Preface to the papers on Small Area Estimation

Small area estimation (SAE) has been, and still predominantly is, a very fertile area in official and survey statistics research with important theoretical and applied contributions. In the last decades, an increasing number of National Statistical Institutes and other organisations across the world have recognised the importance of producing small area statistics and their potential use for informing policy decisions. Cutting edge developments in model-based small area methods are used in practice for the production of national statistics. Among many, examples of organisations with research interests in small area estimation include the US Bureau of Census, the UK Office for National Statistics, The World Bank, the Statistical Office of Italy, the Central Bureau of Statistics in Holland, Statistics Canada, the Australian Bureau of Statistics, the Brazilian Statistical Office, the National Council for the Evaluation of Social Development Policy (Consejo Nacional de Evaluacion de la Politica de Desarrollo Social) in Mexico and the Ministry for Social Development (Ministerio de Desarrollo Social) in Chile.

Over time, users' needs have surpassed the limits of what can be achieved with traditional SAE methods. For example, in addition to simple linear parameters like averages and proportions, users request the estimation of more complex indicators such as geographically disaggregated measures of deprivation and inequality. In addition, the availability of what is known as big data, e.g. satellite and mobile phone data, and probabilistically linked administrative and survey data, has created new methodological challenges. Meeting the increasing complexity of users' needs requires new specialised methodology and software that extends beyond conventional survey operations. This has created new research opportunities and the need for closer collaboration between researchers and practitioners for transferring research into practice and hence, maximising the impact of research.

The present themed papers in this issue include a selection on SAE. The call for papers was published at the first Latin American International Statistical Institute satellite meeting on small area estimation that took place at the Pontificia Universidad Católica de Chile in Santiago, Chile in August 2015. This conference was part of a series of scientific meetings devoted solely to small area estimation. Starting with the 2001 meeting in Maryland, USA, SAE conferences have taken place in Jyvaskyla, Finland (2005), Pisa, Italy (2007), Elche, Spain (2009), Rhine (river cruise) in Germany (2009), Trier, Germany (2011), Bangkok, Thailand (2013), Poznan, Poland (2014), Santiago, Chile (2015), Maastricht, Netherlands (2016) and Paris, France (2017). The next conference will take place in Shanghai, China, June 16-18, 2018.

The call for papers for this issue attracted many high quality submissions. All manuscripts went through the full peer review process of the journal after which a total of 11 papers were selected for publication. The selected manuscripts put forward new frequentist and Bayesian methodologies on a range of topics including robust and non-parametric methods, poverty mapping, measurement error models and time series models with new tools for model testing and selection also being proposed. The papers include applications in ecology, economics, health and medicine, transportation and education.

Summary of the papers

A major criticism of the use of model-based methods in survey estimation is their reliance on assumptions that are hard to check and satisfy in practice. On the other hand, in many applications, the use of models is deemed necessary for improving the precision of estimates. In recent years part of the small area literature has focused on developing small area methods that are robust to departures from the model assumptions (e.g. Ghosh et al., 2008; Chambers et al., 2014). Papers that appear in this special issue continue this tradition by considering extensions of popular small area models, which assume Gaussian distributions for the model error terms e.g., the Fay-Herriot model (Fay and Herriot, 1979). The extensions allow for alternative distributions, which are potentially more suitable for particular types of survey data like, for example, business survey data. Moura, Neves and Silva present small area models with skew-normal and skew-*t* distributions for skewed survey data, and apply the models to business surveys. Ferrante and Pacei extend a normal, multivariate Fay-Herriot area-level model by allowing the random effects and the sampling errors to have skew-normal distributions. The model is also applied to business survey data for estimating value added and labour costs. A different form of robustness against failure of the model assumptions is obtained by the use of non-parametric models. Wagner, Münnich, Hill, Stoffels and Udelhoven study the use of non-parametric small area models that use shape-constrained penalised B-splines and apply these models for estimating timber reserves in the German region of Rhineland-Palatinate.

Two of the papers that appear in this issue propose new methodologies for estimating nonlinear parameters in particular, income deprivation (poverty) and inequality indicators. This topic generated considerable debate in the small area literature. Marhuenda, Molina, Morales and Rao extend the methodology proposed in Molina and Rao (2010) by proposing an Empirical Best (EB) predictor under the two-fold nested error regression model and apply the methodology for estimating gender-specific poverty rates in counties of the Spanish region of Valencia. The methodology proposed in this paper makes the application of the EB methodology more realistic in situations where survey data are collected via multi-stage clustered designs. Dash and Chambers, on the other hand, focus on an alternative poverty mapping methodology that has been extensively used by the World Bank (Elbers et al., 2003) and propose robust Mean Squared Error estimators for poverty estimates produced by the use of this method.

The paper by Schmid, Bruckschen, Salvati and Zbiranski presents one of the first attempts to use big data sources -mobile data in their application- as covariate information in area level models.

The authors apply the model for deriving gender-specific, sub-national benchmarked literacy rates in Senegal. We expect that in the near future, incorporating such sources of data as covariates when producing small area estimates will become common practice. Despite the challenges and the open research questions, it is encouraging to see a methodology that enables us to do this.

Another area of research in small area estimation with renewed interest is whether or how to account for measurement errors in model covariates. This topic is of particular interest since the ease of access to regularly updated covariate information from large surveys is an important advantage compared to access to Census and administrative micro-data, which is difficult due to confidentiality constraints. Moreover, covariates derived from big data (e.g. mobile data), as in the paper by Schmid et al., are also likely to be affected by measurement error although quantification of the measurement error in this case is challenging. SAE methods must account for the measurement error in covariates obtained from survey data. In the present special issue, Arima, Bell, Datta, Franco and Liseo consider a multivariate Fay-Herriot model where the covariates are assumed to be subjected to observation errors. The authors develop Markov chain Monte Carlo methodology which is applied for estimating U.S. county level poverty rates for school-aged children. Similar in spirit but using a unit-level model is the paper by Maples, which presents a methodology that allows for the use of covariate data coming from a large independent survey. The methodology is applied for deriving small area estimates of disability.

A topic that until recently received relative little attention is model selection and testing. (e.g. Datta et al., 2011; Pfeffermann, 2013). Lombardia, Lopez-Vizcaino and Rueda propose a mixed generalised Akaike Information Criterion (AIC) for model selection. The method is compared to alternative methods, including the conditional AIC criterion, using simulations and real labour market and health data. On a related topic, Torkashvanda, Jafari Jozani and Torabi propose clustering of small areas based on the Euclidean distance between covariates, and propose a statistical test to investigate the homogeneity of between clusters variance components.

Finally, Bollineni-Balabay, van den Brakel, Palm, and Boonstra present a paper on another topic that has attracted interest in the small area estimation literature, namely borrowing strength over time. The paper compares state space models (estimated by use of the Kalman filter combined with a frequentist approach to hyper-parameter estimation), with multilevel time series models fitted under the hierarchical Bayesian framework. The application of the methods is to data collected in the Dutch Travel Survey, which has small sample sizes and discontinuities caused by survey redesigns.

We thank all the authors for their willingness to submit papers for this issue and for respondingto requests for revisions in a timely manner. We are particularly grateful to the Joint Editor foroverseeing our work, as Guest Associate Editors, with admirable patience. We also thank thereferees. Without their constructive comments this issue would not be possible.

Guest Editors
Nikos Tzavidis, University of Southampton
Li-Chun Zhang, University of Southampton and Statistics Norway
Danny Pfeffermann, Southampton and Hebrew Universities and Central Bureau of Statistics, Israel
Partha Lahiri, University of Maryland

References

Chambers, R., H. Chandra, N. Salvati, and N. Tzavidis (2014). Outlier robust small area estimation. *Journal of the Royal Statistical Society Series B 76 (1)*, 47–69.

Datta, G. S., P. Hall, and A. Mandal (2011). Model selection by testing for the presence of small-area effects, and application to area-level data. *Journal of the American Statistical Association 106 (493)*, 362–374.

Elbers, C., J. Lanjouw, and P. Lanjouw (2003). Micro-level estimation of poverty and inequality. *Econometrica 71 (1)*, 355–364.

Fay, R. E. and R. A. Herriot (1979). Estimation of income for small places: An application of james-stein procedures to census data. *Journal of the American Statistical Association 74 (366)*, 269–277.

Ghosh, M., T. Maiti, and A. Roy (2008). Influence functions and robust bayes and empirical bayes small area estimation. *Biometrika 95 (3)*, 573–585.

Molina, I. and J. N. K. Rao (2010). Small area estimation of poverty indicators. *The Canadian Journal of Statistics 38 (3)*, 369–385.

Pfeffermann, D. (2013). New important developments in small area estimation. *Statistical Science 28 (1)*, 40–68.