

Discussion of “Small area estimation: its evolution in five decades”, by Malay Ghosh

Yan Li¹

Prof. Ghosh leads us step gradually into the realm of small area estimation (SAE) through the evolution of SAE for the past five decades, introducing various SAE methods of synthetic estimators, composite estimators, and model-based estimators for small area parameters, mean squared error approximations, adjustment methods of benchmarking and transformation, etc. The paper broadens and deepens our understanding of different perspectives of the SAE and provides a few illustrative real-life applications. It is a great review paper for general audience, especially for our graduate students in survey statistics and related areas, who wish to have a snapshot of the SAE research.

Prof. Ghosh focuses his review on the inferential aspects of the two celebrated small area models ----- the Fay-Herriot (FH) area model and the unit level nested error regression (NER) model. In the implementation of these models, variable selection plays a vital role and my discussion centers around this topic, which complements Professor Ghosh’s paper.

There is a vast literature on variable selection, a subtopic of model selection. We refer to the Institute of Mathematical Statistics Monograph edited by Lahiri (2001) for different approaches and issues in model selection and the book by Jiang and Nguyen (2015) for model selection methodology especially designed for mixed models. Variable selection methods for general linear mixed model can be, of course, applied to select variables for the FH and NER models as they are special cases of the general linear mixed model. Many data analysts not familiar with mixed models, however, use software meant for linear regression models to select variables. This approach may result in loss of efficiency in variable selection. Lahiri and Suntornchost (2015) and Li and Lahiri (2019) proposed simple adjustment methods so that the data users can select reasonable models by calculating their favorite variable selection criteria, such as AIC, BIC, Mallows’s C_p , and adjusted R^2 , which are developed for standard linear regression model assuming independent identically distributed (*iid*) errors. The goal of the two

¹ Joint Program in Survey Methodology and Department of Epidemiology and Biostatistics, University of Maryland, College Park, USA. E-mail: yli6@umd.edu. ORCID: <https://orcid.org/0000-0001-8241-7464>.