

Estimation for Small Areas and Domains: Model Assisted Methods

Risto Lehtonen
University of Helsinki
risto.lehtonen@helsinki.fi

Abstract

In the past few years, research in statistical methods for the estimation of statistics for small population subgroups or domains (with small domain sample size) has received considerable attention. In the literature, these methods fall under the headline of small area estimation (SAE). The paper by Pfeffermann (2013) provides an up-to-date review. Major European research projects in small area estimation have been implemented; examples are the EURAREA, AMELI and SAMPLE projects funded under the European Union Framework Programmes (FP). A series of international SAE conferences have been organized; the next will be held in Poland in 2014. SAE methods are being implemented in various application areas including for example social and economic statistics, health sciences and forestry inventories. Methods also have been investigated in the context of the latest German Population Census.

There are two major approaches in small area estimation: model-based methods and design-based methods. Lehtonen and Veijanen (2009) summarize developments in design-based SAE methodology and Datta (2009) covers model-based SAE. In both design-based and model-based SAE, models such as generalized linear mixed models (GLMM) and unit-level auxiliary information from the population play an important role. Model-based methods rely exclusively on the specified model and therefore, the estimators of domain parameters such as totals tend to be design biased, and the bias does not necessarily vanish with increasing the domain sample size. Design-based estimators for domain totals, on the other hand, are (nearly) design unbiased by the construction principle. But design-based estimators tend to have poor accuracy (measured with the MSE) in domains whose sample size is small or very small whereas model-based estimators can have good accuracy in such domains. However, the accuracy of model-based estimators can be poor if the bias component dominates the MSE. So, there are pros and cons in both approaches.

In the talk I discuss recent developments in design-based model-assisted methods for the estimation of totals for population domains. The methods include variants of generalized regression (GREG) estimators (Särndal, Swensson and Wretman 1992) aimed to the estimation of domain totals (Lehtonen, Särndal and Veijanen 2003 and 2005; Lehtonen and Veijanen 2009) and extensions of model calibration methods (Wu and Sitter 2001; Särndal 2007) to small area estimation (Lehtonen and Veijanen 2012). Assisting models include logistic fixed-effects and mixed models. Accuracy of model-assisted SAE methods is compared with certain model-based methods that use mixed models (Rao 2003). Availability of unit-level auxiliary information is assumed. Statistical properties (bias and accuracy) are investigated by Monte Carlo experiments using artificially generated data and real population data maintained by Statistics Finland.

References

- Datta, G. S. (2009). Model-based approach to small area estimation. Chapter 32 in *Sample Surveys: Inference and Analysis* (D. Pfeffermann and C. R. Rao, eds.). *Handbook of Statistics* 29B 251–288. North-Holland, Amsterdam.
- Lehtonen, R., Särndal, C.-E. and Veijanen, A. (2003). The effect of model choice in estimation for domains, including small domains. *Survey Methodology*, 29, 33–44.
- Lehtonen, R., Särndal, C.-E. and Veijanen, A. (2005). Does the model matter? Comparing model-assisted and model-dependent estimators of class frequencies for domains. *Statistics in Transition*, 7, 649–673.

- Lehtonen, R. and Veijanen, A. (2009). Design-based methods of estimation for domains and small areas. Chapter 31 in *Sample Surveys: Inference and Analysis* (D. Pfeffermann and C. R. Rao, eds.). *Handbook of Statistics* 29B 219–249. Amsterdam: Elsevier.
- Lehtonen, R. and Veijanen, A. (2012). Small area poverty estimation by model calibration. *Journal of the Indian Society of Agricultural Statistics* 66, 125–133.
- Pfeffermann, D. (2013). New important developments in small area estimation. *Statistical Science* 28, 40–68.
- Rao, J.N.K. (2003). *Small Area Estimation*. Hoboken: John Wiley & Sons, Inc.
- Särndal, C.-E. (2007). The calibration approach in survey theory and practice. *Survey Methodology* 33, 99–119.
- Särndal, C.-E., Swensson, B. and Wretman, J.H. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.
- Wu, C. and Sitter, R.R. (2001). A model-calibration approach to using complete auxiliary information from survey data. *Journal of the American Statistical Association*, 96, 185–193.