Small Area Estimation in the Presence of Constraints

A dissertation submitted for the Degree of Doctor of Philosophy

R.A.B. Abeygunawardana

Faculty of Science

University of Colombo

January, 2019

Abstract

Surveys are traditionally used for estimating population parameters. These surveys are mainly focused on estimating parameters for relatively large populations. These estimates are used for representing not only the whole population but also subpopulations such as sub-areas or socio-economic groups in the main population. Although the overall sample is representative of the whole population, there may not be sufficient number of elements in the sample that represent various subpopulations. A domain or subpopulation is called a small area, if the domain specific sample is small. There is a huge demand from public and private sectors to have precise and accurate estimates for small areas in the populations. The requirements of estimates for small areas may occur once the survey has been conducted. Although a census can be conducted to get more accurate estimates for any small areas it is costly and time consuming. Also, it is not possible to conduct surveys every time to get all necessary estimates. Small Area Estimation (SAE) techniques play the main role to provide more precise estimates for small areas. SAE uses available auxiliary information to improve the precision of estimates for small areas.

SAE has two main branches called direct and indirect estimation. Direct estimation uses design based or domain based approaches while indirect estimation use implicit or explicit models to get precise estimates. Implicit models commonly use traditional demographic models or indirect domain estimates. However explicit models use basic area level or basic unit level models. In this research new methodologies are developed for the basic unit level models. There are two approaches used in the field of SAE such as frequentist approach and Bayesian approach. Among them, in this study frequentist approach is used to develop new methodologies for SAEs. Linear Mixed Models (LMM) is commonly used to get the fixed and random effects into the model. In the SAE, the effect due to subpopulations or small areas is considered as the random effect while the effect due to the auxiliary variables is considered as the fixed effect. A commonly used SAE model which is a special case of LMM known as components of variance model or nested error regression model is considered in this study. Researchers have developed various methodologies to obtain parameters of LMM to have improved SAEs using least squares method, maximum likelihood method, Bayesian methods and other methods. Constrained estimation techniques are rarely used to obtain the model parameters. However there may be situations where prior knowledge on model parameters are available. The prior knowledge may be based on past statistical information or some existing theory. There is a research gap in the use of prior knowledge in SAEs. This study focuses on least squares method with constraints to get model parameters.

Equality Constrained Least Squares (ECLS) estimates and Inequality Constrained Least Squares (ICLS) estimates are used to obtain the estimates for the parameters in the model. Then these new estimates are used to get improved statistics for small areas. The improvements of estimates are mainly calculated using the Mean Squared Errors (MSE) of the predictors. In this study the following improvement of the small area estimates were achieved.

1. New univariate SAE with equality and inequality constraints were proposed.
2. New bivariate SAE with equality and inequality constraints were proposed.
3. New general classes of predictors of SAEs with ECLS and ICLS estimates were proposed.
4. New estimates were theoretically compared with existing results using MSE.
5. ECLS and ICLS estimates and MSEs of SAEs are numerically compared using a data set.
6. Basic properties such as unbiasedness and MSE of new small area statistics are studied.