

STATISTICS IN TRANSITION new series, September 2017
Vol. 18, No. 3, pp. 369–371, DOI 10.21307/stattrans-2016-075

FROM THE EDITOR

This issue is composed of eleven articles grouped in four sections, as follows: the first group of papers, devoted to *sampling methods and estimation* issues consists of three papers; the next four papers constitute the section of *research articles*, followed by *other articles* (containing just one conference paper), and the whole issue closes with *research communicates and letters* containing two papers of the nature of 'work in progress'. They are briefly characterized below.

In the first paper, *Population Variance Estimation Using Factor Type Imputation Method*, **Ranjita Pandey** and **Kalpana Yadav** propose a variance estimator based on factor type imputation in the presence of non-response. They describe properties of the new estimators along with their optimality conditions. The proposed classes of factor type ratio estimators are shown to be more efficient than some of the existing estimators – such as the usual unbiased estimator of variance, ratio-type, dual to ratio type and ratio cum dual to ratio estimators. Their performances in terms of relative efficiencies are illustrated with simulated and real data sets. In particular, one of the proposed estimators is shown to perform best from the point of view of increasing efficiency (but all the three proposed FT type estimators are the best estimators in the sense of having the largest PRE).

Krzysztof Beck's paper, *Bayesian Model Averaging and Jointness Measures: Theoretical Framework and Application to the Gravity Model of Trade* discusses the Bayesian model averaging (BMA) along with the benefits due to combining the knowledge generated through the analysis of different models. The BMA structure is described together with its most important statistics (prior parameter proposals, prior model size distributions, and also the jointness measures). Its application is illustrated with the gravity model of trade, where determinants of trade are chosen from the list of nine different variables. It enabled the identification of four robust determinants: geographical distance, real GDP product, population product and real GDP per capita distance. All variables, except for population product, have coefficient signs predicted by the theory. For instance, the complementary relationship between real GDP product and population product allowed one to explain the negative sign of the population product coefficient.

In the paper *On Asymmetry of Prediction Errors in Small Area Estimation*, **Tomasz Żądło** starts with an observation that the mean squared error (MSE), which reflects only the average prediction accuracy, is insufficient and even inadequate as a measure of overall quality since we are interested not only in the

average but in the whole distribution of prediction errors. Therefore, the author proposes to use an alternative measure of prediction accuracy in the context of small area estimation, taking into account a modified version of the empirical best predictor based on a generalization of the predictor presented by Molina and Rao (2010). The generalization results from the assumption of a longitudinal model and possible changes of the population and subpopulations in time. The considerations are supported by results of the real data application.

The second part of this issue begins with an article by **Mirosław Krzyśko** and **Lukasz Smaga**, entitled *An Application of Functional Multivariate Regression Model to Multiclass Classification*. The authors propose the scale response functional multivariate regression model based on possible functions representation of functional predictors and regression coefficients. The proposed functional multivariate regression model is employed to multiclass classification for multivariate functional data. Computational experiments performed on real data sets demonstrate the effectiveness of the proposed method for classification for functional data.

Tomasz Górecki's and **Maciej Łuczak's** article on *Stacked Regression with a Generalization of the Moore-Penrose Pseudoinverse* is devoted to the problem of making an optimal selection among available methods of classification. The authors propose a combined method that allows one to consolidate information from multiple sources in a better classifier. They discuss the stacked regression (SR) as a way of forming linear combinations of different classifiers toward improved accuracy of classification through employing the Moore-Penrose (MP) pseudoinverse to find the solution to a system of linear equations. Due to the computational difficulty with a greater number of features, they propose a genetic approach to handle the problem. Experimental results on various real data sets demonstrate that the improvements are efficient and that this approach outperforms the classical SR method, providing a significant reduction in the mean classification error rate.

In the next article, *An Additive Risks Regression Model for Middle-Censored Lifetime Data*, **P. G. Sankaran** and **S. Prasad** discuss the middle-censoring data problem arising in situations where the exact lifetime of study subjects becomes unobservable, and whether it happens to fall in a random censoring interval. The authors propose a semiparametric additive risks regression model for analysing middle-censored lifetime data arising from an unknown population. They estimate regression parameters and the unknown baseline survival function by two different methods – the first method uses the martingale-based theory, and the second method is an iterative method. The finite sample behaviour of the estimators is assessed through simulation studies, and the utility of the model with a real life data set is demonstrated in the conclusions.

The article by **Luboš Marek**, **Stanislava Hronová** and **Richard Hindls**, *Option for Predicting the Czech Republic's Foreign Trade Time Series as Components in Gross Domestic Product*, analyses the time series data for the

foreign trade of the Czech Republic (CR), and the issue of predictions in such series using the SARIMA and transfer-function models. The authors' goal is to propose models suitable for describing the time series of the exports and imports of goods and services from/to the CR and to subsequently use these models for predictions in quarterly estimates of the gross domestic product's component resources and utilization. They suggest a class of models with time lag as suitable, allowing for making predictions in the time series of the CR exports and imports several months ahead.

In the next article, *Subjective approach to assessing poverty in Poland – implications for social policy*, **Leszek Morawski** and **Adrian Domitrz** discuss the effect of adopting a particular weigh system in constructing equivalised household income, such as based on the OECD recommendations concerning such scales. Poland is an interesting case for applying an alternative, subjective approach to calculating equivalent scales due to relatively large average size of households. The overall poverty rates for the two approaches are not distinctly different but they lead to significantly different distributions of poverty. For instance, the subjective approach suggests that one-person households and not large families should be considered most exposed to risk of material poverty. Since the relative positions of different policy-relevant groups of households in the distribution of income differ significantly, the respective programs of social transfers may need to be revised in order to be better targeted.

The section *other articles* includes a conference paper (presented at the Multivariate Statistical Analysis conference held in Łódź, 2016) by **Marek Walesiak** and **Andrzej Dudek**, *Selecting the Optimal Multidimensional Scaling Procedure for Metric Data with R environment*. The authors start with an observation that the main decision problem of multidimensional scaling (MDS) procedure for the metric measurement data consists in making selection of the method of normalization of the values of the variables and of distance measure, and finally of a MDS model. The article proposes a solution that allows choosing the optimal multidimensional scaling procedure out of 18 normalization methods included in the analysis and of 5 distance measures for 3 types of MDS models using two criteria: Kruskal's Stress-1 fit measure and Hirschman-Herfindahl HHI index. An empirical example provides illustration of the proposed procedure.

Finally, there are two articles in the last section, *research communicates and letters*. In *Sample Allocation in Estimation of Proportion in a Finite Population Divided into Two Strata* **Wojciech Zieliński** and **Dominik Sieradzki** discuss the problem of estimating a proportion of objects with a particular attribute in a finite population. Classical estimator is compared with the estimator, which uses the information that the population is distributed among two strata. In the numerical example it was shown that variance of stratified estimator may be smaller by one-fourth compared to variance of classical estimator.

Remarks of the Estimation of Position Parameters by **Czesław Domański** concludes the issue. The author puts under reconsideration the classic problem of the level of accuracy of estimation of random variable parameter due to the lack of an unambiguous procedure to determine the scope of the distance between the value of an estimator and the real value of parameter. Some suggestions on how to deal with the situation when an obtained interval is too wide are provided.

Włodzimierz Okrasa

Editor