Statistica Sinica **35** (2025), 1-22
doi:https://doi.org/10.5705/ss.202023.0098

AN UNBIASED PREDICTOR FOR SKEWED RESPONSE VARIABLE WITH MEASUREMENT ERROR IN COVARIATE

Sepideh Mosaferi*, Malay Ghosh and Shonosuke Sugasawa

University of Massachusetts Amherst, University of Florida and Keio University

Abstract: We introduce a new small area predictor when the Fay-Herriot normal error model is fitted to a logarithmically transformed response variable, and the covariate is measured with error. This framework has been previously studied by Mosaferi, Ghosh and Steorts (2023). The empirical predictor given in their manuscript cannot perform uniformly better than the direct estimator. Our proposed predictor in this manuscript is unbiased and can perform uniformly better than the one proposed in Mosaferi, Ghosh and Steorts (2023). We derive an approximation of the mean squared error (MSE) for the predictor. The prediction intervals based on the MSE suffer from coverage problems. Thus, we propose a non-parametric bootstrap prediction interval which is more accurate. This problem is of great interest in small area applications since statistical agencies and agricultural surveys are often asked to produce estimates of right skewed variables with covariates measured with errors. With Monte Carlo simulation studies and two Census Bureau's data sets, we demonstrate the superiority of our proposed methodology.

Key words and phrases: Bayes estimator, prediction interval, transformation.

1. Introduction

Small area estimation concerns producing estimates or predictions of means, totals or quantiles for each of a finite collection of geographic regions, where there are a small number of sampled units in each individual region (area). Classical models used in small area estimation take the form of mixed linear models that result from the concatenation of a model for error in direct sample-based estimators for each area and an additional model that connects areas through the use of covariates and area-specific random effects.

These *linking models* take the direct estimators to be linear combinations of covariates and random effects. We focus here on what is called the *area level model* (Ghosh and Rao, 1994; Pfefferman, 2013; Rao and Molina, 2015, Chap. 4) which uses covariates at the level of the areas. Recently, Mosaferi, Ghosh and Steorts (2023) proposed a model of the Fay-Herriot type and developed an empirical predictor for small area quantities that they are right skewed.

*Corresponding author. E-mail: smosaferi@umass.edu