STATISTICS IN TRANSITION new series, March 2023
Vol. 24, No. 2, pp. VII–XII

From the Editor

The March issue of Statistics in Transition new series presents readers with a set of twelve articles written by twenty-seven authors from ten countries (in order of appearance): Turkey, Nigeria, India, Poland, Pakistan, Sweden, Palestine, Ethiopia, Morocco, and Canada. This diversity is also reflected in the thematic variety of the real-life issues raised in these texts.

An attentive reader may notice that despite being the first issue published this year, it has the number 2. This is due to the fact that it was previously issued (as the first) of a special issue, prepared jointly with the Statistics of Ukraine, entitled *A New Role for Statistics*, which is devoted to the challenges – and ways of overcoming them – in the functioning of the official statistics system in the conditions of the war in Ukraine, along with often original solutions (methodological and organizational) in order to maintain the continuity of the production of necessary data and important indicators.

Research articles

In the first paper, **Ipek Deveci Kocakoç** and **Istem Köymen Keser** discuss ***Outlier detection based on the functional coefficient of variation***. The aim of the article is twofold: to show that the functional coefficient of variation is more sensitive to abrupt changes than the functional standard deviation and to propose the utilisation of the functional coefficient of variation as an outlier detection tool. Several simulation trials have shown that the coefficient of the variation function allows the effects of outliers to be seen explicitly. The coefficient of variation function is proposed as an outlier identification method in this study. The CV function is a better descriptive statistics for determining abrupt changes than standard deviation. The availability of the first and second derivatives of the CV function also strengthens its utilization. In the case of outliers in the data set, it is also proven to be a useful statistic. By using the one-out method, outlier curves can easily be detected among others. Therefore, the CV function may be utilized in outlier detection as a confirmatory and complementary method to different outlier detection methods such as outliergram and functional boxplot.

Nureni Olawale Adeboye, Sakinat Oluwabukonla Folorunso, Olawale Victor Abimbola, and **Rasaki Yinka Akinbo** present ***Modelling the volatility of African capital markets in the presence of the COVID-19 pandemic: evidence from five emerging economies in Africa***. The study employed Exponential Generalised Autoregressive Conditional Heteroscedasticity (EGARCH) procedures to develop

stock volatility models for the pre- and COVID-19 era. The Fixed-Effects Two Stage Least Square (TSLS) technique was utilised to establish an empirical relationship between capital market volatility and the COVID-19 occurrence based on equity market indices and COVID-19 reported cases of five emerging African economies: Nigeria, Egypt, South Africa, Gabon and Tanzania. The stock series was made stationary at the first order differencing and the model results indicated that the stock volatility of all the countries responded sharply to the outbreak of COVID-19 with the average stock returns of Nigeria and Gabon suffering the most shocks. Through empirical analysis, this article has exemplified and emphasized the impact of COVID-19 on the volatility of stock markets within the African continent.

The next aticle, by **Vivek Verma, Dilip C. Nath,** and **S. N. Dwivedi** entitled ***Bayesian estimation of fertility rates under imperfect age reporting***, outlines the application of the Bayesian method of parameter estimation to situations where the probability of age misreporting is high, leading to transfers of an individual from one age group to another. An essential requirement for Bayesian estimation is prior distribution, derived for both perfect and imperfect age reporting. As an alternative to the Bayesian methodology, a classical estimator based on the maximum likelihood principle has also been discussed. As evident from the obtained results, even with inaccuracy in age reporting, the Bayesian technique has been found most promising for estimating TMFR, and obtained Bayes' estimates are more precise and reliable than those obtained using the maximum likelihood procedure. Apart from the estimation of transition probabilities, the Bayesian technique has been found to be more useful in estimating the pattern of fertility rates even in situations where there is inaccuracy in age reporting.

Adam Chwila's paper ***The prediction of new COVID-19 cases in Poland with machine learning models*** proposes several possible machine learning approaches to forecasting new confirmed COVID-19 cases, including the LASSO regression, Gradient Boosted (GB) regression trees, Support Vector Regression (SVR), and Long-Short Term Memory (LSTM) neural network. The above methods are applied in two variants: to the data prepared for the whole Poland and to the data prepared separately for each of the 16 voivodeships (NUTS 2 regions). The learning of all the models has been performed in two variants: with the 5-fold time-series cross-validation as well as with the split into the single train and test subsets. The computations in the study used official statistics from government reports from the period of April 2020 to March 2022. The machine learning models can help not only successfully predict different COVID-19 characteristics in the short-term periods, but also explain the factors that have the highest impact on the predictions for considered datasets.

Muhammad Nadeem, Khadija Noreen, H. M. Kashif Rasheed, Rashid Ahmed, and **Mahmood Ul Hassan** discuss ***New generators for minimal circular generalised neighbour designs in blocks of two different sizes***. The authors desribed that the minimal neighbour designs (NDs) are used when a response of a treatment (direct effect) is affected by the treatment(s) applied in the neighbouring units. Minimal generalised NDs are preferred when minimal NDs cannot be constructed. Through the method of cyclic shifts (Rule I), the conditions for the existence of minimal circular generalised NDs are discussed, in which v/2 unordered pairs do not appear as neighbours. Certain generators are also developed to obtain minimal circular generalised NDs in blocks of two different sizes, where k2 = 3, 4 and 5. All these designs are constructed using i sets of shifts for k1 and two for k2.

Abdulhakeem Eideh's article ***On representativeness, informative sampling, nonignorable nonresponse, semiparametric prediction and calibration*** focuses on modelling framework of the semi-parametric prediction of a finite population total while specifying the probability distribution of the response units under informative sampling and nonignorable nonresponse. This is the most general situation in surveys and other combinations of sampling informativeness and response mechanisms can be considered as special cases. Furthermore, based on the relationship between response distribution and population distribution, the authors introduce a new measure of the representativeness of a response set and a new test of nonignorable nonresponse and informative sampling, jointly. Finally, a calibration estimator is obtained when the sampling design is informative and the nonresponse mechanism is nonignorable. The paper is purely mathematical and focuses on the role of informativeness of sampling design and informativeness of nonresponse in adjusting various predictors for bias reduction.

Zofia Zielińska-Kolasińska and **Wojciech Zieliński** in their paper ***A new confidence interval for the odds ratio*** deal with the problem of interval estimation of the odds ratio. The authors propose a new confidence interval for the odds ratio which is based on the exact distribution of the sample odds ratio, hence it works for large as well as for small samples. The coverage probability of that confidence interval is at least the nominal confidence level, in contrast to the asymptotic confidence intervals known in the literature. The information on the sample sizes and the sample odds ratio is sufficient for constructing the new confidence interval. Unfortunately, no closed formulae for the ends of the confidence interval are available. However, for given nA, nB and observed cOR the ends may be easily numerically computed with the aid of the standard software such as R, Mathematica, etc.

In the paper ***Determinants of livestock products export in Ethiopia***, **Ermyas Kefelegn** proposes a procedure for identifying the determinants of the export of

Ethiopian livestock products using vector autoregressive and vector error correction models. Multivariate time series is used to model the association between the products of the Ethiopian livestock export included in the study. Vector autoregressive and vector error correction models are used for modelling and inference. The results indicated the existence of a long term correlation between the volume of live animals, meat and leather exports. The volume of meat export is significantly affected by a lag occurring in the export of live animals in the short-run. The quarterly data from 2002 to 2017 were tested for seasonality and results revealed that all of the series were not affected by periodicity. Moreover, unit root tests show that all four series were non-stationary in level, but stationary after first differencing. The long-run equation shows that the volume of live animals export has a positive long-run relationship with the volume of meat export.

The paper ***On some efficient classes of estimators using auxiliary attribute*** by **Shashi, Bhushan** and **Anoop Kumar** considers some efficient classes of estimators for the estimation of population mean using known population proportion. The classical ratio, and regression estimators suggested by Naik and Gupta (1996) and Abd-Elfattah et al. (2010) estimators are identified as the members of the suggested class of estimators. The expressions of bias and mean square errors are derived up to first-order approximation. The authors have introduced various novel classes of estimators by combining difference estimator with Srivastava, Walsh and Log type estimators and Srivastava type estimator with Walsh type estimator for estimating the population mean of study variable utilizing the information on an auxiliary attribute and compared them with the relevant contemporary estimators till date. The proposed classes of estimators are recommended for the estimation of population mean when information is available in the form of auxiliary attribute.

El Moury Ibtissam, Mohamed Hadini, Adil Chebir, Ben Ali Mohamed and **Echchelh Adil** present ***Proposal of a causal model measuring the impact of an ISO 9001 certified Quality Management System on financial performance of Moroccan service-based companies***. The paper's goal is to test and validate a causal model designed to measure the performance of an ISO 9001 certified Quality Management System (QMS) and its impact on a company's financial performance. By means of this causal analysis (model), the study examines the relationship between: QMS and the financial performance of 41 companies based in Morocco, the management responsibility process and all the QMS processes, the management resources process and all the QMS processes, and the organisational and financial performance of the studied companies. All of the considered firms are part of the service industry and range from medium-sized to large companies. The data gathered in this study have been instrumental in devising actionable insights.

In the next paper, **Emilia Tomczyk** discusses ***Dynamics of survey responses before and during the pandemic: entropy and dissimilarity measures applied to business tendency survey data***. The author starts with the question how to verify whether the pandemic of 2020–2022 can be seen as just another contraction phase. Entropy and dissimilarity measures are employed to study the characteristics of the expectations and assessments expressed in the business tendency survey of Polish manufacturing companies. The empirical results show that the dynamics of the manufacturing sector data, particularly as far as general economic conditions are concerned, set the pandemic period apart. The economic consequences of the COVID-19 pandemic expressed in business tendency surveys tend to be unfavourable, but the statistical properties or the degree of the concentration of respondents' answers do not correspond closely either to the expansion or contraction phases of the business cycle.

Research Communicates and Letters

In the *Research Communicates and Letters* section an article by **Leonardo Campanelli,** entitled ***Breaking Benford's law: a statistical analysis of COVID-19 data using the Euclidean distance statistic*** analyses the COVID-19 weekly case counts by country provided by the World Health Organization, (updated to December 20, 2021) using the Euclidean distance statistical test of Benford's law. The null hypothesis that the first-digit distribution of those counts follows Benford's distribution is tested using weekly confirmed cases instead of daily ones following the requirement of having counts that extended over many orders of magnitudes so to improve the compliance of the data sets with Benford's law. For the same reason daily and weekly death counts were not considered. Also, cumulative cases were not considered as their numbers flatten (especially at the end of a 'wave'), thus distorting relative digit frequencies.

Włodzimierz Okrasa
Editor