A Bayesian Reweighting Technique for Simulating Small Area Micropopulation Datasets in Australia

Dr Azizur Rahman
School of Computing and Mathematics
The Charles Sturt University
Boorooma Street, Wagga Wagga
NSW 2678, AUSTRALIA

Email: azrahman@csu.edu.au
Ph/Fx: +61 2 6933 4744

Abstract

Creating a reliable synthetic micropopulation dataset at small area level is challenging in the microsimulation modelling approach of small area estimation. Although a range of methods are in use for generating spatial microdata, none of these methods can consider the entire scenario of the overall population at small area level. As a result, a newly generated micropopulation dataset from those approaches often leads to inconsistent data for many small areas and to validate the outputs of model built on such a synthetic microdata is also difficult. In this paper, we propose a novel approach of Bayesian reweighting tools for generating synthetic spatial micropopulation data at statistical local area levels in Australia.

The new system uses MCMC simulation with a joint posterior density based iterative algorithm. Results demonstrate that the new method takes consideration of the complete scenario of micropopulation data units at small area level, and it can produce statistically reliable small area estimates and their variance estimations as well as credible intervals.

Brief description of the paper:

Context:
Policy makers want accurate estimates at the local level for making effective policy decisions on various issues such as housing stress, poverty, social disadvantage, health, and various assistance programs. Small area statistics can be produced by small area estimation (SAE) techniques. Nowadays indirect modelling approaches of SAE such as spatial microsimulation modelling (SMM) have received much attention (Rahman, 2008a). Typically, indirect SAE is the process of using statistical models and/or geographic models to link survey outcomes to a set of predictor variables known for small areas, in order to predict small area-level estimates (Rahman et al. 2010). However, building a spatial microsimulation model is very difficult for many reasons. The creation of reliable spatial microdata is still challenging due to a lack of mathematically sound reweighting algorithms (Rahman, 2011). In addition, there has not been much research on testing the statistical significance of the microsimulation model

outputs yet, and deriving estimates of how reliable these outputs may be (Rahman and Harding 2012; Rahman et al. 2013). Thus, a new reweighting tool for generating synthetic spatial micropopulation data at small area level is crucial.

Basic theory and significance:

As for any area being sampled, a finite population usually has two parts - observed units in the sample called data and unobserved sampling units in the population. The main challenge we found in this approach of microdata simulation is to establish the linkage of observed data to the unobserved sampling units in the small area population. Essentially it is a prediction problem, where a modeller tries to find a probability distribution of unobserved responses using the observed sample and the auxiliary data. The Bayesian methodology (see, Ericson 1969; Lo 1986; Little 2007; Aitkin 2008; Rahman 2008b) can deal with such a prediction problem. The new method is principally based on the Bayesian prediction theory and Markov Chain Monte Carlo (MCMC) method of iteration.

This new microdata simulation technique is a probabilistic approach which is quite different from the deterministic approach used in GREGWT and the intelligent searching tool *simulated annealing* used in combinatorial optimisation. The method can adopt the generalised regression model operated in the GREGWT algorithm to link observed units in the sample and unobserved units in the population. In contrast, from the viewpoint of the combinatorial optimisation reweighting method, this new system uses the MCMC simulation with a posterior density based iterative algorithm. As the Bayesian joint posterior probabilities of the parameters for the observed sample units and unobserved population units are estimated through MCMC method, the proposed microdata simulation methodology is somewhat linked with a chain Monte Carlo sampling. However, it is rather different from the multiple imputation technique advanced by Rubin (1987) and other researchers. The basic computation process of the new proposed approach is predominantly associated with a prediction distribution of unobserved population units given the sample units.

References:

Aitkin, M. 2008, Applications of the Bayesian bootstrap in finite population inference, *Journal of Official Statistics*, vol. 24, no. 1, pp. 21-51.

Ericson, W.A. 1969, 'Subjective Bayesian models in sampling finite populations ', *Journal of the Royal Statistical Society: Series B*, vol. 31, no. 2, pp. 195-233.

Little, R. 2007, An objective Bayesian view of survey weights, *O'Bayes 07*, http://3w.eco.uniroma1.it/OB07/papers/little.ppt, Accessed 27 June 2008.

Lo, A.Y. 1986, 'Bayesian statistical inference for sampling a finite population', *The Annals of Statistics*, vol. 14, no. 3, pp. 1226-1233.

Rahman, A. 2008a, A review of small area estimation problems and methodological developments, *Online Discussion Paper - DP66*, NATSEM, University of Canberra.

Rahman, A. 2008b, *Bayesian predictive inference for some linear models under Student-t errors*, Saarbrucken, VDM Verlag.

Rahman, A. and Harding, A. (2012), A new analysis of the characteristics of households in housing stress: results and tools for validation, *Paper presented at the 6th Australasian Housing Researchers' Conference 2012 (AHRC12)* The University of Adelaide, Adelaide, South Australia, pp. 1-23 (February 8 – 10);

Rahman, A., 2011. Small area housing stress estimation in Australia: Microsimulation modelling and statistical reliability, Ph.D. Thesis, University of Canberra, Canberra.

Rahman, A., Harding, A., Tanton, R. and Liu, S. (2010), "Methodological issues in spatial microsimulation modelling for small area estimation", *The International Journal of Microsimulation* 3(2), pp. 3-22.

Rahman, A., Harding, A., Tanton, R. and Liu, S. (2013), "Simulating the characteristics of populations at the small area level: New validation techniques for a spatial microsimulation model in Australia", *Computational Statistics & Data Analysis*, 57 (1), pp. 149-165.

Rubin, D.B. 1987, *Multiple imputation for nonresponse in surveys*, New York, John Wiley and Sons.