sciendo

From the Editor

The passing quarter was full of a series of events of great importance for the scientific life of the global community of statisticians. Two of them – the meeting of IAOS/International Association for Official Statistics and the 3rd Congress of Polish Statistics on the occasion of the 110th anniversary of the Polish Statistical Association/PSA – took place at the same time and in the same venue, in Krakow's Convention Center, on April 25–28. Like previous congresses, this one also gave the opportunity to award the highest distinction, which the Polish Statistical Association awards to distinguished persons for an extraordinary contribution to the development of statistical sciences, which is the medal of Jerzy Spława-Neyman. This time, the Neyman medal was awarded to Prof. Partha Lahiri, Danny Pfeffermann and Wlodzimierz Okrasa. Along with congratulations to the laureates, their short bionotes are included in the first section of this issue.

The June issue presents a set of 12 articles – there are 8 manuscripts in Research papers section, 3 conference papers from the 39th Multivariate Statistical Analysis 2021, which took place in November 2021 in Lodz, Poland provided as others articles, and 1 paper published in the Research Communciates&Letters part. Our authors come from Egypt, Saudi Arabia, Pakistan, USA, Poland, India, Nigeria, Malesia, Algeria, and Italy. We are pleased to be recognized by such a respectable group of scientists.

Research articles

In the first paper, **Abdelfattah Mustafa A.** and **M. I. Khan** discuss ***The length-biased power hazard rate distribution: some properties and applications.*** The authors show that this distribution reports an extension of several probability distributions, namely: exponential, Rayleigh, Weibull, and linear hazard rate. The procedure of maximum likelihood estimation was taken for parameters and derived. The applicability of the model was explored by three real data sets. Also, to examine the performance of the technique, a simulation study is extracted. The superiority of the new model has been exhibited by some real data sets. It has been seen that Power Hazard Rate Distribution can adequately provide better fits than other models.

The article entitled ***Jackknife winsorized variance estimator under imputed data*** prepared by **Muhammad Umair SohaiL, Fariha Sohil, Javid Shabbir,** and **Sat Gupta** show the problem of missing and extreme values for the estimation of population variance. The presence of extreme values either in the study variable, or the auxiliary

variable, or in both of them, can adversely affect the performance of the estimation procedure. The authors have considered three different situations for the presence of extreme values and also have considered jackknife variance estimators for the population variance by handling these extreme values under stratified random sampling. Bootstrap technique ABB was carried out to understand the relative relationship more precisely. The authors also modified the linearized version of the jackknife variance estimator suggested by Rao (1996) for the precise estimation of winsorized variance, which is helpful with computer programs that use linearized methods for the estimation of variance. The stratified sampling scheme was discussed as it is commonly used in large scale socio-economic surveys.

Maciej Jewczak and **Magdalena Brudz** in their manuscript *Socio-economic development and quality of life of Nuts-2 units in the European Union* examine the level of socio-economic development and quality of life in the European Union in the years 2004 and 2018. The analyses were conducted for a rarely used level of spatial data aggregation, i.e. for NUTS-2 units, but they cover only those European regions that were EU members in 2004. As the primary research tool, the two-dimensional development matrix was adopted, which enabled the verification of the hypothesis regarding the convergence of synthetic measures that indicate the levels of socio-economic development and quality of life in the EU regions. For these indices, the development matrix was also used to identify the strengths and weaknesses as well as the opportunities and threats for selected spatial units, and, at the same time, to estimate the rates of change of the socio-economic development and quality of life levels. Depending on the criteria considered, the most common methods for determining the degree of the advancement of life quality or socio-economic development include taxonomical techniques and analyses of potential, which are based mainly on objective data sourced from official registers. A very important fact from this study is that the scientific analysis covered data at the regional level, while most studies focus only on quality of life or socio-economic development at the macro level.

Arora S., Mahajan K. K., and **Jangra V.** present *A Bayesian estimation of the Gini index and the Bonferroni index for the Dagum distribution with the application of different priors.* The Bayesian estimators and highest posterior density credible intervals were obtained for two popular inequality measures, viz. the Gini index and the Bonferroni index in the case of the Dagum distribution. The study has considered the informative and non-informative priors, i.e. the Mukherjee-Islam prior and the extension of Jeffrey's prior, respectively, under the presumption of the Linear Exponential (LINEX) loss function. The authors have carried out a Monte Carlo simulation study in order to obtain the relative efficiency of both the Gini and Bonferroni indices while taking into consideration different priors and loss functions.

It was observed that Mukherjee-Islam prior performs better than the extension of Jeffrey's prior in terms of having smaller estimated loss. It was also observed that the LINEX loss function results in smaller loss as compared to squared error loss function (SELF) for small, medium and large sample sizes irrespective of the choice of prior. The expected loss decreases as the sample size increases.

The paper by **Sanusi Alhaji Jibrin** and **Rosmanjawati Abdul Rahman** entitled ***ARFURIMA models: simulations of their properties and applications*** defines the Autoregressive Fractional Unit Root Integrated Moving Average (ARFURIMA) model for modelling ILM time series with fractional difference value in the interval of $1<dd<2$. The performance of the ARFURIMA model was examined through a Monte Carlo simulation. Also, some applications were presented using the energy series, bitcoin exchange rates and some financial data to compare the performance of the ARFURIMA and the Semiparametric Fractional Autoregressive Moving Average (SEMIFARMA) models. The presented simulations studies confirmed superiority of the ARFURIMA over the ARIMA in simulating nonstationary and the FURI series and thus proved the ILM properties of the ARFURIMA model and its large sample properties too. Some applications of the model were presented and further confirmed a better fit of the ARFURIMA compared to the SEMIFARMA model.

Abdelmalek Gagui's and **Abdelhak Chouaf's** article ***On the nonparametric estimation of conditional hazard estimator in the single functional index*** characterises the conditional hazard estimator of a real response where the variable is given a functional random variable (i.e. it takes values in an infinite-dimensional space). The authors focus on the functional index model as a good compromise between nonparametric and parametric models to prove the asymptotic normality of the proposed estimator under general conditions and in cases where the variables satisfy the strong mixing dependency. The means of the kernel estimator method, based on a single-index structure, were used. A simulation of the proposed methodology has shown that it is efficient for large sample sizes. It was also shown that the estimator provides good predictions under this model. In non-parametric functional statistics, the semi-metric of the projection type is very important for increasing the concentration property. The functional index model is a special case of this family of semi-metrics because it is based on the projection on a functional direction, which is important for the implementation of the method in practice.

In the next paper **Mateusz Borkowski** focuses on ***Institutional equilibrium in EU economies in 2008 and 2018: SEM-PLS models*** to identify the strength and direction of the development of the relationship between formal and informal institutions and to assess the institutional equilibrium of modern economies. The article presents a comprehensive model of the institutional structure and a unique method of

measuring institutional equilibrium. The structural equations modelling based on partial least squares (SEM-PLS) was applied. The study included 27 EU economies and the research period covered the years 2008 and 2018. The results of the study demonstrate that the quality of informal institutions strongly, positively determines the quality of formal institutions. The conducted analyses indicate that modern economies are diversified in terms of the quality of informal and formal institutions and, consequently, in institutional equilibrium.

Sakshi Kaushik, Alka Sabharwal, and **Gurprit Grover** present the manuscript entitled ***Extracting relevant predictors of the severity of mental illnesses from clinical information using regularisation regression models.*** The authors describe the relevant predictors of the severity of mental illnesses (measured by psychiatric rating scales) from a wide range of clinical variables consisting of information on both laboratory test results and psychiatric factors. The laboratory test results collectively indicate the measurements of 23 components derived from vital signs and blood tests results for the evaluation of the complete blood count. The 8 psychiatric factors known to affect the severity of mental illnesses are considered, viz. the family history, course and onset of an illness, etc. Retrospective data of 78 patients diagnosed with mental and behavioural disorders were collected from the Lady Hardinge Medical College & Smt. S.K, Hospital in New Delhi, India. The observations missing in the data were imputed using the non-parametric random forest algorithm. This paper adds to the literature of medical research aimed at identifying the biomarkers for diagnosis and predictors of the severity status of mental disorders, and should be helpful in developing valid and efficient approaches to diagnose the disorders at an early stage The clinicians can use the relevant factors to build a profile of the patient and his needs, and also effective strategies for treatment planning.

Other articles

39th Multivariate Statistical Analysis 2021, Lodz. Conference Papers

The section starts with the paper prepared by **Czesław Domański** and **Robert Kubacki** entitled ***Regression model of water demand for the city of Lodz as a function of atmospheric factors.*** The authors presented the results of the work on a statistical model which determined the influence of individual atmospheric factors on the demand for water in the city of Lodz, Poland, in 2010-2019. In order to build the model, the study used data from the Water Supply and Sewage System Company (Zakład Wodociągów i Kanalizacji Sp. z o.o.) in the city of Lodz complemented with data on weather conditions in the studied period. The analysis showed that the constructed models make it possible to perform a forecast of water demand depending on the expected weather conditions. The relation between daily weather variables and water

use in the city of Lodz, Poland were examined. It was confirmed that the maximum daily temperature is a good predictor of water demand, and that holidays are significant in decreasing the water demand. Moreover, wind speed is a good predictor of water demand. It is likely that higher wind speed increases evaporation of water, which induces a cooling effect and thus decreases daily water consumption. Together, all these variables explain between 65% of the variations in the city of Lodz.

Stefano Bonnini's and **Getnet Melak Assegie's** article evaluates ***Advances on permutation multivariate analysis of variance for big data.*** Due to the gap in the literature about combined permutation tests, in particular for big data with a large number of variables*,* a Monte Carlo simulation study was carried out to investigate the power behaviour of the tests, and the application to a real case study was performed to show the utility of the method. It was provided that among the distribution free solutions to the multivariate analysis of variance in the family of combined permutation tests, the method based on the Tippet combination is in general preferable, especially if there is no preventive information about the possible percentage of variables (or marginal distributions) under the alternative hypothesis. Instead of the Tippett combination, the Fisher rule can be applied when the percentage is close to 100%. The Liptak combination seems to be non-convenient in general. This methodological tool is an important and useful solution of testing problems for big data, especially when the number of variables is very large and the sample sizes are small. The usefulness and the effectiveness of the method is confirmed by the application to the case study concerning the survey on the organizational well-being at the University of Ferrara.

The paper by **Tadeusz Bednarski**, **Piotr B. Nowak,** and **Magdalena Skolimowska-Kulig** examines ***Scaled Fisher consistency for the partial likelihood estimation in various extensions of the Cox model.*** The Cox proportional hazards model has become the most widely used procedure in survival analysis, and the theoretical basis of the original model has been developed in various extensions. The authors have investigated the accuracy of inference based on the primer Cox model in the existence of unobserved heterogeneity, that is, when the data generating mechanism is more complex than presumed and described by the kind of an extension of the Cox model with undefined frailty. It was shown that the conventional partial likelihood estimator under the considered extension is Fisher-consistent up to a scaling factor, provided symmetry-type distributional assumptions on covariates. The results of simulation experiments that reveal an exemplary behaviour of the estimators were presented.

Research Communicates and Letters

The Research Communicates & Letters section presents a paper prepared by **Hemlata Joshi, S., Azarudheen, M. S. Nagaraja,** and **C. Singh** entitled ***On the quick estimation of probability of recovery from COVID-19 during first wave of epidemic***

in India: a logistic regression approach*. Due to the fact that the COVID-19 became a threat all across the world with the new cases every day, and there is still a difficult situation with no effective medicine, it is very important to know if a patient with COVID-19 is going to recover or die. The study is based on the situation in India and the data published by the Ministry of Health and Family Welfare of India were used for the empirical analysis. The manuscript shows a model that has been developed to estimate the probability of recovery of a patient based on the demographic characteristics, as most of the Indian population is living in poor hygienic conditions. The probability model is developed using the indirect method of estimation based on some demographic factors, and it was found that the probability of recovery from coronavirus disease is statistically the same in both males and females.

Włodzimierz Okrasa
Editor