

IMPROVED MODEL-ASSISTED ESTIMATION VIA PROBABILITY THRESHOLDING

Xianpeng Zong, Geoffrey K. F. Tso and Guohua Zou*

*Beijing University of Technology, City University of Hong Kong
and Capital Normal University*

Abstract: In survey sampling, model-assisted approach is often used to improve the precision of survey estimators when auxiliary information is available. Generally, the model-assisted estimators are nonlinear functions of some classical Horvitz–Thompson estimators constructed via inverse probability weighting, which are seriously affected by the heterogeneous inclusion probabilities. In this paper, we improve the classical model-assisted estimation via probability thresholding, and propose the improved linear and nonparametric model-assisted estimators for finite populations. The proposed estimators are shown to be asymptotically design unbiased and design consistent. The corresponding design mean squared errors and their estimators are also derived. We theoretically prove that the new model-assisted estimators are asymptotically not worse than the commonly used model-assisted estimators. Two simulation examples and an empirical application indicate good finite sample performance of the proposed estimators.

Key words and phrases: Horvitz–Thompson estimator, model-assisted estimation, probability thresholding, superpopulation model, survey sampling.

1. Introduction

In survey sampling, three kinds of frameworks are used to make statistical estimation and inference: design-based approach, model-based approach, and model-assisted approach. Generally, the last one is better than the first two when auxiliary information is available. Model-assisted method provides a convenient framework that uses a superpopulation model to describe the relationship between the variable of interest and the auxiliary variables. A model-assisted estimator improves the precision of the traditional survey estimators when the model is correct, and maintains desirable properties such as asymptotic design unbiasedness and design consistency when the model is incorrect. In the past decades, various superpopulations are considered. For instance, Särndal, Swenson and Wretman (1992) detailed linear model-assisted estimation that assumes the superpopulation models are ratio or linear models. Based on the hypothesis of nonparametric model, Breidt and Opsomer (2000) used local polynomial regression to construct a nonparametric model-assisted estimator. Breidt et al. (2007)

*Corresponding author. E-mail: ghzou@amss.ac.cn