Robust Small Area Estimation: Methods, Theory, Applications, and Open Problems

Sylvia Harmening

Department of Statistics and Econometrics, Otto Friedrich University Bamberg, Germany,
sylvia.harmening@uni-bamberg.de

Small area estimation (SAE) has become increasingly important in both research and practical applications. A small area refers either to a geographic area or to a subpopulation for which direct estimates are unreliable because of limited domain-sample sizes. SAE addresses this issue by borrowing strength - that is, by using auxiliary information through statistical modeling to improve estimation accuracy. However, this improvement involves a trade-off, which forms the central theme of the book. The authors address a crucial gap between classical SAE models and the growing demand for reliable estimation in the presence of model misspecification, outliers, and complex data structures. Intended for researchers and graduate students in statistics, data science, and related fields, the book also offers practical guidance for practitioners, including those working in government and public sector organizations.

The book comprises seven well-structured chapters that progress logically from fundamental concepts to cutting-edge developments. Each chapter reinforces its theoretical discussions with illustrative examples, simulation experiments, or case studies based on real data.

Chapter 1 'Small Area Estimation: A Brief Overview' presents the motivation and fundamental ideas behind SAE. It outlines key estimation strategies, including direct and indirect estimation methods such as the Fay-Herriot model and the nested-error regression (NER) model, and provides an overview of available software packages.

Chapter 2 'SAE Methods Built on Weaker Assumptions' examines how SAE methods and mean squared prediction error (MSPE) estimation can be developed with fewer or less restrictive statistical assumptions, making them more robust to model misspecification. It introduces techniques such as the robust empirical Bayes estimator, the regression average, non-Gaussian mixed models and heteroscedastic NER models. The chapter also includes simulation studies and practical examples, such as grape production and income data.

Chapter 3 'Outlier Robustness' explores methods that make SAE robust to outliers. Robust techniques like the robust EBLUP, M-quantile regression, and density power divergence, which reduce outlier influence while keeping efficiency, are introduced. Detecting and adjusting for outliers to improve prediction accuracy are also covered.

Chapter 4 'Observed Best Prediction' introduces a method designed to make SAE more robust to model misspecification. Unlike traditional EBLUP, observed best prediction (OBP) estimates model parameters by minimizing the observed mean squared prediction error, giving more weight to areas with high sampling variance leading to predictions that remain reliable even when the assumed model is partly wrong. Moreover, the observed best selective predictor (OBSP), which combines variable se-

lection and parameter estimation, and the compromised best predictor (CBP), which can be described as a weighted average of the EBLUP and the OBP, are explored.

Chapter 5 'More Flexible Models' addresses advanced SAE methods that use semi-parametric, non-parametric, and functional models to reduce dependence on parametric assumptions. These models employ tools like splines, kernel functions, and functional mixed-effects models to capture complex, nonlinear relationships between variables. By allowing model flexibility, they improve robustness against model misspecification and perform better even for time-series data.

Chapter 6 'Model Selection and Diagnostics' discusses how to choose and validate models in SAE. It reviews classical and modern selection tools such as information criteria, fence methods, and shrinkage selection. The chapter also explains diagnostic techniques like (robust) goodness-of-fit tests and the tailoring method to detect violations of the model assumptions.

Chapter 7 'Other Topics' discusses several additional topics connected to robust SAE. It covers benchmarking, Bayesian, and machine learning methods (like mixed-effects random forests, neural networks, and gradient boosting), as well as approaches for handling missing data and classified mixed model prediction. The chapter concludes by discussing new challenges - such as Big Data, data quality, and privacy protection (differential privacy) - and calls for future SAE methods that remain robust in modern data environments.

The book is an impressive and timely contribution to the literature on SAE, skillfully combining a deep theoretical framework, modern methodological advances, and practical insights for real-world applications. A notable strength of the book lies in its integration of theory and applications. The authors devote substantial attention to real-world examples, including data on income and poverty, agricultural yields, and health indicators. Although the book is mainly aimed at researchers and graduate students, this hands-on approach is particularly valuable for practitioners, helping them understand and adopt robust SAE methods with greater ease.