Statistica Sinica **35** (2025), 1-23
doi:https://doi.org/10.5705/ss.202023.0045

SMALL AREA ESTIMATION USING EBLUPS UNDER THE NESTED ERROR REGRESSION MODEL

Ziyang Lyu* and A.H. Welsh

University of New South Wales and Australian National University

Abstract: Estimating characteristics of domains (referred to as small areas) within a population from a sample survey of the population is an important problem in survey statistics. In this paper, we consider model-based small area estimation under the nested error regression model. We discuss the construction of mixed model estimators (empirical best linear unbiased predictors, EBLUPs) of small area means and the conditional linear predictors of small area means. Under the asymptotic framework of increasing numbers of small areas and increasing numbers of units in each area, we establish asymptotic linearity results and central limit theorems for these estimators which allow us to establish asymptotic equivalences between estimators, approximate their sampling distributions, obtain simple expressions for and construct simple estimators of their asymptotic mean squared errors, and justify asymptotic prediction intervals. We present model-based simulations that show that in quite small, finite samples, our mean squared error estimator performs as well or better than the widely-used Prasad and Rao (1990) type estimators and is much simpler, so is easier to interpret. We also carry out a design-based simulation using real data on consumer expenditure on fresh milk products to explore the design-based properties of the mixed model estimators. We explain and interpret some surprising simulation results through analysis of the population and further design-based simulations that highlight important differences between the model- and design-based properties of mixed model estimators in small area estimation.

Key words and phrases: Increasing area size asymptotics, indirect estimator, mean squared error estimation, mixed model estimator, model-based prediction, prediction intervals

1. Introduction

Estimates of characteristics such as means and totals for areas, domains or clusters within a population (all referred to as areas) obtained from sample survey data are widely used for resource allocation in social, education and environmental programs, and as the basis for commercial decisions. Direct estimates which use only data specific to an area, can have large standard errors because of relatively small area-specific sample sizes. Small area estimation is concerned with producing more reliable estimates with valid measures of

*Corresponding author. E-mail: ziyang.lyu@unsw.edu.au