STATISTICS IN TRANSITION new series, June 2019
Vol. 20, No. 2, pp. I−IV

FROM THE EDITOR

Let me start this issue with a good news about our journal. On behalf of the Editorial Board and the Editorial Office, I am pleased to share with our partners and supporters – authors, reviewers and readers – information about the continuous growth of the *Statistics in Transition new series* in terms of the scores and recognition by the prestigious indexation bases. For instance, according to the new CiteScore metrics released by Scopus for the year 2018, our journal has obtained 0, 27 (compared to 0, 21 in 2017). Even more, the journal has recently been included into the New Scimago Journal & Country Rank – a portal developed from the information contained in the Scopus database. We are thankful to everyone who contributed to these achievements in any way – they all count!

A set of eleven papers published in this issue of the *Statistics in Transition new series* presents a heterogeneous body of works – from statistical and econometric research articles through the conference papers from the 2nd Congress of Polish Statistics (Warsaw, July 2018) and a paper from the 27th Conference on Classification and Data Analysis (Ciechocinek, September 2018) to a research communicate.

The main part which contains research papers starts with an article ***Statistical inference of exponential record data under Kullback-Leiber divergence measure*** by **Raed R. Abu Awwad, Ghassan K. Abufoudeh** and **Omar M. Bdair.** Using one parameter exponential record data, the authors conduct statistical inferences (maximum likelihood estimator and Bayesian estimator) for the suggested model parameter. They also aim to predict the future (unobserved) records and to construct their corresponding prediction intervals based on observed set of records. In the estimation and prediction processes, the square error loss and the Kullback-Leibler loss functions are employed. Numerical simulations were conducted to evaluate the Bayesian point predictor for the future records. Subsequently, data analyses involving the times (in minutes) to breakdown of an insulating fluid between electrodes at voltage 34 kV have been performed to show the performance of the methods developed on estimation and prediction.

Anoop Chaturvedi's and **Sandeep Mishra's** paper ***Generalized Bayes Estimation of Spatial Autoregressive Models*** deals with the spatial autoregressive (SAR) models, which are widely used in spatial econometrics for analysing spatial. The authors derive a Generalized Bayes estimator for estimating the parameters of a SAR model for data involving spatial autocorrelation structure. The admissibility and minimaxity properties of the estimator have been discussed. For investigating the finite sample behaviour of the estimator, the results of a simulation study have been presented, and the approach was applied to demographic data on total fertility rate for selected

Indian states - an improvement over the usual least squares estimator was demonstrated for a wide range of the parametric settings.

Jana Cibulková, Zdenek Šulc, Sergej Sirota, and Hana Rezankova discuss ***The effect of binary data transformation in categorical data clustering*** in the paper focused on hierarchical clustering of categorical data. They compare two approaches, the first of which embraces performing a binary transformation of the categorical variables into sets of dummy variables and then use the similarity measures suited for binary data, while the second uses similarity measures developed for the categorical data. The comparison of these two approaches is performed on generated datasets with categorical variables and the evaluation is done using both the internal and the external evaluation criteria. In conclusion, the authors demonstrate that the binary transformation is not necessary in the process of clustering categorical data since the second approach yields comparable results under the less demanding conditions.

Grażyna Dehnel's and **Marek Walesiak's** paper ***A comparative analysis of economic efficiency of medium-sized manufacturing enterprises in districts of Wielkopolska province using the hybrid approach with metric and interval-valued data*** describes a hybrid approach to evaluating economic efficiency of medium-sized manufacturing enterprises (employing from 50 to 249 people) in districts of Wielkopolska province, using metric and interval-valued data. The authors employ an approach that combines multidimensional scaling with linear ordering. The analysis was based on data prepared in a two-stage process. First, a data set of observations was obtained for metric variables describing economic efficiency of medium-sized manufacturing enterprises. Next, the unit-level data were aggregated at district level and turned into two types of data: metric and interval-valued data. The analysis of interval-valued data was carried out using symbolic-to-classic and symbolic-to-symbolic approaches, and the results of the two approaches were compared. [The calculations were made with scripts prepared in the R environment.]

Adedayo A. Adepoju's and **Tayo P. Ogundunmade's** paper ***Economic growth and its determinants: a cross-country evidence*** presents evidence from a panel of 126 countries, over the time period of 2010 to 2014, that economic growth is dependent on various factors. The authors found that government expenditure control, reduced inflation and increased trade openness are the factors that boost the economic growth of a country. Significant evidence is seen for government consumption, fiscal policy and trade openness. No significant relationship has been observed between exchange rate and economic growth, whereas unemployment influences output for African countries. The cross regional analysis of Asian, European, African, Caribbean and American countries gives specific determinants for these regions. Fiscal balance has shown a consistent positive relationship with economic growth throughout the analyses. Fiscal balance and unemployment rate played their role in the growth of African countries. Inflation rates and increased openness were significant for some regions. Exchange rate did not return significant coefficients for any of the sub-regions. Government consumption, trade openness, policy interest rate and industrial production rate showed significant effect for different regions of the world.

In the article ***Application of the strategy combining monetary unit sampling and a Horvitz–Thompson estimator of error amount in auditing – results of a simulation study,*** **Bartłomiej Janusz** discusses a possible alternative for testing audit populations with high error rates under a pragmatic assumption that auditors need information on the performance of different statistical methods when applied to audit populations. A strategy combining systematic Monetary Unit Sampling and confidence intervals for the total error based on the Horvitz-Thompson estimator with normality assumption was checked – including its reliability and efficiency – using real and simulated data sets. It was shown that, for the majority of populations, the interval coverage rate was lower than the assumed confidence level. In most cases confidence intervals were too wide to be of practical use to auditors. Confidence intervals tended to become wider as the observed error rate increased. Tests disclosed that the distribution of the Horvitz-Thompson estimator was not normal. A detailed analysis of the distributions of the error amount in the examined real audit populations is also given.

The next paper ***Estimation of Energy Intensity in Indian Iron and Steel Sector: A Panel Data Analysis*** by **Anukriti Sharma, Hiranmoy Roy** and **Narendra Nath Dalei** presents results of the empirical estimation of the energy intensity of Indian Iron and Steel sector accounting for the impact of ECA (Energy Conservation Act, 2001) and PAT (Perform, Achieve and Trade mechanism), Phase-I in dummy variable form. The results indicate that the decline in energy consumption in this sector until 2011 can also be attributed to Energy Conservation Act implemented in the year 2001 along with other factors. The authors conclude that ECA has a significant impact on reduction of energy intensity of the steel firms. PAT does not seem to have a considerable impact on energy intensity alone but in the years where both PAT and ECA are prevalent, i.e. from 2012 to 2015, there seems to be a significant impact of around 0.050 reduction in energy intensity, as accounted for by different models in this paper. In addition, the empirical results suggest that profit margin intensity was found to be negatively related to energy intensity implying more profitable firms invest more in energy efficiency.

The next articles are based on the conference presentations.

In the paper ***Variable selection in multivariate functional data classification*** by **Tomasz Górecki, Mirosław Krzyśko,** and **Waldemar Wołyński,** a new variable selection method is considered in constructing a classification using multivariate functional data approach. The variable selection is a dimension reduction method, which leads to the replacement of the high-dimensional vector process by a low-dimensional vector with a comparable classification error. Various classifiers appropriate for functional data are used. The proposed variable selection method is based on functional distance covariance (dCov) and the Hilbert-Schmidt Independent Criterion (HSIC), which are discussed in the literature. The method employed is a modified version of the procedure described by Kong et al. (2015), and the proposed methodology is illustrated with a real data. The authors consider this approach as an alternative to other variable selection methods.

Roman Zmyślony's and **Arkadiusz Kozioł's** paper ***Testing hypotheses about structure of parameters in models with block compound symmetric covariance structure*** deals with testing the hypotheses of the so-called structured mean vector and the structure of a covariance matrix. To this aim, Jordan algebra properties are used and tests based on best quadratic unbiased estimators (BQUE) are constructed. For convenience coordinate-free approach is employed as a tool for characterization of best unbiased estimators and testing hypotheses. To obtain the test for mean vector, linear function of mean vector with the standard inner product in null hypothesis is changed into equivalent hypothesis about some quadratic function of mean parameters (it is shown that both hypotheses are equivalent and testable). In both tests the idea of the positive and negative part of quadratic estimators is applied to get the test statistics which have F distribution under the null hypothesis. Finally, power functions of the obtained tests are compared with other known tests like LRT or Roy test. For some set of parameters in the model the presented tests have greater power than the above mentioned tests.

The paper by **Barbara Pawełek, *Extreme gradient boosting method in the prediction of company bankruptcy*** discusses the use of machine learning methods to predict company bankruptcy. Comparative studies carried out on selected methods to determine their suitability for predicting company bankruptcy have demonstrated high levels of prediction accuracy for the extreme gradient boosting method in this area. This method is resistant to outliers and relieves the researcher from the burden of having to provide missing data. The special aim of this study is to assess how the elimination of outliers from data sets affects the accuracy of the extreme gradient boosting method in predicting company bankruptcy, with intention to show the advantages of application of the extreme gradient boosting method in bankruptcy prediction based on data free from the outliers. The research was conducted using 64 financial ratios for the companies operating in the industrial processing sector in Poland. The research results indicate that it is possible to increase the detection rate for bankrupt companies by eliminating the outliers reported for companies which continue to operate as a going concern from data sets.

The issue concludes with a research communicate*, **Efficient two-parameter estimator in linear regression,*** by **Ashok V. Dorugade,** who discusses two-parameter estimators in linear model with multicollinearity. The author proposes an alternative efficient two-parameter estimator along with examination of its properties and the results of its comparison with the OLS estimator, as well as the ordinary ridge regression (ORR) estimators. Also, using the mean squared error criterion the proposed estimator performs more efficiently than OLS estimator, ORR estimator and other reviewed two-parameter estimators. A numerical example and simulation study are finally conducted to illustrate the superiority of the proposed estimator.

Włodzimierz Okrasa

Editor