodeships according to estimated distribution characteristics, including the level of poverty and income inequality, using hierarchical cluster analysis methods.

The novelty of the paper is the segmentation of Polish voivodeships combined with the parametric approach based on the density curves of their income distributions.

2. Methodological remarks

2.1. Income distribution characteristics

Let $Y > 0$ be a random variable representing gross or net incomes as well as taxes. Let $F(y)$ be its cumulative distribution function (cdf), $f(y)$ the corresponding density function (pdf) and $F^{-1}(p) = \inf\{y: F(y) \geq p\}$ its quantile function for $0 < p < 1$.

Income distribution characteristics comprise many popular descriptive statistical measures, frequently applied in numerous statistical analyses, including measures of central tendency, dispersion and shape, but a special attention is paid to the measurement of distribution inequality. Income inequality refers to the degree of difference in earnings among various individuals or segments of a population.

Measures of inequality, also called concentration coefficients, are widely used to study income, welfare and poverty issues. Amongst numerous inequality measures the most popular is the Gini index, defined in terms of the Lorenz curve (Gini, 1912;1914). In its over 100-year history, the index has had numerous formulas and interpretations (see, e.g.: Jędrzejczak, 2011).

In practice, inequality coefficients, including the **Gini index**, are usually estimated from empirical data coming from representative samples. One can estimate the value of the Gini index from the survey data using the following formula (1), proposed by Fei et al. (1978):

$$\hat{G} = \frac{2\sum_{i=1}^{n}\left(w_i y_{(i)} \sum_{j=1}^{i} w_j\right) - \sum_{i=1}^{n} w_i y_{(i)}}{\left(\sum_{i=1}^{n} w_i\right)\sum_{i=1}^{n} w_i y_{(i)}} - 1, \tag{1}$$

where: $y_{(i)}$ – household incomes in a non-descending order,

w_i – survey weight for i-th economic unit,

$\sum_{j=1}^{i} w_j$ – rank of i-th economic unit in n-element sample.

Formula (1) can be applied for individual data coming from random samples, including the samples obtained within the Household Budget Survey, as it incorporates survey weights available for each statistical unit and is constructed on the basis of their inclusion probabilities. These weights make it possible to adjust for unequal selection probabilities and non-response bias.

Another type of inequality indices are the measures based on distribution quantiles or the estimators of these quantiles – these measures are very popular due to their simplicity and straightforward economic interpretation. In this class the most useful measures are: quintile and decile dispersion ratios (Panek, 2011) and the reciprocal of

the decile dispersion ratio called **the dispersion index for the end portions of the distribution**. The estimator of this index is:

$$\widehat{K}_{1:10} = \frac{\sum_{i \in GD_1} y_i}{\sum_{i \in GD_{10}} y_i} \tag{2}$$

where GD_j is *j*-th decile group.

This index (2) takes values from the interval [0,1] and if it is closer to 1, the inequality is lower (mean incomes in the extremal decile groups are the same).

A natural consequence of income inequality can be material poverty and social exclusion. Starting with the seminal publication of Sen (1976), numerous poverty measures have been proposed and analyzed but in practice the most popular are poverty indexes belonging to the class FGT (Foster–Greer–Thorbecke**),** which addresses three basic aspects of this phenomenon: poverty incidence, depth and severity (see: Panek, 2011). In this class, the most popular index is undoubtfully at-risk-of-poverty rate also called **poverty head-count ratio**, which can be estimated from the *n*-element sample by means of the following formula:

$$\widehat{W}_p = \frac{\sum_{i=1}^{n} I_i w_i}{\sum_{i=1}^{n} w_i}, \tag{3}$$

where: I_i – indicator function taking value 1 when *i*-th household's equivalent income is below a poverty line, and taking value 0 in the opposite situation,

w_i – survey weight for *i*-th economic unit.

The poverty threshold is usually assumed as 60% of the national median equivalent income - this approach is used in EU-SILC, the European Union Statistics on Income and Living Conditions, anchored by Eurostat. In some publications of Statistics Poland one can find another relative poverty line - 50% of mean equivalent income. In this study we utilize the Eurostat poverty threshold equal to $y^* = 0.6Me$, where *Me* is established as the estimate of median equivalent household income for the whole country.

Poverty gap index provides additional information regarding the distance of households from the poverty line. This measure captures the mean aggregate income or consumption shortfall relative to the poverty line across the whole population. It is obtained by adding up all the shortfalls of the poor and dividing the total by the number of the poor.

The estimator of the poverty gap index, which incorporates sampling weights w_i, is the following:

$$\widehat{PG}_p = \frac{\sum_{i=1}^{n_p} ((y^* - y_i)/y^*) w_i}{\sum_{i=1}^{n_p} w_i}, \tag{4}$$

where: y_i is household equivalent income,

y^*denotes poverty line (poverty threshold),

n_p is the number of poor households,
w_i accounts for the survey weight for i-th household.

The index (4) can be interpreted as minimum income that should have been transferred from the rich to the poor households to totally cancel poverty.

For many practical applications it seems reasonable to approximate empirical income data by means of a well-fitted theoretical model.

Since Pareto proposed his first income distribution model in 1896, many economists and mathematicians have tried to describe empirical distributions by simple mathematical formulas with a small number of parameters. These formulas can be useful for many reasons. Firstly, applying a theoretical model simplifies the analysis, because different distribution characteristics can be performed by the same parameters. Secondly, a theoretical model well fitted to the data can be used to the prediction of wage and income distributions in different divisions. Moreover, approximating empirical wage and income distributions using theoretical curves can smooth out irregularities resulting from imperfect data collection methods, which is often the case with income data. After decades of applications to real data in different divisions, we can propose the Dagum model as an excellent candidate to model income distributions of males and females (see: Jędrzejczak, Pekasiewicz, 2020).

The Dagum distribution is based on both empirical and stochastic foundations, similarly to the Pareto model (Dagum, 1977), but in contrast to the Pareto it fits population income quite well throughout the entire domain and has been successfully applied for many countries all over the world. Therefore, in the case of different subpopulations, e.g. regions or social groups, fitting the income distribution by means of the Dagum distribution can also be applied.

The probability density function of the Dagum distribution is given by (Kleiber and Kotz, 2003, Jędrzejczak and Pekasiewicz, 2020):

$$f(x) = \frac{ap}{b^{ap}} x^{ap-1} \left(1 + \left(\frac{x}{b}\right)^{a}\right)^{-1-p} \quad \text{for } x > 0, \tag{5}$$

and the corresponding cdf takes the form:

$$F(x) = \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-p} \quad \text{for } x > 0. \tag{6}$$

where $a > 0$ is scale parameter, $b > 0$ and $p > 0$ are shape parameters.

Using a fit to the Dagum distribution, the Gini index can be determined from the formula:

$$G = \frac{\Gamma(\hat{p})\Gamma\left(2\hat{p}+\frac{1}{\hat{a}}\right)}{\Gamma(2\hat{p})\Gamma\left(\hat{p}+\frac{1}{\hat{a}}\right)} - 1 \quad \text{for } \hat{a} > 1, \tag{7}$$

and poverty head-count ratio from the formula:

$$\widehat{W} = \hat{F}(y^*) \tag{8}$$

Deciles necessary to determine the decile groups in the definition of the dispersion index for the end portions of the distribution (see: formula (2)) are determined based on the quantile function of the Dagum distribution.

Any measure of deviation, based on a synthesis of the absolute or squared differences between observed and estimated frequencies obtained for class intervals, can be a candidate for assessing the goodness of fit. We will evaluate the following goodness-of-fit measures based on this approach: the Mortara index (A_1) and the Pearson index (A_2), together with the coefficient of distribution similarity *Wp,* also called an overlap measure (Gastwirth, 1976).

The coefficient of distribution similarity can be obtained from grouped data by means of the following formula:

$$Wps = \sum_{j=1}^{s} \min(w_j, \widehat{w}_j), \tag{9}$$

where: *s* is the number of class intervals, w_j and $\widehat{w}_j$ denote empirical and theoretical frequencies, respectively.

The other consistency measures mentioned above take the form (Zenga et al., 2010):

$$A_1 = \frac{1}{n}\sum_{j=1}^{s} |n_j - \hat{n}_j|, \tag{10}$$

$$A_2 = \sqrt{\frac{1}{n}\sum_{j=1}^{s} \frac{(n_j - \hat{n}_j)^2}{\hat{n}_j}}, \tag{11}$$

where n_j and $\hat{n}_j$denote the empirical and theoretical frequencies for class intervals.

2.2. Methods of cluster analysis adopted in the study

The segmentation of Polish voivodeships from the point of view of theoretical income distribution parameters estimated on the basis of the HBS 2003-2011 was proposed in Jędrzejczak, Kubacki (2017). Brzezińska (2018) applied correspondence analysis and the Ward hierarchical method to a multivariate analysis of poverty in Poland, based on the reports on economic poverty published by Statistics Poland in 2015.

In the paper, the hierarchical cluster analysis methods were used to perform the segmentation of Polish voivodeships in terms of selected income distribution characteristics, concerning income inequality and poverty indices.

At the first stage, the clusterSim package (Walesiak, Dudek, 2020) was applied to determine the optimal clustering procedure. In our study, complete-linkage clustering with the general distance measure $GDM1$ (Walesiak, 2016) turned out to be optimal, as it maximizes the silhouette index considered the best measure of classification quality. The procedure was finally implemented by adopting interval or ratio scales for measuring variables and taking into account the number of clusters no greater than 7, after a prior normalization of variables using classical standardization.

The general distance measure $GDM1$ for metric data is given by the following equation:

$$GDM1 = d_{ik} = \frac{1}{2} - \frac{\sum_{j=1}^{m}(z_{ij}-z_{kj})(z_{kj}-z_{ij}) + \sum_{j=1}^{m}\sum_{\substack{l=1 \\ l\neq i,k}}^{n}(z_{ij}-z_{lj})(z_{kj}-z_{lj})}{2\left[\sum_{j=1}^{m}\sum_{l=1}^{n}(z_{ij}-z_{lj})^2 * \sum_{j=1}^{m}\sum_{l=1}^{n}(z_{kj}-z_{lj})^2\right]^{\frac{1}{2}}} \quad (12)$$

where:
z_{ij} (z_{kj}, z_{lj}) – i-th (k-th, l-th) normalized observation on j-th variable.
i, k, l – the numbers of the objects; $i, k, l \in \{1, \ldots, n\}$
j – the number of the variable; $j \in \{1, \ldots, m\}$

The silhouette index (SI), applied to assess the general quality of clustering, is defined as (Kaufman, Rousseeuw, 1990; Dudek, 2020):

$$SI = \frac{1}{n}\sum_{i=1}^{n}\frac{b(i)-a(i)}{\max\{a(i);\ b(i)\}} \quad (13)$$

where:
n – number of objects in dataset,
$a(i)$ - average distance from object i to other objects belonging to cluster P_r, (object i belongs to cluster P_r),
$b(i)$ - minimum of average distance from object i to other objects belonging to cluster P_s (object i does not belong to cluster P_s).

Therefore,

$$a(i) = \sum_{k\in\{P_r\setminus i\}} d_{ik}/(n_r - 1) \quad (14)$$

$$b(i) = \min_{s\neq r}\left\{\sum_{k\in P_s} d_{ik}/n_s\right\} \quad (15)$$

where:
$\{d_{ik}\}$ – distance matrix,
n_r – number of objects in cluster P_r,
n_s – number of objects in cluster P_s.

The silhouette index values, defined by formula (13), range from -1 to +1 (Kaufman, Rousseeuw, 1990). The SI score from 0.70 to 1.00 indicates very good clustering results with well-separated clusters. Values between 0.50 to 0.70 identify moderate clustering quality with some clusters well separated and the others overlapping. If the SI index is between 0.25 and 0.50, one can recognize poor clustering performance, where many objects are misclassified or clusters overlap. The SI score below 0.25 indicates a very poor clustering quality, suggesting that no substantial structure has been found.

Calculations were performed using the R environment (clusterSim package) and the STATISTICA PL v. 13.3 statistical package.

3. Results

All the calculations were based on the random sample coming from the Household Budget Survey (HBS) 2021 provided by Statistics Poland and being the main source of information on income of the population of households. The variable of interest was household equivalent income, with an equivalence scale established as the LIS (Luxembourg Income Study) scale (square root of the number of household members). The calculations were carried out taking into account the sampling weights (for details see: GUS, 2024).

In the empirical analysis, which adopted a complete-linkage clustering of the Polish regions, we included the following variables, which in our opinion comprise basic statistical characteristics of income distributions:

- Average equivalent income,
- Poverty head-count ratio,
- Poverty gap index,
- Gini index,
- Dispersion index for the end portions of the distribution.

The above-mentioned variables reflect various aspects of income distribution, i.e. its central tendency, dispersion and shape, with particular emphasis on how total income in a voivodeship is divided among households. The choice of the variables was supported by the considerations presented in the Introduction regarding possible (and sometimes ambiguous) interactions between the level of average income, inequality and poverty.

Based on complete-linkage clustering with the general distance measure (12), the optimal number of clusters obtained was 5. The value of the silhouette index equal to 0.64 confirms that a meaningful group structure was found.

Figure 1 presents the dendrogram obtained for the Polish voivodeships. The layout of the dendrogram tells us which voivodeships are most similar to each other. "Branch" height indicates how similar or different they are: the greater the height, the greater difference. The red line illustrates the division of the dendrogram into 5 clusters.

The clusters created as a result of multivariate statistical analysis reveal hidden similarities and differences between voivodships that may be the basis for interesting statistical and economic conclusions. The regions constituting the clusters are not always adjacent to each other, which is a natural consequence of the procedure based on economic rather than geographical characteristics. Cluster no. 1 comprises the Dolnośląskie and Mazowieckie voivodeships, cluster no. 2 the Opolskie, Pomorskie and Wielkopolskie voivodeships, cluster no. 3 the Kujawsko-Pomorskie, Lubelskie, Łódzkie, Podlaskie and Warmińsko-Mazurskie voivodeships, cluster no. 4 the Lubuskie, Małopolskie and Śląskie voivodeships, and cluster no. 5 the Podkarpackie, Świętokrzyskie and Zachodniopomorskie voivodeships.

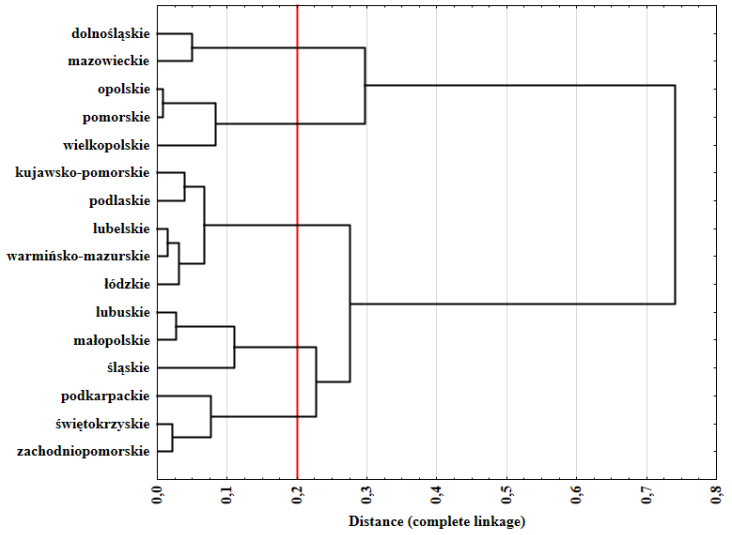


Figure 1: Dendrogram obtained for voivodeships

Source: authors' calculations

The identified clusters are visualized in Figure 2.

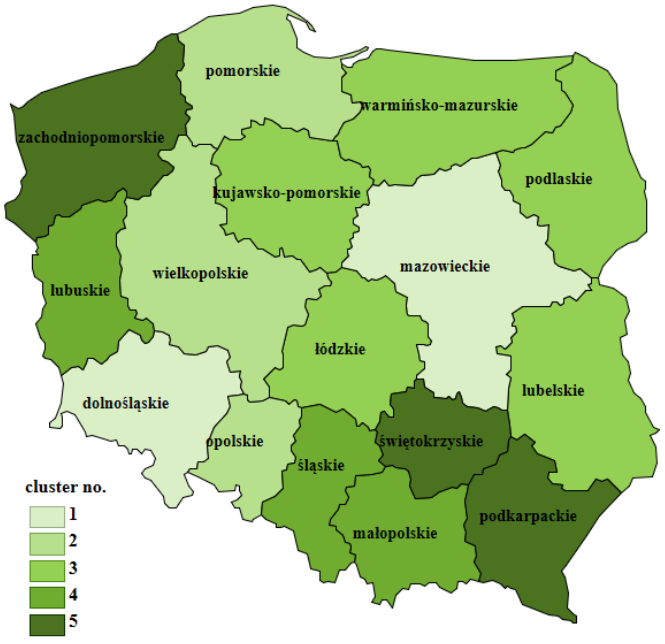


Figure 2: Visualization of the resulting clusters

Source: authors' calculations

The preliminary description of the clusters with regard to the analyzed variables is illustrated in a heatmap (Figure 3). In addition, descriptive statistics for the obtained clusters are presented in Table 1 and Figure 4.

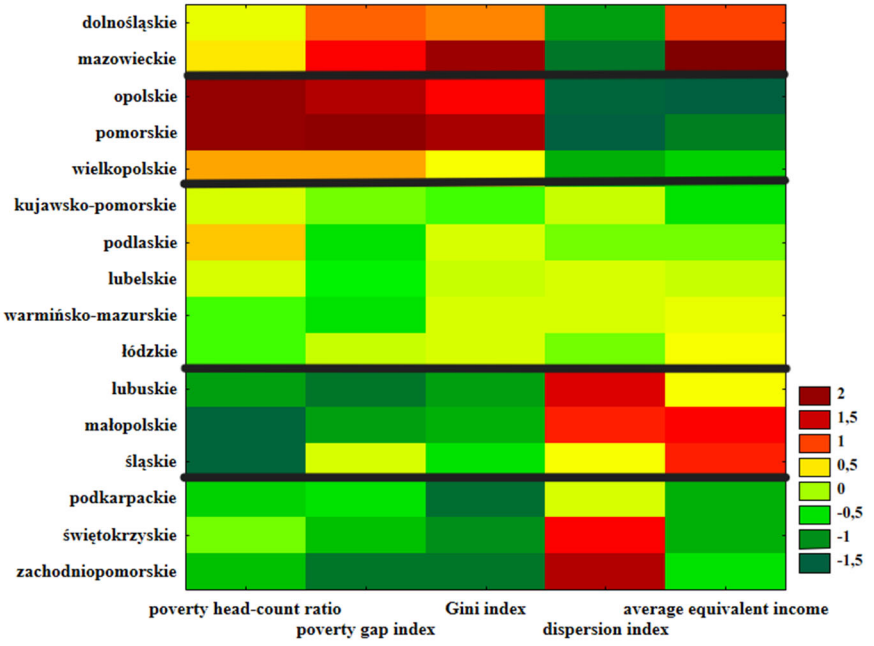


Figure 3: Heatmap – cluster characteristics

Source: authors' calculations

The legend of the heatmap (Figure 3) shows the boundaries of the intervals based on the standardized values of the variables taken into account (average equivalent income, poverty head-count ratio, poverty gap index, Gini index and dispersion index). The light colors indicate close-to-average values while dark red or dark green correspond to very high or very low levels of the variables for the selected regions, respectively.

The voivodeships in cluster no. 1 are characterized by moderate values of the poverty head-count ratio, high values of the poverty gap index and the Gini index, a low dispersion index and high average equivalent income.

The voivodeships in cluster no. 2 are described by high values of the poverty head-count ratio, the poverty gap index and the Gini index, as well as a low dispersion index and low average equivalent income.

The voivodeships in cluster no. 3 are characterized by an average level of all the indicators studied, except for the Wielkopolskie Voivodeship, which has higher values of the poverty head-count ratio and the poverty gap index compared to other voivodeships in this cluster.

The voivodeships in cluster no. 4 are described by low values of the poverty head-count ratio, poverty gap index and the Gini index, high dispersion index and high average equivalent income.

The voivodships included in cluster no. 5 are characterized by low values of poverty head-count ratio, the poverty gap index and the Gini index, as well as a fairly high dispersion index and low average equivalent income.

The conclusions drawn from the heatmap are supported by the values of the descriptive statistics presented in Table 1 and Figure 4.

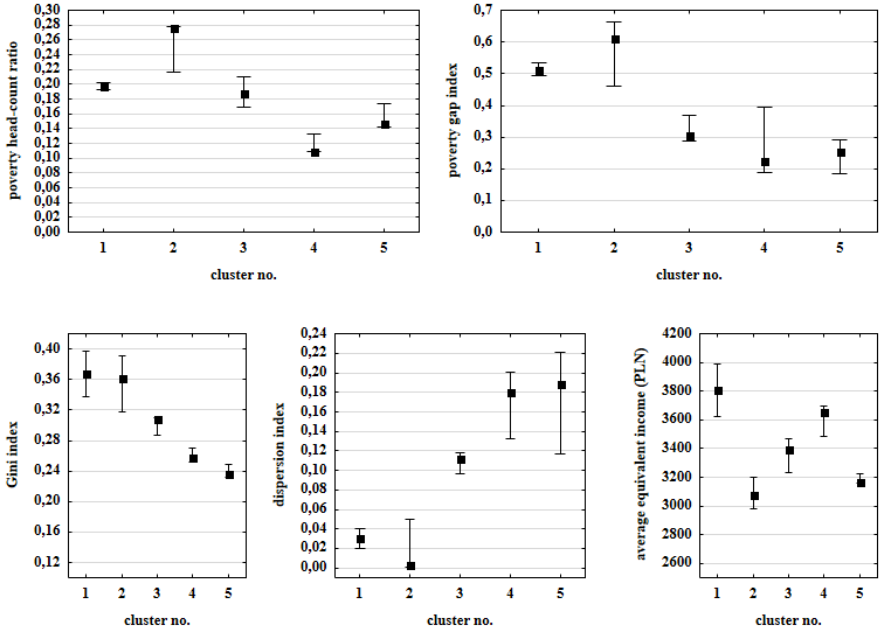


Figure 4: Comparison of clusters by poverty level and income inequality (median with range (min-max))

Source: authors' calculations

On average, the highest values of the poverty head-count ratio, the poverty gap index and the Gini index are observed for the voivodeships in cluster no. 2 (Opolskie, Pomorskie and Wielkopolskie voivodships); these voivodeships are also characterized by the lowest equivalent income on average.

On average, the highest values of the Gini index with a relatively high poverty gap index, are observed for voivodeships belonging to cluster no. 1 (Dolnośląskie and Mazowieckie voivodeships); these voivodeships are also characterized by the highest average equivalent income.

Table 1: Cluster characteristics – descriptive statistics

Variable	Measure	Cluster no.				
		1	2	3	4	5
		n=2	n=3	n=5	n=3	n=3
Poverty head-count ratio	mean	0.20	0.26	0.19	0.12	0.15
	SD	0.01	0.04	0.02	0.01	0.02
	CV	3.5%	13.7%	9.4%	11.1%	11.0%
Poverty gap index	mean	0.51	0.58	0.32	0.27	0.24
	SD	0.03	0.11	0.04	0.11	0.06
	CV	6.0%	18.1%	11.2%	40.7%	22.8%
Gini index	mean	0.37	0.36	0.30	0.26	0.24
	SD	0.04	0.04	0.01	0.01	0.01
	CV	11.5%	10.3%	3.1%	3.7%	3.7%
Dispersion index	mean	0.03	0.02	0.11	0.17	0.18
	SD	0.01	0.03	0.01	0.03	0.05
	CV	48.3%	156.4%	9.9%	20.4%	30.5%
Average equivalent income	mean	3808.76	3088.70	3380.18	3612.22	3185.19
	SD	258.74	109.09	95.87	114.84	37.94
	CV	6.8%	3.5%	2.8%	3.2%	1.2%

Note: SD - standard deviation; CV – coefficient of variation

Source: authors' calculations

The voivodeships in cluster no. 3 (Kujawsko-Pomorskie, Lubelskie, Łódzkie, Podlaskie and Warmińsko-Mazurskie) are characterized by an average level of all the examined variables.

The voivodeships belonging to cluster no. 4 (Lubuskie, Małopolskie and Śląskie voivodeships) and cluster no. 5 (Podkarpackie, Świętokrzyskie and Zachodniopomorskie voivodeships) are characterized by, on average, the lowest values of the poverty head-count ratio, the poverty gap index and the Gini index, as well as the highest values of the dispersion index. The compared clusters are distinguished from each other by the level of average equivalent income (high in cluster no. 4 and low in cluster no. 5).

The approximation of income distributions for voivodeships, carried out by means of the Dagum model, and the application of formulas (7)-(8) did not affect the classification of voivodeships, however, it smoothed out irregularities and shed light on differences in the location and shape of the compared subpopulations. Figure 5 shows the empirical and theoretical distributions (determined by formula (5)) for voivodeships representing individual clusters. Table 2 contains the parameters of the distributions adjusted to the empirical data, while Table 3 values of fit measures calculated on the basis of formulas (9)-(11).

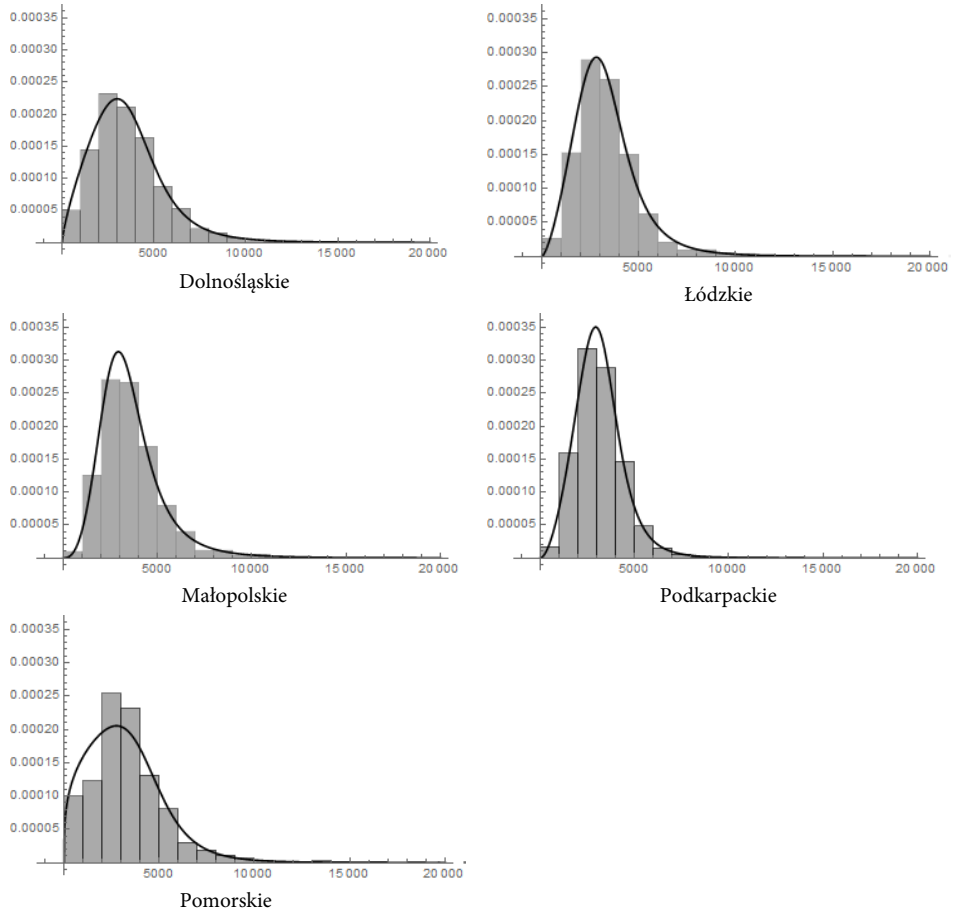


Figure 5: Income distributions in selected Polish voivodeships and fitted Dagum distributions

Source: authors' calculations

Table 2: Dagum distribution parameters for cluster representatives

Voivodeship	Dagum distribution parameters		
	a	***b***	***p***
Dolnośląskie	4.402	4667.61	0.403
Łódzkie	4.456	3775.38	0.554
Małopolskie	4.093	3456.44	0.868
Podkarpackie	5.815	3719.32	0.469
Pomorskie	4.964	4981.48	0.265

Source: authors' calculations

Table 3: Goodness-of-fit measures for cluster representatives

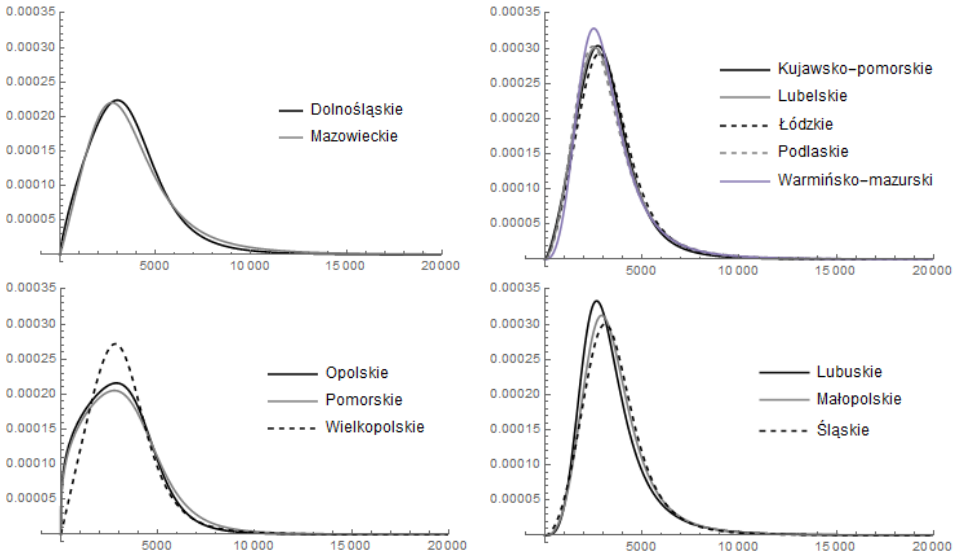
Voivodeship	Goodness-of-fit measure		
	Wps	A_1	A_2
Dolnośląskie	0.996	0.008	0.437
Łódzkie	0.984	0.031	1.572
Małopolskie	0.995	0.008	0.659
Podkarpackie	0.959	0.082	0.127
Pomorskie	0.930	0.139	0.228

Source: authors' calculations

On the basis of the goodness-of-fit measures presented in Table 3, it can be noted that the consistency of the empirical and theoretical distributions is satisfactory – the similarity coefficient of the distributions generally exceeds 0.95, i.e. the compared empirical and theoretical frequencies are very similar. The high goodness-of-fit of empirical and theoretical distributions is also confirmed by the low values of the A1 and A2 measures. The worst results were obtained for the Pomorskie voivodeship, which can also be seen in Figure 5.

Both the values of goodness-of-fit measures and the plots of empirical and theoretical distributions suggest good consistency of the income distributions of the representatives of each cluster with the Dagum model. Analogous results were obtained for the other voivodeships.

The plots of the Dagum densities estimated for the voivodeships constituting individual clusters are presented in Figure 6.



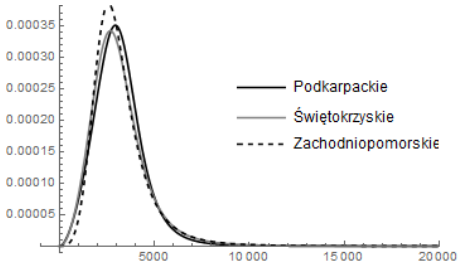


Figure 6. Dagum distributions estimated for voivodeships within the obtained clusters

Source: authors' calculations

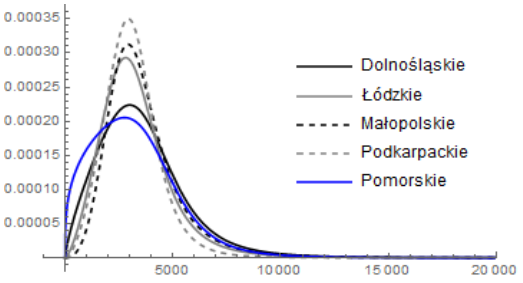


Figure 7: Dagum distributions estimated for cluster representatives

Source: authors' calculations

Figure 6 shows that the differences between the estimated income distributions for the members of each cluster are relatively small, but the differences between the income distributions for voivodeships from different clusters are substantial, as can be seen in Figure 7.

4. Conclusion

The aim of the paper was to reveal regional differences in income distributions in Poland, with a special attention paid to income inequality and poverty. The statistical analysis, based on individual data from the Household Budget Survey, applied selected non-parametric and parametric methods for estimating inequality and poverty measures, as well as the complete-linkage clustering for grouping the Polish voivodeships. The approximation of empirical distributions performed using the three-parameter Dagum model turned out to be an appropriate tool for both parametric estimation and visualization of the results. The estimated distribution characteristics enabled the effective segmentation of voivodeships, which was additionally confirmed by very similar Dagum density curves obtained for each cluster. On the contrary, the

differences between clusters are substantial. For the richest cluster, which consists of Dolnośląskie and Mazowieckie voivodeships, the highest poverty gap has been observed, which can be connected with discrepancies located at the left tail of income distributions. The most difficult situation was observed for the cluster containing: Opolskie, Pomorskie and Wielkopolskie - the cluster represents highest values of the average poverty head-count ratio, the poverty gap index and the Gini index and is also characterized by the lowest equivalent income. In the cluster including: Podkarpackie, Świętokrzyskie and Zachodniopomorskie, the poverty indices are small due to relatively small overall inequality, even if the average income level for this cluster is low.

The results of the calculations confirmed the assumption that income distributions in Poland differ significantly across regions. The obtained clusters allowed detecting groups of regions that may require separate social policies aimed at increasing household income, or rather at reducing income inequality.

References

Barca, F., (2009). An agenda for a reformed cohesion policy: A place-based approach to meeting European Union challenges and expectations, Independent Report prepared at the request of Danuta Hübner. *Commissioner for Regional Policy, EU Commission*, Brussels.

Brzezińska, J., (2018). Statistical Analysis of Economic Poverty in Poland Using R. *Econometrics. Advances in Applied Data Analysis*, 22(2), pp. 45–53.

Bonesmo Fredriksen, K., (2012). Income Inequality in the European Union. *OECD Economics Department Working Papers*, n. 952, *OECD Publishing*, Paris.

Dagum, C., (1977). A New Model of Personal Income Distribution: Specification and Estimation. *Economie Appliquee*, 30, pp. 413-437.

Dańska-Borsiak, B., (2024). Ocena adekwatności PKB per capita jako miary poziomu życia w powiatach. *Wiadomości Statystyczne. The Polish Statistician*, 69(7), pp. 1–21.

Deininger, K., Squire, L., (1998). New Ways of Looking at Old Issues: Inequality and Growth. *Journal of Development Economics,* 57(2).

Dudek, A., (2020). Silhouette Index as Clustering Evaluation Tool, (in) Jajuga K. et al. (eds.), Classification and Data Analysis, Studies in Classification, Data Analysis, and Knowledge Organization. *Springer*, pp. 19–33.

Eurostat, (2014). *Statistics in Focus 12/2014*. https://ec.europa.eu/eurostat/statistics-explained/.

Fei J., Ranis G., Kuo S., (1978). Growth and the Family Distribution of Income by Factor Components. *Quarterly Journal of Economics*, 92, pp. 17–53.

Gastwirth, J. L., (1975). Statistical Measures of Earning Differentials. *American Statistician*, 29, pp. 32–35.

Gini, C., (2012). Variabilità e mutabilità: contributo allo studio delle distribuzioni e delle relazioni statistiche. *Studi Economico-Giuridici*, *Facoltà di Giurisprudenza della Regia Universit`a di Cagliari*, anno III, parte II, Cuppini, Bologna.

Gini, C., (1914). Sulla Misura Della Concentrazione e Della Variabilita` dei Caratteri, [w:] *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti. Anno Accademico 1913–1914*, Tomo LXXIII – Parte Seconda.

GUS, (2024|). Budżety gospodarstw domowych w 2023 roku. *Zakład Wydawnictw Statystycznych*, Warszawa. https://stat.gov.pl/obszary-tematyczne/warunki-zycia/dochody-wydatki-i-warunki-zycia-ludnosci/budzety-gospodarstw-domowych-w-2023-roku,9,22.html.

Li, H., Squire, L. and Zou H., (1998). Explaining International and Intertemporal Variations in Income Inequality. *Economic Journal*, 108 (446), pp. 26–43.

OECD, (2008). Growinq Unequal: Income Distribution and Poverty in OECD Countries. *OECD Publishing*, Paris. https://www.oecd-ilibrary.org/social-issues-migration-health/growing-unequal_9789264044197-en.

OECD, (2011). Divided We Stand. Why Inequality Keeps Rising. *OECD Publishing*, Paris. https://www.oecd-ilibrary.org/social-issues-migration-health/the-causes-of-growing-inequalities-in-oecd-countries_9789264119536-en.

OECD, (2015). In It Together: Why Less Inequality Benefits All. *OECD Publishing*, Paris. https://www.oecd-ilibrary.org/employment/in-it-together-why-less-inequality-benefits- all_9789264235120-en.

OECD, (2024). Income and wealth inequalities, [in:] *Society at a Glance 2024: OECD Social Indicators.*

Jędrzejczak, A., (2011). Metody analizy rozkładów dochodów i ich koncentracji. *Wydawnictwo Uniwersytetu Łódzkiego*, Łódź.

Jędrzejczak, A., (2015). Regional Income Inequalities in Poland and Italy. *Comparative Economic Research*, 18(4), pp. 27–45.

Jędrzejczak, A., Kubacki, J., (2017). Analiza rozkładów dochodu rozporządzalnego według województw z uwzględnieniem czasu, [in:] Klasyfikacja i analiza danych.

Teoria i zastosowanie. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu. Taksonomia*, 29(469), pp. 69–81.

Jędrzejczak, A., (2023). Klasyczne i nieklasyczne metody analizy nierówności dochodowych. *Wydawnictwo Uniwersytetu Łódzkiego*, Łódź.

Jędrzejczak, A., Pekasiewicz, D., (2020). Teoretyczne rozkłady dochodów gospodarstw domowych i ich estymacja. *Wydawnictwo Uniwersytetu Łódzkiego*, Łódź.

Jędrzejczak, A., Pekasiewicz, D., (2022). Nierównomierność ekwiwalentnych dochodów gospodarstw domowych w województwie łódzkim. *Wiadomości Statystyczne*, 67(6), pp. 29–51.

Kaufman, L., Rousseeuw, P. J., (1990). Finding groups in data: an introduction to cluster analysis. *John Wiley*, New York.

Kleiber, C., Kotz, S., (2003). Statistical Size Distributions in Economics and Actuarial Sciences. *Wiley*, Hoboken.

Malina, A., (2020). Analiza przestrzennego zróżnicowania poziomu rozwoju społeczno-gospodarczego województw Polski w latach 2005–2017. *Nierówności Społeczne a Wzrost Gospodarczy*, No. 61(1), pp. 138–155.

Panek, T., (2011). Ubóstwo, wykluczenie społeczne i nierówności. Teoria i praktyka pomiaru. *Oficyna Wydawnicza SGH*, Warszawa.

Sen, A., (1976). Poverty: An Ordinal Approach to Measurement. *Econometrica*, 44(2), pp. 219–231.

Walesiak, M., Dudek, A., (2020). The Choice of Variable Normalization Method in Cluster Analysis, In Soliman KS (ed.), *Education Excellence and Innovation Management: A 2025 Vision to Sustain Economic Development During Global Challenges*, pp. 325–340, ISBN 978-0-9998551-4-1.

Walesiak, M., (2016). Uogólniona miara odległości GDM w statystycznej analizie wielowymiarowej z wykorzystaniem programu R. *Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu*, Wrocław.

Zenga, M. M., Pasquazzi, L. and Zenga, M., (2010). Rapporto n. 188: First applications of a new three-parameter distribution for non-negative variables. Mediolan: *Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali, Università degli Studi di Milano Bicocca*.

STATISTICS IN TRANSITION new series, September 2025
Vol. 26, No. 3, pp. 155–180, https://doi.org/10.59139/stattrans-2025-032
Received – 09.11.2023; accepted – 30.04.2025

Modelling the asymmetric relationship between energy and CO_2 emissions in the Visegrad group: empirical evidence from a panel NARDL approach

Błażej Suproń[1]

Abstract

The study aims to understand the impact of renewable energy consumption, non-renewable energy consumption, and economic growth on carbon dioxide (CO_2) emissions per capita (measured in metric tonnes) in the Visegrad countries between 1991 and 2021. Using a Nonlinear Autoregressive Distributed Lag (NARDL) model for panel data, the research captures both long-term dependencies and short-term dynamics. The results show that a reduction in CO_2 emissions yielded by a significant long-term decrease in non-renewable energy consumption is proportionally larger than the increase in the emissions caused by the growth in the consumption of such energy. What is more, GDP growth in the V4 countries increases CO_2 emissions, but GDP decline contributes significantly more to the reduction in emissions. In contrast, renewable energy consumption consistently reduces CO_2 emissions over the long term, with no significant asymmetry detected. In the short term, both economic growth and non-renewable energy consumption increase CO_2 emissions. The error correction factor suggests a rapid adjustment of CO_2 emissions towards a long-term equilibrium, with a decreasing trend over time. These results have some policy implications, i.e. they suggest that for the V4 countries, increasing investment in technologies and solutions that improve energy efficiency will be crucial for reducing non-renewable energy consumption and, consequently, CO_2 emissions, without negatively impacting economic growth. Additionally, stable and long-term promotion of renewable energy should be a policy priority in order to effectively contribute to emission reductions.

Key words: renewable energy, non-renewable energy, CO_2 emissions, growth, NARDL panel.

1. Introduction

Energy is a fundamental production factor in modern economies, without which no economic activity is possible (Stern 2019). The efficient use of energy resources is a key competitive advantage for individual economies, influencing their efficiency and

[1] Department of Economics, West Pomeranian University of Technology in Szczecin, Szczecin, Poland. E-mail: bsupron@zut.edu.pl. ORCID: https://orcid.org/0000-0002-7432-1670.

cost intensity and therefore their economic development (Fanchi 2023). While energy is a key driver of economic growth, its production, particularly from conventional sources, has significant environmental implications (Ndoricimpa 2017). Since the Industrial Revolution, energy production, particularly from coal and oil, has contributed significantly to the emission of greenhouse gases, especially carbon dioxide (CO_2), which numerous studies have identified as a primary driver of global warming (Al-Ghussain 2019). The excessive release of greenhouse gases has resulted in a global temperature increase of 1.1°C, with further rises occurring at a rate of 0.2°C per decade (Pörtner et al. 2022).

The economic implications of climate change have become a key topic of discussion at the political and scientific levels. It is imperative that global economies pursue sustainable solutions in energy production and consumption (Czyżewski, Polcyn, and Brelik 2022). The need to improve environmental conditions and halt global warming has driven the global community to adopt regulations aimed at reducing the use of fossil fuels in energy production and thereby cutting CO_2 emissions. This effort began with the signing of the Kyoto Protocol in 1997, followed by the Paris Agreement in 2015 (Flanker 2016). Furthermore, in 2019, the European Union initiated the European Green Deal, which aims to halve CO_2 emissions by 2030 and achieve climate neutrality by 2050 (Samper, Schockling, and Islar 2021). The indicated actions have contributed to the green transformation, which can be defined as the integration of economic growth with environmental stewardship to ensure a high quality of life for current and future generations, while effectively and rationally utilizing available resources in line with civilizational progress (Cheba et al. 2022).

Achieving the stated goals while maintaining conditions for economic growth requires replacing non-renewable energy sources with renewable ones, which is part of the energy transformation process (Adedoyin et al. 2021). For countries with a significant reliance on fossil fuels in their energy mix, a transition to renewable energy sources will be particularly significant. The Visegrad Group is an example. It comprises the Czech Republic, Hungary, Poland and Slovakia. These countries have undergone substantial economic transformations, shifting to market economies while experiencing significant changes in their production structures and numerous economic shocks (Ambroziak et al. 2021). Despite experiencing substantial economic growth, these nations still rely significantly on fossil fuels, including coal, for energy production (Brodny and Tutak 2021). Consequently, the energy transformation process represents a significant challenge for these countries, given the substantial investments required to increase the share of renewable energy sources in their energy mix (Marzec and Zioło 2016). As highlighted by Bigos (2017), Poland and the Czech Republic are facing significant challenges in their energy sectors due to their substantial reliance on coal for electricity generation and heating. Slovakia faces greater challenges in adapting its industrial sector, while Hungary aims to achieve greater energy independence from

fossil fuels, including variable gas, by 2030. However, the impact of potential energy transformation on economic growth and CO_2 emissions in these countries is still not fully understood, according to Suproń and Myszczyszyn (2023).

A growing body of research from a range of regions and countries indicates that renewable energy can play a role in reducing CO_2 emissions while supporting economic growth (Alper and Oguz 2016; Bhuiyan et al. 2022). However, the results are not always definitive. For example, Antonakakis et al. (2017) found that, in a sample of 106 countries, there is no definitive evidence that the use of renewable energy promotes growth more efficiently and sustainably for the environment. Many previous studies have concentrated on the linear relationships between energy, economic growth and environmental pollution (Mardani et al. 2019). However, Khan i Sun (2024) posits that assuming a linear relationship is a flawed approach that leads to oversimplified conclusions. This is because it assumes that variables change proportionally, which may not fully account for external shocks. It is important to note that many economic relationships based on data represent non-linear relationships, which reflect the nature of economic events (Chauvet and Jiang 2023).

Furthermore, research in environmental economics is increasingly emphasising the non-linearity of relationships between CO_2 emissions, energy production and economic factors (Iorember, Usman, and Jelilov 2019; Kirikkaleli, Abbasi, and Oyebanji 2023; Akadiri and Adebayo 2022). As Kouton (2019) notes, asymmetries emerge due to a range of economic and natural shocks, including technological changes, policy shifts, economic policies, and even natural disasters. Employing non-linear models facilitates a more comprehensive examination of time series, particularly during periods when such shocks have occurred.

Over the past 30 years, the Visegrad Group countries have faced a series of significant internal and external challenges. These include economic transformation, accession to the EU, the financial crisis, the Eurozone crisis and the impact of the global pandemic. As a result, economic data for these countries is characterized by a few structural breaks and non-linear stochastic processes. Additionally, some studies suggest that the relationships between CO_2 emissions and the economy and energy sector in these countries may exhibit non-linear U- or N-shaped patterns (Suproń 2024; Supron and Myszczyszyn 2023).

Considering the above rationale and the V4 countries' commitments to achieving climate neutrality, the aim of the study was to determine the short- and long-term impacts of renewable and non-renewable energy consumption, as well as economic growth, on CO_2 emissions in the V4 countries, incorporating non-linear effects. Based on the study objectives and existing literature, the following research hypotheses were formulated:

Hypothesis H1: There is a negative and asymmetric relationship between renewable energy consumption and CO_2 emissions in the Visegrad countries in the long term.

Hypothesis H2: There is a positive and asymmetric relationship between non-renewable energy consumption and CO_2 emissions in the Visegrad countries in the long term.

Hypothesis H3: Economic growth has a positive and asymmetric impact on CO_2 emissions in the Visegrad countries in the long term.

The choice of the research method was based on an analysis of the existing literature on the subject. Many studies analyzing the relationship between energy consumption, production, and CO_2 emissions use panel data, including econometric techniques such as linear VAR, ARDL, and the non-linear generalized method of moments (GMM) (Masron and Subramaniam 2018; Rasoulinezhad and Saboori 2018; Antonakakis, Chatziantoniou, and Filis 2017).

This study employs a modern estimation method based on the NARDL model, which differs from previous research in this field. The choice of this methodological approach is driven by the fact that traditional panel methods often encounter issues related to heteroskedasticity and cross-sectional dependence in time series data. These issues limit the applicability of traditional panel methods and necessitate compromises in the estimation process. The use of nonlinear methods addresses these methodological challenges and yields more accurate results. Additionally, NARDL allows for the modelling of long-term effects for both positive and negative deviations from equilibrium, as well as nonlinear short-term effects, which are often overlooked in traditional panel regression models (Shin, Yu, and Greenwood-Nimmo 2014; B. Li et al. 2023). Furthermore, these models are resilient to structural shifts in the data, a common challenge with economic indicators for Central and Eastern European countries, which have undergone substantial economic transformation and encountered numerous internal and external disruptions over the past three decades.

The novel aspects of this study can be summarized as follows. Firstly, this study employs a contemporary methodology based on the NARDL model, offering new insights that differ from those of earlier research. Secondly, the findings related to Central and Eastern European countries, particularly the Visegrad Group, have significant implications for countries that may undergo similar transformation processes in the future. Finally, to the best of our knowledge, no previous studies have used NARDL methodology to study the V4 group with relationships of energy, GDP and CO_2.

The remainder of this paper is structured as follows. Section 1 provides a review of the recent literature. Section 2 presents the methodology and the empirical model. Section 3 describes the data and the results of the empirical analysis. Section 4 provides a discussion to the results of the study. The final section provides a summary.

2. Literature review

There is a large body of theoretical and empirical research focusing on the relationship between energy consumption, economic growth, and CO_2, in both highly developed and developing economies (Bąk and Cheba 2023). The literature can be divided into three main streams of research: the first focuses on the causal relationship between energy consumption and CO_2, the second focuses on the relationship between production and air pollution. The third combines both, providing a unified framework for identifying the links between energy consumption, CO_2 emissions and economic growth (Antonakakis, Chatziantoniou, and Filis 2017).

Based on ample empirical evidence, existing aggregate reviews of the literature indicate the existence of unidirectional and bidirectional positive relationships between economic growth and CO_2, (Omri 2014; Ozturk 2010; Tugcu, Ozturk, and Aslan 2012). In contrast, some studies to date have not confirmed a significant interaction between economic growth and higher CO_2. However, it should be pointed out that individual studies differ both in the selection of countries, the study periods, the length of the time series and the methodology used (Haberl et al. 2020).

As previously mentioned, the second part of the research focused on the topics of energy consumption and production and the impact of the indicated variables on CO_2. Considering the previous research results in this area, most studies confirmed a positive unidirectional relationship between increased CO_2, and electricity production. (Aziz et al. 2022). Some authors have conducted more detailed analyses comparing the impact of renewable and non-renewable energy on economic growth. Based on the evidence, it has been indicated that renewable energy consumption promotes CO_2, but only after it has exceeded a certain share in the energy mix (Shahbaz & Sinha, 2019; Tiba & Omri, 2017). Thus, according to the existing scientific consensus, renewable energy appears to be an effective instrument for sustainable decarbonization (Bourcet 2020).

The rapid development of econometric methods has meant that previous studies have been conducted for both individual countries and groups of countries. The use of panel methods has contributed to the development of new results and conclusions from the estimation of broad data sets. A detailed characterization of recent studies using panel methods is presented in Table 1.

Table 1: A review of recent panel data studies and their results in the impact of economic growth and energy consumption on CO_2 emissions

Source	Country (region)	Causality	Model	Period
Allard et al. (2018)	74 countries	REW→ CO_2 (-)	PQARDL	1994-2012
Anwar et al. (2021)	ASEAN countries	REW→ CO_2 (-) NREW→ CO_2 (+)	FMOLS / DOLS	1980-2013
Armeanu et al. (2017)	EU Countries	REW→ GDP (+)	PVAR	2003-2014

Table 1: A review of recent panel data studies and their results in the impact of economic growth and energy consumption on CO_2 emissions (cont.)

Source	Country (region)	Causality	Model	Period
Bekun et al. (2019)	EU countries	REW→ CO_2 (-) NREW→ CO_2 (+)	ARDL-PMG	1996-2014
Ben Jebli et al. (2020)	102 countries	REW→ CO_2 (-)	GMM	1990-2015
Bhattacharya et al. (2017)	85 countries	NREW→ CO_2 (+) REW→ CO_2 (-)	FMOLS	1991-2012
Busu & Nedelcu (2021)	EU countries	REW→ CO_2 (+)	OLS panel	2000-2019
Cai et al. (2018)	G7 countries	REW→ CO_2 (-)	ARDL-PMG	1965-2015
Charfeddine & Kahia (2019)	MENA countries	REW→ CO_2 (-)	PVAR	1980-2015
Chen et al. (2022)	97 countries	REW→ CO_2 (-)	DPTRM	1995-2015
Cialani (2017)	150 Countries	CO_2 ⇔ GDP	ECM panel	1960-2008
Gozgor et al. (2018)	OECD Countries	NREW→ GDP (+) REW→ GDP (+)	ARDL	1990 - 2013
Inglesi-Lotz & Dogan (2018)	Africa countries	NREW→ CO_2 (+) REW→ CO_2 (-)	DOLS	1980-2011
Ito (2017)	42 developed countries	REW→ GDP (-)	OLS panel	2002-2011
Lazăr et al. (2019)	CEE Countries	GDP→ CO_2 (+)	FMOLS	1996-2015
Li et al. (2020)	Post-Communist Economies	GDP→ CO_2 (+)	OLS panel	1996-2018
Ma et al. (2021)	Germany and France	REW→ CO_2 (-) NREW→ CO_2 (+)	FMOLS	1995-2015
Muço et al. (2021)	European transition economies	GDP→ CO_2 (+) REW→ GDP (+)	PVAR	1990-2018
Ozcan and Ozturk (2019)	Emerging countries	REW ≠ GDP (+)	Causality panel	1990-2016
Pope et al. (2019)	EU Countries	REW→ GDP (+)	PVECM	1990-2014
Papież et al., (2019)	EU Countries	REW≠GDP NREW ≠GDP	PVAR	1995-2015
Saqib et al. (2022)	GCC	NREW→ CO_2 (+)	OLS panel	1993-2019
Suproń (2024)	Visegrad countries	NREW → CO_2 (+) GDP → CO_2 (+)	Asymmetric FGLS	1991-2021
Chen et al. (2022)	24 countries	REW→ CO_2 (-) NREW→ CO_2 (+)	ARDL-PMG	1995-2014
Zhang and Liu (2019)	NSEA-10 countries	REW→ CO_2 (-) NREW→ CO_2 (-)	FMOLS	1995-2014

REW - use/consumption of renewable energy; NREW - use/consumption of non-renewable energy; CO_2 - carbon dioxide emissions; GDP - economic growth; (+) positive impact; (-) negative impact.

Source: author's study.

As suggested by the summary presented in Table 1 on the impact of renewable and non-renewable energy consumption and economic growth on CO_2, some results show consensus, while others often generate contradictory conclusions. These discrepancies may be due to the different data used, different econometric methods and the quality of the time series, among other things (Antonakakis, Chatziantoniou, and Filis 2017; Stern and Common 2001; Yang and Zhao 2014).

Despite the wide range of research in this area, there still seems to be a research gap regarding the Visegrad countries. According to the literature analysis presented by Suproń & Myszczyszyn (2023) a panel study examining the relationship between energy consumption from different sources and economic growth on CO_2 emissions in the V4 countries has not yet been carried out.

Developments in econometrics, newer estimators and models make it possible to provide new evidence in many of the areas studied. An analysis of the existing literature indicates that the relationship between energy consumption and CO_2 emissions has not been extensively investigated using non-linear ARDL-PMG models. To date, this model has been used to establish relationships between CO_2, and labor force (Naseer et al. 2022) and foreign investment (Deng, Liu, and Sohail 2022) in the BRICS countries. There have also been studies using the NARDL panel model to analyze the asymmetric impact of education on CO_2 in MINT countries (Ahmed et al. 2022) and the ICT sector on environmental pollution in GCC countries (Saqib, Duran, and Hashmi 2022).

Although the impact of renewable and non-renewable energy and economic growth on CO_2 emissions has been extensively studied, significant research gaps remain, particularly regarding the use of non-linear methods for V4 economies. Non-linear analysis is crucial in this context, as it enables a more detailed examination of the relationship between changes in independent variables and CO_2 emissions (Ahmed et al. 2022). Despite this, the current literature does not confirm the existence of long-term asymmetric relationships between energy production from renewable and non-renewable sources, economic growth, and CO_2 emissions in the Visegrad countries.

3. Data and methodology

This study examined the relationship between renewable (REW) and non-renewable (NREW) electricity consumption and economic growth (GDP), as well as carbon dioxide (CO_2) emissions. The research was conducted for the V4 countries, with a time series covering the years from 1991 to 2021. Table 2 sets out the full characteristics of the time series studied. The study used a panel data set, with the variables transformed to the form of natural logarithms.

Table 2: Data, units and data sources

Variable	Full name	Unit	Source
CO_2	Carbon dioxide (CO_2) emissions per capita	Metric tonnes per capita	World Development Indicators (WDI)
GDP	GDP per capita	GDP per capita in constant 2015 US dollars	World Development Indicators (WDI)
REW	Electricity consumption from renewable sources	Tonnes of oil equivalent (per capita)	EEA / Eurostat
NREW	Electricity consumption from non-renewable sources	Tonnes of oil equivalent (per capita)	EEA / Eurostat

Source: World Bank Data and Eurostat Database[2].

In this study, the NARDL methodology proposed by Shin et al. (2014) is employed to determine the relationship between CO_2 emissions, renewable energy consumption, non-renewable energy consumption, and economic growth. The NARDL-PMG model based on the ARDL-PMG model, which is applied to panel data, provides short- and long-term parameter estimation and support for integrated variables at both I(0) and I(1) levels (Pesaran, Shin, and Smith 2001). NARDL models are also effective for small samples and robust to structural breaks in the data. The NARDL approach is more flexible in relation to the dynamics of cointegration between variables. In addition, NARDL models allow the assessment of both long- and short-term effects of the independent variables on the dependent variable with consideration of asymmetric effects. Considering the purpose of the study and the literature, the basic specification of the model was defined as follows:

$$\Delta CO_{2_t} = f(GDP, REW, NREW) \tag{1}$$

To drop serial correlation and heteroskedasticity, the model was transformed to a log-linear form:

$$\Delta CO_{2_t} = \beta_0 + \beta_{1t} lnGDP_t + \beta_{2t} lnREW_t + \beta_{3t} lnNREW_t + \varepsilon_t \tag{2}$$

[2] WDI dataset: https://databank.worldbank.org/metadataglossary/world-development-indicators/series/EN.ATM.CO2E.PC, accessed on 31.05.2023; Eurostat dataset: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Renewable_energy_statistics, accessed on 31.05.2023.

The variables lnGDP and lnREW have been transformed into subtotals according to the assumptions of the NARDL model to highlight their asymmetric effects. This allows the study to consider both their positive and negative effects on CO_2 emissions. The subtotals of the variables studied are presented below in the form of equations (Shin, Yu, and Greenwood-Nimmo 2014):

$$\sum_{n=1}^{t} \Delta lnGDP_{it}^{-} = \sum_{n=1}^{t} \min(\Delta \ln GDP_{it}^{-}, 0) \cap \sum_{n=1}^{t} \Delta lnNREW_{it}^{-} = \sum_{n=1}^{t} \min(\Delta lnNREW_{it}^{-}, 0) \tag{3}$$

$$\sum_{n=1}^{t} \Delta lnGDP_{it}^{+} = \sum_{n=1}^{t} \max(\Delta \ln GDP_{it}^{+}, 0) \cap \sum_{n=1}^{t} \Delta lnNREW_{it}^{+} = \sum_{n=1}^{t} \max(\Delta \ln NREW_{it}^{+}, 0)$$

where GDP^{+} and $NREW^{+}$ indicate positive changes in the series, GDP^{-} and $NREW^{-}$ indicate negative changes in the series.

The full panel NARDL model including asymmetric effects and short- and long-term impacts is presented below:

$$\begin{aligned} \Delta \ln CO_{2,it} = \delta &+ \sum_{p=1}^{n1} \delta_{1p} \Delta \ln CO_{2,it-p} + \sum_{p=0}^{n2} \delta_{2p} \Delta lnGDP_{it-p}^{+} \\ &+ \sum_{p=0}^{n3} \delta_{3p} \Delta \ln GDP_{it-p}^{-} \sum_{p=0}^{n4} \delta_{4p} \Delta \ln NREW_{it-p}^{+} + \sum_{p=0}^{n5} \delta_{5p} \Delta lnNREW_{it-p}^{-} \\ &+ \sum_{p=0}^{n6} \delta_{6p} \Delta lnREW_{it-p} + \varphi_1 lnCO_{2,it-1} + \varphi_2 lnGDP_{it-1}^{+} + \varphi_3 lnGDP_{it-1}^{-} \\ &+ \varphi_4 lnNREW_{it-1}^{+} + \varphi_5 lnNREW_{it-1}^{-} + \varphi_6 lnREW_{it-1} + \varepsilon_t \end{aligned} \tag{4}$$

where i is the cross-sectional dimension, t is the time dimension, p is the corresponding order of delay, δ are the parameters of the short-term relation, and φ the parameters of the long-term relationship.

The study used an estimation based on the Pooled Mean Group (PMG) estimator. The selection of the optimal number of delays in the model, on the other hand, was determined according to the Akaike Information Criterion (AIC). The PMG estimator provides greater reliability for cross-country data, as it takes into account regional specificities and allows for a better interpretation of the long-run equilibrium (Attiaoui and Boufateh 2019).

In addition, PMG allows for heterogeneity in short-run coefficients, while long-run coefficients can be identical and homogeneous for the entire panel area (Pesaran, Shin, and Smith 1999). Testing for asymmetry was carried out using the Wald test and the x^2

statistic (Andrews 1987; Shin, Yu, and Greenwood-Nimmo 2014). Because all variables tested are stationary at level I (1) and the cross-section of the panel tested was four objects, cointegration tests were applied Kao (1999). In order to strengthen the results obtained, causality was tested using the Dumitrescu-Hurlin panel pairwise procedure (Dumitrescu and Hurlin 2012).

4. Research results

4.1. Exploratory Data Analysis

Table 3 presents the fundamental descriptive statistics for the surveyed data for the years 1991 and 2021. The Visegrad Group countries (Czech Republic, Hungary, Poland, and Slovakia) experienced notable shifts in carbon dioxide (CO_2) emissions, per capita GDP, and energy consumption from renewable and non-renewable sources between 1991 and 2021. There was a notable decrease in CO_2 emissions per capita over the period in question. In 1991, the average CO_2 emissions were 9.36 metric tonnes per capita. By 2021, this figure decreased to 6.84 metric tonnes. The reduction in the standard deviation from 3.07 to 2.65 tonnes indicates a more uniform distribution of emissions across the region, likely reflecting improvements in environmental policies and industrial technologies. GDP per capita grew considerably. In 1991, the average GDP per capita was $7,284.27, rising to $17,286.35 by 2021. Despite this growth, the reduction in the standard deviation from $2,330.20 to $2,003.16 indicates a reduction in income inequality within the region, although significant disparities in living standards between countries persist.

There was an increase in the consumption of electricity from non-renewable sources, with the average rising from 0.36 to 0.47 tonnes of oil equivalent per capita. This was accompanied by a rise in the standard deviation, which increased from 0.11 to 0.16 tonnes. This indicates an increasing demand for energy and the potential for delays in the transition to renewable sources. In contrast, there was a notable increase in electricity consumption from renewable sources, with the average rising from 0.01 to 0.09 tonnes of oil equivalent per capita.

While the contribution of renewable energy to overall consumption remains relatively modest, the decline in the standard deviation from 0.01 to 0.02 tonnes suggests a more consistent uptake of renewable technologies across the countries. In conclusion, the period from 1991 to 2021 saw significant developments in the Visegrad Group countries, including reductions in CO_2 emissions and economic growth. The increased consumption of renewable energy reflects an ongoing energy transition, although challenges remain with rising energy use from non-renewable sources.

Table 3: Descriptive statistics

Variable	1991				2021			
	Mean	Min	Max	SD	Mean	Min	Max	SD
CO_2	9.36	6.47	13.65	3.07	6.84	4.25	9.87	2.65
GDP	7284.27	4743.75	10304.87	2330.20	17286.35	15485.47	19715.97	2003.16
NREW	0.36	0.25	0.51	0.11	0.47	0.32	0.68	0.16
REW	0.01	0.00	0.03	0.01	0.09	0.06	0.11	0.02

Source: author's calculations.

Figures 1–3 present main time series data for the V4 countries. The CO_2 emissions per capita have declined in all four countries since 1991 (Fig. 1), reflecting technological progress and reduced industrial intensity. The Czech Republic and Hungary showed steady decreases, with minor fluctuations. Poland and Slovakia also reduced emissions, despite periods of stagnation. Renewable energy consumption per capita increased in all countries (Fig. 2), most notably in Slovakia. Growth in the Czech Republic and Poland was moderate; Hungary accelerated after 2010. Non-renewable energy consumption per capita remained relatively stable (Fig. 3), with slight increases in the Czech Republic and Poland, stagnation in Hungary, and fluctuations in Slovakia.

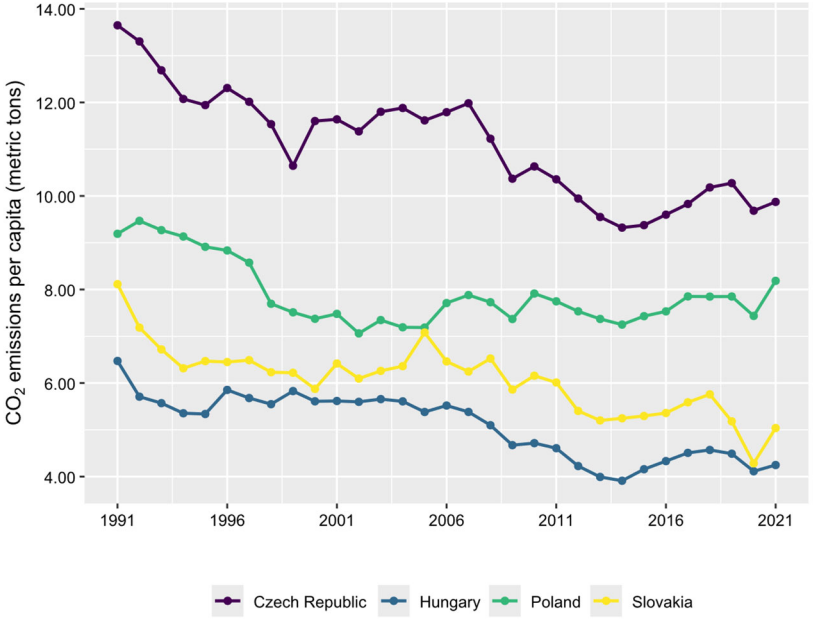


Figure 1: CO_2 emissions per capita in the study countries 1991–2021

Source: WDI database.

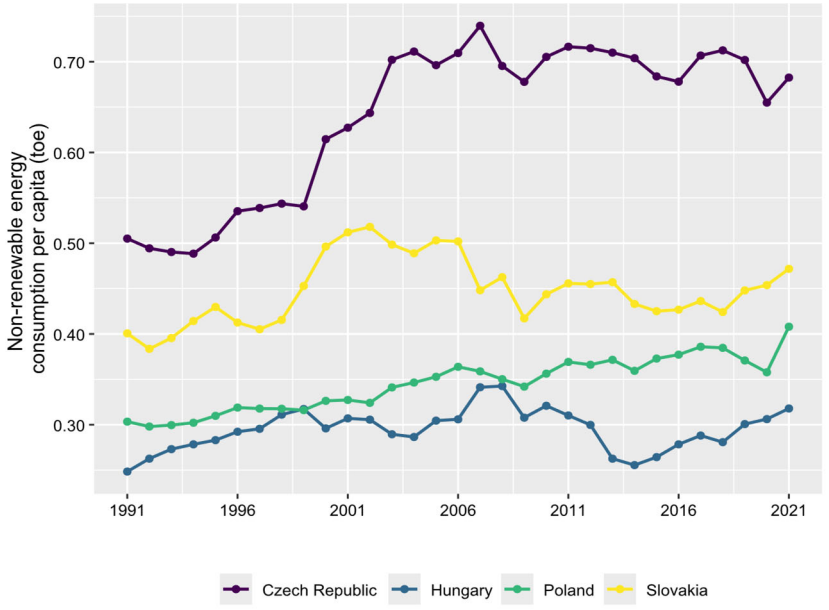


Figure 2: Electricity consumption from non-renewable sources in the study countries 1991–2021

Source: EEA/Eurostat database.

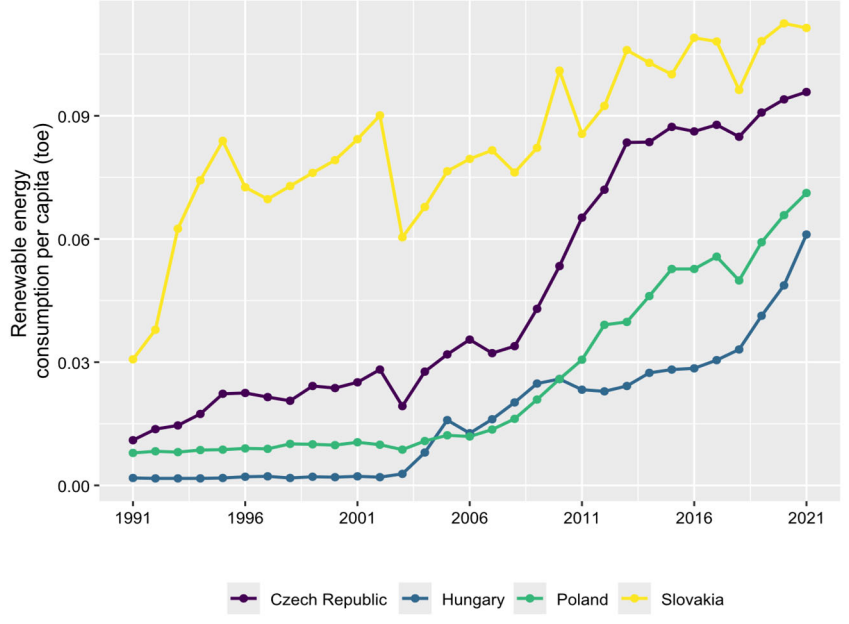


Figure 3: Electricity consumption from renewable sources in the study countries 1991–2021

Source: EEA/Eurostat database.

4.2. Model estimation result

The next step of the study was to conduct unit root tests. Three different types of tests were used for this purpose. The first was the unit root test of the second generation CIPS panel (Pesaran 2007) and the first generation test – Augmented Dickey–Fuller test (Dickey and Fuller 1979). The results of the tests are shown in Table 4. The tests performed indicate that all variables are stationary at the first difference, I (1). The results confirm that the tested variables meet the assumptions of the NARDL model.

Table 2: Panel data unit root tests

Variable	I (0) Level		I (1) First difference	
	CIPS	ADF	CIPS	ADF
$\ln CO_2$	-2.14	-0.98	-5.64*	- 7.24*
lnGDP	-1.54	0.75	-3.61*	-5.85*
lnNREW	-2.08	-0.26	-4.95*	-6.68*
lnREW	-1.23	0.94	-3.36*	-6.38*

*The significance of the coefficients is indicated by an asterisk in the tables, where ***, **, * denotes 5%, 1%, and 0.1% significance level, respectively.*

Source: author's calculations.

Table 5 shows the results of the Wald test for the long-run coefficients and diagnostic tests. The results indicate that there is significant asymmetry in the long run for the GDP and NREW variables at the 10% significance level, while for the REW variable, the presence of statistically significant asymmetry could not be confirmed. Therefore, parameters for the REW variable were estimated using a linear method. The Kao cointegration tests (Kao 1999) for the panel data is also statistically significant at the 5% level. The result of the test performed confirms that there is a sustained strong relationship between the study variables in the long term.

Table 5: NARDL model asymmetry test results and diagnostics

Variable	Statistics	Value	Probability
lnGDP	F	4.594	0.034
	$\chi 2$	4.595	0.032
lnNREW	F	2.833	0.095
	$\chi 2$	2.834	0.092
lnREW	F	1.182	0.279
	$\chi 2$	1.183	0.276
Kao-cointegration test	t	-2.371	0.009
Log likelihood			257.1

Source: author's calculations.

The results of the estimation of the parameters of the long-run model are presented in Table 6. All the results obtained are statistically significant at the 10% significance level, allowing inference from the estimated model. The NARDL results indicate the existence of asymmetry in the long term regarding the impact of economic growth and energy production from non-renewable sources on CO_2 emissions in the studied countries. According to the obtained parameters, a 1% increase in non-renewable energy consumption leads to a 0.26% increase in CO_2 emissions, while a 1% decrease in consumption generates a 0.63% decrease in CO_2 emissions. In contrast, a 1% increase in GDP translates into a 0.21% increase in CO_2 emissions, while a 1% decrease in GDP results in a 1.02% decrease in CO_2 emissions. The results also show that the increase in energy consumption from renewable sources in the long term does not show significant asymmetry. Its increase by 1% is associated with a decrease in CO_2 emissions of 0.05%.

Table 6: NARDL long-run results

Variable	Coefficient	Standard error	t-statistics	Probability
lnREW	-0.045	0.016	-2.831	0.005
φ^+ lnNREW	0.269	0.158	1.700	0.092
φ^- lnNREW	0.637	0.098	6.496	0.000
φ^+ lnGDP	0.208	0.062	3.340	0.001
φ^- lnGDP	1.022	0.339	3.014	0.003
Const.	1.442	0.098	14.588	0.000

Source: author's calculations.

Table 7 shows the results of the estimation of the short-term model. The results confirm the impact of economic growth and energy consumption from non-renewable sources on CO_2. However, no statistically significant short-term correlations between renewable energy consumption were confirmed. As a result of the model estimation, an error correction factor (ECT T-1) of -0.61 was also estimated. The negative and statistically significant correction factor is in line with the convergence requirements and further confirms the existence of long-term cointegration. In interpreting the ECT coefficientt-1, it is important to point out the relatively rapid adjustment of CO_2, to long-run equilibrium (within about 1.5 years) in the face of shocks, and the general downward trend in the long run in the countries studied.

Table 7: NARDL short-run results

Variable	Coefficient	Standard error	t-statistics	Probability
ECTt-1	-0.614	0.312	-1.970	0.032
$\Delta lnCO_2$ t-1	0.168	0.186	0.905	0.368
$\Delta lnCO_2$ t-2	-0.063	0.137	-0.459	0.647
$\Delta lnCO_2$ t-3	-0.118	0.226	-0.521	0.604
ΔlnREW	-0.029	0.041	-0.700	0.486
ΔlnREW t-1	-0.038	0.026	-1.421	0.159
ΔlnREW t-2	0.012	0.038	0.304	0.762
ΔlnREW t-3	-0.003	0.036	-0.080	0.936
ΔlnNREW	0.355	0.190	1.868	0.065
ΔlnGDP	0.512	0.258	1.985	0.050
ΔlnGDP t-1	0.083	0.195	0.428	0.670
ΔlnGDP t-2	0.205	0.049	4.138	0.000
ΔlnGDP t-3	0.236	0.342	0.691	0.491

Source: author's calculations.

In the final stage of the study, a causality test was conducted based on the Dumitrescu & Hurlin panel data test (2012). The test results indicate the existence of unidirectional causality from lnGDP to $lnCO_2$, from lnNREW to $lnCO_2$, from lnREW to $lnCO_2$, and from lnGDP to lnREW. Additionally, causality is observed from lnGDP to lnNREW, at the 10% significance level. Thus, according to the results of the causality test, all the variables tested have an impact on CO_2, which aligns with the model estimation results obtained. Furthermore, the causality test indicates that economic growth in the V4 countries induces energy consumption from both renewable and non-renewable sources.

Table 8: Results of Pairwise Dumitrescu-Hurlin panel causality tests

Cause → Effect	W-Stat.	Zbar-Stat.	Prob.
lnGDP → $lnCO_2$	4.211	1.698	0.090
$lnCO_2$ → lnGDP	1.286	-0.750	0.453
lnNREW → $lnCO_2$	2.514	1.851	0.081
$lnCO_2$ → lnNREW	2.475	0.245	0.806
lnREW → $lnCO_2$	4.244	1.726	0.084
$lnCO_2$ → lnREW	2.571	0.326	0.745
lnNREW → lnGDP	3.499	1.102	0.270
lnGDP → lnNREW	5.531	2.802	0.005
lnREW → lnGDP	1.521	-0.553	0.580
lnGDP → lnREW	12.200	8.224	0.000
lnREW → lnNREW	1.297	-0.741	0.459
lnNREW → lnREW	2.836	0.547	0.584

Source: author's calculations.

5. Discussion

In summary, the results of the study provide significant insights into the relationships between economic growth, renewable and non-renewable energy consumption, and CO_2 emissions in the V4 countries. The study indicates that there are both symmetric and asymmetric relationships between CO_2 emissions, renewable energy consumption, and economic growth. Notably, the impact of asymmetry on CO_2 emissions was confirmed only for the long term, while in the short term, all relationships were symmetric. These results demonstrate that asymmetric relationships between CO_2 emissions, energy consumption, and economic growth become more apparent over longer periods, reflecting long-term structural processes within the economy. In the short term, responses are often quicker and more uniform, leading to symmetric relationships.

However, over a longer time horizon, factors such as technology adaptation, changes in energy policies, and shifts in economic structure can result in more complex, asymmetric interactions. Additionally, the analysis of short-term asymmetric relationships would require higher frequency data (quarterly or monthly) for all variables studied, which are currently unavailable. In such cases, it would be possible to account for potential discrete non-linear dependencies.

The study indicates that the long-term asymmetric impact on CO_2 emissions is primarily driven by non-renewable energy consumption and economic growth. Specifically, a GDP decrease leads to a significantly larger reduction in emissions than a GDP increase causes a rise. Similarly, a decrease in non-renewable energy consumption results in a much greater drop in CO_2 emissions than an increase causes a rise.

Combining these findings, it can be noted that further economic development in the V4 countries may not have such a negative impact on the environment. This is associated with shifting growth towards more efficient energy use, developing less emissions-intensive economic sectors, and adopting ecological energy sources. According to the results, a fundamental issue for the V4 countries is their heavy reliance on non-renewable sources compared to renewable sources. The asymmetric relationship suggests that a significant reduction in emissions could be achieved either by curbing economic activity or by decreasing non-renewable energy consumption. Since causality tests indicate that economic growth in the V4 countries is driven by increased energy consumption, ensuring sustainable economic development can only be achieved by replacing non-renewable sources with renewable ones. The study indicates that only in the long term does an increase in renewable energy usage have a negative and symmetric effect on CO_2 emissions.

These observations align with findings by Papież et al. (2019), who noted that the positive economic aspects associated with renewable energy consumption become

apparent only when the share of renewable energy in the energy mix is significantly increased. Similarly, Ben Jebli et al. (2020) found that renewable energy positively impacts economic growth and reduces CO_2 emissions in higher-income countries. Therefore, it appears crucial in this context that efforts to reduce CO_2 emissions in the V4 countries may follow only two pathways: first, by merely reducing non-renewable energy consumption, which could result in decreased economic activity, or second, by increasing the use of renewable energy, thereby ensuring sustainable economic development. The significance of renewable energy consumption's impact only in the long term suggests that the energy transformation process requires a substantial period to achieve its goals. In summary, the results obtained confirm hypotheses H2 and H3 and partially verify hypothesis H1 concerning the effect of renewable energy on CO_2 reduction, but only in a linear manner.

It is not straightforward to compare these results with those presented in earlier studies for three reasons. Firstly, to date, no study using the NARDL-PMG method has fully addressed the variables examined in this study. Secondly, the group of countries analyzed here does not overlap with those in previous studies. The V4 countries have mostly been included in broader sets of countries studied (e.g. all EU countries). Thirdly, this study employed the widest possible time range for the time series analyzed, which allowed for more precise results compared to previous research. Nevertheless, when limiting the comparison to studies that have investigated linear and non-linear relationships between energy consumption, economic growth, and CO_2 emissions, the results obtained are somewhat like those of previous studies.

In reference to studies specifically concerning the V4 group, the results obtained in this research align with those of Supron and Myszczyszyn (2023), which indicate a negative impact of non-renewable energy on environmental pollution and a positive effect of renewable energy. Similarly, the study by Litavcová and Chovancová (2021), which did not cover all V4 countries, also partially confirms these findings, particularly in the context of CO_2 emissions. This suggests that higher energy consumption leads to increased emissions but is necessary to sustain economic growth.

However, with some caution when considering studies using non-linear methods, including NARDL, that involve different countries, there are notable similarities. Kirikkaleli et al. (2023) and Akadiri & Adebayo (2022) confirm an asymmetric relationship between economic growth and CO_2 emissions in India, where a decrease in GDP results in a greater reduction in emissions than an increase in GDP leads to an increase in emissions. In contrast, the results of this study differ from those obtained for GCC countries by Islam & Rahaman (2023), who, using the NARDL-PMG model, did not confirm a non-linear but only a linear relationship between economic growth and CO_2 emissions in the long term. Additionally, studies using linear panel models by Li et al.

(2014), Lazăr et al. (2019) and Muço et al. (2021) confirmed a positive linear relationship between economic growth and CO_2 emissions.

Bekun et al. (2019), using the ARDL-PMG model, confirmed that an increase in renewable energy consumption contributes to a decrease in CO_2 emissions, while non-renewable energy consumption results in an increase in CO_2 emissions in the long term within EU countries, consistent with the results of this study. Furthermore, Chen et al. (2022), who studied a larger number of countries using non-linear methods (GMM), obtained similar results for developed countries, finding that a 1% increase in consumption from renewable sources results in a 0.04% decrease in CO_2 emissions, while a 1% increase from non-renewable sources results in a 0.6% increase. These findings are comparable to those obtained in this study.

6. Conclusions

The results provide robust evidence of a persistent and long-term relationship between renewable and non-renewable energy consumption, economic growth and CO_2 emissions in the V4 countries. The estimated model indicates that changes in these variables have a lasting impact on CO_2 emissions in this region. The study also confirmed that there is a significant asymmetry in the long-term relationship between non-renewable energy consumption and CO_2 emissions. This suggests that efforts to reduce non-renewable energy consumption may have a more significant impact on emission reductions. Also, economic growth shows an asymmetric impact on CO_2 emissions in the V4 countries, where a decrease in economic growth has a greater impact on CO_2 reduction than its increase. In view of this, energy efficiency improvements will be necessary to reduce CO_2 emissions without negative effects on economic growth. Investment in technologies and solutions for more efficient energy use should therefore be a priority in the V4 countries. Unlike non-renewable energy, renewable energy consumption does not show a significant asymmetry in the long-term relationship with CO_2. This suggests that policies promoting the use of renewable energy sources can contribute to reducing CO_2 emissions in the long term. At the same time, increasing the use of renewable energy should occur in a stable and long-term manner.

This study, like any empirical work, has certain limitations that could guide future research. First, environmental pollution is represented by CO_2 emissions. Future research could expand the analysis of the impact of energy consumption on the environment by including other greenhouse gases resulting from human economic activities (such as sulfur hexafluoride, carbon monoxide, and nitrous oxide). This would allow for more precise information and enable a comparative analysis of the effects of different types of gases. Second, this study considers economic activity as the total output expressed through GDP. To increase accuracy, future studies could focus on examining

the asymmetric impact of individual economic sectors on environmental pollution. Finally, the study focuses on a narrow geographical area, encompassing the V4 countries. In future research, to draw broader conclusions, the presented model could be applied to a larger number of countries, covering the entire EU. This could also involve using classification methods and analyzing countries in groups based on characteristics such as the level of economic development and renewable energy usage.

Moreover, the NARDL model itself has certain limitations due to the way data is estimated. The model yields the best results when using relatively long time series with high frequency. As environmental and energy data for many countries have only been collected since the mid-1990s, trade-offs between measurement accuracy and the number of variables and lags are necessary. In addition, many important indicators have an even shorter measurement history. As the number of data periods increases, future studies could introduce additional variables into the model, such as innovation, research and development and social factors. This would increase the breadth and depth of knowledge, making it more detailed and accurate.

References

Adedoyin, F. F., Ozturk, I., Bekun, F. V., Agboola, P. O. and Agboola M. O., (2021). Renewable and Non-Renewable Energy Policy Simulations for Abating Emissions in a Complex Economy: Evidence from the Novel Dynamic ARDL. *Renewable Energy,* 177 (November), pp. 1408–20. https://doi.org/10.1016/j.renene.2021.06.018.

Ahmed, N., Sheikh A. A., Hassan, B., Khan S. Nawaz, Borda R. Cosio, Huamán, J. M. C. and Senkus, P., (2022). The Role of Educating the Labor Force in Sustaining a Green Economy in MINT Countries: Panel Symmetric and Asymmetric Approach. *Sustainability,* 14(19), pp. 12067. https://doi.org/10.3390/su141912067.

Akadiri, S. S., Adebayo, T. S., (2022). Asymmetric Nexus among Financial Globalization, Non-Renewable Energy, Renewable Energy Use, Economic Growth, and Carbon Emissions: Impact on Environmental Sustainability Targets in India'. *Environmental Science and Pollution Research,* 29(11), pp. 16311–23. https://doi.org/10.1007/s11356-021-16849-0.

Al-Ghussain, L., (2019). Global Warming: Review on Driving Forces and Mitigation. *Environmental Progress & Sustainable Energy,* 38(1), pp. 13–21. https://doi.org/10.1002/ep.13041.

Allard, A., Takman J., Gazi Salah Uddin and Ali Ahmed, (2018). The N-Shaped Environmental Kuznets Curve: An Empirical Evaluation Using a Panel Quantile Regression Approach. *Environmental Science and Pollution Research* 25(6), pp. 5848–61. https://doi.org/10.1007/s11356-017-0907-0.

Alper, A., Oguz O., (2016). The Role of Renewable Energy Consumption in Economic Growth: Evidence from Asymmetric Causality. *Renewable and Sustainable Energy Reviews* 60 (July), pp. 953–59. https://doi.org/10.1016/j.rser.2016.01.123.

Ambroziak, Ł., Chojna, J., Miniszewski, M., Strzelecki, J., Szpor, A., Śliwowski, P., Święcicki I. and Wąsiński M., (2021). Visegrad Group - 30 years of transformation, integration and development. Warsaw: Polish Economic Institute. https://pie.net.pl/grupa-wyszehradzka-30-lat-transformacji-integracji-i-rozwoju/.

Andrews, Donald W. K., (1987). Asymptotic Results for Generalized Wald Tests. *Econometric Theory,* 3(3), pp. 348–58. https://doi.org/10.1017/S0266466600010434.

Antonakakis, N., Chatziantoniou, I. and Filis, G., (2017). Energy Consumption, CO_2 Emissions, and Economic Growth: An Ethical Dilemma. *Renewable and Sustainable Energy Reviews,* 68 (February), pp. 808–24. https://doi.org/10.1016/j.rser.2016.09.105.

Anwar, A., Siddique, M., Eyup, D. and Sharif, A., (2021). The Moderating Role of Renewable and Non-Renewable Energy in Environment-Income Nexus for ASEAN Countries: Evidence from Method of Moments Quantile Regression. *Renewable Energy,* 164, pp. 956–67. https://doi.org/10.1016/j.renene.2020.09.128.

Armeanu, D. S., Vintilă, G. and Gherghina, S. C., (2017). Does Renewable Energy Drive Sustainable Economic Growth? Multivariate Panel Data Evidence for EU-28 Countries. *Energies,* 10(3). https://doi.org/10.3390/en10030381.

Attiaoui, I., Boufateh T., (2019). Impacts of Climate Change on Cereal Farming in Tunisia: A Panel ARDL–PMG Approach. *Environmental Science and Pollution Research,* 26(13), pp. 13334–45. https://doi.org/10.1007/s11356-019-04867-y.

Aziz, S., Maltese, I., Marcucci, E., Gatta, V., Benmoussa, R. and Irhirane El Hassan, (2022). Energy Consumption and Environmental Impact of E-Grocery: A Systematic Literature Review. *Energies,* 15(19), p. 7289. https://doi.org/10.3390/en15197289.

Bąk, I., Cheba, K., (2023). Green Transformation: Applying Statistical Data Analysis to a Systematic Literature Review. *Energies,* 16(1), p. 253. https://doi.org/10.3390/en16010253.

Bekun, F. V., Adewale, Alola A. and Sarkodie, S. A., (2019). Toward a Sustainable Environment: Nexus between CO_2 Emissions, Resource Rent, Renewable and Nonrenewable Energy in 16-EU Countries. *Science of The Total Environment,* 657 (March), pp. 1023–29. https://doi.org/10.1016/j.scitotenv. 2018.12.104.

Ben J., M., Farhani, S. and Guesmi, K., (2020). Renewable Energy, CO_2 Emissions and Value Added: Empirical Evidence from Countries with Different Income Levels.

Structural Change and Economic Dynamics, 53 (June), pp. 402–10. https://doi.org/10.1016/j.strueco.2019.12.009.

Bhattacharya, M., Awaworyi Churchill, S. and Paramati, S. R., (2017). The Dynamic Impact of Renewable Energy and Institutions on Economic Output and CO_2 Emissions across Regions. *Renewable Energy,* 111 (October), pp. 157–67. https://doi.org/10.1016/j.renene.2017.03.102.

Bhuiyan, M. A., Zhang, Q., Khare V., Mikhaylov, A., Pinter, G. and Huang, X., (2022). Renewable Energy Consumption and Economic Growth Nexus—A Systematic Literature Review. *Frontiers in Environmental Science,* 10 (April). https://doi.org/10.3389/fenvs.2022.878394.

Bigos, K., (2017). Zrównoważony rozwój w krajach Grupy Wyszehradzkiej: ujęcie teoretyczne i empiryczne. *International Entrepreneurship Review,* 3(3), pp. 107–26.

Bourcet, C., (2020). Empirical Determinants of Renewable Energy Deployment: A Systematic Literature Review. *Energy Economics,* 85 (January), p. 104563. https://doi.org/10.1016/j.eneco.2019.104563.

Brodny, J., Tutak, M., (2021). The Comparative Assessment of Sustainable Energy Security in the Visegrad Countries. A 10-Year Perspective. *Journal of Cleaner Production,* 317 (October), p. 128427. https://doi.org/10.1016/j.jclepro.2021.128427.

Busu, M., Nedelcu, A. C., (2021). Analyzing the Renewable Energy and CO_2 Emission Levels Nexus at an EU Level: A Panel Data Regression Approach. *Processes,* 9(1), p. 130. https://doi.org/10.3390/pr9010130.

Cai, Y., Sam, C. Y. and Chang, T., (2018). Nexus between Clean Energy Consumption, Economic Growth and CO_2 Emissions. *Journal of Cleaner Production,* 182 (May), pp. 1001–11. https://doi.org/10.1016/j.jclepro.2018.02.035.

Charfeddine, L., Kahia, M., (2019). Impact of Renewable Energy Consumption and Financial Development on CO_2 Emissions and Economic Growth in the MENA Region: A Panel Vector Autoregressive (PVAR) Analysis. *Renewable Energy,* 139 (August), pp. 198–213. https://doi.org/10.1016/ j.renene.2019.01.010.

Chauvet, M., Jiang, C., (2023). Nonlinear Relationship between Monetary Policy and Stock Returns: Evidence from the U.S. *Global Finance Journal,* 55 (February), p. 100796. https://doi.org/10.1016/j.gfj.2022.100796.

Cheba, K., Bąk, I., Szopik-Depczyńska, K. and Ioppolo, G., (2022). Directions of Green Transformation of the European Union Countries. *Ecological Indicators,* 136 (March), p. 108601. https://doi.org/10.1016/j.ecolind. 2022.108601.

Chen, C., Pinar, M. and Stengos, T., (2022). Renewable Energy and CO_2 Emissions: New Evidence with the Panel Threshold Model. *Renewable Energy,* 194 (July), pp. 117–28. https://doi.org/10.1016/j.renene.2022.05.095.

Cialani, C., (2017). CO_2 Emissions, GDP and Trade: A Panel Cointegration Approach'. *International Journal of Sustainable Development & World Ecology,* 24(3), pp. 193–204. https://doi.org/10.1080/13504509.2016.1196253.

Czyżewski, B., Polcyn, J. and Brelik, A., (2022). Political Orientations, Economic Policies, and Environmental Quality: Multi-Valued Treatment Effects Analysis with Spatial Spillovers in Country Districts of Poland. *Environmental Science & Policy,* 128 (February), pp. 1–13. https://doi.org/10.1016/j.envsci. 2021.11.001.

Deng, Z., Liu, J., and Sohail, S., (2022). Green Economy Design in BRICS: Dynamic Relationship between Financial Inflow, Renewable Energy Consumption, and Environmental Quality. *Environmental Science and Pollution Research,* 29(15), pp. 22505–14. https://doi.org/10.1007/s11356-021-17376-8.

Dickey, D. A., Fuller, W., A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association,* 74(366a), pp. 427–31. https://doi.org/10.1080/01621459.1979. 10482531.

Dumitrescu, E.-I., Hurlin C., (2012). Testing for Granger Non-Causality in Heterogeneous Panels. *Economic Modelling,* 29(4), pp. 1450–60. https://doi.org/ 10.1016/j.econmod.2012.02.014.

Fanchi, J. R., (2023). *Energy In The 21st Century: Energy In Transition*. World Scientific.

Flanker, R., (2016). The Paris Agreement and the New Logic of International Climate Politics. *International Affairs,* 92(5), pp. 1107–25. https://doi.org/ 10.1111/1468-2346.12708.

Gozgor, G., Lau, Chi Keung M. and Lu, Z., (2018). Energy Consumption and Economic Growth: New Evidence from the OECD Countries. *Energy,* 153 (June), pp. 27–34. https://doi.org/10.1016/j.energy.2018.03.158.

Haberl, H., Wiedenhofer, D., Virág, D., Kalt, G., Plank, B., Brockway, P., Fishman, T., et al., (2020). A Systematic Review of the Evidence on Decoupling of GDP, Resource Use and GHG Emissions, Part II: Synthesizing the Insights. *Environmental Research Letters,* 15(6), p. 065003. https://doi.org/10.1088/ 1748-9326/ab842a.

Inglesi-Lotz, R., Dogan, E. (2018). The Role of Renewable versus Non-Renewable Energy to the Level of CO_2 Emissions a Panel Analysis of Sub- Saharan Africa's Big 10 Electricity Generators. *Renewable Energy,* 123 (August), pp. 36–43. https:// doi.org/10.1016/j.renene.2018.02.041.

Iorember, P. T., Usman, O. and Jelilov, G., (2019). Asymmetric Effects of Renewable Energy Consumption, Trade Openness and Economic Growth on Environmental Quality in Nigeria and South Africa. MPRA Paper. 2019. https:// mpra.ub.uni-muenchen.de/96333/.

Islam, Md. S., Rahaman Sk Habibur, (2023). The Asymmetric Effect of ICT on CO_2 Emissions in the Context of an EKC Framework in GCC Countries: The Role of Energy Consumption, Energy Intensity, Trade, and Financial Development. *Environmental Science and Pollution Research,* 30(31), pp. 77729–41. https://doi.org/10.1007/s11356-023-27590-1.

Ito, K., (2017). CO_2 Emissions, Renewable and Non-Renewable Energy Consumption, and Economic Growth: Evidence from Panel Data for Developing Countries. *International Economics,* 151 (October), pp. 1–6. https://doi.org/10.1016/j.inteco.2017.02.001.

Kao, C., (1999). Spurious Regression and Residual-Based Tests for Cointegration in Panel Data. *Journal of Econometrics*, 90(1), pp. 1–44. https://doi.org/10.1016/S0304-4076(98)00023-2.

Khan, A., Sun, C., (2024). The Asymmetric Nexus of Energy-Growth and CO_2 Emissions: An Empirical Evidence Based on Hidden Cointegration Analysis. *Gondwana Research,* 125 (January), pp. 15–28. https://doi.org/10.1016/j.gr.2023.07.012.

Kirikkaleli, D., Abbasi K. R. and Oyebanji, M. O., (2023). The Asymmetric and Long-Run Effect of Environmental Innovation and CO2 Intensity of GDP on Consumption-Based CO_2 Emissions in Denmark. *Environmental Science and Pollution Research,* 30(17), pp. 50110–24. https://doi.org/10.1007/ s11356-023-25811-1.

Kouton, J., (2019). The Asymmetric Linkage between Energy Use and Economic Growth in Selected African Countries: Evidence from a Nonlinear Panel Autoregressive Distributed Lag Model. *Energy Economics,* 83 (September), pp. 475–90. https://doi.org/10.1016/j.eneco.2019.08.006.

Lazăr, D., Minea, A. and Purcel, A.-A., (2019). Pollution and Economic Growth: Evidence from Central and Eastern European Countries. *Energy Economics,* 81 (June), pp. 1121–31. https://doi.org/10.1016/j.eneco. 2019.05.011.

Li, B., Saif Ur Rahman, Sahar Afshan, Azka Amin and Younas S., (2023). Energy Consumption and Innovation-Environmental Degradation Nexus in BRICS Countries: New Evidence from NARDL Approach Using Carbon Dioxide and Nitrous Oxide Emissions. *Environmental Science and Pollution Research,* 30(53), pp. 113561–86. https://doi.org/10.1007/s11356-023-29927-2.

Li, Jin Hua, Yu Jie Yang, Bo Wen Li, Wen Jin Li, Gang Wang and Johannes M. H. Knops, (2014). Effects of Nitrogen and Phosphorus Fertilization on Soil Carbon Fractions in Alpine Meadows on the Qinghai-Tibetan Plateau. *PLOS ONE* 9 (7): e103266. https://doi.org/10.1371/journal.pone.0103266.

Li, R., Jiang, H., Sotnyk, I., Kubatko, O. and Ismail Almashaqbeh, Y. A., (2020). The CO_2 Emissions Drivers of Post-Communist Economies in Eastern Europe and Central Asia. *Atmosphere,* 11(9), p. 1019. https://doi.org/10.3390/ atmos11091019.

Litavcová, E., Chovancová, J., (2021). Economic Development, CO_2 Emissions and Energy Use Nexus-Evidence from the Danube Region Countries. *Energies,* 14(11), p. 3165. https://doi.org/10.3390/en14113165.

Ma, X., Ahmad, N. and Oei, Pao-Yu, (2021). Environmental Kuznets Curve in France and Germany: Role of Renewable and Nonrenewable Energy. *Renewable Energy,* 172 (July), pp. 88–99. https://doi.org/10.1016/j.renene.2021.03.014.

Mardani, A., Streimikiene, D., Cavallaro, F., Loganathan, N. and Khoshnoudi M., (2019). Carbon Dioxide (CO_2) Emissions and Economic Growth: A Systematic Review of Two Decades of Research from 1995 to 2017. *Science of The Total Environment,* 649 (February), pp. 31–49. https://doi.org/10.1016/j.scitotenv.2018.08.229.

Marzec, L., Zioło M., (2016). Zróżnicowanie odnawialnych źródeł energii w krajach grupy wyszehradzkiej. *Metody Ilościowe w Badaniach Ekonomicznych,* 17(2), pp. 75–85.

Masron, T. Ariffin, Subramaniam, Y., (2018). The Environmental Kuznets Curve in the Presence of Corruption in Developing Countries. *Environmental Science and Pollution Research,* 25(13), pp. 12491–506. https:// doi.org/10.1007/s11356-018-1473-9.

Muço, K., Valentini, E. and Lucarelli, S., (2021). The Relationships between GDP Growth, Energy Consumption, Renewable Energy Production and CO_2 Emissions in European Transition Economies. *International Journal of Energy Economics and Policy,* 11(4), pp. 362–73.

Naseer, S., Song Huaming, Supat Chupradit, Adnan Maqbool, Nik Alif Amri Nik Hashim and Hieu Minh Vu, (2022). Does Educated Labor Force Is Managing the Green Economy in BRCS? Fresh Evidence from NARDL-PMG Approach. *Environmental Science and Pollution Research,* 29(14), pp. 20296–304. https://doi.org/10.1007/s11356-021-16834-7.

Ndoricimpa, A., (2017). Threshold Effects of Inflation on Economic Growth: Is Africa Different? *International Economic Journal,* 31(4), pp. 599–620.

Omri, A., (2014). An International Literature Survey on Energy-Economic Growth Nexus: Evidence from Country-Specific Studies. *Renewable and Sustainable Energy Reviews,* 38 (October), pp. 951–59. https://doi.org/10.1016/j.rser.2014.07.084.

Ozcan, B., Ozturk I., (2019). Renewable Energy Consumption-Economic Growth Nexus in Emerging Countries: A Bootstrap Panel Causality Test. *Renewable and Sustainable Energy Reviews,* 104 (April), pp. 30–37. https:// doi.org/10.1016/ j.rser.2019.01.020.

Ozturk, I. (2010). A Literature Survey on Energy–Growth Nexus. *Energy Policy* 38(1), pp. 340–49. https://doi.org/10.1016/j.enpol.2009.09.024.

Papież, M., Śmiech, S. and Frodyma, K., (2019). Effects of Renewable Energy Sector Development on Electricity Consumption – Growth Nexus in the European Union. *Renewable and Sustainable Energy Reviews* 113 (October), p. 109276. https:// doi.org/10.1016/j.rser.2019.109276.

Pesaran, M. Hashem, (2007). A Simple Panel Unit Root Test in the Presence of Cross-Section Dependence. *Journal of Applied Econometrics,* 22(2), pp. 265–312. https:// doi.org/10.1002/jae.951.

Pesaran, M. Hashem, Shin, Y. and Smith, R. J., (2001). Bounds Testing Approaches to the Analysis of Level Relationships. *Journal of Applied Econometrics* 16(3), pp. 289–326. https://doi.org/10.1002/jae.616.

Pesaran, M. Hashem, Shin, Y. and Smith, R. P., (1999). Pooled Mean Group Estimation of Dynamic Heterogeneous Panels. *Journal of the American Statistical Association,* 94(446), pp. 621–34. https://doi.org/10.1080/01621459.1999. 10474156.

Pörtner, H.-O., Roberts, D. C., Poloczanska, E. S., Mintenbeck K., Tignor, M., Alegría, A., Craig, M., et al., (2022). IPCC, 2022: Summary for Policymakers.

Rasoulinezhad, E., Saboori, B., (2018). Panel Estimation for Renewable and Non-Renewable Energy Consumption, Economic Growth, CO_2 Emissions, the Composite Trade Intensity, and Financial Openness of the Commonwealth of Independent States. *Environmental Science and Pollution Research,* 25(18), pp. 17354–70. https://doi.org/10.1007/s11356-018-1827-3.

Samper, J. A., Schockling, A. and Islar, M., (2021). Climate Politics in Green Deals: Exposing the Political Frontiers of the European Green Deal. *Politics and Governance,* 9(2), pp. 8–16. https://doi.org/10.17645/pag.v9i2.3853.

Saqib, N., Duran, I. A. and Hashmi, N., (2022). Impact of Financial Deepening, Energy Consumption and Total Natural Resource Rent on CO_2 Emission in the GCC Countries: Evidence from Advanced Panel Data Simulation. *International Journal of Energy Economics and Policy,* 12(2), pp. 400–409. https://doi.org/10.32479/ ijeep.12907.

Shahbaz, M., Sinha, A., (2019). Environmental Kuznets Curve for CO_2 Emissions: A Literature Survey. *Journal of Economic Studies,* 46(1), pp. 106–68. https:// doi.org/10.1108/JES-09-2017-0249.

Shin, Y., Yu, B. and Greenwood-Nimmo, M., (2014). Modelling Asymmetric Cointegration and Dynamic Multipliers in a Nonlinear ARDL Framework. In *Festschrift in Honor of Peter Schmidt: Econometric Methods and Applications*, edited by Robin C. Sickles and William C. Horrace, pp. 281–314. New York, NY: Springer. https://doi.org/10.1007/978-1-4899-8008-3_9.

Stern, D. I., (2019). Energy and Economic Growth. In *Routledge Handbook of Energy Economics*, pp, 28–46. Routledge.

Stern, D. I., Common, M. S., (2001). Is There an Environmental Kuznets Curve for Sulfur? *Journal of Environmental Economics and Management,* 41(2), pp. 162–78. https://doi.org/10.1006/jeem.2000.1132.

Suproń, B., (2024). Assessing the Long-Term Asymmetric Relationship between Energy Consumption and CO_2 Emissions: Evidence from the Visegrad Group Countries. *Economics and Business Review,* 10(1). https://doi.org/10.18559/ebr. 2024.1.1082.

Suproń, B., Myszczyszyn, J., (2023). Impact of Renewable and Non-Renewable Energy Consumption and CO_2 Emissions on Economic Growth in the Visegrad Countries. *Energies,* 16(20), p. 7163. https://doi.org/10.3390/en16207163.

Supron, B., Myszczyszyn, J., (2023). Impact on Economic Growth in the Visegrad Countries of Renewable and Non-Renewable Energy Consumption and CO_2 Emissions. Preprints. https://doi.org/10.20944/preprints202309.1036.v1.

Tiba, S., Omri, A., (2017). Literature Survey on the Relationships between Energy, Environment and Economic Growth. *Renewable and Sustainable Energy Reviews,* 69 (March), pp. 1129–46. https://doi.org/10.1016/j.rser.2016.09.113.

Tugcu, C. T., Ozturk, I. and Aslan, A., (2012). Renewable and Non-Renewable Energy Consumption and Economic Growth Relationship Revisited: Evidence from G7 Countries. *Energy Economics,* 34(6), pp. 1942–50. https://doi.org/10.1016/j.eneco.2012.08.021.

Yang, Z., Zhao, Y., (2014). Energy Consumption, Carbon Emissions, and Economic Growth in India: Evidence from Directed Acyclic Graphs. *Economic Modelling,* 38 (February), pp. 533–40. https://doi.org/10.1016/j.econmod. 2014.01.030.

Zhang, S., Liu, X., (2019). The Roles of International Tourism and Renewable Energy in Environment: New Evidence from Asian Countries. *Renewable Energy* 139 (August), pp. 385–94. https://doi.org/10.1016/j.renene.2019.02.046.

STATISTICS IN TRANSITION new series, September 2025
Vol. 26, No. 3, pp. 181–194, https://doi.org/10.59139/stattrans-2025-033
Received – 05.02.2025; accepted – 24.04.2025

Multivariate two-sample permutation test with directional alternative for categorical data

Stefano Bonnini[1], **Michela Borghesi**[2]

Abstract

This paper presents a distribution-free test, based on the permutation approach, on treatment effects with a multivariate categorical response variable. The motivating example is a typical case-control biomedical study, performed to investigate the effect of the treatment called "assisted motor activity" (AMA) on the health of comorbid patients affected by "low back pain" (LBP), "hypertension" and "diabetes". Specifically, the goal was to test whether the AMA determines an improvement in the functionality and the perceived health status of patients. Two independent samples (treated and control group) were compared according to 13 different binary or ordinal outcomes. The null hypothesis of the test consists in the equality in the distribution of the multivariate responses of the two groups, whereas under the alternative hypothesis, the health status of the treated patients is better. The approach proposed in this work is based on the Combined Permutation Test (CPT) method, which is suitable for analyzing multivariate categorical data in the presence of confounding factors. A stratification of the groups and intra-stratum permutation univariate two-sample tests are conducted to avoid the potential confounding effects. P-values from the partial tests are combined using the CPT approach to create a suitable test statistic for the overall problem.

Key words: nonparametric statistics, permutation test, multivariate statistics, categorical data.

1. Introduction

This study involves applying a distribution-free test using a permutation approach to address a multivariate biostatistical problem. The main goal of this work is to test the effect of "assisted motor activity" (AMA) on the health of patients affected by "low back

[1] Department of Economics and Management, University of Ferrara, Ferrara, Italy.
E-mail: stefano.bonnini@unife.it. ORCID: https://orcid.org/0000-0002-7972-3046.
[2] Department of Economics and Management, University of Ferrara, Ferrara, Italy.
E-mail: michela.borghesi@unife.it. ORCID: https://orcid.org/0000-0003-1872-5766.

pain" (LBP), "hypertension" and "diabetes". AMA is a therapy that focuses on targeted physical exercises to regain mobility restrictions resulting from various causes.

In the literature, some studies were conducted on the effect of motor activity on patients affected by low back pain, especially on proving that treatments based on specific physical exercises can improve the health of patients (Ferreira et al., 2007; Gordon and Bloxham, 2016). This type of analysis almost always consists of a randomized controlled trial and parametric methods are used which often require large sample sizes to obtain reliable results (Macedo et al., 2012; Macedo et al., 2014).

This study is a case-control experiment that compares two independent samples. The samples consist of a treated group of 27 patients (group 1) and a control group of 20 patients (group 0). At time t_0 (prior to treatment), there is no significant difference in health status between the two groups. A comparison is made at time t_1, following the treatment for group 1. Health status is assessed using 13 different binary or ordinal outcomes. The null hypothesis states that the distribution of multivariate responses is equal for groups 1 and 0, while the alternative hypothesis suggests that the treated patients have a better health status. This alternative hypothesis is directional, indicating a multivariate stochastic dominance issue for ordinal variables. Since being over 60 years old and having LBP could be a risk factor, a confounding factor s is established, where $s = 1$ if the patient is over 60 years old and has LBP, and $s = 0$ otherwise.

The methodology outlined in this study is grounded in the Combined Permutation Test (CPT), which is effective for handling multivariate categorical data and addressing the issue of confounding factors (Pesarin and Salmaso, 2010; Bonnini et al., 2014). The combined permutation testing approach has been successfully applied in a wide range of contexts (Alibrandi et al., 2022). Specifically, it has seen extensive use in empirical research (Toma et al., 2017), involving both numerical and categorical variables (Bonnini et al., 2023), large-scale data scenarios (Simon and Tibshirani, 2012), and regression analyses (Bonnini and Borghesi, 2022). Furthermore, this method has proven effective in testing both directional and non-monotonic hypotheses (Bonnini et al., 2024a; Stute et al., 1998), in the analysis of count data (Bonnini et al., 2024b), and in numerous other applications. For more detailed information on the methodology of the nonparametric combination, refer to Bonnini et al. (2024).

To mitigate confounding effects by comparing comparable patients regarding the confounder, we implement stratification of the groups and conduct intra-stratum permutation univariate two-sample tests. Considering there are 13 components in the multivariate response and 2 strata, the total number of partial tests amounts to 26. Utilizing the CPT methodology to combine the p-values from these partial tests yields a test statistic that is appropriate for the overall analysis. From the application point of

view, the aim is to evaluate if AMA leads to enhancements in the functionality and perceived health condition of patients with comorbidities, using a significance level $\alpha = 0.05$.

Section 2 is dedicated to the description of the statistical problem and Section 3 includes a description of the proposed methodological solution. In Section 4 a simulation study is carried out, Section 5 is dedicated to the case study and the results are reported and commented in the conclusions (Section 6).

2. Statistical problem

Let $X_{1,sv}$ and $X_{0,sv}$ represent the v-th outcome or, equivalently, the v-th component of the multivariate response, in the stratum s for the treated and the control group respectively, with $s = 0,1$ and $v = 1, \dots, 13$. The partial problem related to the v-th outcome and the stratum s consists of testing

$$H_{0,sv}: X_{1,sv} =^{d} X_{0,sv} \tag{1}$$

versus the alternative hypothesis

$$H_{1,sv}: X_{1,sv} >^{d} X_{0,sv}, \tag{2}$$

where $=^{d}$ denotes equality in distribution and $>^{d}$ indicates stochastic dominance. Such hypotheses may be written as

$$H_{0,sv}: F_{1,sv}(x) = F_{0,sv}(x)\ \forall x \tag{3}$$

and

$$H_{1,sv}: F_{1,sv}(x) \leq F_{0,sv}(x)\ \forall x \text{ and } \exists x | F_{1,sv}(x) < F_{0,sv}(x)\,, \tag{4}$$

where $F_{j,sv}(x)$ represent the cumulative distribution function of $X_{j,sv}$, with $j = 0,1$.

Under the null hypothesis, for the v-th outcome, both the intra-stratum partial null hypothesis $H_{0,1v}$ and $H_{0,0v}$ are true, thus $H_{0,v}: H_{0,1v} \cap H_{0,0v}$. Similarly, $H_{1,v}: H_{1,1v} \cup H_{1,0v}$, with the same notation. Hence, the overall null and alternative hypothesis of the problem can be denoted by

$$\begin{cases} H_0: \bigcap_{v=1}^{13} H_{0,v} \\ H_1: \bigcup_{v=1}^{13} H_{1,v} \end{cases} \tag{5}$$

For problems of stochastic dominance for ordered categorical variables, the literature offers quite a long list of exact and approximate solutions (see Agresti and Klingenber, 2005; Han et al., 2004; Hirotsu, 1986; Loughin and Scherer, 1998; Loughin,

2004; Lumely, 1996; Nettleton and Banerjee, 2000). For the univariate case, most of the methodological solutions are based on the restricted maximum likelihood ratio test (see Cohen et al., 2000; Silvapulle and Sen, 2005; Wang, 1996). Asymptotic null distributions of the test statistic are mixtures of chi-squared, implying that the mixture's weights depend on the unknown population distribution. In general, this is a complex problem with no easy solution especially among the parametric methods and in particular the likelihood approach.

Nonparametric solutions are proposed by Pesarin (1994), Brunner and Munzel (2000), Pesarin (2001), Troendle (2002), Pesarin and Salmaso (2006), Agresti (1999), and Arboretti and Bonnini (2009). The presented problem can be solved by performing $13 \times 2 = 26$ partial permutation tests (Anderson-Darling type test statistics) and combining the p-values first with respect to the strata (within each variable) and then with respect to the variables (Pesarin, 2010; Pesarin and Salmaso, 2010; Bonnini et al., 2014).

3. Methodological solution

The absolute frequency of the number of statistical units on which such a category is observed (i.e. the j-th ordered category) within the stratum s for the v-th variable in the treated and control group is $f_{1j,sv}$ and $f_{0j,sv}$ respectively. For that reason, the cumulative frequencies in the two groups can be denoted by $F_{1j,sv} = \sum_{r=1}^{j} f_{1r,sv}$ and $F_{0j,sv} = \sum_{r=1}^{j} f_{0r,sv}$ respectively. For the partial test concerning the null and the alternative hypothesis, the Anderson-Darling type test statistic is used

$$T_{sv} = \sum_{j=1}^{k_v-1} \left(F_{0j,sv} - F_{1j,sv}\right)\left[F_{\cdot j,sv}\left(n_s - F_{\cdot j,sv}\right)\right]^{-0.5}, \tag{6}$$

where k_v is the number of ordered categories of the v-th variable, $F_{\cdot j,sv} = F_{0j,sv} + F_{1j,sv}$ and $n_s = F_{\cdot k_v,sv}$. In order to address the testing concerning the v-th variable, specifically testing $H_{0,v}$ against $H_{1,v}$, a first-level combination of the significance level functions of the partial tests of the two strata may be applied. If $L_{sv}(t_{sv}) = P[T_{sv} \geq t_{sv}|\boldsymbol{X}]$ is the significance level function for the s-th stratum and the v-th variable given the observed dataset $\boldsymbol{X}$, for any $t_{sv} \in \mathbb{R}$, according to the permutation distribution, a suitable combined test statistic for the v-th variable is

$$T'_v(t_{1v}, t_{0v}) = \max\left[\left(1 - L_{1v}(t_{1v})\right)\left(1 - L_{0v}(t_{0v})\right)\right], \tag{7}$$

for any couple of values $(t_{1v}, t_{0v}) \in \mathbb{R}^2$, where $L_{sv}(t_{sv}) = P[T_{sv} \geq t_{sv}|H_0, \boldsymbol{X}]$ and $\boldsymbol{X}$ is the observed dataset.

Similarly, to solve the general multivariate problem, a second-level combination concerning the variables may be carried out. Let $L'_v(t'_v) = P[T'_v \geq t'_v|\boldsymbol{X}]$ denote the

significance level function of T'_v for any $t'_v \in \mathbb{R}$. The second-level combined test statistic is the following:

$$T''_v(t'_1, \ldots, t'_{13}) = \max\left[\left(1 - L'_1(t'_1)\right), \ldots, \left(1 - L'_{13}(t'_{13})\right)\right]. \quad (8)$$

In the end, H_0 is rejected if the p-value of the combined test is less than or equal to the significance level $\alpha = 0.05$, specifically if $L''(t''_{obs}) \leq \alpha$, where $L''(t'') = P[T'' \geq t''|\boldsymbol{X}]$ with $t'' \in \mathbb{R}$.

Probabilities and p-values are calculated based on the null permutation distributions derived from permuting the rows of $\boldsymbol{X}$, as the condition of exchangeability holds under the null hypothesis.

Generally, in the case of a multivariate two-sample permutation test, the general CPT procedure in the presence of $q \geq 2$ response variables concern testing $H_0: \bigcap_{j=1}^{q} H_0^{(j)}$ against the alternative hypothesis $H_1: \bigcup_{j=1}^{q} H_1^{(j)}$. The decision rule consists in the rejection of H_0 for large values of T, as follows:

$$\hat{\lambda}_{j,0} = \hat{P}\left(T_j \geq T_{j,obs} | H_0^{(j)}\right) = \hat{L}_j\left(T_{j,obs}\right) \quad (9)$$

$$\hat{\lambda}^*_{j,b} = \hat{P}\left(T_j \geq T^*{}_{j,b} | H_0^{(j)}\right) = \hat{L}_j\left(T^*{}_{j,b}\right). \quad (10)$$

An alternative resampling strategy (based on Glivenko-Cantelli theorem) consists in carrying out B (e.g. 10000) independent random samplings from the permutation space (CMC method).

Finally, the combined p-value can be computed in the following way:

$$T^*_{comb,b} = \psi\left(\lambda^*_{1,b}, \ldots, \lambda^*_{q,b}; w_1, \ldots, w_q\right)$$

with $\psi: (0,1)^{2q} \to \mathbb{R}$ a suitable combining function, that must satisfy the following conditions:

1. ψ is monotonic decreasing with respect to each argument: $\psi\left(\ldots, \lambda_{j,b}, \ldots\right) > \psi\left(\ldots, \lambda'_{j,b}, \ldots\right)$ if $\lambda_{j,b} < \lambda'_{j,b}$,
2. upper and lower bound: $\psi\left(\ldots, \lambda_{j,b}, \ldots\right) \to \bar{\psi}$ if $\lambda_{j,b} \to 0$ and $\psi\left(\ldots, \lambda_{j,b}, \ldots\right) \to \underline{\psi}$ if $\lambda_{j,b} \to 1$,
3. limited acceptance region: $\forall \alpha \in (0,1)\ t_{comb,\alpha} < \bar{\psi} < \infty$.

In the literature, one of the most commonly used combining functions is the combination function of Tippett: $\psi_T = \max_j\left(1 - \lambda_j\right)$.

4. Simulation study

In order to study the power behavior of the proposed methodological approach, consisting of combined permutation tests, a Monte Carlo simulation study is carried out. It mainly requires exchangeability under H_0 and it is distribution-free. It is suitable for both two-sided and one-sided alternatives.

The application of the test, the data random generations, and the comparative performance assessments were carried out through original *R* scripts created by the authors. For the evaluation of the methods, according to the null permutation distribution of the test statistics, 1000 random permutations were carried out. For each setting, 1000 datasets (replicates) were randomly generated.

Sample data were randomly simulated from normal q-variate distributions with null mean vector and variance-covariance matrix Σ and then transformed into q-variate categorical data. Formally, we have:

$$\Sigma = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}$$

where ρ represents the Pearson correlation and the covariance between each couple of components of the multivariate normal random variable.

Let z_{ti} denote the observation related to the t-th component observed on the i-th statistical unit, with $t = 1, \dots, q$ and $i = 1, \dots, n$. The rule of the transformation into categorical variables is as follows:

$$x_{ti} = \begin{cases} \text{category 1 if } z_{ti} \leq -1.5 \\ \text{category 2 if } -1.5 < z_{ti} \leq 0 \\ \text{category 3 if } 0 < z_{ti} \leq 1.5 \\ \text{category 4 if } z_{ti} > 1.5 \end{cases}$$

Therefore, the setting parameters in the simulations are the following:

- n: sample size of the control and the treated group,
- q: number of response variables,
- δ: mean value of each underlying normal random variable for the treated group, i.e. for each component of the q-variate response (for the control group the mean is zero),
- ρ: Pearson correlation index that represents the dependence between each couple of the k components of the multivariate response.

The nominal significance level chosen for the simulations, under all the settings considered in the study, is $\alpha = 0.05$.

Looking at Table 1, we can say that the proposed solution, under the null hypothesis, respect the nominal α level. This means that the test is well approximated.

Table 1: Rejection rates of the proposed solution under H_0, with $\delta = 0$, $\rho = 0.30$ and $\alpha = 0.05$

n	q	Rejection rates
10	5	0.004
100		0.004
10	10	0.008
100		0.002

Source: authors' analysis.

Furthermore, in Figure 1, the consistency of the proposed methodology is shown. The greater the sample size n, the greater the power of the proposed nonparametric procedure. The power tends to 1 when n diverges.

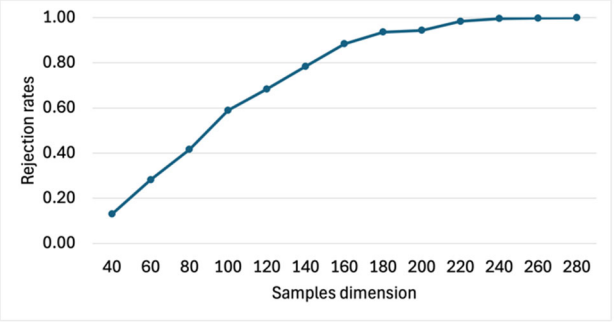


Figure 1: Rejection rates of the proposed test under H_1 as a function of n, with $q = 5$, $\delta = 0.05$, $\rho = 0.30$ and $\alpha = 0.05$

Source: authors' analysis.

The Combined Permutation Tests are exact, unbiased and consistent. CPTs are suitable when the problem can be broken down into several sub-problems (multi-aspect, multivariate, multistrata, multiple comparisons, …) with possible different types of alternatives. Finally, they are suitable for complex alternatives (stochastic ordering, umbrella alternatives, multivariate one-sided test, …), for multivariate numeric, categorical, and mixed variables.

5. Case study

The case study concerns a test on the effect of "assisted motor activity" (AMA) on the health of patients affected by "low back pain" (LBP), "hypertension" and "diabetes".

The health status is measured according to 13 different binary or ordinal outcomes. The components of the ordinal multivariate response variable are listed below:

- X_1 – Perceived Health: Self-assessment on the general state of health (1-bad, 2-neutral, 3-good),
- X_2 – Moderate Activity: Self-assessment on the ability to perform moderate physical activity (1-no, 2-partial, 3-yes),
- X_3 – Stair Climbing: Self-assessment on the difficulty in stair climbing (1-yes, 2-partial, 3-no),
- X_4 – Physical Performance: Physical performance lower than expected in the last month (1-yes, 2-no),
- X_5 – Activity Limitations: Need to limit some types of activity in the last month (1-yes, 2-no),
- X_6 – Emotional State: Physical performance lower than expected due to emotional state in the last month (1-yes, 2-no),
- X_7 – Mind Concentration: Decrease of mind concentration in the last month due to emotional state (1-yes, 2-no),
- X_8 – Pain: Difficulty in daily activities due to pain in the last month (1-very much, 2-somewhat, 3-undecided, 4-not much, 5-not at all),
- X_9 – Calm and Serenity: Frequency of calm and serenity in the last month (1-never, 2-rarely, 3-every once in a while, 4-sometimes, 5-always),
- X_{10} – Full of Energy: Frequency of feeling full of energy in the last month (1-never, 2-rarely, 3-every once in a while, 4-sometimes, 5-always),
- X_{11} – Discouraged and Sad: Frequency of feeling discouraged and sad in the last month (1-always, 2-sometimes, 3-every once in a while, 4-rarely, 5-never),
- X_{12} – Social Activities: Frequency of negative effects of health and emotional state on social activities in the last month (1-always, 2-sometimes, 3-every once in a while, 4-rarely, 5-never),
- X_{13} – Stress Level: Self-assessment of the level of stress (1-very high, 2-high, 3-average, 4-moderate, 5-low, 6-very low).

In this application, the null hypothesis of no effect of AMA on the health of treated patients (t.p.) is tested against the alternative hypothesis of positive effect of AMA on the health of treated patients. In formula we have the following:

$$H_0: [health_1 =^d health_0] \cap [activ_1 =^d activ_0] \cap \ldots \cap [stress_1 =^d stress_0]$$
$$\equiv H_0: \bigcap_{v=1}^{13} H_{0,v}$$

vs

$$H_1: [health_1 >^d health_0] \cup [activ_1 >^d activ_0] \cup \ldots \cup [stress_1 >^d stress_0]$$
$$\equiv H_0: \bigcup_{v=1}^{13} H_{1,v}.$$

Looking at Table 1, the joint frequency distribution of treatment and confounding factors are reported. It is evident that, considering the stratified methodology, we deal with small sample sizes. Hence, the need to use a non-parametric approach becomes crucial.

Table 2: Joint frequency distribution of treatment and confounding factor

Stratum	Group		Total
	0	**1**	
Over 60 with LBP (s=1)	10	7	**17**
Oher patients (s=0)	10	20	**30**
Total	**20**	**27**	**47**

Source: authors' analysis.

Each partial test can be broken down into two sub-partial tests, i.e. one for each stratum. Let $s = 1$ and $s = 0$ denote the stratum of over 60 with LBP and other patients respectively:

$$H_0: \bigcap_{v=1}^{13} [H_{0,v0} \cap H_{0,v1}] \equiv H_0: \bigcap_{v=1}^{13} \bigcap_{s=0}^{1} H_{0,vs}$$

vs

$$H_0: \bigcup_{v=1}^{13} [H_{1,v0} \cup H_{1,v1}] \equiv H_0: \bigcup_{v=1}^{13} \bigcup_{s=0}^{1} H_{1,vs}.$$

Summarizing the characteristics of the problem, we are in the presence of a multivariate test with 13 marginal variables, categorical data, a directional alternative hypothesis (stochastic dominance), a stratified test to avoid confounding effects and small sample sizes.

The application of the described method leads an overall p-value of 0.019. Thus, we have empirical evidence in favor of the hypothesis of significant effect of AMA on the health of patients (i.e. significant effect of the treatment).

To attribute the overall significance to some specific outcomes (second level partial tests) we need to adjust the partial p-values to control the Family Wise Error Rate (FWER), i.e. the probability of wrong rejection of one or more partial null hypotheses.

According to the Bonferroni-Holm procedure the adjusted p-values are computed as follows:

$$\tilde{p}_{(k)} = \max_{i \le k} \left\{ \min\left[1, (13 + 1 - i) p_{(i)}\right] \right\}, \tag{11}$$

where $p_{(1)} \le p_{(2)} \le \cdots \le p_{(q)}$ are the sorted p-values of the partial tests.

Table 2: Adjusted p-values of the partial tests (significance in bold)

Test	Variable	Adjusted p-value
T_1	Perceived Health	0.361
T_2	Moderate Activity	0.357
T_3	Stair Climbing	0.232
T_4	Physical Performance	0.444
T_5	Activity Limitations	0.444
T_6	Emotional State	0.444
T_7	Mind Concentration	0.304
T_8	Pain	0.444
T_9	Calm and Serenity	0.444
T_{10}	Full of Energy	0.444
T_{11}	Discouraged and Sad	**0.019**
T_{12}	Social Activities	0.444
T_{13}	Stress Level	0.258

Source: authors' analysis.

According to Table 2, the positive effect of the treatment on the health of patients can only be attributed to a decrease in sadness and discouragement.

6. Conclusions

This work aims to test the null hypothesis of no effect of AMA on the health of treated patients versus the alternative hypothesis of a positive effect of AMA on the health of treated patients. The state of health is a multivariate categorical variable and for such a problem of stochastic dominance, the combined permutation test method is a valid solution. The proposed approach represents a valid solution for the described testing problem, whose complexity is due to the multivariate nature of the response, the categorical data, the directional alternative hypothesis, the presence of confounders and the small sample sizes. The application to the case study provides an overall p-value of the CPT equal to 0.019, which provides empirical evidence in favor of the hypothesis of the significant effect of AMA on the health of patients. Furthermore, looking at the partial p-values, the positive effect of the treatment on the health of patients can only

be attributed to a decrease in sadness and discouragement. The good power behavior and in particular the consistency of the proposed test are proved in the simulation study.

The proposed method offers several promising directions for future research. First, it could be extended to accommodate longitudinal multivariate categorical data, allowing for the analysis of repeated measures or follow-up studies while preserving the nonparametric nature of the test. Moreover, integration with hierarchical or mixed-effects frameworks would enable its application in multi-center studies or clustered data settings. Another valuable advancement would be the development of software packages to facilitate wider adoption of the method in applied research. Robust handling of missing data through imputation strategies combined with permutation logic also represents a key area for enhancement. Furthermore, the extension of the method to comparisons involving more than two groups could broaden its applicability in complex experimental designs. In terms of application, the approach may prove particularly useful in the evaluation of public policy interventions using categorical indicators in socio-economic studies. Additional relevant fields include educational research (e.g. assessing student outcomes through ordinal rating scales), marketing and consumer behavior (e.g. preference or satisfaction surveys), and environmental epidemiology, where categorical health outcomes are compared across groups exposed to different environmental conditions. These areas all stand to benefit from a flexible, robust, and distribution-free testing procedure such as the one proposed in this study.

References

Alibrandi, A.; Giacalone, M.; Zirilli, A., (2022). Psychological stress in nurses assisting Amyotrophic Lateral Sclerosis patients: A statistical analysis based on Non-Parametric Combination test. *Mediterr. J. Clin. Psychol.*, 10.

Agresti, A., (1999). Modelling ordered categorical data: recent advances and future challenges. *Statistics in medicine*, 18(17-18), pp. 2191–2207.

Agresti, A., Klingenberg, G. B., (2005). Multivariate tests comparing binomial probabilities, with application to safety studies for drugs. *J. R. Stat. Soc.*, Ser. C 4, pp. 691–706

Arboretti Giancristofaro, R., and Bonnini, S., (2009). Some new results on univariate and multivariate permutation tests for ordinal categorical variables under restricted alternatives. *Statistical Methods and Applications*, 18, pp. 221–236.

Bonnini, S., Borghesi, M., (2022). Relationship between Mental Health and Socio-Economic, Demographic and Environmental Factors in the COVID-19 Lockdown Period—A Multivariate Regression Analysis. *Mathematics*, 10, pp. 3237.

Bonnini, S., Borghesi, M., and Giacalone, M., (2023). Advances on multisample permutation tests for "V-shaped" and "U-shaped" alternatives with application to Circular Economy. *Ann. Oper. Res.*, 342, pp. 1655–1670.

Bonnini, S., Borghesi, M., and Giacalone, M., (2024a). Simultaneous marginal homogeneity versus directional alternatives for multivariate binary data with application to circular economy assessments. *Appl. Stoch. Models Bus. Ind.*, 40, pp. 389–407.

Bonnini, S., Borghesi, M., and Giacalone, M., (2024b). Semi-parametric approach for modeling overdispersed count data with application to industry 4.0. *Socio-Econ. Plan. Sci.*, 95, p. 101976.

Bonnini, S., Corain, L., Marozzi, M., and Salmaso, L., (2014). *Nonparametric Hypothesis Testing. Rank and Permutation Methods with Applications in R*. Wiley, Chichester.

Bonnini, S., Assegie, G. M., and Trzcinska, K., (2024). Review about the Permutation Approach in Hypothesis Testing. *Mathematics*, 12, pp. 2617–1.

Brunner, E., Munzel, U., (2000). The nonparametric Behrens–Fisher problem: asymptotic theory and small-sample approximation. *Biom. J.*, 42, pp. 17–25

Cohen, A., Kemperman, J. H. B., Madigan, D., and Sakrowitz, H. B., (2000). Effective directed tests for models with ordered categorical data. *Aust. N. Z. J. Stat.*, 45, pp. 285–300

Ferreira, M. L., Ferreira, P. H., Latimer, J., Herbert, R. D., Hodges, P. W., Jennings, M. D., ... and Refshauge, K. M., (2007). Comparison of general exercise, motor control exercise and spinal manipulative therapy for chronic low back pain: a randomized trial. *Pain*, 131(1-2), pp. 31–37.

Gordon, R., Bloxham, S., (2016, April). A systematic review of the effects of exercise and physical activity on non-specific chronic low back pain. In *Healthcare* (Vol. 4, No. 2, p. 22). MDPI.

Han, K. E., Catalano, P. J., Senchaudhuri, P., and Metha, C., (2004). Exact analysis of dose-response for multiple correlated binary outcomes. *Biometrics,* 60, pp. 216–224

Hirotsu, C., (1986). Cumulative chi-squared statistic as a tool for testing goodness-of-fit. *Biometrika,* 73, pp. 165–173

Loughin, T. M., (2004). A systematic comparison of methods for combining p-values from independent tests. *Comput. Stat. Data Anal.*, 47, pp. 467–485

Loughin, T. M., Scherer, P. N., (1998). Testing for association in contingency tables with multiple column responses. *Biometrics*, 54, pp. 630–637

Lumely, T., (1996). Generalized estimating equations for ordinal data: a note on working correlation structures. *Biometrics*, 52, pp. 354–361

Macedo, L. G., Latimer, J., Maher, C. G., Hodges, P. W., McAuley, J. H., Nicholas, M. K., ... and Stafford, R., (2012). Effect of motor control exercises versus graded activity in patients with chronic nonspecific low back pain: a randomized controlled trial. *Physical therapy*, 92(3), pp. 363–377.

Macedo, L. G., Maher, C. G., Hancock, M. J., Kamper, S. J., McAuley, J. H., Stanton, T. R., ... and Hodges, P. W., (2014). Predicting response to motor control exercises and graded activity for patients with low back pain: preplanned secondary analysis of a randomized controlled trial. *Physical therapy*, 94(11), pp. 1543–1554.

Marcus, R., Eric, P., and Gabriel, K. R., (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63(3), pp. 655–660

Nettleton, D., Banerjee, T., (2000). Testing the equality of distributions of random vectors with categorical components. *Comput. Stat. Data Anal.*, 37, pp. 195–208

Pesarin, F., (1994). Goodness-of-fit testing for ordered discrete distributions by resampling techniques. *Metron,* 52, pp. 57–71.

Pesarin, F., (2001). *Multivariate Permutation Tests with Applications in Biostatistics*. Wiley: Chichester.

Pesarin, F, Salmaso, L., (2006). Permutation tests for univariate and multivariate ordered categorical data. *Austrian Journal of Statistics,* 35, pp. 315–324.

Pesarin F., Salmaso L., (2010). *Permutation Tests for Complex Data. Theory, Applications, and Software*. Wiley, Chichester.

Silvapulle, M. J., Sen, P. K., (2005). *Constrained Statistical Inference, Inequality, Order, and Shape Restrictions*. Wiley, New York

Simon, N., Tibshirani, R., (2012). A permutation approach to testing interactions in many dimensions. *arXiv*, arXiv, 1206.6519.

Stute, W., Thies, S., and Zhu, L., (1998). Model checks for regression: An innovation process approach. *Ann. Statist*., 26, pp. 1916–1934.

Toma, P., Miglietta, P. P., Zurlini, G., Valente, D., and Petrosillo, I., (2017). A non-parametric bootstrap-data envelopment analysis approach for environmental policy planning and management of agricultural efficiency in EU countries. *Ecol. Indic*., 83, pp. 132–143.

Troendle, J. F., (2002). A likelihood ratio test for the nonparametric Behrens–Fisher problem. *Biom. J.,* 44(7), pp. 813–824

Wang, Y., (1996). A likelihood ratio test against stochastic ordering in several populations. *J. Am. Stat. Assoc*., 91, pp. 1676–1683

STATISTICS IN TRANSITION new series, September 2025
Vol. 26, No. 3, pp. 195–208, https://doi.org/10.59139/stattrans-2025-034
Received – 10.12.2024; accepted – 24.04.2025

Normality tests for transformed large measured data: a comprehensive analysis

Abu Feyo Bantu[1], **Andrzej Kozyra**[2], **Józef Wiora**[3]

Abstract

In statistical analysis, evaluating the normality of large datasets is crucial for validating parametric tests, particularly in areas such as Global Navigation Satellite System (GNSS) measurements, where data often exhibit non-normal characteristics resulting from their variability and errors. This research aims to transform the measured GNSS data and to assess the effectiveness of transformation methods in achieving normality. Techniques like logarithmic, quantile and rank-based Inverse Normal Transformation (INT) were evaluated using visual methods (histograms, Q-Q plots), descriptive statistics (skewness, kurtosis) and statistical tests, including Kolmogorov-Smirnov (KS), Anderson-Darling (AD), Lilliefors (LF), D'Agostino K-squared (DA), Shapiro-Wilk (SW), Jarque-Bera (JB), Cramér-von Mises (CM), and Pearson Chi-square (Chi2) tests. The sensitivity of these tests to deviations from normality was assessed through the Receiver Operating Characteristic (ROC) analysis and the Area Under the Curve (AUC) values at a significance level of 0.1, using Monte Carlo (MC) simulations across the varying sample sizes. The results showed that untransformed latitude data consistently failed normality tests, while transformed data displayed normal characteristics. The rank-based INT showed superior effectiveness, influenced by the original distribution and characteristics of the dataset. The findings underscore the importance of tailored transformations in large-scale data applications, enhancing the accuracy and applicability of parametric statistical methods in geospatial and other industrial domains.

Key words: GNSS, ROC, normality test, statistical analysis.

1. Introduction

In statistical analysis, assessing the normality of datasets is essential, as many parametric tests rely on the assumption of normally distributed data as described in (Tabachnick et al., 2019). In highly competitive manufacturing industries, pilot runs typically involve very small sample sizes to accelerate the launch of new products. While these small datasets may approximate normal distributions due to their limited variability, large datasets, such as those collected from Global Navigation Satellite Systems (GNSS), often deviate significantly from normality due to various sources of error (Li et al., 2016). Yap and Sim (2011) highlighted that these non-normal characteristics, influenced by factors such as atmospheric

[1]Department of Measurements and Control Systems, Silesian University of Technology, Gliwice, Poland. E-mail: abu.feyo.bantu@polsl.pl. ORCID: https://orcid.org/0000-0001-6463-9864.
[2]Department of Measurements and Control Systems, Silesian University of Technology, Gliwice, Poland. E-mail: andrzej.kozyra@polsl.pl. ORCID:https://orcid.org/0000-0003-2645-3537.
[3]Department of Measurements and Control Systems, Silesian University of Technology, Gliwice, Poland. E-mail: jozef.wiora@polsl.pl. ORCID: https://orcid.org/0000-0002-8450-8623.

disturbances, multipath effects, and satellite geometry, pose challenges to the application of traditional parametric methods. Fault detection in GNSS systems is essential for ensuring reliability; however, traditional methods often assume Gaussian noise, which limits their effectiveness in real-world scenarios where measurement noise deviates from normality. Yan (2024) proposed a jackknife-based test statistic for fault detection in linearized pseudorange positioning systems, which does not rely on specific noise distribution assumptions. This method, combined with hypothesis testing using the Bonferroni correction, enhances robustness against non-Gaussian noise, making it particularly suitable for GNSS applications.

While such robust techniques improve fault detection and identify sources of errors in GNSS applications, the assumption of normality in large datasets remains a debated topic. Wilcox (2010) argues that, based on the central limit theorem, data naturally tend toward a normal distribution as sample size increases. This suggests that normality assumptions hold in large datasets, regardless of the method used to assess normality. However, Demir (2022) challenges this notion; it is not true: since the number of data is large, the data will not always have a normal distribution, particularly in the presence of measurement errors and non-Gaussian noise.

To address this limitation, data transformation techniques are widely used to modify data distributions and achieve a closer approximation to normality. According to Osborne (2010), these methods are crucial for fulfilling statistical assumptions, enhancing effect sizes, and preparing datasets for comprehensive analysis. Huang et al. (2023) emphasizes that data transformation is a commonly employed technique for normalizing data to enhance statistical modeling. The study also outlines various transformation methods, such as logarithmic, log-sinh, Box-Cox, Yeo-Johnson, and square root approaches, each designed to address different types of non-normality. For instance, the Box-Cox transformation is a flexible family of power transformations that adjusts data to stabilize variance and reduce skewness. Peterson (2021) further elaborates on the implementation of the Box-Cox transformation, emphasizing its ability to select an optimal parameter (λ) that maximizes normality. Similarly, the Yeo-Johnson transformation extends the Box-Cox method to accommodate negative values, making it more versatile for diverse datasets (Cai and Xu, 2024).

Despite the availability of these transformation techniques, selecting the most appropriate method for large, real-world datasets remains challenging due to their varying characteristics and sources of variability (Khatun, 2021). Normality tests such as Shapiro-Wilk, Anderson-Darling, and D'Agostino's K-squared provide insights into dataset distributions but offer limited guidance on effectively transforming data to satisfy normality assumptions (Razali and Wah, 2011). Moreover, studies like Ogaja (2022) highlight the importance of preprocessing steps in GNSS data analysis, where raw observables undergo rigorous processing to estimate geodetic parameters. Barba et al. (2021) demonstrate the use of adapted R packages to analyze GNSS time series, focusing on displacement velocities, noise levels, and temporal forecasts. Their work underscores the necessity of identifying and addressing outliers, gross errors, and noise to ensure reliable interpretation of GNSS data.

This research aims to transform the measured data and evaluate the efficiency of transformation methods in the normality of large GNSS datasets. It compares log, quantile, and rank-based INT transformations to provide insights for improving the reliability of parametric statistical methods in measurement-driven fields. We evaluated data normality using

graphical, descriptive, and statistical methods, including p-value tests, to examine the effects of transformations. The transformed data were then compared with untransformed data to assess improvements in normality.

The remainder of this paper is as follows: Section 2 reviews the relevant data transformation and normality tests. Section 3 introduces our methods and data collection techniques. Section 4 presents the results and discussion through graphical representations, descriptive and statistical analysis, and evaluation of normality test performance. Finally, Section 5 concludes the study.

2. Theory

2.1. Data transformation techniques

Raymaekers and Rousseeuw (2024) describe data transformation as a technique for handling non-Gaussian data by preprocessing variables to approximate normality. This process ensures that the transformed data are nearly normal at their core, while a few outliers may still deviate. Here is a breakdown of common techniques, including their formula as shown in Table 1.

Table 1: Data transformation techniques

Transformation techniques	Formula	When to use
Log Transform	$y = \log(x)$	When data are positively skewed.
Quantile Transform	$y_i = \text{Quantile}(x_i)$	When you want to map data to a target distribution (e.g. normal).
Rank-Based INT	$y_i = \Phi^{-1}\left(\frac{r_i}{N+1}\right)$	When data are highly non-normal or non-monotonic.

where y_i is the transformed data point after mapping, x_i is the original data point, Φ^{-1} is the inverse CDF of the normal distribution, N is the total number of data points, and r_i is the rank of the i-th data point in the dataset.

Log transformation is a widely used statistical method primarily applied to reduce data skewness and mitigate variability caused by outliers. It is commonly believed to enhance normality by making the data distribution more closely resemble a normal distribution (Sun and Xia, 2024). As described by Ghasemi and Zahediasl (2012), it involves taking the logarithm of each data value, which compresses the range of large values while preserving the relationships between data points.

According to Pham (2021), quantile transformation is a statistical technique that converts data from its original distribution to a target distribution. This approach effectively normalizes non-normal data and mitigates issues caused by outliers and heavy tails.

As explained by McCaw et al. (2020), the rank-based inverse normal transformation (INT) is commonly applied to highly skewed data and non-normally distributed traits. INT is a non-parametric mapping that replaces sample quantiles by quantiles from the standard normal distribution. After INT, the marginal distribution of any continuous outcome is asymptotically normal. INT has the effect of symmetrizing and concentrating the residual

distribution around zero. After applying data transformations to the dataset, normality was evaluated using various testing techniques to assess how well the data conform to a normal distribution.

2.2. Normality tests

A normality test is a statistical procedure used to determine whether a dataset follows a normal distribution. Since many statistical methods rely on the assumption of normality in population data, it is crucial to verify whether this assumption holds before applying such methods (Kwak and Park, 2019). There are three major methods to check the normality: graphical, descriptive, and statistical (Zygmonta, 2023).

Graphs allow the easy assessment of major data departures from normality (Obilor and Amadi, 2018). The normal Quantile-Quantile (Q-Q) plot is a widely used and successful method for determining data normality. As explained by Stine (2017), it validates the assumption of normality in a sample dataset. It compares the two datasets to see if they have the same distribution. A histogram is a bar graph representing the frequency distribution of a dataset. It divides data into bins and counts the number of observations in each bin. For a normally distributed dataset, it exhibits a bell-shaped curve, symmetric around the mean. If the dataset is non-normal, the histogram deviates from the bell-shaped curve.

The second method is descriptive analysis, which uses skewness (S) and kurtosis (K) to evaluate the shape of data distributions. This is one of the most commonly employed techniques for this purpose. Skewness S evaluates the asymmetry of a probability distribution, indicating how much the data deviate from a symmetrical distribution. A normal distribution, which is symmetrical and bell-shaped, has zero skewness. A positively skewed distribution ($S > 0$) has a long right tail, while a negatively skewed one ($S < 0$) has a long left tail with most values concentrated on the right (Tabachnick et al., 2019).

Kurtosis K is a statistical measure that describes the shape of a probability distribution by analysing its tails and peak relative to a normal distribution. According to Tabachnick et al. (2019), positive kurtosis indicates a distribution with heavier tails and a sharper peak, whereas negative kurtosis reflects lighter tails and a flatter peak. Excess kurtosis, calculated as $K-3$, compares the kurtosis of a given distribution to that of a normal distribution. There are three main types of kurtosis: mesokurtic $(K = 3, \text{excess kurtosis} = 0)$, representing a normal distribution; leptokurtic $(K > 3, \text{excess kurtosis} > 0)$, indicating a distribution with heavier tails and a sharper peak; and platykurtic $(K < 3, \text{excess kurtosis} < 0)$, which corresponds to a distribution with lighter tails and a flatter peak than a normal distribution.

The third technique is statistical, and p-values of tests such as Kolmogorov-Smirnov (KS), Anderson-Darling (AD), Lilliefors (LF), D'Agostino's (DA), Shapiro-Wilk (SW), Jarque-Bera (JB), Cramér-von Mises (CM), and Pearson's Chi-square (Chi2).

The KS test compares the empirical Cumulative Distribution Function (CDF) of a sample to the CDF of a reference distribution (e.g. normal). The test statistic, as described in Kumbhar et al. (2024), is determined by the maximum absolute difference between the two CDFs. The Anderson-Darling test assesses whether a sample comes from a specified distribution. It leverages the principle that, under the assumption that the data originate from the hypothesized distribution, their CDF should follow a uniform distribution (Kwak and Park,

2019). The AD test assesses normality by comparing the test statistic to a critical value (cv) at a given significance level. It is a variation of the KS test, which applies additional weighting to emphasize differences in the tails of the distribution. The LF test is a variation of the KS test; it corrects for the fact that the parameters (mean and variance) of the normal distribution are estimated from the sample, which is calculated as in (Uyanto, 2022). The test statistic is the maximal absolute difference between empirical and hypothetical CDF. The DA test checks the S and K of the data and compares it to a normal distribution. It is used to assess if a dataset deviates from normality in terms of its symmetry (skewness) and peakness (kurtosis), with the detailed formula for its calculation provided in (D'Agostino, 2017).

The SW test evaluates how well the sample data fits a normal distribution by comparing the sample to the expected values under normality, with its statistical formula provided in (Kwak and Park, 2019).

The JB test is employed to verify the normality of a data set before applying standard statistical tests such as the t-test, z-test, or F-test. For a detailed explanation of the formula and parameters, refer to (Aslam et al., 2021). It evaluates S and K in the data, comparing them to their expected values under a normal distribution.

The CVM test measures the difference between the empirical distribution function of a sample and the CDF of the normal distribution. It is particularly useful when analyzing the tails of the probability density function (PDF). For a detailed description of its equation (Von Mises, 2014). The Chi2 test compares the observed frequencies in bins with the expected frequencies for a normal distribution. The original test by Pearson was designed to see whether an observed set of frequency O was in agreement with a multinomial distribution with parameters $m, p_1, \ldots, p_k$. This is done by calculating the expected frequency $E_i = mp_i$ and the test statistic $X = \sum(O-E)^2/E$ (Rolke and Gongora, 2021).

Different normality tests often produce different results; some tests reject the null hypothesis of normality while others fail to reject it (Demir, 2022). Therefore, it is better to crosscheck the normality of the data using different techniques. The statistical value and median p-value are the essential elements of the comparison tests (Khatun, 2021). The statistical value measures the difference between the observed and normal distributions. The median p-value for a normality test is defined as:

$$p_{\text{median}} = \begin{cases} p\left(\frac{n+1}{2}\right) & \text{if } n \text{ is odd,} \\ \dfrac{p\left(\frac{n}{2}\right) + p\left(\frac{n}{2}+1\right)}{2} & \text{if } n \text{ is even.} \end{cases} \tag{1}$$

It is the middle value of the p-values obtained from repeated tests on different samples $p_1, p_2, \ldots, p_n$, where p_i represents the p-value from the i-th normality test. It is a non-parametric measure of the central tendency of the p-values (Tabachnick et al., 2019). It helps to assess the normality by considering its skewness and kurtosis. By analyzing these p-values at 5% and 10% significance levels and varying sample sizes, we can gauge how well the data conform to a normal distribution. A p-value below 0.05 suggests rejecting the null hypothesis (H0), indicating non-normality, while a high p-value suggests insufficient evidence to reject H0, indicating possible normality.

After conducting normality tests, the performance of these tests was evaluated using the ROC curve with Area Under the Curve (AUC) values, offering insight into their ability to detect normality in both the original and transformed data. ROC is a graphical tool used to assess the diagnostic ability of a binary classifier, such as normality tests in statistics (Patrício et al., 2017). It is a plot of the True Positive Rate (TPR) versus the False Positive Rate (FPR) at various threshold settings. The TPR is:

$$\text{TPR} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Negative (FN)}} \tag{2}$$

and FPR is:

$$\text{FPR} = \frac{\text{False Positive (FP)}}{\text{False Positive (FP)} + \text{True Negative (TN)}} \tag{3}$$

so to plot the ROC curve, we need to compute TPR and FPR for multiple threshold values. The AUC summarizes the entire location of the ROC curve rather than depending on a specific operating point. It is an effective and combined measure of sensitivity and specificity that describes the inherent validity of diagnostic tests (Nahm, 2022). The AUC value ranges from 0 to 1:

- AUC = 1: Perfect classification,
- AUC = 0.5: No discrimination (i.e. random guessing),
- AUC < 0.5: Worse than random guessing.

A higher AUC indicates better performance of the normality test, as it means the test can better distinguish between normal and non-normal data. It can be approximated by the trapezoidal rule as follows:

$$\text{AUC} = \sum_{i=1}^{n-1} 2\left(\text{FPR}_i - \text{FPR}_{i-1}\right) \cdot \left(\text{TPR}_i + \text{TPR}_{i-1}\right) \tag{4}$$

where FPR_i and TPR_i are the values at each threshold.

3. Methods

To collect GNSS sample data, we deployed an L76X GPS module in a fixed outdoor location. This module, known for its compact design and high positioning accuracy, was connected to a Raspberry Pi, which served as the control system. The Raspberry Pi handled module configuration, data logging, and real-time monitoring, ensuring continuous and reliable data collection throughout the experiment.

Over a continuous period of 96 hours, the GPS module recorded a substantial amount of GNSS data, generating a total of over two million lines formatted in the National Marine Electronics Association (NMEA) standard. This widely used format in GNSS systems organizes navigation and positional data into standardized sentences, making it ideal for subsequent analysis. From the entire recorded dataset, `$GPGGA` sentences, with a specific

type of NMEA message were extracted. These sentences contain key positional and signal quality parameters, including fix quality, satellite count, horizontal dilution of precision (HDOP), and latitude / longitude, all of which are essential for evaluating the GNSS performance. A total of 270 791 rows of `$GPGGA` sentences were isolated from the raw dataset for detailed analysis. Python scripts were utilized to efficiently filter, parse, and structure the extracted data into a log file.

4. Results and discussion

4.1. Graphical data analysis

Figure 1 highlights the analysis of untransformed latitude data. The histogram on the left shows a positively skewed distribution, deviating from the normal curve, indicating non-normality. The Q-Q plot on the right further confirms this by showing significant deviations of data points from the reference line, particularly in the tails. These results emphasize the challenges of applying parametric statistical methods to non-normal latitude data.

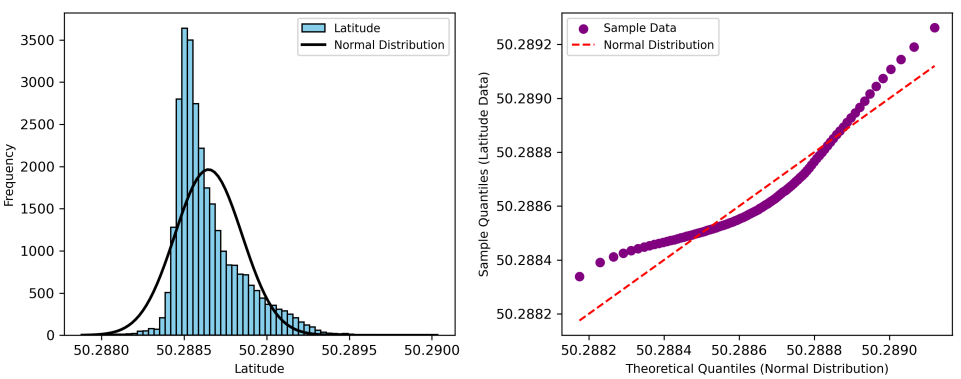


Figure 1: Distribution of untransformed GNSS latitude data histogram plot (left) and Q-Q plot (right)

Figure 2 demonstrates the success of a rank-based inverse normal transformation (INT) in normalizing latitude data. The histogram shows a symmetric, bell-shaped distribution, while the Q-Q plot confirms that the transformed data points closely align with those of a standard normal distribution. These results indicate the INT effectiveness in making the data suitable for parametric statistical methods.

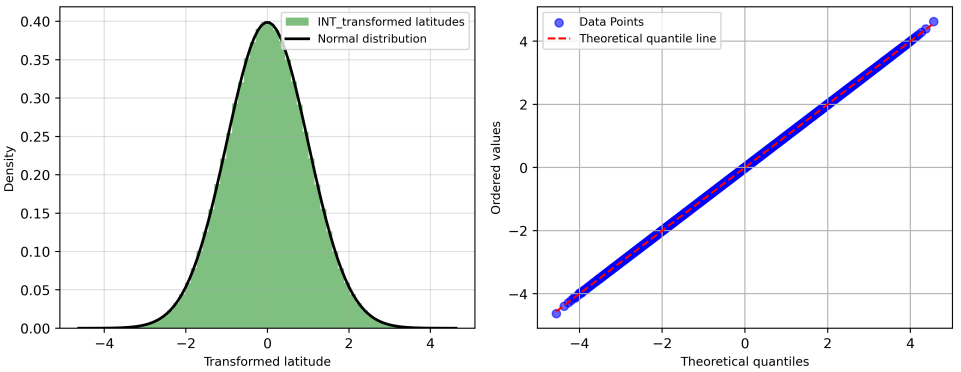


Figure 2: Distribution of transformed GNSS latitude data histogram plot (left) and Q-Q plot (right)

4.2. Descriptive data analysis (skewness and kurtosis)

Table 2 provides a statistical comparison of skewness and kurtosis values between transformed and untransformed datasets, applying rank-based INT, quantile transformation, and log transformation. This comparison underscores the effectiveness of these methods in addressing distributional issues such as skewness and kurtosis, which are pivotal for accurate statistical analysis and enhancing model performance.

Table 2: Comparison of transformed and untransformed descriptive data statistics

Test Type	Transformed data			Untransformed data
	Rank-based INT	Quantile Transform	Log Transform	
Skewness	0.000001	0.019956	1.241654	1.241669
Kurtosis	-0.000215	0.116283	1.961968	1.962022

The rank-based INT transformation achieves a skewness of 0.000 001, effectively neutralizing asymmetry in the data. This indicates an almost perfect symmetric distribution. The quantile transformation reduces skewness to 0.019 956, which is close to zero, reflecting minimal asymmetry. Both the log transform and untransformed data retain significant skewness: 1.241 654 and 1.241 669, respectively. This suggests the data are positively skewed and remain far from a symmetric distribution.

In the case of kurtosis, rank-based INT yields a kurtosis value of -0.000 215, achieving a near-normal distribution by minimizing the impact of extreme values. The quantile transformation also reduces kurtosis to 0.116 283, demonstrating a flattened distribution compared to the untransformed data. However, the log transform shows kurtosis values of 1.961 968, nearly similar to the untransformed data 1.962 022, indicating a heavy-tailed distribution persists.

Overall, the rank-based INT transformation is the most effective method for addressing distributional issues in the dataset, achieving near-normal distribution properties with minimal skewness and kurtosis. Quantile Transformation also performs well but is less ef-

fective than INT. Log transformation retains significant skewness and kurtosis, making it less suitable for analyses requiring normality.

4.3. Statistical and p-value of tests

Table 3 compares statistical test results for transformed and untransformed data to assess normality at a significance level of $\alpha = 0.05$. The untransformed data exhibit significant deviation from a normal distribution, as indicated by high statistical values and consistently low p-values (0.000) across most tests.

Table 3: Comparison of transformed and untransformed statistics and p-value

Test Type	Transformed data		Untransformed data	
	Statistic value	p-value	Statistic value	p-value
Kolmogorov-Smirnov (KS)	0.001	0.964	0.133	0.000
Anderson-Darling (AD)	0.020	0.787(cv)	$9.58\text{x}10^{3}$	0.787(cv)
Lilliefors (LF)	0.000	0.990	0.132	0.001
D'Agostino's K-squared (DA)	0.000	1.000	$5.60\text{x}10^{4}$	0.000
Shapiro-Wilk (SW)	1.000	1.000	0.897	0.000
Jarque-Bera (JB)	0.001	1.000	$1.13\text{x}10^{5}$	0.000
Cramér-von Mises (CM)	0.004	1.000	$9.03\text{x}10^{4}$	0.000
Pearson's Chi-square (Chi2)	$4.12\text{x}10^{5}$	0.000	$5.77\text{x}10^{5}$	0.000

The transformed data consistently produce statistical values near 0.000 and p-values greater than $\alpha = 0.05$, indicating a failure to reject the H0. However, the Chi2 test reports a high statistical value and a p-value below $\alpha = 0.05$, suggesting the H0 is rejected. At a significance level of $\alpha = 0.05$, the AD test's critical value (0.787) exceeds the calculated statistical value (0.020), indicating a failure to reject the H0. All tests effectively validated the normality of the transformed data, confirming no significant deviations from normality, except for the Chi2 test. The p-values of all tests are 0.000, except for the AD test, where the critical value (0.787) is lower than the corresponding statistical value, indicating a rejection of H0 and significant deviations from normality in the case of untransformed data.

4.4. Performance evaluation of normality tests

4.4.1 Comparison of ROC for untransformed latitudes

From Figure 3, the ROC curves are clustered around the diagonal line, with AUC values close to 0.5 for most tests. This suggests that at smaller sample sizes (45, 55, 65, 75, 85, 95, 105), the tests struggle to differentiate between normality and deviations from it, exhibiting limited sensitivity and specificity.

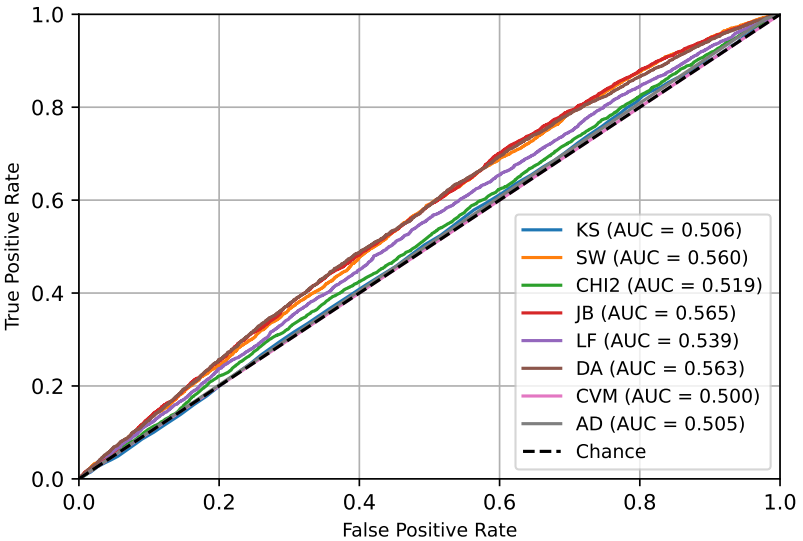


Figure 3: ROC analysis for untransformed GNSS latitude data with a small sample size

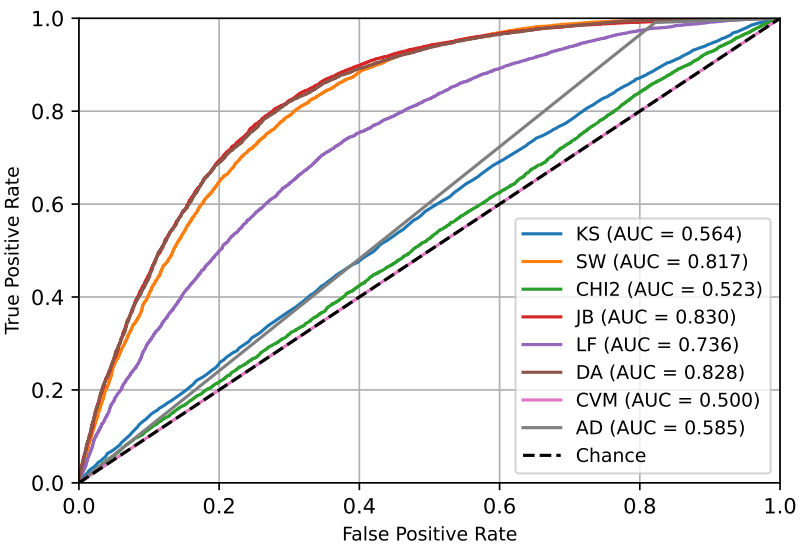


Figure 4: ROC analysis for untransformed GNSS latitude data with a large sample size

Figure 4 describes ROC curve analysis of tests at 0.1% significance level. At larger sample sizes (250, 350, 450, 550, 650, 750, 850, 950, 1050), tests like SW, JB, and DA demonstrate enhanced discriminatory power, becoming more effective at detecting deviations from normality. This improvement is likely due to the increased statistical power associated with larger datasets.

4.4.2 Comparison of ROC for transformed latitudes

Figure 5 and Figure 6 demonstrate the ROC curves and corresponding AUC metrics for various normality tests applied to transformed latitude data for small and large sample sizes at the significance level of 0.1%. From Figure 5, all AUC values are approximately between 0.500 and 0.511, which indicates that the performance of the normality tests is nearly equivalent to random classification. This suggests that the transformations applied to the latitude data were successful in normalizing the data, making it indistinguishable from a truly normal distribution at smaller sample sizes.

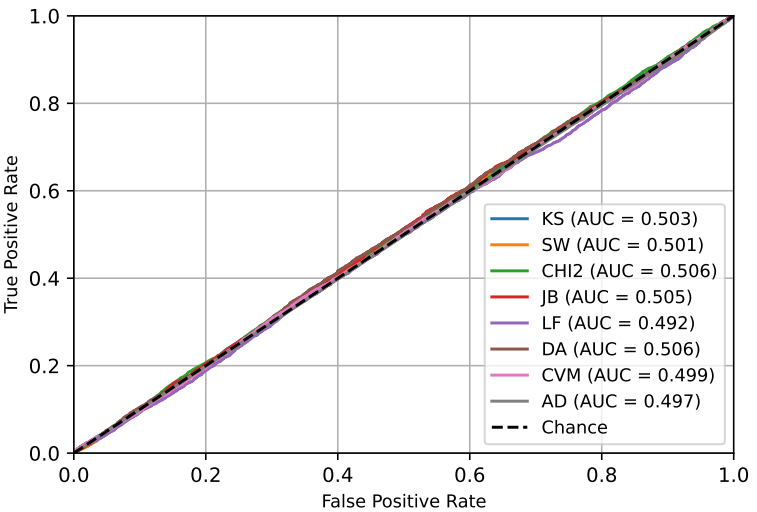


Figure 5: ROC analysis for transformed GNSS latitude data with a small sample size

From Figure 6, all tests have AUC values near 0.5, suggesting that the tests are performing no better than random classification in distinguishing normality from non-normality in the transformed data. This indicates no significant separation between the true positive rate (sensitivity) and the false positive rate. For larger sample sizes, the transformations applied to the latitude data effectively normalize the dataset. If the transformed data closely resemble the synthetic normal data, the tests may struggle to identify significant differences, leading to a high number of false positives and false negatives. This ultimately results in a lower AUC score. An AUC near 0.5 suggests that the test is essentially making random guesses due to the indistinguishability of the data.

A moderate AUC (0.6 - 0.7) implies that the transformation of latitude data into a normal distribution is not entirely successful. Conversely, a high AUC > (0.8) indicates that the transformation has not completely normalized the data, allowing for detectable differences.

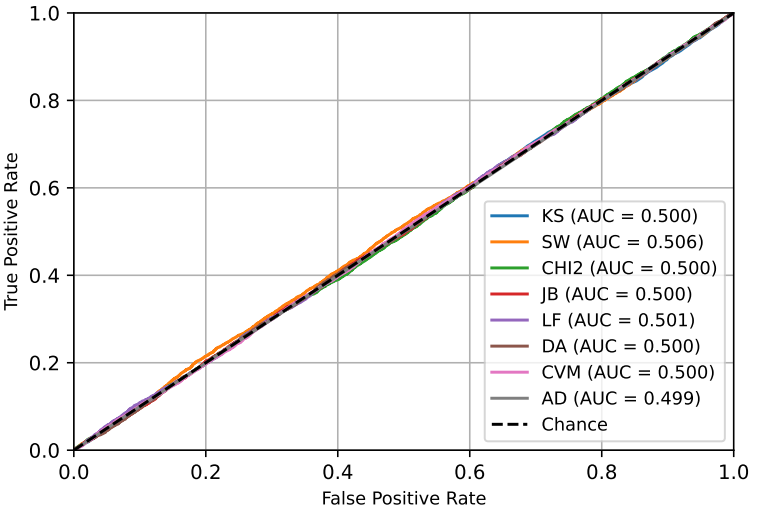


Figure 6: ROC analysis for transformed GNSS latitude data with a large sample size

5. Conclussion

The study emphasizes the importance of assessing and transforming large datasets, such as GNSS measurements, to ensure normality for the validation of parametric statistical tests. The untransformed GNSS latitude data were identified as non-normal using various visual and statistical tests, including histograms, Q-Q plots, skewness, kurtosis, and statistical tests: KS, AD, DA, SW, JB, CVM, Chi2, and LF. Among the transformation techniques, the rank-based Inverse Normal Transformation (INT) demonstrated high effectiveness in enhancing data normality, as validated by various testing methods. The efficiency of statistical tests' was assessed using ROC and AUC analysis, which successfully categorized untransformed data as non-normal and transformed data as normal. These findings underscore the necessity of using tailored transformation methods in large-scale data applications, particularly in geospatial and industrial fields, to enhance the reliability and applicability of parametric statistical methods.

Acknowledgements

This research was funded by the Polish Ministry of Science and Higher Education, grant number 02/050/BKM-24/0042 (AF) and 02/050/BK-24/0032 (JW, AK).

References

Aslam, M., Sherwani, R. A. K. and Saleem, M., (2021). Vague data analysis using neutrosophic Jarque-Bera test. *Plos one*, 16(12), e0260689.

Barba, P., Rosado, B., Ramírez-Zelaya, J. and Berrocoso, M., (2021). Comparative anal-

ysis of statistical and analytical techniques for the study of GNSS geodetic time series. *Engineering Proceedings*, 5(1), p. 21.

Cai, J. and Xu, X., (2024). Bayesian analysis of mixture models with yeo-johnson transformation. *Communications in Statistics-Theory and Methods*, 53(18), pp. 6600–6613.

D'Agostino, R. B., (2017). Tests for the normal distribution. *in* Goodness-of-fit-techniques, Routledge, pp. 367–420.

Demir, S., (2022). Comparison of normality tests in terms of sample sizes under different skewness and kurtosis coefficients. *International Journal of Assessment Tools in Education*, 9(2), pp. 397–409.

Ghasemi, A., and Zahediasl, S., (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, 10, pp. 486–489.

Huang, Z., Zhao, T., Lai, R., Tian, Y. and Yang, F., (2023). A comprehensive implementation of the log, Box-Cox and log-sinh transformations for skewed and censored precipitation data. *Journal of Hydrology*, 620, pp. 129347.

Khatun, N., (2021). Applications of normality test in statistical analysis. *Open Journal of Statistics*, 11(1), pp. 113–122.

Kumbhar, D. D., Kumar, S., Dubey, M., Kumar, A., Dongale, T. D., Pawar, S. D. and Mukherjee, S., (2024). Exploring statistical approaches for accessing the reliability of y2o3-based memristive devices. *Microelectronic Engineering*, 288, p. 112166.

Kwak, S. G. and Park, S. H., (2019). Normality test in clinical research. *Journal of Rheumatic Diseases*, 26(1), pp. 5–11.

Li, D.-C., Wen, I.-H. and Chen, W.-C., (2016). A novel data transformation model for small data-set learning. *International Journal of Production Research*, 54(24), pp. 7453–7463.

McCaw, Z. R., Lane, J. M., Saxena, R., Redline, S. and Lin, X., (2020). Operating characteristics of the rank-based inverse normal transformation for quantitative trait analysis in genome-wide association studies. *Biometrics*, 76(4), pp. 1262–1272.

Nahm, F. S., (2022). Receiver operating characteristic curve: overview and practical use for clinicians. *Korean journal of anesthesiology*, 75(1), pp. 25–36.

Obilor, E. I. and Amadi, E. C., (2018). Test for significance of Pearson's correlation coefficient. *International Journal of Innovative Mathematics, Statistics & Energy Policies*, 6(1), pp. 11–23.

Ogaja, C. A., (2022), GNSS data processing. *in* Introduction to GNSS Geodesy: Foundations of Precise Positioning Using Global Navigation Satellite Systems, Springer, pp. 119–134.

Osborne, J., (2010). Improving your data transformations: Applying the Box-Cox transformation. *Practical Assessment, Research, and Evaluation*, 15(1).

Patrício, M., Ferreira, F., Oliveiros, B. and Caramelo, F., (2017). Comparing the performance of normality tests with roc analysis and confidence intervals. *Communications in Statistics: Simulation and Computation*, 46, pp. 7535–7551.

Peterson, R. A., (2021). Finding optimal normalizing transformations via best normalize. *R Journal*, 13(1).

Pham, L., (2021). Frequency connectedness and cross-quantile dependence between green bond and green equity markets. *Energy Economics*, 98, p. 105257.

Raymaekers, J. and Rousseeuw, P. J., (2024). Transforming variables to central normality. *Machine Learning*, 113(8), pp. 4953–4975.

Razali, N. M. and Wah, Y. B., (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, 2, pp. 13–14.

Rolke, W. and Gongora, C. G., (2021), A chi-square goodness-of-fit test for continuous distributions against a known alternative. *Computational Statistics*, 36(3), pp. 1885–1900.

Stine, R. A., (2017). Explaining normal quantile-quantile plots through animation: the water-filling analogy. *The American Statistician*, 71(2), pp. 145–147.

Sun, J. and Xia, Y. (2024). Pretreating and normalizing metabolomics data for statistical analysis. *Genes & Diseases*, 11(3), p. 100979.

Tabachnick, B. G., Fidell, L. S. and Ullman, J. B., (2019). *Using Multivariate Statistics*, 7th ed., Pearson.

Uyanto, S. S., (2022). An extensive comparisons of 50 univariate goodness-of-fit tests for normality. *Austrian Journal of Statistics*, 51(3), pp. 45–97.

Von Mises, R., (2014). *Mathematical theory of probability and statistics*, Academic press.

Wilcox, R. R., (2010). *Fundamentals of modern statistical methods: Substantially improving power and accuracy*, Vol. 249, Springer.

Yan, P., (2024). Jackknife test for faulty GNSS measurements detection under non-gaussian noises. *in* 'Proceedings of the 37th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2024)', pp. 1619–1641.

Yap, B. W. and Sim, C. H., (2011). Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation*, 81, pp. 2141–2155.

Zygmonta, C. S., (2023). Managing the assumption of normality within the general linear model with small samples: Guidelines for researchers regarding if, when and how. *The Quantitative Methods for Psychology*.

STATISTICS IN TRANSITION new series, September 2025
Vol. 26, No. 3, pp. 209–221, https://doi.org/10.59139/stattrans-2025-035
Received – 10.12.2024; accepted – 24.04.2025

A minimum variance unbiased estimator of finite population variance using auxiliary information

Archana Panigrahi [1], **Priyaranjan Dash**[2], **Gopabandhu Mishra**[3]

Abstract

A class of estimators of finite population variance (S_y^2) using auxiliary information has been proposed under simple random sampling without replacement (SRSWOR) scheme. An attempt has been made to derive the minimum variance unbiased estimator of finite population variance from the proposed class of unbiased estimators. The efficiency of the class of estimators under optimality is compared with the usual unbiased estimator (t_{V0}=s_y^2), ratio type estimator (t_{VR}), product type estimator{\ (t}_{VP}), regression type estimator {(t}_{Vlr}), exponential ratio type estimator{\ (t}_{VER}), exponential product type estimator {\ (t}_{VEP}), and ratio-in-regression estimator (t_s), both theoretically and empirically under general conditions and under bivariate normality. The proposed class of estimator performs better than these estimators under certain realistic conditions. The proposed class of estimators is generalized for the case of multi-auxiliary variables.

Key words: simple random sampling without replacement (SRSWOR), unbiased estimator, auxiliary variable, population variance, efficiency.

1. Introduction

Estimation of finite population variance draws attention of many researchers due to its wide spread applications in various fields like portfolio analysis, asset evaluation, stock market analysis in finance; quality control problems, traffic controls, opinion polls, biostatistics, logistics analysis and many more. So, researchers have been paying specific attention towards estimating the finite population variance $\left(S_y^2\right)$ with a greater precision for many decades. Consider a finite population of N $(< \infty)$ units and Y_i being

[1] Ravenshaw University, Cuttack, Odisha, India. E-mail: archanapanigrahi10@gmail.com. ORCID: https://orcid.org/ 0009-0006-3349-9524.

[2] Utkal University, Vanivihar, Bhubaneswar, Odisha, India. E-mail: prdashjsp@gmail.com. ORCID: https://orcid.org/ 0000-0001-6561-4095.

[3] Utkal University, Vanivihar, Bhubaneswar, Odisha, India. E-mail: gmishrauu@gmail.com. ORCID: https://orcid.org/ 0000-0003-3959-2333.

the value of the study variable y for the i-th unit of the population $(i = 1(1)N)$. We denote

$$S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2 \tag{1}$$

as the finite population variance. In order to estimate S_y^2, we draw a sample of size n from the population of size N using simple random sampling without replacement (SRSWOR) scheme. Suppose y_i is the value of the study variable y for the i-th unit of the sample $(i = 1(1)n)$.

To start with a very simple popular estimator of finite population variance, we consider the sample variance

$$t_{V0} = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 = s_y^2, \tag{2}$$

which is an unbiased estimator of S_y^2 and its variance is given by

$$V(t_{V0}) = \theta S_y^4[\beta_{04} - 1], \tag{3}$$

where $\theta = \frac{1}{n} - \frac{1}{N}$, $\beta_{04} = \frac{\mu_{04}}{\mu_{02}{}^2}$, $\mu_{rs} = \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^r(Y_i - \bar{Y})^s$

and μ_{rs} is called the (r,s)-th order bivariate central moment of (x,y) for the population.

In order to improve the present estimator by exploiting the information on an auxiliary variable x (if available), Isaki (1983) proposed a ratio estimator for finite population variance using advance knowledge of population variance of auxiliary variable S_x^2 as

$$t_{VR} = s_y^2\frac{S_x^2}{s_x^2} \tag{4}$$

The estimator t_{VR} is a biased estimator of S_y^2 and its MSE up to $o(n^{-1})$ is given by

$$MSE(t_{VR}) = \theta S_y^2[\beta_{04} + \beta_{40} - 2\beta_{22}], where\ \beta_{40} = \frac{\mu_{40}}{\mu_{20}{}^2}, \beta_{22} = \frac{\mu_{22}}{\mu_{20}\,\mu_{02}}. \tag{5}$$

It may be seen that t_{VR} is more efficient than t_{V0} if

$$\S > \frac{1}{2}\frac{C.V.(s_x^2)}{C.V.(s_y^2)}, \tag{6}$$

where § denotes the correlation coefficient between s_x^2 and s_y^2. Further, $C.V.(s_x^2)$ and $C.V.(s_y^2)$ denote the coefficient of variation of s_x^2 and s_y^2 respectively. Thus, t_{VR} is

conditionally more efficient than the usual variance estimator t_{V0}, which does not use any auxiliary information.

Following Isaki (1983), a product estimator for S_y^2 is suggested as

$$t_{VP} = s_y^2 \frac{S_x^2}{S_x^2}. \tag{7}$$

It is also a biased estimator of S_y^2 and its MSE up to $o(n^{-1})$ is given by

$$MSE(t_{VP}) = \theta S_y^4[\beta_{04} + \beta_{40} + 2\beta_{22} - 4] \tag{8}$$

Isaki (1983) also suggested a regression type estimator for estimating S_y^2, which is given by

$$t_{VLR} = s_y^2 + \widehat{B}(S_x^2 - s_x^2) \tag{9}$$

where $B = \frac{Cov(s_x^2, s_x^2)}{V(s_x^2)}$ and $\widehat{B}$ is the estimated value of B.

This estimator t_{VLR} is also a biased estimator of S_y^2 and its MSE up to $o(n^{-1})$ is given by

$$MSE(t_{VLR}) = \theta S_y^4[\beta_{04} - 1]\,[1 - \S^2] \tag{10}$$

Singh et al. (2009) suggested an exponential ratio-type and an exponential product-type estimator for S_y^2, which are given by

$$t_{VER} = s_y^2 \exp\left(\frac{S_x^2 - s_x^2}{S_x^2 + s_x^2}\right) \tag{11}$$

and

$$t_{VEP} = s_y^2 \exp\left(\frac{S_x^2 + s_x^2}{S_x^2 - s_x^2}\right). \tag{12}$$

respectively. Both these estimators are biased estimators of S_y^2 and their MSE up to $o(n^{-1})$ are given by

$$MSE(t_{VER}) = \theta S_y^4\left[\beta_{04} + \frac{1}{4}\beta_{40} - \beta_{22} - \frac{1}{4}\right] \tag{13}$$

and

$$MSE(t_{VEP}) = \theta S_y^4\left[\beta_{04} + \frac{1}{4}\beta_{40} + \beta_{22} - \frac{9}{4}\right] \tag{14}$$

respectively.

Swain (2015) suggested a class of ratio-in-regression estimators for estimating the finite population variance $\left(S_y^2\right)$ as

$$t_s = ws_y^2 + (1-w)s_y^2\left(\frac{S_x^2}{s_x^2}\right) \tag{15}$$

The value of w for which the above class of estimators attains a minimum variance is given by

$$w_{opt} = \beta_{40} - \beta_{22}.$$

Using this value of w in t_s, we get the optimum estimator (estimator attaining a minimum variance) for this class as

$$t_{s(opt)} = (\beta_{40} - \beta_{22})s_y^2 + (1 - \beta_{40} + \beta_{22})s_y^2\left(\frac{S_x^2}{s_x^2}\right). \tag{16}$$

The mean square error of this estimator is given by

$$MSE\left(t_{s(opt)}\right) = \theta S_y^4\left[(\beta_{04}-1) - \frac{(\beta_{22}-1)^2}{\beta_{40}-1}\right]. \tag{17}$$

For a brief review of literature about the variance estimation, one can go through Evans (1951), Wakimoto (1971), Liu (1974), Das and Tripathi (1978), Srivastava and Jhajj (1980), Wu (1982), Upadhyaya and Singh (1983), Singh (1986), Singh et al. (1988), Prasad and Singh (1990,1992), Swain and Mishra (1992, 1994a,b), Singh and Biradar (1994), Cebrian and Garcia (1997), Kadilar and Cingi (2007), Shabbir and Gupta (2007), Dubey and Sharma (2008), Yadav et al. (2013), Singh and Solanki (2013a,b), Singh et al. (2014), Dash and Sunani (2019) and many more.

2. Proposed class of unbiased estimators

A class of estimators for estimating finite population variance (S_y^2) in the presence of auxiliary information is proposed by taking the linear combination of the sample variance of study variable (s_y^2) and the sample variance of auxiliary variable (s_x^2). The proposed class of estimators of S_y^2 is given by

$$t_{VM} = \lambda_0 s_y^2 + \lambda_1 s_x^2, \tag{18}$$

where λ_0 and λ_1 are two suitable constants or statistics. For different values of λ_0 and $\lambda_1$one can get different estimators. The following table presents some particular estimators deduced from the proposed class of estimator for different suitably chosen values of λ_0 and λ_1 .

Table 1: Some Deduced Estimators

λ_0	λ_1	**Deduced estimators**
1	0	$t_{V0} = s_y^2$ [Usual unbiased estimator]
$\frac{S_x^2}{s_x^2}$	0	$t_{VR} = s_y^2 \frac{S_x^2}{s_x^2}$ [Isaki (1983)]
$\frac{s_x^2}{S_x^2}$	0	$t_{VP} = s_y^2 \frac{s_x^2}{S_x^2}$
1	$\frac{(S_x^2 - s_x^2)}{s_x^2}$	$t_{Vlr} = s_y^2 + \hat{B}(S_x^2 - s_x^2)$ [Isaki (1983)]
$\exp\left[\frac{S_x^2 - s_x^2}{S_x^2 + s_x^2}\right]$	0	$t_{VER} = s_y^2 \exp\left[\frac{S_x^2 - s_x^2}{S_x^2 + s_x^2}\right]$, [Singh et al. (2009)]
$\exp\left[\frac{s_x^2 + S_x^2}{s_x^2 - S_x^2}\right]$	0	$t_{VEP} = s_y^2 \exp\left[\frac{S_x^2 + s_x^2}{S_x^2 - s_x^2}\right]$ [Singh et al. (2009)]
$\beta_{40} - \beta_{22}$	$(1-\lambda_0)\frac{s_y^2}{s_x^2}\frac{S_x^2}{s_x^2}$	$t_{s(opt)} = \lambda_0 s_y^2 + (1-\lambda_0)\, t_{VR}$, [Swain, 2015]

3. Variance of the proposed class of estimators

To construct a class of unbiased estimators, we have to choose the constants λ_0 and λ_1 such that the class of estimators in (18) is unbiased for S_y^2 and we shall attempt to identify minimum variance unbiased estimator in this class. The condition for t_{VM} to be unbiased for S_y^2 is given by

$$(\lambda_0 - 1)S_y^2 + \lambda_1 S_x^2 = 0 \tag{19}$$

The variance of the estimator t_{VM} is

$$V(t_{VM}) = \lambda' S_2 \lambda, \tag{20}$$

where

$$\lambda = \begin{bmatrix}\lambda_0 \\ \lambda_1\end{bmatrix} \quad \text{and} \quad S_2 = \begin{bmatrix}S_{00} & S_{01} \\ S_{10} & S_{11}\end{bmatrix},$$

$$S_{00} = V(s_y^2) = \theta S_y^4[\beta_2(y) - 1], \quad S_{11} = V(s_x^2) = \theta S_x^4[\beta_2(x) - 1]$$

$$S_{10} = S_{01} = \mathrm{Cov}(s_x^2, s_y^2) = \theta S_x^2 S_y^2[\beta_{22}(x,y) - 1].$$

The optimum value of λ satisfying the unbiasedness condition (19) and leading to a minimum variance of t_{VM} is

$$\lambda_{opt} = S_y^2 \frac{S_2^{-1} Q_2}{Q_2' S_2^{-1} Q_2} = \lambda^*, \tag{21}$$

where S_2^{-1} is inverse of matrix S_2, $Q_2 = \begin{bmatrix}S_y^2 \\ S_x^2\end{bmatrix}$, Q_2' is the transpose of Q_2.

Hence, after expansion, we get

$$\lambda^* = \frac{S_y^2\begin{bmatrix}S_y^2 S_{11} - S_x^2 S_{10} \\ -S_y^2 S_{10} + S_x^2 S_{00}\end{bmatrix}}{S_y^4 S_{11} - 2S_y^2 S_x^2 S_{10} + S_x^4 S_{00}} \tag{22}$$

From (22) we get optimum values of λ_0 and λ_1, which are denoted by λ_0^* and λ_1^* where

$$\lambda_0^* = S_y^2 \frac{(S_y^2 S_{11} - S_x^2 S_{10})}{S_y^4 S_{11} - 2S_y^2 S_x^2 S_{10} + S_x^4 S_{00}} = \frac{\beta_{40} - \beta_{22}}{\beta_{40} - 2\beta_{22} + \beta_{04}}, \tag{23}$$

$$\lambda_1^* = S_y^2 \frac{S_x^2 S_{00} - S_y^2 S_{10}}{S_y^4 S_{11} - 2S_y^2 S_x^2 S_{10} + S_x^4 S_{00}} = \frac{S_y^2}{S_x^2} \frac{\beta_{04} - \beta_{22}}{\beta_{40} - 2\beta_{22} + \beta_{04}} \tag{24}$$

are called optimality conditions.

The minimum variance unbiased estimator of population variance is given by

$$t_{VM(o)} = \lambda_0^* s_y^2 + \lambda_1^* s_x^2, \tag{25}$$

Using the values of λ_0^* and λ_1^* in equation (20), the minimum(optimum) variance of t_{VM} is given by

$$V\left(t_{VM(o)}\right)_{opt} = \frac{S_y^4}{Q_2' S_2^{-1} Q_2} = S_y^4 \frac{S_{00} S_{11} - S_{10}^2}{S_y^4 S_{11} - 2S_y^2 S_x^2 S_{10} + S_x^4 S_{00}} \tag{26}$$

$$= \theta S_y^4 \frac{[\beta_{40} - 1][\beta_{04} - 1]\,[1 - \S^2]}{\beta_{40} - 2\beta_{22} + \beta_{04}}. \tag{27}$$

4. Comparison of efficiency

The members of the suggested class of estimators t_{VM} are unbiased estimators of the finite population variance S_y^2. Hence, this class of estimators t_{VM} in its optimal case is compared with several other estimators in terms of efficiency only. The comparison of the efficiency of suggested class of estimators t_{VM} under optimality with other estimators are made under the following two conditions:

(a) under general condition and

(b) under bivariate normality condition.

I. The suggested class of estimators t_{VM} under optimality $\left(t_{VM(o)}\right)$ is always more efficient than the usual unbiased estimator t_{V0} .

II. The class of estimators t_{VM} under optimality $\left(t_{VM(o)}\right)$ is more efficient than t_{VR} if

Case (a): $[\beta_{40} - 2\beta_{22} + \beta_{04}]^2 > [\beta_{40} - 1][\beta_{04} - 1]\left[1 - \S^2\right]$ (28)

Case (b): $4(1 - \rho^2)^2 > [1 - \S^2]$, (29)

where, ρ is the correlation coefficient between the study variable y and the auxiliary variable x.

III. The class of estimators t_{VM} under optimality $\left(t_{VM(o)}\right)$ is more efficient than t_{VP} if

Case(a): $[\beta_{40} - 2\beta_{22} + \beta_{04}]^2 - 4[\beta_{22} - 1]^2 > [\beta_{40} - 1][\beta_{04} - 1][1 - \S^2]$ (30)

Case(b): $4(1 - 2\rho^2) > [1 - \S^2]$ (31)

IV. The class of estimators t_{VM} under optimality $\left(t_{VM(o)}\right)$ is more efficient than t_{Vlr} if

Case(a): $$\left[\beta_{04}+1\right] > 2\beta_{22} \tag{32}$$

Case(b): $$\rho^2 < \frac{1}{2} \tag{33}$$

V. The class of estimators t_{VM} under optimality $\left(t_{VM(o)}\right)$ is more efficient than t_{VER} if

Case(a): $$[\beta_{40}-2\beta_{22}+\beta_{04}]\left[\beta_{40}-2\beta_{22}+\beta_{04}-\frac{1}{4}\right] > [\beta_{40}-1][\beta_{04}-1][1-\S^2] \tag{34}$$

Case(b): $$(1-\rho^2)(15-16\rho^2) > 4[1-\S^2] \tag{35}$$

VI. The class of estimator t_{VM} under optimality $\left(t_{VM(o)}\right)$ is more efficient than t_{VEP} if

Case(a): $$[\beta_{40}-2\beta_{22}+\beta_{04}]\left[\frac{1}{4}\beta_{40}-2\beta_{22}+\beta_{04}-\frac{9}{4}\right] > [\beta_{40}-1][\beta_{04}-1][1-\S^2] \tag{36}$$

Case(b): $$(1-\rho^2)(15-16\rho^2) > 4[1-\S^2] \tag{37}$$

VII. The class of estimator t_{VM} under optimality $\left(t_{VM(o)}\right)$ is more efficient than $t_{s(opt)}$ if

Case(a): $$(\beta_{04}-1)(\beta_{40}-1)-(\beta_{22}-1)^2[\beta_{40}-2\beta_{22}+\beta_{04}] > [\beta_{40}-1]^2[\beta_{04}-1][1-\S^2] \tag{38}$$

Case(b): $$2(1-\rho^4)(1-\rho^2) > [1-\S^2] \tag{39}$$

5. Empirical Study

To study the performance of the proposed class of estimators numerically, we consider 16 populations collected from different literature as described in Table 2 (8 populations having positive correlation between x and y) and in Table 3 (8 populations having negative correlation between x and y). The source of populations, description of the nature of x and y, N, n, ρ, β_{40} and β_{04} for these populations are presented in these tables. We have considered the usual unbiased estimator ($t_{V0}=s_y^2$), ratio type estimator (t_{VR}) due to Isaki (1983), product type estimator (t_{VP}), regression type estimator (t_{Vlr}) due to Isaki (1983), exponential ratio type estimator (t_{VER}) due to Singh et al. (2009), exponential product type estimator (t_{VEP}) due to Singh et al. (2009), Swain (2015) ratio-in-regression estimator ($t_{s(opt)}$) in order to compare with the proposed class of estimators under optimality, i.e. the minimum variance unbiased estimator $\left(t_{VM(o)}\right)$ of the proposed class. Assuming simple random sampling without replacement, the relative efficiency of different estimators with respect to t_{V0} is compiled in Table 3 (for the cases in which the correlation coefficient is positive) and Table 4 (for the cases in which the correlation coefficient is negative).

Table 2: Description of the Populations ($\rho > 0$)

Pop^n No.	Source	x	y	N	ρ	β_{40}	β_{04}
1	Jobson (1992), p .674	Percentage of White People in Population	Total Mortality Rate	80	0.18	1.93	3.261
2	Draper & Smith (1966), p. 366	Amount of Tricalcium Silicate	Amount of Tricalcium Aluminate	13	0.23	1.677	3.075
3	Black (2009), p.517	Job Satisfaction	Advancement Opportunities	19	0.26	2.737	1.889
4	Stevens (2009), p. 81	Knowledge	Instructor Evaluation	32	0.28	3.099	2.511
5	Draper & Smith (1966), p.352	Pounds of Crude Glycerin Made	Pounds of Steam used monthly	25	0.31	3.053	2.278
6	Murthy (1967), p. 91	Holding Size (in acres)	Cultivated Area (in acres)	36	0.37	17.47	3.183
7	Cochran (1953), p. 113	Size of large US Cities in 1920	Size of large US Cities in 1930	49	0.4	7.504	38.91
8	Black (2009), p. 517	Relationship with Supervisor	Advancement Opportunities	19	0.4	2.654	1.889

Table 3: Description of the Populations ($\rho < 0$)

Pop^n No.	Source	x	y	N	ρ	β_{40}	β_{04}
1	Black (2009), p. 511	Time Period	Total No. of Savings Organization	13	-0.17	1.786	1.602
2	Maddala (1988), p. 197	Year	Prices Received by Farmer for Food	20	-0.21	1.794	2.81
3	Draper & Smith (1966), p. 352	Number of Start-ups	Average Atmospheric Temperature	25	-0.24	4.075	1.532
4	Härdle & Hlávka (2007), p.335	Marks for Car Safety	Marks for Car Economy	24	-0.27	3.346	2.684
5	Maddala (1988), p. 197	Year	Food Prod. per Capita	20	-0.34	1.794	2.453
6	Black (2009), p. 567	Nuclear Electricity Production (in billion kwh)	Dry Gas Prod. (in trillions of cubic feet)	26	-0.4	1.579	3.012
7	Black (2009), p.567	Dry Gas Prod. (in trillions of cubic feet)	Fuel Rate for Automobile (in miles/ gallon)	26	-0.42	1.599	3.012
8	Härdle & Hlávka (2007), p. 339	Expenditure on Meat	Expenditure on Wine	12	-0.44	2.056	1.837

Table 4: PRE of Estimators with respect **to t_{V0}** (when **ρ** > 0)

Pop^n No.	N	n	Estimators					
			t_{V0}	t_{VR}	t_{VER}	t_{Vlr}	$t_{s(opt)}$	$t_{VM(o)}$
1	80	32	100	65.403	86.088	100.84	124.56	374.65
2	13	6	100	57.714	77.839	114.46	251.23	607.36
3	19	8	100	28.694	57.014	103.74	114.56	184.99
4	32	13	100	37.636	67.504	101.32	126.19	193.88
5	25	10	100	53.811	97.319	109.54	110.44	126.71
6	36	15	100	11.286	32.855	100.33	105.30	117.8
7	49	20	100	85.608	96.046	100	264.98	680.84
8	19	8	100	28.613	56.103	105.73	137.15	198.64

Table 5: PRE of Estimators with respect **to t_{V0}** (when **ρ** < 0)

Pop^n No.	N	n	Estimators					
			t_{V0}	t_{VR}	t_{VER}	t_{Vlr}	$t_{s(opt)}$	$t_{VM(o)}$
1	13	6	100	33.164	59.471	110.71	122.31	135.27
2	20	8	100	131.08	129.57	135.18	269.34	651.64
3	25	10	100	14.945	41.679	100.04	102.23	118.96
4	24	10	100	47.729	83.372	101.61	149.31	196.28
5	20	8	100	56.408	80.016	102.4	165.42	247.35
6	26	11	100	63.243	82.057	108.09	197.85	373.49
7	26	11	100	72.962	90.02	100.45	254.66	413.27
8	12	5	100	42.041	72.777	100.28	141.64	170.44

6. Extension to the case of multi-auxiliary variables

The proposed class of unbiased estimators can be extended to the case of multi-auxiliary variables when the information on population variances $s_{x_1}^2, s_{x_2}^2, \ldots, s_{x_k}^2$ for $k-$ auxiliary variables is available. The class of unbiased-estimators of S_y^2 is

$$\mathrm{T} = \lambda_0 \mathrm{s}_\mathrm{y}^2 + \lambda_1 \mathrm{s}_{\mathrm{x}_1}^2 + \cdots + \lambda_\mathrm{k} \mathrm{s}_{\mathrm{x}_\mathrm{k}}^2, \tag{40}$$

where the $(k+1)$ constants λ_i $(i = 0,1,2,\cdots,k)$ are chosen so that T will be unbiased for S_y^2. So, we have

$$\mathrm{E}(\mathrm{T}) = \mathrm{S}_\mathrm{y}^2 \quad \text{i.e. } \lambda' \, \mathrm{Q}_{k+1} = \mathrm{S}_\mathrm{y}^2 \tag{41}$$

and the variance of the class of estimators T is

$$\mathrm{V}(\mathrm{T}) = \lambda' \mathrm{S}_{k+1} \lambda, \tag{42}$$

where $\lambda' = [\lambda_0, \lambda_1, \cdots, \lambda_\mathrm{k}]$ is a matrix of order $1 \times (k+1)$ and $\mathrm{Q}_{k+1}{}' = \left[S_y^2, S_{x_1}^2, \cdots, S_{x_k}^2\right]$ is a matrix of order $1 \times (k+1)$; $\mathrm{S}_{k+1} = [S_{ij}]$ is the dispersion

matrix of order $(k+1)$, such that $S_{00} = Var(s_y^2)$, $S_{0j} = Cov\left(s_y^2, s_{x_j}^2\right) = S_{j0}$, $S_{ij} = Cov\left(s_{x_i}^2, s_{x_j}^2\right) = S_{ji}$ $j = 1, \cdots, k, i = 1,2, \cdots, k$.

The value of λ which minimises (41) subject to the unbiasedness condition (40) is found to be

$$\lambda_{opt} = \mathrm{S_y^2} \frac{S_{k+1}^{-1} \mathrm{Q}_{k+1}}{\mathrm{Q}_{k+1}{}' S_{k+1}^{-1} \mathrm{Q}_{k+1}} = \lambda^*(say), \tag{43}$$

where S_{k+1}^{-1} is inverse of the matrix S_{k+1} and $\lambda^{*\prime} = [\lambda_0^*, \lambda_1^*, \cdots, \lambda_k^*]$ are the optimum values of λ'. From the above conditions, we get the minimum variance unbiased estimator of this class as

$$\mathrm{T}_{opt} = \lambda_0^* \mathrm{s_y^2} + \lambda_2^* \mathrm{s_{x_1}^2} + \cdots + \lambda_k^* \mathrm{s_{x_k}^2}. \tag{44}$$

The optimum variance of T_{opt} is given by

$$\mathrm{V}(\mathrm{T}_{opt}) = \frac{\mathrm{S_y^4}}{\mathrm{Q}_{k+1}{}' S_{k+1}^{-1} \mathrm{Q}_{k+1}} \tag{45}$$

7. Conclusion

1) In Table 4, for populations 1 to 8 the percent relative efficiency of the minimum variance unbiased estimator $(\mathrm{t_{VM(o)}})$ of the proposed class is maximal.
2) In Table 5, for populations 1 to 8 the percent relative efficiency of the minimum variance unbiased estimator $(\mathrm{t_{VM(o)}})$ of the proposed class is maximal.

When the study variable and auxiliary variable are positively correlated, the proposed estimator is more efficient than the ratio type estimator $(\mathrm{t_{VR}})$, exponential ratio type estimator $(\mathrm{t_{VER}})$, regression type estimator $(\mathrm{t_{Vlr}})$ and ratio-in-regression estimator $(\mathrm{t_{s(opt)}})$. We may write the efficiency (E) in descending order

$$\mathrm{E}(\mathrm{t_{VM(o)}}) \geq \mathrm{E}(\mathrm{t_{s(opt)}}) \geq \mathrm{E}(\mathrm{t_{Vlr}}) \geq \mathrm{E}(\mathrm{t_{VER}}) \geq \mathrm{E}(\mathrm{t_{VR}}).$$

When the study variable and auxiliary variable are negatively correlated, the proposed estimator is more efficient than the product type estimator $(\mathrm{t_{VP}})$, exponential product type estimator $(\mathrm{t_{VEP}})$, regression type estimator $(\mathrm{t_{Vlr}})$ and ratio-in-regression estimator $(\mathrm{t_{s(opt)}})$. We may write the efficiency (E) in descending order

$$\mathrm{E}(\mathrm{t_{VM(o)}}) \geq \mathrm{E}(\mathrm{t_{s(opt)}}) \geq \mathrm{E}(\mathrm{t_{Vlr}}) \geq \mathrm{E}(\mathrm{t_{VEP}}) \geq \mathrm{E}(\mathrm{t_{VP}}).$$

References

Black, K., (2009). Business Statistics for Contemporary Decision Making, 6th Edition. *John Wiley and Sons*, Inc., New York, USA.

Cebrian, A. A., Garcia, M. R., (1997). Variance estimation using auxiliary information: An almost unbiased multivariate ratio estimator. *Metrika*, Vol. 45, pp. 171–178.

Cochran, W. G., (1977). Contribution to Theory of Sampling, 3rd Edition. *Wiley Eastern Limited*, India.

Das, A. K., Tripathi, T. P., (1978). Use of auxiliary information in estimating the finite population variance. *Sankhya C*, Vol. 40, pp. 139–148.

Dash, P. R., Sunani, K., (2019). Estimation of Finite Population Variance Using A Family of Exponential Estimators. *International Journal of Research and Analytical Reviews*, Vol. 6(6), pp. 1200–1206.

Draper, N. R., Smith, H., (1966). Applied Regression Analysis, 1st Edition, John Wiley and Sons Inc., New York, USA.

Dubey, V., Sharma, H. K., (2008). On estimating population variance using auxiliary information. *Statist. Trans. New Ser*., 9(1), pp. 7–18.

Evans, D., (1951). On the variance estimators of SD and Variance. *Jour. Amer. Stat. Assoc.,* 46, pp. 220–224.

Hardle, W., Hlavka, Z., (2007). Multivariate Statistics: Exercise and Solutions. *Springer Science+Business Media*, LLC, New York, USA.

Isaki, C. T., (1983). Variance estimation using auxiliary information. *Jour. Amer. Stat. Assoc.*, Vol. 78(381), pp. 117–123.

Jobson, J. D., (1992). Applied Multivariate Data Analysis Volume II: Categorical and Multivariate Methods. *Springer Science+Business Media*, LLC, New York, USA.

Kadilar, C., Cingi, H., (2007). Improvement in Variance Estimation in Simple Random Sampling. *Communications in Statistics-Theory & Methods*, Vol. 36(11), pp. 2075–2081.

Liu, T. P., (1974). A generalized unbiased estimator for the variance of a finite population. *Sankhya C*, Vol. 36, pp. 23–32.

Maddala, G. S., (1992). Introduction to Econometrics, 2nd Edition. *Macmillan Publishing Company*, New York, USA.

Murthy, M. N., (1967). Sampling Theory and Methods. *Statistical Publishing Society*, Calcutta, India.

Prasad, B., Singh, H. P., (1990). Some improved ratio-type estimators of finite population variance in sample surveys. *Communications in Statistics-Theory & Methods*, Vol. 19(3), pp. 1127–1139.

Prasad, B., Singh, H. P., (1992). Unbiased estimators of finite population variance using auxiliary information in sample surveys. *Communications in Statistics-Theory & Methods*, Vol. 21(5), pp. 1367–1376.

Shabbir, J., Gupta, S., (2007). On improvement in variance estimation using auxiliary information. *Communications in Statistics-Theory & Methods*, Vol. 36(12), pp. 2177–2185.

Singh, H. P., Biradar, R. S., (1994). Estimation of finite population variance using auxiliary information. *Jour. Ind. Soc. Statist. Operas. Res.*, Vol. 15(1–4), pp. 47– 63.

Singh, H. P., Upadhyaya, L. N. and Namjoshi, U. D., (1988). Estimation of finite population variance. *Current Science*, Vol. 57, pp. 1331–1334.

Singh, H. P., (1986). A note on the estimation of variance of sample mean using the knowledge of coefficient of variation in normal population. *Communications in Statistics-Theory & Methods*, Vol. 15(12), pp. 3737–3746.

Singh, H., Singh, A. and Solanki, R., (2014). Estimation of finite population variance using auxiliary information in sample surveys. *Statistica*, Vol. 74(1), pp. 99-116.

Singh, H. P. Solanki, R. S. (2013a). A new procedure for variance estimation in simple random sampling using auxiliary information. *Statistical Papers*, Vol. 54(2), pp. 479–497.

Singh, H. P. Solanki, R. S., (2013b). Improved estimation of finite population variance using auxiliary information. *Communications in Statistics-Theory & Methods*, Vol. 42(15), pp. 2718–2730.

Singh, R., Chauhan, P., Sawan, N. and Smarandache, F., (2009). Improved exponential estimator for population variance using two auxiliary variables. *Octogon Mathematical Magazine*, Vol. 17(2), pp. 667–674.

Srivastava, S. K., Jhajj, H. S., (1980). A class of estimators using auxiliary information for estimating finite population variance. *Sankhya C*, Vol. 42, pp. 87–96.

Stevens, J. P., (2009). Applied Multivariate Statistics for the Social Sciences, 5th Edition. *Routledge*, New York, USA.

Swain, A. K. P. C., Mishra, G., (1994a). Estimation of population variance under unequal probability sampling. *Sankhya B*, Vol. 56, pp. 374–384.

Swain, A. K. P. C., Mishra, G., (1992). Unbiased estimators of finite population variance using auxiliary information. *Metron*, L-n, 3–4, pp. 201–216.

Swain, A. K. P. C., Mishra, G., (1994b). Limiting distribution of the ratio estimator of finite population variance. *Sankhya B*, Vol. 56, pp. 11–17.

Swain, A. K. P. C., (2015). Generalized Estimator of Finite population Variance. *Journal of Statistical Theory and Applications*, Vol. 14(1), pp. 45–51.

Upadhyaya, L. N., Singh H. P., (1983). Use of auxiliary information in the estimation of population variance. *Mathematical Forum*, Vol. VI(2), pp. 33–36.

Wakimoto, K., (1971). Stratified random sampling (I): Estimation of population variance. *Ann. Inst. Statist. Math.*, Vol. 23, pp. 233–252 .

Wu, C. F., (1982). Estimation of variance of the ratio estimator. *Biometrika*, Vol. 69(1), pp. 183–189.

Yadav, R., Upadhyaya, L. N., Singh, H. P., Chatterjee, S., (2013). A generalized family of transformed ratio-product estimators for variance in sample surveys. *Communications in Statistics-Theory & Methods*, Vol. 42(10), pp. 1839–1850.

STATISTICS IN TRANSITION new series, September 2025
Vol. 26, No. 3, pp. 223–227

About the Authors

Adeboye Nureni is a statistician and data analyst. A graduate of Statistics with certificates in National Diplomas, B.S., M.S. and Ph.D. degrees in Statistics as well as PGD in Planning, Research and Statistics. He is also a certified data scientist with a Coursera certificate from Johns Hopkins University, Baltimore, United States. He has many articles in reputable journals and co-authored 6 published books. He is currently a lecturer in the Statistics Department, Osun State University, Osogbo, Nigeria and his research activities are mainly in time-series econometrics, official statistics, biostatistics and machine learning. He is a member of both international and Nigerian statistical bodies, a member of Data Science Nigeria, a Research Fellow of the SINRHD, and an executive member of the International Association of Official Statistics.

Adesina Olumide is a Senior Lecturer in Statistics and Data Science. His research interests are mathematical statistics, computational statistics, and data science (machine learning and deep learning). He has published more than 60 research papers in international/national journals and conferences. Dr Olumide Adesina is a fellow of the Royal Statistical Society. He is also an active member of other scientific professional bodies.

Alzoubi Loai is a Full Professor of Statistics at Al al-Bayt University, Jordan. His research interests are statistical inference, sampling and distribution theory in particular. Professor Alzoubi has published more than 40 research papers in international/national journals.

Bantu Abu received a B.S. degree in Electrical and Computer Engineering and an M.S. degree in Control and Instrumentation Engineering in 2014 and 2018 from Jimma University, Ethiopia. From 2014 to 2023, he served as a lecturer in the Department of Electrical and Computer Engineering at Jimma University. In parallel, he held engineering roles at the Ethiopian Electric Utility, where he gained practical experience in power distribution systems and instrumentation, contributing to national infrastructure projects for over two years. Currently, he is pursuing a Ph.D. in Automatic Control, Electronics, and Computer Science at the Silesian University of Technology in Gliwice, Poland. His doctoral research focuses on normality tests, statistical modelling of non-Gaussian noise, and the application of advanced statistical methods in engineering diagnostics and uncertainty evaluation.

Bonnini Stefano is an Associate Professor of Statistics at the University of Ferrara, Italy. His main research interests are multivariate statistics, permutation tests, and composite indicators. He has published more than 100 research papers in international journals and conferences, and ten books. He is a member of the Italian Statistical Society.

Borghesi Michela is a Research Fellow in Statistics at the University of Ferrara, Italy. She received her Ph.D. in "Economics and Management of Innovation and Sustainability". Her main research interests concern: nonparametric statistics, multivariate analysis, and complex test of hypotheses. She is a member of the Italian Statistical Society. She is the author of some papers published in international scientific journals and has taken part as a speaker at some international scientific conferences.

Brania Krzysztof holds two Master's degrees: Financial Mathematics (Jagiellonian University) and Management (AGH University, Kraków), a Bachelor's in Computer Technology and Econometrics (AGH University, Kraków), and a Postgraduate Certificate in Valuation & Risk (Middlesex University, London). He has been a Research and Teaching Assistant at AGH University in Kraków since 2020, lecturing in applied mathematics, statistics, probability, financial mathematics, and risk management. He brings over 12 years of quantitative finance experience, developing risk models, high-frequency trading tools, and data-driven solutions at leading banks (HSBC, UBS) and fintech/trading firms (OANDA, OSTC, Star Beta). His research focuses on time series analysis, stochastic processes, financial econometrics, and machine learning applications in quantitative finance.

Dash Priyaranjan is a Professor and Head of the Department at Post Graduate Department of Statistics, Utkal University, Odisha, India. He has more than 25 years of teaching experience at UG and PG levels. His area of interest is survey sampling and stochastic processes. He has published more than 50 research papers in international/national journals and conferences. He has also published five books and is a life member of many scientific professional bodies.

Djuraidah Anik is a Professor of Statistics at the School of Data Science, Mathematics, and Statistics, IPB University (Bogor Agricultural University), Indonesia. She holds a Ph.D. in Statistics from the Postgraduate School of IPB University. Her research interests encompass robust frontier models, spatial and spatio-temporal statistics, statistical downscaling, and predictive modeling. In addition to her academic work, she serves as a reviewer for several leading international journals in statistics.

Dutta-Powell Ravi is an independent researcher from Australia. He has an interest in data analytics, statistics, and behavior science. He has conducted field and online experiments on consumer behavior, and his current research focus is on understanding

the strengths and limitations of practical applications of Benford's Law. He has a number of publications in a range of fields, including consumer policy, financial behavior, and environmental regulation.

Gurgul Henryk is a Full Professor at AGH University of Krakow, Department of Applications of Mathematics in Economics. His research interests include econometrics, financial econometrics, economic growth, globalization, energy economics, international financial markets, input-output models. Author of more than 250 publications. Visiting Professor at universities in Graz and Klagenfurt (Austria), Erlangen-Nuernberg, Greifswald, Ilmenau, Saarbruecken (Germany), Trieste and Siena (Italy), Valencia (Spain), Koper (Slovenia), Joensuu (Finland). Honors: awarded the most prestigious prize for economics in Poland "Bank Handlowy w Warszawie S.A. Award" in 2007, the Medal of Honour, University of Graz, Austria, in 2015. In 2025, he received the First Degree Prize in the Professor Władysław Takliński Prize Competition for teaching achievements. ISI Elected Member.

Hardan Abdullah S. is an Associate Professor of accounting. His research interests are financial accounting, managerial accounting and financial analysis. Dr. Hardan has published more than fifteen peer-reviewed research papers in international/national journals. He is also a peer reviewer for numerous scholarly articles. Dr. Hardan has supervised over eighteen dissertations and theses.

Irfan Mohd is a Senior Researcher in the Department of Mathematics National Institute of Technology Raipur, Chhattisgarh, India. His research interests are mathematical and computational statistics, Bayesian analysis, censoring schemes, reliability theory, survival analysis, distribution theory, competing risk models and data science. Mr. Irfan has published more than 7 research papers in international/national journals and conferences.

Jędrzejczak Alina is a Full Professor of Social Sciences in the discipline of economics and finance. She is currently running the Department of Statistical Methods at the Faculty of Economics and Sociology, University of Lodz. Simultaneously she holds the position of an expert in the Centre of Mathematical Statistics at the regional Statistical Office in Lodz. Her main areas of interest include: statistical analysis of income distribution, methods of income inequality measurement, small area estimation. She is an active member of the Polish Statistical Association and the Italian Statistical Society (Società Italiana di Statistica).

Mishra Gopabandhu is a Former Professor at the Department at Post Graduate Department of Statistics, Utkal University, Odisha, India. He has more than 35 years of

teaching experience at PG levels. His area of interest is survey sampling. He has published more than 75 research papers in international/national journals and conferences. He is the life member of many scientific professional bodies.

Misztal Małgorzata (PhD in Economics) is working in the Department of Statistical Methods at the University of Łódź. Her research interests focus on multivariate statistics and data mining, methods of incomplete data analysis and the application of statistical methods in medical diagnostics, natural sciences and social sciences. She is the author of numerous research papers in national and international journals and conference publications.

Morawski Leszek is an Associate Professor of economics and finance. His research interests include the analysis of income inequality, family economics, and the use of tax and benefit microsimulation models to policy evaluation. He has participated in the development of such models for Poland. Professor Morawski has published over 50 scientific articles in international and national journals and at conferences.

Panigrahi Archana is an Assistant Professor in Statistics at Ravenshaw University, Cuttack. Her research interest is sample survey methods. She has published 6 research papers in international and national journals. She has presented 11 research papers in international and national conferences. She has been honored with "Ramanujan Puzzle Awards", 2012, by Odisha Mathematical Society and "Pranakrushna Parija Popular Science Book Award", 2013, by Odisha Bigyan Academy, Government of Odisha, for her book "Mana Chhuan Ganita" in Odia. She has published 38 popular articles in "Dream 2047", published by the Department of Science and Technology, Government of India, and also written popular articles relating to science and mathematics, published by state of Odisha and by Government of India.

Pekasiewicz Dorota is an Associate Professor at the University of Lodz (Department of Statistical Methods). Her research interests concentrate on non-classical methods of estimation and verification of statistical hypotheses. Her works focus primarily on order statistics and its application in statistical inference, sequential tests and the methods of analyzing household income (including inequality, poverty, wealth). She is the author or co-author of more than 70 publications, including four monographs and two textbooks.

Pranata Ismail is an employee at Statistics Indonesia, specializing in data management. He has recently earned his master's degree in Statistics and Data Science from IPB University. His research interests include economic statistics, statistical inference, and data analysis. He has published a scientific article in Barekeng: Journal of Mathematics and Its Application, titled "Robust Stochastic Production Frontier to Estimate Technical Efficiency of Rice Farming in Sulawesi Selatan."

Sharma Anup Kumar is an Assistant Professor in the Department of Mathematics National Institute of Technology Raipur, Chhattisgarh, India. His research interests are computational statistics, sampling theory, Bayesian analysis and data science. Professor Anup has published 30 research papers in international/national journals and conferences. Professor Anup is an active member of many scientific professional bodies.

Suproń Błażej is an Assistant Professor at the Faculty of Economics at the West Pomeranian University of Technology in Szczecin, Poland. His research interests lie in data science and econometrics, with a particular focus on advanced Bayesian modelling, autoregressive models, machine learning methods, and the connection between ecological and transport economics. He primarily uses quantitative research methods in his studies.

Wiora Józef received the M.S. degree in Electronics in 2000, the Ph.D. degree in Automatic Control and Robotics in 2006, and the D.S. degree in Automation, Electronics, and Electrical Engineering in 2019, all from the Silesian University of Technology (SUT), Gliwice, Poland. He is currently an Associate Professor with the Department of Measurements and Control Systems at SUT. His research interests include measurement uncertainty evaluation, physical system modelling, applications of fractional-order calculus, and microprocessor-based technologies.