STATISTICS IN TRANSITION new series and *SURVEY METHODOLOGY*
Joint Issue: Small Area Estimation 2014
Vol. 17, No. 1, pp. 3–6

FROM THE GUEST EDITORS (PART 2)

The second part of this Joint Issue of Statistics in Transition and Survey Methodology includes seven articles. These two issues have been split according to which guest editors have been looking after the articles. They are not necessarily sequenced according to the themes that appeared in the original Conference programme.

The first paper, by Erciulescu and Fuller, presents a small area procedure where the mean and variance of an auxiliary variable are subject to estimation error. They consider fixed and random specifications for these auxiliary variables. Their study was motivated by a situation where the sample used for small area estimation was a subsample of a larger survey. The larger survey furnished estimates of the distribution of the auxiliary variables. They demonstrate that efficiency gains associated with the random specification for the auxiliary variable measured with an error can be obtained. They propose a parametric bootstrap procedure for the mean squared error of the predictor based on a logit model. The resulting bootstrap procedure has a smaller bootstrap error than a classical double bootstrap procedure with the same number of samples.

The second paper, by Münnich, Burgard, Gabler, Ganninger and Kolb, develops a sampling design that can support accurate estimation for the 2011 German Census. In contrast to carrying out a classical census, a register-assisted census, using population register data and an additional sample, was implemented. The main objective of the census was to produce the total population counts at fairly low levels of geography. Ralf Münnich et al. provide an overview of how the sampling design recommendations were set up to fulfill legal requirements and to guarantee an optimal, yet flexible, source of information. Small area methods, as well as traditional methods, were used to produce these counts. Empirical results of the small area estimation are presented.

The next three papers present developments in small area estimation methodology and practical application in various fields of empirical research and statistics production, including poverty research and fisheries statistics. The first paper, by Guadarrama, Molina and J. N. K. Rao, provides a review on methods for the estimation of poverty indicators for small areas, including design-based direct estimation and a number of model-based small area estimation methods: the Fay-Herriot area level model, the World Bank poverty mapping method (the ELL method) and three Bayesian variants previously published by the authors. These are the empirical best/Bayes (EB) and hierarchical Bayes (HB) methods and a Census EB method providing an extension of the EB method. While the

Fay-Herriot method employs area-level data, the other methods require unit-level auxiliary information. The ELL, EB, Census EB and HB methods rely on statistical data infrastructures where access to unit-level records of population units taken for example from administrative registers and population censuses is available for research and statistics production. This option is becoming frequently met in an increasing number of countries and much of current small area research is conducted under this assumption. The list of advantages and disadvantages, reported for each of the methods, appears helpful for practitioners facing the challenge of choosing a small area method for a particular estimation task. Statistical properties (bias and accuracy) of methods are assessed empirically by model-based simulation experiments with unit-level synthetic data following a nested error model, throwing further light on the methodological summaries of the methods. Extensive simulation scenarios of varying complexity include informative sampling and a nested error model with outliers; these scenarios in particular are important for practical purposes. For practical application, it is important that also situations are considered where some of the underlying assumptions of the methods do not hold, which is often the case in practice. The conclusions drawn by the authors on the relative performance of the methods are useful for researchers and practitioners.

Because of its applicability in various data infrastructures, the Fay-Herriot model has been widely used in small area estimation purposes all over the world and new developments are often needed to extend the method for practical situations at hand. A robust hierarchical Bayesian approach for the Fay-Herriot area-level model is presented in the second paper, written by Chakraborty, Datta and Mandal. The starting point is the authors' observation on a possible poor performance of the standard Fay-Herriot area-level model in the presence of outliers. The new method is aimed for cases where extreme values are met for some of the random effects of small area means, causing problems in the standard Fay-Herriot procedure under normality assumptions of the random effects. The authors propose a two-component normal mixture model, which is based on noninformative priors on the model variance parameters, regression coefficients and the mixing probability. The method is aimed as an alternative to a scale mixture of normal distributions with known mixing distribution for the random effects. The authors apply their method to real data of US Census Bureau for poverty rate estimation at county level. The results indicate that probabilities of having large random effects are expected to be low for most areas but can be large for some areas, thus calling for attention to handle the possible heterogeneity of the data. Simulation studies based on artificially generated data are conducted to assess the performance of the proposed method against the standard Fay-Herriot model. In the first set of experiments, the authors verify the robustness of the proposed method to outliers in the cases considered. In further simulations, the authors show that their method tends to perform better than the Fay-Herriot method when the possibility of presence of outliers is high, and performs similarly in situations where outliers are not expected. In their concluding notes the authors provide a useful discussion on the possible causes of exceptionally large random effects for certain areas, calling for a careful specification of the linking model and the choice of the explanatory (auxiliary) variables.

The third paper, by Hernandez-Stumpfhauser, Breidt and Opsomer, provides a refinement of the Fay-Herriot approach for a particular small area estimation problem. The authors consider a practical problem of developing a new weighting procedure for a regular fisheries survey in the United States on recreational fishing in saltwater. For the estimation of the recreational catch, fishing catch per trip is estimated from one survey and the number of fishing trips from another survey. Data from these two surveys are combined to estimate recreational fishing catch in 17 US states. For weighting procedure, estimates are needed for the fraction of fishermen who leave the fishing site during a prespecified time interval on a selected day. The distribution of daily departure times is needed within spatio-temporal domains subdivided by mode of fishing. Direct estimates could be obtained but they are not sufficient because of a large number of estimation domains, causing very small (even zero) domain sample sizes. The authors develop a small area estimation solution based on the Fay-Herriot approach. More specifically, the authors show that with a certain hierarchical model formulation that is slightly more complex as the standard mixed model, fast and accurate model selection procedure based on variational/Laplace approximation to the posterior distribution can be implemented for the particular estimation problem considered. Even if the underlying linear mixed model can be complex involving fixed and random effects for the states, waves and fishing modes and interaction terms, the method can serve as a cost-effective alternative to the computationally more demanding MCMC sampler. By empirical comparison of MCMC and the proposed variational/Laplace approaches using real data, the authors show that the results are essentially identical, thus motivating the use of the method in practice.

The production of small area statistics by national statistical agencies and international statistical institutes is becoming more and more important for societal planning and evaluation and the allocation of public funds to regional areas and other population subgroups. In the next paper, Kordos presents a personal view on the development of certain aspects of small area estimation methodology and practice in the context of official statistics. The author first summarizes the main approaches in small area estimation with some historical remarks. He continues by discussing the important issue of the use of administrative records in official statistics production and as auxiliary information in the construction of estimators for various regional indicators. The author presents a summary of international conferences on small area estimation organized in past years, covering a period from 1985. Further, he presents a review of selected international small area estimation programs and research projects on small area estimation. A special property of these research activities is that they are conducted in cooperation with research communities on small area estimation and actors whose responsibility is in the production of official small area statistics. The interaction has proven fruitful in motivating ongoing research and development in small area estimation methodology and for boosting the implementation of methods in regular official statistics production. This aspect might well be taken as the main message of the paper by Kordos.

In many national statistical institutes, the design-based approach has offered the prevailing paradigm in official statistics production for decades. Good reasons are the ability of the approach to provide estimates having favorable statistical properties such as design-unbiasedness, which is often appreciated by the clients, and the availability of powerful statistical procedures and tools that use effectively the auxiliary information supplied in various forms. Calibration techniques and generalized regression estimation are examples of such methods. While relative standard errors of design-based estimates can be sufficiently small for population domains whose sample size is large, this is not necessarily the case for small domains. It is in this field of action where model-based small area estimation is challenging the design-based approach. In the final paper, Hidiroglou and Estevao present an empirical assessment of selected design-based methods against some existing model-based small area estimation methods, considered at Statistic Canada. Traditional design-based estimators include the Horvitz-Thompson estimator, two variants of calibration estimators and a modified regression estimator. A synthetic estimator and the standard EBLUP and its variant called pseudo-EBLUP represent model-based methods. The relative performance of the methods is assessed in design-based simulation experiments, where in addition to "ideal" conditions also misspecified models are considered. The relative performance of the methods differs depending on whether the model holds or not. Of the traditional design-based estimators, the domain-specific calibration estimator and the modified regression estimator indicate the best efficiency. The model-based small area estimators tend to outperform the design-based methods in efficiency, especially for small domains. As expected, the model-based methods can suffer from large design bias in cases where the model is misspecified.

Several persons (in addition to the Editor and Guest Editors) have served as reviewers of papers published in this thematic issue of the journal. We acknowledge the efforts of F. Jay Breidt, Isabel Molina, Domingo Morales, Ari Veijanen, Mamadou Diallo and Jon Rao: their encouraging and productive comments directly contributed to the quality of the papers.

Risto Lehtonen and **Graham Kalton**

Guest Editors