

STATISTICS IN TRANSITION new series, December 2016
Vol. 17, No. 4, pp. 585–588

FROM THE EDITOR

This fourth issue of the volume 17 of the *Statistics in Transition new series* completes the annual (quarterly) editorial cycle for the year 2016. This provides an opportunity for the Journal Editor to express deep gratefulness, also on behalf of all the editorial bodies – Editorial Board, Associate Editors, and the Editorial Office – to its all collaborators and contributors: authors and peer-reviewers, technical staff and other supporters for making it – actually a product of joint efforts of all persons involved – to be appearing on the regular basis while playing a role in the advancement of science of data production and use across disciplines and sectors. Special thanks go to the experts who have served as reviewers, devoting generously their time and expertise in order to ensure high scientific and quality standards of the papers published over the passing year - theirs names are listed in the acknowledgements. This is also an opportunity to acknowledge the help and advice we have been obtaining from the members of the aforementioned panels. And to announce an intention to include some new names from among the actively supporting us experts to the panel of Associate Editors, along with an appreciation of cooperation with our journal over many years of few persons who will leave the panel with the new year.

This issue has customary structure, being composed of the three major sections. But I am pleased to announce an innovation to be introduced with the next issue, which will rely on adding a new section 'Research Communicates'. It is envisioned as a possibility for presenting new conceptual or empirical results of an ongoing research work in the form of not fully completed research paper, sometimes even in its embryo stage, however, advanced enough to pass the peer-reviewing process.

The first section, devoted to sampling and estimation issues, contains two papers. It begins with **J. Subramani's** paper *New Median Based Ratio Estimator For Estimation of the Finite Population Mean*, which deals with a new median based ratio estimator in the absence of an auxiliary variable. The bias and mean squared error of the proposed median based ratio estimator are obtained. The performance of the median based ratio estimator is compared with that of the SRSWOR sample mean, ratio estimator and linear regression estimator for certain natural population. It is shown from the numerical comparisons that the proposed median based ratio estimator outperforms the SRSWOR sample mean, ratio estimator and also the linear regression estimator.

In the second paper **Housila P. Singh** and **Surya K. Pal** propose *A New Family of Estimators of the Population Variance Using Information on Population Variance of Auxiliary Variable in Sample Surveys*. Starting with

observation that there are several recently proposed classes of estimators (e.g., due to Sharma and Singh (2014) or Singh and Pal (2016)) the authors emphasize the need for checking their properties. They provide asymptotic expressions of bias and mean squared error (*MSE*) of the suggested family of estimators. Asymptotic optimum estimator (*AOE*) in the family of estimators is also identified. Some subclasses of estimators of the proposed family of estimators have been identified and some theoretical comparisons among the estimators are discussed in the paper.

The research articles section starts with **Elzbieta Golata's** paper *Shift in methodology and population census quality*. The shift in methods embraces a move in conducting population census from a conventional enumeration through a sample survey and a mixed approach to administrative data. This paper compares two censuses which were conducted in Poland in 2002 and 2011; each of them presents by itself different case of the traditional method (2002 census) and the combined approach (2011 census), respectively. The quality of census data is discussed with essential aims and objectives to provide reliable information on the population age and sex structure in detailed territorial division. Quality assessment is provided for the whole country and at regional level. It starts with the consideration of coverage errors using multiple sources of data and non-matching methods - in particular: demographic analysis based on previous censuses, vital statistics and a comparison with other existing sources. Different cross-sections by sex, age and place of residence are considered. The questions of adequacy and divergence are discussed in the substantive terms.

Olusanya Elisa Olubusoye, Grace Oluwatoyin Korter and Afees Adebare Salisu present their results on *Modelling Road Traffic Crashes Using Spatial Autoregressive Model with Additional Endogenous Variable*. The authors construct a model based on a linear cross sectional Spatial Autoregressive (SAR) framework with additional endogenous variables, exogenous variables and SAR disturbances. The focus is on Road Traffic Crashes (RTC) in Oyo state, Nigeria. The number of RTC in each Local Government Area (LGA) of the state is the dependent variable. A weights matrix, travel density, land area and major road length of each LGA were used as exogenous variables and population was the IV. The objective was to determine the hotspots and examine whether the number of RTC cases in a given LGA is affected by the number of RTC cases of neighbouring LGAs and an instrumental variable. The hotspots include Oluyole, Ido, Akinyele, Egbeda, Atiba, Oyo East, and Ogbomosho South LGAs. The study concludes that the number of RTC in a given LGA is affected by the number of RTC in contiguous LGAs. The authors address some policy implication of their results, such as that road safety and security measures must be administered simultaneously to LGAs with high concentration of RTC and their neighbours to achieve significant remedial effect.

Kumar Prabhash, Vijay M Patil, Vanita Noronha, Amit Joshi and Atanu Bhattacharjee in the paper *Bayesian Accelerated Failure Time and its*

Application in Chemotherapy Drug Treatment Trial propose an alternative to the Cox proportional hazards model (CPH), which is normally applied in clinical trial data analysis. Since this model can generate severe problems with breaking the proportion hazard assumption, the authors present an accelerated failure time (AFT) instead. The model can be used through consideration of different covariates of interest and random effects in each section. The model is simple to fit by using OpenBugs software and is revealed to be good for the Chemotherapy data. However, other model comparison tools can be used to compare the models in different computational platforms.

Henryk Gurgul's and Marcin Suder's paper *Calendar And Seasonal Effects of Size of Withdrawals From ATMs Managed by Euronet* analyses the calendar effects on withdrawals from Automated Teller Machines (ATMs), daily data, managed by the Euronet network for the period from January 2008 to March 2012. They focus on the identification of specific calendar and seasonal effects in the ATM cash withdrawal series of the company in the Polish provinces of Lesser Poland (Małopolska) and Subcarpathian (Podkarpackie). The results of the analysis show that withdrawals depend strongly on the day of the week. On Fridays more cash is withdrawn than on other days, and Saturdays and Sundays are the days of the week with the lowest level of withdrawals. In a month, it can be seen that cash withdrawals take place more often in the second and in the last weeks of the month. This observation suggests that withdrawals from ATMs can be related to the profile of wage withdrawals. In Poland in the public sector wages are paid at the beginning of the month, and in the private sector at the end of the month. The time series of withdrawals also reflect seasonality. The largest amounts are withdrawn in July, August and December. Reason for the increased demand for cash are the summer holidays and the Christmas season. The results reflect consumer habits which show substantial calendar and seasonal effects.

A set of the next three papers is a selection of the articles based on presentations given at the 34th Multivariate Statistical Analysis Conference being held at the University of Lodz in October 2015.

The paper by **Marta Dziechciarz-Duda and Józef Dziechciarz**, *The Identification of Training Needs For Human Capital Quality Improvement in Poland – a Statistical Approach* addresses the issues of practical importance concerning methods of assessing efficiency of the Competency Development Programme launched recently by the Ministry of Science and Higher Education. It embraces allocation of additional financial means for activities meant to prepare students with the so-called soft skills necessary in scientific careers and on the labour market. Courses developing skills such as team work ability, leadership, creativity, independent thinking and innovative approach to problem solving will be financed. A thorough analysis of needs starts with existing databases describing the quality of human capital in Poland in order to identify those competencies that graduates of universities are missing. The paper discusses possible statistical tools applicable for that purpose, from the simple descriptive statistics to advanced multivariate statistical analysis.

In the next paper, *Interval Estimation of Higher Order Quantiles. Analysis of Accuracy of Selected Procedures* by **Dorota Pekasiewicz** selected nonparametric and semiparametric estimation methods of higher orders quantiles are considered. The construction of nonparametric confidence intervals is based on order statistics of appropriate ranks from random samples or from generated bootstrap samples. Semiparametric bootstrap methods are characterized by double bootstrap simulations. The values of bootstrap sample below the prearranged threshold are generated by the empirical distribution and the values above this threshold are generated by the distribution based on the asymptotic properties of the tail of the random variable distribution. The results of the study allow one to draw conclusions about the effectiveness of the considered procedures and to compare these methods.

Grażyna Dehnel's paper *M-Estimators in business statistics* discusses one of the techniques that is meant to deal with outlying observations, namely M-estimation, from the evaluative perspective. Until recently the implementation of robust regression methods, such as M-estimation or MM-estimation, was limited due to their iterative nature. With advances in computing power and the growing availability of statistical packages, such as R and SAS, Stata, the applicability of robust regression methods has increased considerably. The M-estimation method is being assessed using data from a survey of small and medium-sized businesses. The comparison involves nine *M-estimators*, each based on a different weighting function. For instance, the largest gain in efficiency and robustness of *M-estimators* was obtained when Talworth's and Tukey's functions were used.

The last article, *Informative versus Non-Informative Prior Distributions and their Impact on the Accuracy of Bayesian Inference* by **Wioletta Grzenda** discusses the benefits arising from the use of the Bayesian approach to predictive modelling, along with exemplification using a linear regression model and a logistic regression model. The impact of informative and non-informative prior on model accuracy is systematically examined and compared. The data from the Central Statistical Office of Poland on unemployment by individual districts in Poland are used, and Markov Chain Monte Carlo methods (MCMC) was employed in modelling. These results indicate that the accuracy of models estimated with informative a priori distributions is higher. Therefore, when additional out-of-sample knowledge is available, the appropriate selection of a priori distribution can improve the accuracy of regression and classification models.

Włodzimierz Okrasa

Editor