

## FROM THE EDITOR

Three kinds of papers included in this issue are arranged in a conventional way, starting with sampling methods and estimation issues followed by research papers and other articles, the last being a selection of papers from two conferences: *Multivariate Statistical Analysis 2015*, held in Łódź and *Classification and data analysis – theory and applications, 2015*, held in Gdańsk. Altogether eleven regular articles and a book review constitute this issue.

The first article, *Small Area Estimation of Income under Spatial SAR Model* by **J. Kubacki** and **A. Jędrzejczak** presents the method of hierarchical Bayes (HB) estimation under small area models with spatially correlated random effects and a spatial structure implied by the Simultaneous Autoregressive (SAR) process. Calculations are based on the concept of sampling from a posterior distribution under generalized linear mixed models implemented in WinBUGS software along with estimation for small areas by means of the HB method in the case of known model hyperparameters using data from household income data. A three-stage procedure which employs Balanced Repeated Replication, bootstrap and Generalized Variance Function, and additional simulations were conducted to show the influence of the spatial autoregression coefficient on the estimation error reduction. For high spatial correlation between domains, a noticeable MSE reduction showed that the HB-SAR method is in general more efficient than the traditional spatial EBLUP technique.

**R. Shanker's** paper *Sujatha distribution and its applications* proposes a new one-parameter lifetime distribution named “Sujatha Distribution” with an increasing hazard rate for modelling lifetime data. The author presents its first four moments and expressions for coefficient of variation, skewness, kurtosis and index of dispersion. Various mathematical and statistical properties of the proposed distribution are discussed including its hazard rate function, mean residual life function, stochastic ordering, mean deviations, Bonferroni and Lorenz curves, and stress-strength reliability. The applications and goodness of fit of the distribution is also discussed using three real lifetime data sets, and the fit has been compared with one-parameter lifetime distributions including Akash, Shanker, Lindley and exponential distributions.

The next paper, *Estimation of mean on the basis of conditional simple random sample* by **J. L. Wywiał**, discusses a problem of estimation of the population mean in a finite and fixed population on the basis of the conditional simple random sampling design dependent on order statistics (quantiles) of an auxiliary variable. The well-known Horvitz-Thompson and ratio type estimators as well as the sample mean are taken into account under the conditional simple random sampling designs. In conclusions following the results of empirical

analysis the author stresses that under some additional conditions the proposed estimation strategies based on the conditional simple random sample are in general more accurate than the mean from the simple random sample drawn without replacement. In particular, it can be expected that the sampling design might be useful in the case when there are censored observations of the auxiliary variable and in the presence of outliers.

The research papers section opens with an article by **N. M. Al-Kandari** and **P. Lahiri** on *Prediction of a function of misclassified binary data*. The authors start with an interesting observation that the naive predictor, which ignores the misclassification errors, is unbiased even if total misclassification error is high as long as the probabilities of false positives and false negatives are identical. Other than this case, the bias of the naive predictor depends on the misclassification distribution and the magnitude of the bias can be high in certain cases. They correct the bias of the naive predictor using a double sampling idea where both inaccurate and accurate measurements are taken on the binary variable for all the units of a sample drawn from the original data using a probability sampling scheme. Using this additional information and design-based sample survey theory, the authors derive a bias-corrected predictor and examine the cases where the new bias-corrected predictors can

also improve over the naive predictor in terms of mean square error (MSE). They conclude that given availability of additional data that generates the misclassification error, it may be possible to improve on the proposed method, and this direction they plan to explore in the future.

The next paper, *An extension of the classical distance correlation coefficient for multivariate functional data with applications* by **T. Górecki**, **M. Krzyśko**, **W. Ratajczak** and **W. Wolyński** is devoted to systematic exploration of the relationship between two sets of real variables defined for the same individuals which can be evaluated by a few different correlation coefficients. For the functional data there is one useful tool, canonical correlations. To extend other similar measures to the context of functional data analysis is not a simple task. The authors show how to use the distance correlation coefficient for a multivariate functional case. The approaches discussed are illustrated with an application to some socioeconomic data. The proposed method has been proved to be useful in investigating the correlation between two sets of variables, especially in the context of the hidden structure of the co-dependence between groups (pillars) of variables representing various fields of socio-economic activity.

**A. Szulc's** paper *Changing mortality distribution in developed countries from 1970 to 2010: looking at averages and beyond* starts with applying methods used in income inequality and poverty research to observe changes in life spans distribution in 35 developed countries. The analyses are performed at two levels, using the same methods when possible: i/ taking the countries as the units with a mean length of life being a single parameter representing the distribution, ii/ utilizing the country life tables (taking people as the units) in order to compare other than mean length of life attributes of mortality distribution. Increasing

divergence in the mean length of life across the countries is due to growing distance of the countries below the median, mainly the post-communist ones, to the upper half. The comparisons of the within-country distributions of ages at death by means of the Kullback-Leibler divergence provides similar results. However, poverty and inequality indices calculated at this level yield opposite conclusions. The author concludes that most of the between-country variation might be attributed to the variation in the mean length of life while the changes in within-country inequality reduced this effect.

The third group of articles – all of which are based on conference presentations - starts with *Quality of institutions and total factor productivity in the European Union* by **A. P. Balcerzak** and **M. B. Pietrzak**. The authors examine the relationship between the quality of an institutional system and total factor productivity in the EU countries. The quality of the institutional system is defined from the perspective of incentives that influence the use of the potential of KBE and the method for linear ordering of objects was applied in order to determine the level of effectiveness of the institutional system using data from Fraser Institute. The main hypothesis that the quality of the institutional system in the context of KBE has a significant influence on the level of total factor productivity in the EU was verified through estimation of the parameters of the Cobb-Douglas production function. The calculation was based on Eurostat data. In order to identify the relationship between the quality of the institutional system and the level of TFP a panel model was applied using data for years 2000-2010. In conclusions, the authors report that according to the results of the econometric analysis the expectations concerning the positive influence of the quality of institutions on TFP in the EU countries were confirmed. It was also showed that the new EU member states have high potentials for improving their TFP, given however implementation of effective institutional reforms.

In the paper *Locally regularized linear regression in the valuation of real estate*, **M. Kubus** discusses the valuation of real estate in the comparative approach using a data set with similar properties to the data from sales transactions, within a short period of time. Given the large scope of the data sets, a local regression model was preferred over a global model, and a local feature selection via regularization was employed. The empirical analysis confirmed the effectiveness of such an approach, with special attention being paid to the model quality assessment through cross-validation for estimation of the residual standard error.

**A. Kozera** and **R. Głowicka-Woloszyn** in the article *Spatial autocorrelation in assessment of financial self-sufficiency of communes of Wielkopolska province* employ global and local Moran I statistics using data from two publicly accessible databases - one compiled by the Ministry of Finance (*Indicators for the assessment of financial situation of self-government territorial units*) and the other by the Central Statistical Office (*Local Data Bank*). Calculations were performed in R with packages *spdep*, *maptools* and *shapefiles*. The study demonstrated that the communes of Wielkopolska province of comparable levels of financial self-sufficiency exhibited a moderate tendency to cluster. Clusters of

high levels gathered around larger urban centres, especially around Poznań, while clusters of low levels – in economically underdeveloped agricultural south-eastern and northern part of the province.

The next paper, *Kernel estimation of cumulative distribution function of a random variable with bounded support* by **A. Baszczyńska**, presents an attempt to reduce the so-called boundary effects, which appear in the estimation of certain functional characteristics of a random variable with bounded support. The methods of the cumulative distribution function estimation, in particular the kernel method, as well as the phenomenon of increased bias estimation in boundary region are discussed. Using simulation methods, the properties of the modified kernel estimator of the distribution function are investigated and an attempt to compare the classical and the modified estimators is made.

**S. Wanat, S. Śmiech, and M. Papież** in their article *In search of hedges and safe havens in global financial markets* explore three instrument classes: assets (represented by the S&P500 index), gold and oil prices, and dollar exchange rates. Weekly series of returns of all the instruments from the period January 1995 – June 2015 are analysed. The study is based on conditional correlations between the instruments in different market regimes obtained with the use of copula-DCC GARCH models. It is assumed that different market regimes will be identified by statistical clustering techniques; however, only conditional variances (without conditional covariances) will be taken into account. The author maintain that such an approach has a considerable advantage because it does not require to determine the number of market regimes as it is established by clustering quality measures. The methodology used in the paper makes it possible to treat the relations between instruments symmetrically. In conclusions, they stress that, according to the results obtained in their study, only dollar exchange rates can be treated as a (strong) hedge and a (strong) safe haven for other instruments, while gold and oil are a hedge for assets.

In the book review section, **J. Kordos** presents *Microeconometrics in Business Management* by **J. W. Wiśniewski**. Starting with remark that this book introduces the application of micro-econometric methods for modeling various aspects of economic activity for small- to large-sized enterprises, the reviewer highly recommend the book, especially to experts teaching business management. Basic models used in the modeling of the business (single-equation and multiple-equation systems) are introduced whilst a wide range of economic activity including major aspects of financial management, demand for labour, administrative staff and labour productivity is also explored. The book consists of Preface, Acknowledgments, six chapters which end with Conclusion and Bibliography.

**Włodzimierz Okrasa**

Editor