STATISTICS IN TRANSITION-new series, August 2011
Vol. 12, No. 1, pp. 1—6

FROM THE EDITOR

A set of fourteen articles composing this Journal's issue is arranged in three major groups: comparative surveys - sampling methods and estimation - other articles. A new section is added, conferences, with information about two important international meetings that are under preparation - one organized by the Central Statistical Office of Poland in cooperation with the World Bank, other by the Polish Statistical Association which celebrates in 2012 its one hundredth anniversary.

The two papers included in the first section address similar aspects of comparative surveys in that they both share the focus on comparison of the proposed class of estimators; however, derived from data collected in different ways.

In ***Subsampling the non-respondents in cluster sampling on sampling on two successive occasions,*** **H. P. Singh** and **S. Kumar** discuss the problem of estimation of finite population mean for current occasion in the context of cluster sampling on two successive occasions when there is non-response on both the occasions. Estimators for the current occasion are derived as the particular case when there is non-response on first occasion and second occasion only. A comparison between variances of the estimates is studied, using Hansen and Hurwitz (1946) technique for estimation of population mean for current occasion in the context of cluster sampling over sampling on two successive occasions. Three different possible cases when there is non-response (i) on both the occasions, (ii) only on the first occasion, and (iii) only on the second occasion, are discussed. Authors recommend the approach they developed in this study for when there is a need to correct for non-response in cluster sampling over two occasions. They illustrate the performance of the proposed strategy empirically.

Some rotation patterns in two-phase sampling are analyzed by **G. N. Singh** and **S. Prasad,** focusing on the estimation of population mean on the current occasion using two-phase successive (rotation) sampling on two occasions has been considered. Two-phase ratio, regression and chain-type estimators for estimating the population mean on current (second) occasion have been proposed. Properties of the proposed estimators have been studied and their respective optimum replacement policies are discussed. Estimators are compared with the sample mean estimator, when there is no matching and the natural optimum estimator, which is a linear combination of the means of the matched and unmatched portions of the sample on the current occasion. Results are demonstrated through empirical means of comparison and suitable recommendations are made.

The following six papers compose the sampling methods and estimation section.

Paper by **B. B. Khare** and **R. R. Sinha,** ***Estimation of population mean using multi-auxiliary characters with subsampling the nonrespondents,*** is aimed at suggesting a class of two phase sampling estimators for population mean using multi-auxiliary characters in presence of non-response on study character. The expressions for bias, mean square error and the condition for attaining the minimum mean square error of the proposed class of estimators have been obtained, along with the optimum values of the size of first phase sample, second phase sample and the sub sampling fraction of non-responding group. Those have been determined for the fixed cost and for the specified precision. A comparison of the proposed class of estimators has been carried out with an empirical study.

N. A. Montes De Oca in ***Estimation of average income in Cuban municipalities*** applies a small area statistics approach to capture a new kind of spatial differentiation among the territorial units (municipalities) resulting from recently implemented policies towards increasing economic activity in the country (expanding international tourism and joint ventures, legalization of the possession of US dollars; permitting self-employment, agricultural markets etc.). The focus is on finding small area estimates which are more precise than the direct estimates of monthly mean income for people aged 15 and over at a municipal level (all the 169 Cuban municipalities are included). Though the empirical results obtained so far provide only a rough estimates (given the limited income-relevant data), the estimates are still more precise than the direct estimates for small areas/domains.

In the next article, ***Bootstrap method with calibration for standard error estimators of income poverty measures,*** **A. Zięba-Pietrzak, J. Kordos** and **R. Wieczorkowski** discuss the calibration approach in sample surveys in reference to the Eurostat recommended approach. The authors focus on the indicators of poverty and social exclusion as a major monitoring tool for policy purposes, and they use data from European Statistics on Income and Living Conditions (EU-SILC) for empirical illustration. Given complexity of the EU-SILC sample design, a system of weights for estimates of population parameters and approximate methods of standard error estimation needed to be developed. In the study, the McCarthy and Snowden (1985) bootstrap method for standard errors estimation of income poverty measures is employed. Subsequently, the reweighting of bootstrap weights was applied, and results of such calibration are discussed.

A. Hurairah in ***The beta Pareto distribution,*** introduce a generalization—referred to as the beta Pareto distribution, generated from the logit of a beta random variable. A comprehensive treatment of the mathematical properties of the beta Pareto distribution is given, including expressions for the *kth* moments of the distribution, variance, skewness, kurtosis, mean deviation about the mean, mean deviation about the median, Rényi and Shannon entropyies. Also, the estimation procedures by the methods of moments and maximum likelihood are

provided. It is shown that beta Pareto distributions are the most tractable of all the known distributions out of the certain class of distributions (as introduced by Jones (2004)). According to author's view, the approach presented in this paper can be used as a reference to obtain the corresponding results for other distributions belonging to such a type of generalization.

In the paper ***Robustness of the confidence interval for At-Risk-of-Poverty Rate,*** **W. Zieliński** discusses the problem of robustness of the confidence level of the confidence interval for binomial probability focusing on an earlier (Zieliński, 2009) introduced a nonparametric interval for a*t-risk-of-poverty rate*. Since it appeared that the confidence level of the interval depends on the underlying distribution of the income, for some distributions (e.g. lognormal, gamma, Pareto) the confidence level shown to be smaller than the nominal one. The question about the largest deviance from the nominal level is being extended by the above mentioned issue of robustness. The worst distribution is derived as well as the smallest true confidence level is calculated. Some asymptotic characteristics (sample size tends to infinity) are also specified.

Paper by **J. L. Wywiał,** ***Estimation of domain means on the basis of strategy dependent on depth function of auxiliary variables' distribution*** deals with the problem of estimation of a domain means in a finite and fixed population. It is assumed that observations of a multidimensional auxiliary variable are known in the population. The proposed estimation strategy consists of the Horvitz-Thompson estimator and the non-simple sampling design dependent on a synthetic auxiliary variable observations of which are equal to the values of a depth function of the auxiliary variable distribution. Both spherical and Mahalanobis depth functions are considered. A sampling design is proportionate to the maximal order statistic determined on the basis of the synthetic auxiliary variable observations in a simple sample drawn without replacement. A computer simulation analysis leads to the conclusion that the proposed estimation strategy is more accurate for domain means than the well known simple sample means.

In the *other articles* section included are papers presenting results of methodological investigation or (and) empirical analyses of policy-relevant issues, across a wide spectrum of realms.

In ***Remarks about the generalizations of the Fisher index,*** **J. Białek** discusses the Fisher index (defined as a geometric mean of Laspeyres and Paasche indexes), its theory and methodological characteristics, as the crossing of formulas and weights was considered the most suitable way to derive an "ideal" index formula. Despite that it satisfies most of the postulates of the axiomatic index theory, it has some limitations – e.g., the *time reversibility* test is satisfied only in some specific cases. In order to overcome them author presents a generalized Fisher index and develops some more general class of indexes (including the generalized Fisher index and other indices, like Laspeyres, Paasche, Fisher or Marhall-Edgeworth). As a pragmatic corollary to the employed strategy, author suggests that it is easier to prove that a given index belongs to a particular class than to verify that the axioms in question are satisfied. For

illustrative purposes, the general formula is applied to data from the whole time interval, with intention to show that it can be used in research, for instance, on pension or investment funds.

Paper by **J. Grosman** and **M. Kowerski,** ***Multiple–equation models of ordered dependent variables in exploration of the results of rehabilitation of locomotive organ disorders*** is focused on the factors determining patient's self-service during the admission and release from hospital . A two-equation model of ordered dependent variables is proposed, considered especially useful when the results of rehabilitation of locomotive organ disorders are not described by means of exact values obtained by mechanical measurements, but are described by means of qualitative valuation (ranking) made by a therapist, given that the distances between neighbouring ranks are not known. The advantages of the proposed model were presented on the basis of the results of estimation based on data of 4063 patients of hospitals from Mazowieckie and Warmińsko-Mazurskie provinces. This model allows for simulating the probabilities of the patient's self-service status both at the admission and at the release from a hospital, depending on various factors describing a patient.

The main objective of **T. Khan's** paper ***Identifying an appropriate forecasting model for forecasting total import of Bangladesh*** is to select an appropriate model for time series forecasting of total import (in taka crore) of Bangladesh. The presented study concerns mainly with seasonal autoregressive integrated moving average model (SARIMA), Holt-Winters' trend and seasonal model with seasonality modeled additively, and vector autoregressive model with some other relevant variables. An attempt was made to derive a unique and suitable forecasting model of total import of Bangladesh that would allow for making forecasts with minimum forecasting error. However, more research needs to be done to handle seasonality in a better way and to find more relevant variables that might be useful for forecasting total import of Bangladesh.

B. Liberda and **M. Pęczkowski** ask the question ***Does a change of occupation lead to higher earnings***? in their paper aimed at identifying whether and how the mobility between different types of broadly defined occupation - hired work, self-employment in industry, services and agriculture or social security beneficiaries – affects personal income of individuals. The Markov transition matrices are applied to the panel data from the Polish Household Budget Surveys (30,540 individuals), for years 2007-2008. According to the chief hypothesis, a change of occupation has an effect on individual capability to earn income, controlling for the occupation a person quits and the occupation a person starts, and for age, education level and a character of work (permanent or temporary). The hypothesis was tested using the regression analysis showing that the inter-occupational mobility matters mostly for those quitting hired work, for self-employment, for the better educated, and for respondents above 60 years of age.

In ***Neonatal mortality by gestational age and birth weight in Italy, 1998-2003: a record linkage study,*** **C. Marini** and **A. Nuccitelli** discuss neonatal

mortality rates by gestational age and birth weight category as particularly important indicators of maternal and child health and care quality. Since these specific rates have not been calculated in Italy since 1999, the main aim of their work is to assess the possibility of retrieving information on neonatal mortality by the linkage between records related to live births and records related to infant deaths within the first month of life, with reference to 2003 and 2004 birth cohorts. The focus is on some critical aspects of the most used record linkage approach as specific problems may arise from the choice of records to be linked if there are consistency constraints between pairs (in this context, one death record can be linked to at most one birth record). However, given the actual quality of the starting data, the retrieval of information on neonatal mortality by gestational age and birth weight is limited to Northern Italy. Specific neonatal mortality rates are provided with reference to 2003 and discussed with particular emphasis on quality issues in the data collection processes.

Włodzimierz OKRASA
Editor-in-Chief