

FROM THE EDITOR

This issue of the *Statistics in Transition new series* appears at a very special moment in its history, for two seemingly unrelated reasons which coincide in time: the 25th anniversary of the journal, and 100th anniversary of the Central Statistical Office, renamed recently on Statistics Poland, which sponsors the publication of this international journal of the Polish Statistical Association. As it was announced in the previous issues, there will be a special topical stream within the upcoming 2nd Congress of Polish Statistics (July 10-12, 2018), envisaged as a way to celebrate this outstanding moment, and devoted to discussing the role of such a type of scientific (statistical) journals in promoting statistics as a discipline and as an instrument of creation and sustain community of specialists, and other stakeholders. Without falling into the tone of celebration, it seems worthwhile mentioning here the systematic progression of our journal in terms of its growing visibility in numerous international indexation bases, and of scores (impact factor) obtained from some of the most prestigious ones (for instance, above three times higher Scopus' CiteScore metrics for 2017 compared to 2016). We are totally aware of the fact that the primary source of such recognition was an increasing quality of the journal's articles. I would like to take this opportunity to express my deep appreciation to authors and peer-reviewers for their contributions to our joint efforts towards an excellence. Thanks to them the next quarter of a century of the *Statistics in Transition new series* looks optimistic and worth to support it.

The structure of this issue follows its regular thematic frame, i.e., it contains four sections, starting with *Sampling methods and estimation* and closing up with *Research Communicates and Letters*.

Carl-Erik Särndal's, Imbi Traat's and Kaur Lumiste's paper *Interaction between data collection and estimation phases in surveys with nonresponse* discusses approaches to deal with problems encountered in inference in surveys with nonresponse. The traditional focus on the estimation phase resulted in excelling some methods to reduce the nonresponse bias (propensity weighting and calibrated weighting) while the data collection phase has come into focus only recently. The authors take an integrated view where data collection and estimation are considered together. For a chosen auxiliary vector, they define the concepts incidence and inverse incidence and show their properties and relationship, showing that incidences are used in balancing the response in data collection; and that the inverse incidences are important for weighting adjustment in the estimation.

The paper by **Ceylan Talu Yozgatligil and H. Öztaş Ayhan**, *Univariate sample size determination by alternative components: issues on design efficiency for complex samples* is focused on the sample size determination taking into account some desired objectives: the level of confidence of estimates and the desired precision of the survey results, and the cost of enumeration.

Recently, some international organizations have been using univariate sample size determination approaches for their multivariate sample designs. These approaches also included some design efficiency and error statistics for the determination of the univariate sample sizes. They should be used for determining the survey quality measures after the data collection, not before. The additional components of the classical sample size measure will create selection and representation bias of survey estimates, which is discussed in this article.

Mir Subzar, Showkat Maqbool, Tariq Ahmad Raja, Surya Kant Pal and Prayas Sharma propose new estimators in the paper on *Efficient estimators of population mean using auxiliary information under simple random sampling*. An improved family of estimators for estimation of population mean is developed using the auxiliary information of median, quartile deviation, Gini's mean difference, Downton's Method, Probability Weighted Moments and their linear combinations with correlation coefficient and coefficient of variation. Their performance is analysed by mean square error and bias and compared with the existing estimators in the literature. By this comparison the authors go to conclusion that their proposed family of estimators is more efficient than the estimators offered in the literature. The theoretical results are supported by the empirical study.

O. Olawale Awe's and A. Adedayo Adepoju's article *Modified recursive Bayesian algorithm for estimating time-varying parameters in dynamic linear models* starts with observation that Estimation in Dynamic Linear Models (DLMs) with Fixed Parameters (FPs) has been faced with considerable limitations due to its inability to capture the dynamics of most time-varying phenomena in econometric studies. Since an attempt to overcome this limitation resulted in the use of Recursive Bayesian Algorithms (RBAs) - which also suffers from increased computational problems in estimating the Evolution Variance (EV) of the Time-Varying Parameters (TVPs) - the authors developed an alternative procedure. They propose a modified RBA for estimating TVPs in DLMs with reduced computational challenges.

In the next paper, *Generalized exponential type estimator of population mean in the presence of non-response*, **Siraj Muneer, Javid Shabbir and Alamgir Khalila** propose a class of generalized exponential type estimators to estimate the finite population mean using two auxiliary variables under non-response in simple random sampling. The proposed estimator under non-response in different situations has been studied and gives minimum mean square error as compared to all other considered estimators. Usual exponential ratio type estimator, exponential product type estimator and many more estimators are also identified from the proposed estimator. They use three real data sets to obtain the efficiencies of estimators.

The second section, *Research articles*, starts with **Mauro Mussini's** paper *On measuring polarization for ordinal data: an approach based on the decomposition of the Leti index*. The measurement of polarization for ordinal data - which occurs in the distribution of an ordinal variable - involves the decomposition of the Leti heterogeneity index. The ratio of the between-group component of the index to the within-group component is used to measure the degree of polarization for an ordinal variable. This polarization measure does not

require imposing cardinality on ordered categories to quantify the degree of polarization in the distribution of an ordinal variable. Author addresses the practical issue of identifying groups by using classification trees for ordinal variables. This tree-based approach uncovers the most homogeneous groups from observed data, discovering the patterns of polarization in a data driven way. An application to Italian survey data on self-reported health status is shown.

Ujjwal Das and **Nader Ebrahimi** in the paper entitled ***New method for covariate selection in Cox model*** undertake the problem of selection right predictors starting with discussion of the criterion of penalized regression, known as "least absolute shrinkage and selection operator" (LASSO). The LASSO regression involves a penalizing parameter (commonly denoted by λ) which controls the extent of penalty and hence plays a crucial role in identifying the right covariates. The authors propose an information theory-based method to determine the value of λ in association with the Cox proportional hazards model. Furthermore, an efficient algorithm is discussed in the same context. They demonstrate the usefulness of the proposed method through an extensive simulation study. Finally, the performance of it is compared with existing methods and the algorithm is illustrated using a real data set.

In **Wioletta Grzenda's** and **Ewa Frątczak's** article ***Cohort patterns of fertility in Poland based on staging process – generations 1930-1980*** addressed is the problem of unprecedented changes in the fertility. Currently, the total fertility rate level is very low, about 1.3 children per woman, which is below the replacement level. Many studies have described changes in fertility based on the cross-sectional approach. In the authors' view, the changes of cohort fertility have been described not quite sufficiently. Therefore, they attempt to fill in this gap by the assessment of stochastic fertility tables, calculated for five-year generations of women born in the period 1930-1980. The main goal of this study is to analyse changes in the cohort patterns of female fertility in Poland. During the transformation period in Poland the model of nuclear family changed from two-child model into one-child model, with a high percentage of childless families in the general structure. More recent analysis of 15 Central and East European (CEE) countries, including Poland, confirms such tendencies and shows that despite the growth in fertility rates in the late 2000s, the fertility still remains at a low level.

The section *Other articles* contains a paper based on presentation at the 2018 Multivariate Statistical Analysis Conference in Łódź by **Daniel Kosiorowski, Dominik Mielczarek, Jerzy P. Rydlewski and Małgorzata Snarska**, ***Generalized exponential smoothing in prediction of hierarchical time series***. The authors start with presentation of a grouped functional time series forecasting approach being a combination of individual forecasts obtained using the generalized least squares method. They modify the Shang-Hyndman methodology using a generalized exponential smoothing technique for the most disaggregates functional time series in order to obtain a more robust predictor. They discuss some properties of their proposals based on the results obtained via simulation studies and analysis of real data related to the prediction of demand for electricity in Australia (in 2016).

The final section, *Research Communicates and Letters*, contains two articles. In the first, by **Czesław Stępnia**k, entitled ***On a surprising result of two-candidate election forecast based on the first leadership time***, author presents "a simple but provocative note". He considers an election with two candidates, and under assumption that candidate A was the leader until counting n votes, and he asks the question "How to use this information in predicting the final results of the election?" According to the common belief the final number of votes for the leader should be a strictly increasing function of n . Assuming the votes are counted in random order, it is possible to derive the Maximum Likelihood predictor of the final number of votes for the future winner and loser based on the first leadership time. It appears that this time has little effect on the predicting. The first leadership time is informative for the final results of the election only in the trivial case.

In the next paper, ***The wellbeing effect of community development. Some measurement and modeling issues***, **Włodzimierz Okrasa** and **Dominik Rozkrut** discuss the interconnected methodological tasks, measurement and modeling, in the context of exploration of the cross-level interaction between the local community development and individual wellbeing. The preliminary results illustrate usefulness of an analytical framework aimed to assess an impact of the local development on individual wellbeing through multilevel modeling, accounting for spatial effects. To this aim, a dual measurement system is employed with data from two independent sources: (i) the Local Data Bank (LDB) for calculating a multidimensional index of local deprivation (MILD) and (ii) the Time Use Survey data to construct the U-index ('time of unpleasant state'), considered as a measure of individual wellbeing. Since one of the implications of the main hypothesis on the interaction between community development and individual wellbeing is the importance of 'place' and 'space', a special emphasize has been put on spatial effects, i.e. geographic clusters and spatial associations (autocorrelation, dependence). The evidence that place and space matter for this relationship provides support for validity of both multilevel and spatial approaches (ideally, combined) to this type of problems.

Włodzimierz Okrasa

Editor