ABSTRACT

Title of Dissertation: Robust and Flexible Methods for Small Area Estimation
Yuting Chen
Doctor of Philosophy, 2025

Dissertation Supervised by: Professor Partha Lahiri
Joint Program in Survey Methodology

Sample surveys are widely used to provide estimates for both the overall population and various subpopulations, known as domains, which can be defined by geographic or socio-demographic characteristics. Direct estimators rely solely on domain-specific sample data and are typically design-based, incorporating survey weights and relying on the probability distribution induced by the sampling design for making inferences. Although the total sample size in a survey is typically large, the sample size for specific domains may be small or even zero. When a domain-specific sample is too small to produce direct estimates with adequate precision, the domain is classified as a small domain or small area.

The increasing demand for small area statistics has driven the development of small area estimation (SAE) techniques, which produce reliable estimates for domains with limited or no sample data. This dissertation focuses on enhancing the robustness and flexibility of model-based SAE methodologies by addressing three key challenges: model misspecification, flexible modeling, and uncertainty quantification.

The first study examines the effects of model misspecification on several commonly used small area estimators. The results show that when the underlying model is misspecified, the observed best prediction (OBP) method does not consistently outperform the Empirical Best Linear Unbiased Predictor (EBLUP) in terms of the design-based mean squared prediction error (MSPE), even though OBP being designed to improve design-based MSPE over EBLUP under model misspecification. Both analytical and numerical evidence are provided to show that OBP performs better when using aggregated auxiliary variables compared to using the individual ones. It offers practical insights for handling model misspecification in small area estimation.

The second study develops a framework for predicting complex small area characteristics, which are often nonlinear functions of the study variable for population units, using a nested error regression model with high-dimensional parameters. This study addresses multiple challenges simultaneously. First, it allows both regression coefficients and sampling variances to vary across areas, accommodating heterogeneity and enhancing modeling robustness. Second, we propose a novel algorithm for estimating area-specific model parameters, improving computational efficiency compared to existing algorithms. Third, we introduce a new approach for producing area-specific poverty estimates for out-of-sample areas, yielding less synthetic estimates than existing methods. Design-based simulation studies demonstrate that the proposed method outperforms existing approaches in terms of relative bias and relative root mean squared prediction error. Additionally, the method is applied to household survey data from the 2002 Albania Living Standards Measurement Survey to estimate poverty indicators for Albanian municipalities.

A measurement of any quantity of interest is complete only when accompanied by an evaluation of its uncertainty. The third study advances the theory of parametric bootstrap methods for constructing highly efficient empirical best linear (EBL) prediction intervals for small area

means, incorporating both fixed and random effects. We analytically demonstrate that even when the normality assumption for random effects is relaxed, the proposed EBL prediction interval maintains a second-order coverage error, provided a pivot exists for a suitably standardized random effect when hyperparameters are known. In the absence of a pivot, we find that the order of coverage error of the parametric bootstrap EBL prediction interval is $O(m^{-1})$, and the first-order term is theoretically positive under certain conditions, indicating possible overcoverage of the EBL prediction interval. This characteristic may be advantageous for practitioners who do not account for other properties of prediction intervals. Furthermore, we propose a novel double bootstrap method, which can correct coverage issues in general. Monte Carlo simulations indicate that the proposed single bootstrap method performs well compared to alternative approaches.

Overall, this dissertation provides valuable insights into critical challenges in small area estimation, specifically in model misspecification, flexible modeling, and uncertainty quantification. Future research should explore semi-parametric and nonparametric methods to further enhance the robustness of inference for small areas.

Robust and Flexible Methods for Small Area Estimation

by

Yuting Chen

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2025

Advisory Committee:
Professor Partha Lahiri, Chair/Advisor
Professor Cinzia Cirillo, Dean Representative
Professor Snigdhansu Chatterjee
Professor Brian Kim
Professor Yan Li
Professor Nicola Salvati

Dedication

To my parents, Manxiu Zhou and Xiaoshan Chen.

Acknowledgments

I would like to express my heartfelt gratitude to everyone who has supported and helped me throughout the completion of my dissertation. Without their encouragement, this challenging work would not have been possible.

First and foremost, I owe my deepest thanks to my advisor, Prof. Partha Lahiri. His mentorship has gone beyond the duration of my PhD journey and will continue to guide me throughout my life. As an advisor, he has been consistently positive, encouraging and supportive, which has been invaluable to me as a graduate student facing challenges, big and small. Over the past five years, he has consistently been available for help and advice, offering continuous guidance in my research with his vast knowledge and insightful perspectives. I am profoundly grateful for the time, patience, and care he has devoted to mentoring me, and for connecting me with other researchers, which will undoubtedly benefit my future academic endeavors.

I also wish to express my sincere gratitude to Prof. Jiming Jiang, Prof. Xiaohui Liu, and Prof. Thuan Nguyen. They were the inspiration behind my decision to pursue a PhD in the United States. Their encouragement and support made it possible for me, a student from a small rural village, to imagine the opportunity of pursuing higher education in a foreign country. Their generous guidance on my research and career choices has been invaluable. Without their support, I would not have reached this point in my journey.

I also thank the other members of my committee. Prof. Cinzia Cirillo, Prof. Snigdhansu

Chatterjee, Prof. Brian Kim, Prof. Yan Li, and Prof. Nicola Salvati. Their time and effort in reviewing my work, coupled with their thoughtful and constructive feedback, have been invaluable in shaping this dissertation.

I also express my sincere gratitude to my research collaborator, Dr. Masayo Hirose. During her time at JPSM as a visiting scholar, she has always been generous with her time. There has never been an occasion where I have knocked on her door and she has not given me time. It has been a privilege to work with and learn from such an exceptional scholar.

I am deeply grateful to my parents, whose unwavering support has been a constant source of strength. Although they may not fully understand higher education or life in the United States, they never once discouraged me from pursuing my dreams, no matter how far they took me. My gratitude for them is beyond words.

I also extend my heartfelt thanks to my friends, both within the University of Maryland community and beyond. Thank you for listening to my frustrations, sharing in my achievements, and always being there to celebrate every step of this journey, no matter how big or small. Your support has been immeasurable.

I sincerely appreciate the JPSM staff for creating a welcoming and supportive environment, providing opportunities for growth, and offering their invaluable assistance throughout my time there.

I am also grateful to the iSchool for supporting me as a teaching assistant for four years.

Lastly, I want to express my deepest thanks to my best friend and partner, Hanqing Li. Although he may not always be physically present, his unwavering affection and support make me feel as if he is always by my side. His love has been a guiding force throughout this journey.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ARML	Adjusted Residual Maximum Likelihood
BP	Best Predictor
BLP	Best Linear Predictor
BLUP	Best Linear Unbiased Prediction
BPE	Best Predictive Estimator
CV	Coefficient of Variation
EA	Enumeration Area
EB	Empirical Best
EBL	Empirical Best Linear
EBP	Empirical Best Predictor
EBLUP	Empirical Best Linear Unbiased Predictor
ECDF	Empirical Cumulative Density Function
EFF	Efficiency
ELB	Empirical Linear Best
EUSILC	European Survey on Income and Living Conditions
FGT	Foster–Greer–Thorbecke
HCR	Headcount Ratio
LSMS	Living Standards Measurement Survey
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimator
MSPE	Mean Squared Prediction Error
MSE	Mean Squared Error
NER	Nested Error Regression
NERHDP	Nested Error Regression Model with High-Dimensional Parameter
OBP	Observed Best Prediction

ONS	Office for National Statistics
PG	Poverty Gap
PSU	Primary Sampling Units
RB	Relative Bias
REML	Restricted Maximum Likelihood
RRMSPE	Relative Root Mean Squared Prediction Error
SAIPE	Small Area Income and Poverty Estimates
SAE	Small Area Estimation
SRS	Simple Random Sampling

Chapter 1: Introduction: small area estimation

Sample surveys are widely used in practice to provide estimates not only for the overall population of interest but also for a variety of subpopulations (domains), such as geographic areas and socio-demographic groups. Examples of a geographic domain include a state, county, municipality, school district, etc. On the other hand, a socio-demographic domain may refer to a specific age-gender-race group within a large geographic area. In the context of sample surveys, a domain estimator is referred as "direct" if it is solely based on the domain-specific sample data. A direct estimator may also use known auxiliary information, such as the total of an auxiliary variable. Typically, as direct estimator is "design-based", which makes use of survey weights and the associated inferences are based on the probability distribution induced by the sampling design with the population values held fixed.

A domain (area) is classified as large if the domain-specific sample is large enough to produce direct estimates with an adequate precision under the sample design. A domain is regarded as "small" if the domain-specific sample is not large enough to support direct estimates of adequate precision. Some other terms used to denote a domain with small sample size include "subdomain", "small subgroup" and "minor domain". In some cases, many domains of interest, such as counties, may have no sample units due to the survey design limitations and/or other nonsampling errors such as nonresponse. In this dissertation, the term "small area" is used to

denote any domain for which direct estimates of adequate precision can not be produced.

In recent years, the demand for small area statistics has grown significantly worldwide due to their increasing use in policy formulation, program development, and the allocation of government funds across different geographic regions. For example, the U.S. Census Bureau established the Small Area Income and Poverty Estimates (SAIPE) program to provide annual income and poverty estimates for school districts, counties, and states. The primary objective of the SAIPE program is to support the administration of federal programs and guide the allocation of federal funds to local jurisdictions. Additionally, state and local governments rely on these estimates to distribute resources and manage social programs effectively. For further details on the SAIPE program, we refer readers to Kalton and Citro (2000).

Beyond the United States, international organizations have also recognized the value of small area estimation techniques. The World Bank, for instance, has increasingly applied these methods to estimate population density and monetary poverty in developing countries such as Mexico, Chad, Guinea, Mali, and Niger (Edochie et al., 2024; Newhouse et al., 2022). Similarly, the United Nations launched the SAE4SDG Toolkit in March 2021 to help countries adopt small area estimation techniques for generating more disaggregated data and advancing progress toward the Sustainable Development Goals (SAE4SDG).

1.1 Direct estimation

Let U be a population consisting of N distinct units and y_j be a characteristic of interest for a unit j, $j = 1, \cdots, N$. The common parameters of interest could be the population total, $Y = \sum_{j=1}^{N} y_j$, or the the population mean, $\bar{Y} = Y/N$, under the assumption that measurement

errors are absent. Suppose a sampling design is used to select a sample s of size n from U with probability $p(s)$. For simplicity, we also assume that all elements $j \in s$ can be observed, that is, complete response.

According to Rao and Molina (2015), in the design-based approach, an estimator $\hat{Y}$ of Y is called design-unbiased (or p-unbiased) if the design expectation of $\hat{Y}$ equals Y,

$$\mathrm{E}_p(\hat{Y}) = \sum p(s)\hat{Y}_s = Y, \text{for } Y \in \mathcal{R}, \tag{1.1}$$

where the summation is over all possible samples s under the specified design and $\hat{Y}_s$ is the value of $\hat{Y}$ for the sample s. The design variance of $\hat{Y}$ is denoted as $V_p(\hat{Y}) = \mathrm{E}_p[\hat{Y} - \mathrm{E}_p(\hat{Y})]^2$, and an estimator of $V_p(\hat{Y})$ is denoted as $v(\hat{Y}) = s^2(\hat{Y})$. The variance estimator $v(\hat{Y})$ is p-unbiased for $V_p(\hat{Y})$ if $\mathrm{E}_p[v(\hat{Y})] \equiv V_p(\hat{Y})$. An estimator $\hat{Y}$ is design-consistent if $\hat{Y}$ is p-unbiased and $V_p(\hat{Y})$ tends to zero as the sample size increases.

In this dissertation, we focus on making inferences for small areas. Consider the finite population U of size N partitioned into m small areas of sizes $N_1, \cdots, N_m$. Let U_i, s_i and n_i denote index sets of the finite subpopulation, the sampled units from that finite subpopulation and the corresponding sample sizes, respectively, for each area $i = 1, \cdots, m$, where $n = n_1 + \cdots + n_m$. Note that $n_i = 0$ if an area i is not sampled. Let y_{ij} denote the outcome of the survey variable of interest for the jth unit within ith area. The parameters of interest then could be the m small area subpopulation totals: $Y_i = \sum_{j=1}^{N_i} y_{ij}, i = 1, \cdots, m$. The subpopulation means can be similarly defined as: $\bar{Y}_i = Y_i/N_i$ for m small areas, where N_i may or may not to be known.

Design weights play an important role in constructing direct or designed-based estimators. Let w_{ij} be the sampling weight (inverse of the inclusion probability) of unit j from sampled area

i. The weight w_{ij} can be interpreted as the number of elements in the small area i represented by the sample element j. Thus, the direct estimator of Y_i is the weighted sample total and is expressed as $\hat{Y}_i = \sum_{j \in s_i} w_{ij} y_{ij}$. With $\hat{N}_i = \sum_{j \in s_i} w_{ij}$ as a design-unbiased estimator of the subpopulation size N_i of sampled area i, the estimators of small area means are the form of: $\hat{\bar{Y}}_i = \hat{Y}_i / \hat{N}_i$ for $i = 1, \cdots, m$. Note that if the sampling weights w_{ij} do not depend on the unit j, as it occurs for instance under simple random sampling within areas where $w_{ij} = n_i/N_i$, then $\hat{\bar{Y}}_i$ reduces to the unweighted sample mean: $1/n_i \sum_{j \in s_i} y_{ij}$. We refer readers to Fuller (2011) for more details of sampling statistics.

In practice, one often is interested in making inferences on complicated parameters for small areas, which are nonlinear functions of the outcome variable for the population units. For instance, let E_{ij} be a suitable quantitative measure of welfare for individual j in small area i, such as income or expenditure, and let z be a fixed poverty line. Then, the Foster–Greer–Thorbecke (FGT) poverty measures, introduced by Foster et al. (1984), are expressed as

$$F_{\alpha i} = \frac{1}{N_i} \sum_{j=1}^{N_i} F_{\alpha ij}, \ i = 1, \cdots, m, \tag{1.2}$$

of the values $F_{\alpha ij}$ defined as

$$F_{\alpha ij} = \left(\frac{z - E_{ij}}{z} \right)^{\alpha} I(E_{ij} < z), \ j = 1, \cdots, N_i; \ \alpha = 0, 1, 2, \tag{1.3}$$

where $I(E_{ij} < z) = 1$ if $E_{ij} < z$, otherwise, it equals to 0. When $\alpha = 0$, the FGT measure represents the proportion of individuals living below the poverty line in small area i, commonly referred to as the poverty incidence or headcount ratio. For $\alpha = 1$, the measure, known as the

poverty gap, captures the average relative distance of individuals to the poverty line within the area. When $\alpha = 2$, the measure, termed poverty severity, highlights the inequality among the poor, with higher values of F_{2i} indicating areas experiencing more severe levels of poverty.

Similarly, there are two fundamental types of direct estimators for FGT poverty measures. The first, under simple random sampling within areas, is the unweighted sample mean:

$$\hat{F}_{\alpha i} = 1/n_i \sum_{j \in s_i} F_{\alpha ij}, \ \alpha = 0, 1, 2$$

for all sampled small areas. The second type, which incorporates survey weights, is the weighted sample mean:

$$\hat{F}^w_{\alpha i} = 1/\hat{N}_i \sum_{j \in s_i} w_{ij} F_{\alpha ij}, \ \alpha = 0, 1, 2$$

for all sampled small areas.

As a model-free approach, the design-based estimation can not only incorporate survey design features, such as strata or clusters, but also provide reliable and design-consistent inferences in large samples. However, in practice, the sampling mechanism could be corrupted by nonsampling errors, such as measurement or systematic errors, which makes the design-based method is not applicable, Kalton (1983). When the sample sizes in the subgroups of interest are small or even zero, the designed-based methods would become unreliable and problematic. Additionally, sampling variances for the direct estimates were not available due to the small sample sizes.

1.2 Model-based small area estimation

In making estimates for small areas with adequate level of precision, it is necessary to supplement survey sample data with other relevant information that is often obtained from different administrative and census records as well as various satellite data. In many small area applications, mixed effect model are now routinely used in integrating information from various sources and explaining different sources of errors. Generally, mixed models have both fixed effects and the random effects components. The area-specific random effects can account for between small area variations, not otherwise explained by the fixed effects part of the model. In small area estimation (SAE) literature, these models are generally classified into two broad types: (i) area-level models, which relate small area means to area-specific auxiliary variables; and (ii) unit-level models, which relate unit values of the study variable to unit-specific auxiliary variables.

1.2.1 Area level model

A typical area-level model, introduced by Fay and Herriot (1979), consists of two hierarchical components: the sampling model and the linking model. The sampling model accounts for the sampling error in survey-weighted direct estimates and potential errors from the survey design. The linking model relates the population characteristic to known area-specific auxiliary variables. A key advantage of these models is that they require only aggregate summaries of auxiliary variables from administrative records, recent censuses, and other external sources, thus expanding data availability for model fitting.

For more general applications, let the parameter of interest be a specified function of small area means: $\theta_i = g(\bar{Y}_i)$, for $i = 1, \ldots, m$. For instance, Fay and Herriot (1979) used the logarithm

function. When direct survey estimates $\hat{\bar{Y}}_i$ are available, we define $\hat{\theta}_i = g(\hat{\bar{Y}}_i)$ and assume the following sampling model:

$$\hat{\theta}_i = \theta_i + e_i, \quad i = 1, \ldots, m, \tag{1.4}$$

where the sampling errors e_i are independent and typically assumed to follow normal distributions with mean 0 and variance D_i. The sampling variances D_i are typically assumed to be known. However, in practice, this assumption can be relaxed by first estimating D_i from individual sample data and and then applying a smoothing technique to obtain a more stable estimate (Fay and Herriot, 1979; Ha et al., 2014; Hawala and Lahiri, 2018; Otto and Bell, 1995).

At level 2, we link the parameter θ_i to area-specific auxiliary data $\mathbf{x}_i = (x_{1i}, \cdots, x_{pi})'$ through a linear model

$$\theta_i = \mathbf{x}_i'\boldsymbol{\beta} + v_i, \quad i = 1, \cdots, m \tag{1.5}$$

where $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)'$ is the vector of regression coefficients and v_i's represent area-specific random effects. These random effects are assumed to be independent and identically distributed with mean 0 and variance A. They account for the between area variation beyond what is explained by the auxiliary variables in the model. While normality of v_i is a common assumption, but robust or flexible inference methods that relax this assumption will be explored in later chapters.

Combining (1.4) and (1.5) yields:

$$\hat{\theta}_i = \mathbf{x}_i'\boldsymbol{\beta} + v_i + e_i, \quad i = 1, \cdots, m, \tag{1.6}$$

where v_i and e_i are assumed to be independent. Note that the combined model involves the sampling errors e_i as well as the modeling errors v_i.

Under squared error loss, the best predictor (BP) of θ_i is the conditional expectation of θ_i given the data and model parameters. Distributional assumptions are needed for calculating the BP. With normality assumption on both v_i and e_i, the BP has a closed form given by:

$$\tilde{\theta}_i^B = (1 - B_i)\hat{\theta}_i + B_i\mathbf{x}_i'\boldsymbol{\beta}, \quad i = 1, \cdots, m, \tag{1.7}$$

where $B_i = D_i/(A + D_i)$. Instead, the best linear predictor (BLP) of θ_i always has the explicit form as (1.7).

If A is known, we can obtain the standard weighted least squares estimator of $\boldsymbol{\beta}$, denoting as $\tilde{\boldsymbol{\beta}}(A)$. Replacing $\boldsymbol{\beta}$ in (1.7) by $\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(A)$, a best linear unbiased prediction (BLUP) estimator of θ_i is given by

$$\tilde{\theta}_i^{\text{BLUP}} = (1 - B_i)\hat{\theta}_i + B_i\mathbf{x}_i'\tilde{\boldsymbol{\beta}}. \tag{1.8}$$

Note that BLUP estimators minimize the MSE among the class of linear unbiased estimators and do not dependent on normality of the random effects. Since both $\boldsymbol{\beta}$ and A are often unknown, they must be estimated. Plugging in their estimators obtained via standard methods such as maximum likelihood (ML) or restricted maximum likelihood (REML), yields the empirical best

linear unbiased predictor (EBLUP):

$$\hat{\theta}_i^{\text{EBLUP}} = (1 - \hat{B}_i)\hat{\theta}_i + \hat{B}_i \mathbf{x}_i' \hat{\boldsymbol{\beta}}, \tag{1.9}$$

where $\hat{B}_i = D_i/(\hat{A} + D_i)$ and $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\hat{A})$, and $\hat{A}$ is a consistent estimator of A for large m.

1.2.2 Unit level model

When unit-specific auxiliary data, $\mathbf{x}_{ij} = (x_{ij1}, \cdots, x_{ijp})'$, are available for each population element j in each small area i, a useful approach for inferring small area means $\bar{Y}_i$ is to relate the variable of interest, y_{ij}, to $\mathbf{x}_{ij}$ through a unit-level model. It is often sufficient to assume that only the population means $\bar{\mathbf{X}}_i = (\bar{X}_1, \cdots, \bar{X}_p)'$ are known. A fundamental unit level model, introduced by Battese et al. (1988) and known as the nested error regression (NER) model, is:

$$y_{ij} = \mathbf{x}_{ij}' \boldsymbol{\beta} + v_i + e_{ij}; \quad j = 1, \cdots, N_i, i = 1, \cdots, m. \tag{1.10}$$

It is typically assumed that the random effects v_i's and the sampling errors e_{ij}'s are independent with $v_i \sim N(0, \sigma_v^2)$ and $e_{ij} \sim N(0, \sigma_e^2)$.

If survey sampling follows the assumed model (1.10), then for large N_i, the small area mean simplifies to

$$\bar{Y}_i = \bar{\mathbf{X}}_i' \boldsymbol{\beta} + v_i,$$

noting that $\bar{Y}_i = \bar{\mathbf{X}}_i' \boldsymbol{\beta} + v_i + \bar{E}_i$, where $\bar{E}_i \approx 0$ since $\bar{E}_i$ represents the mean of the N_i sampling

errors e_{ij}. This condition is satisfied under simple random sampling from each area or more generally for sampling designs that use the auxiliary information $\mathbf{x}_{ij}$ in the selection of the samples s_i.

With the assumption of normality for both v_i and e_{ij}, the BP of $\bar{Y}_i$ that minimizes the mean squared error (MSE) is given by:

$$\tilde{\bar{Y}}_i^B = \bar{\mathbf{X}}_i'\boldsymbol{\beta} + \{f_i + (1-f_i)\gamma_i\}(\bar{y}_i - \bar{\mathbf{x}}_i'\boldsymbol{\beta}), \tag{1.11}$$

where $f_i = n_i/N_i$ and $\gamma_i = \sigma_v^2/(\sigma_v^2 + \sigma_e^2/n_i)$. When f_i is negligible, this simplifies to a weighted sum of the sample regression predictor and the synthetic predictor:

$$\tilde{\bar{Y}}_i^B = \gamma_i[\bar{y}_i + (\bar{\mathbf{X}}_i - \bar{\mathbf{x}}_i)'\boldsymbol{\beta}] + (1-\gamma_i)\bar{\mathbf{X}}_i'\boldsymbol{\beta}. \tag{1.12}$$

Notably, this BP expression can be derived without assuming normality, as long as the conditional means of the random effects given the data follow a linear function (Ghosh and Lahiri, 1987). This is also known as the best linear predictor (BLP) of $\bar{Y}_i$. Similarly, when the variance components are known, the BLUP of $\bar{Y}_i$ can be obtained by replacing $\boldsymbol{\beta}$ in (1.11) with its weighted least squares estimator. The empirical BLUP (EBLUP) is then derived by substituting both the unknown regression coefficients and variance components with their respective standard estimators.

The NER model (1.10) may not be suitable for complex survey designs, such as two-stage cluster sampling, as it does not account for random cluster effects. More advanced models, such as the two-fold nested error regression model, have been developed to address these issues (Stukel and Rao, 1999; Torabi and Rao, 2014). Additionally, incorporating survey weights into unit-level

models ensures design consistency and self-benchmarking properties (Lahiri and Mukherjee, 2007; You and Rao, 2002).

1.3 Discussion and outline of the dissertation

We have reviewed the definition of small areas, the demand for small area statistics, traditional direct estimation methods, and model-based small area estimation. Specifically, two fundamental small area models are typically used, both of which rely on assumptions such as the linear relationship between the variable of interest and auxiliary variables, with constant regression coefficients across areas. These models also assume normality for random effects and sampling errors with a common sampling variance across areas. However, these assumptions require justification in applications and may not hold in many practical situations.

From a practical point of view, any proposed model is subject to model misspecification. In Chapter 2, we begin by reviewing the observed best prediction (OBP) method introduced by Jiang et al. (2011) for estimating small area means under a linear mixed model, aiming to reduce the effects of model misspecification. We then investigate the effects of misspecified mean functions and variance components on the predictive performance of various small area estimators, including both OBP and EBLUP under different small area models. Through both analytical and numerical evidence, we demonstrate why OBP, when relying on unit-level auxiliary information, may not necessarily perform better than EBLUP in terms of design-based mean squared prediction error (MSPE). Additionally, we highlight how incorporating area-level auxiliary variables can improve the performance of OBP.

Many measures of poverty and inequality are nonlinear functions of the dependent variable

for population units. This makes many small area estimation methods, typically developed for estimating linear characteristics such as small area means discussed in Subsection 1.2, inapplicable. Moreover, the homogeneity assumption of regression coefficients and sampling variances across areas in the unit-level model (1.10) may not hold in many real-world applications due to variations in socio-economic conditions, data quality, and sampling designs across areas.

In Chapter 3, we develop a framework for small area poverty estimation using the Nested Error Regression Model with High-Dimensional Parameters (NERHDP) introduced by Lahiri and Salvati (2023). This model accounts for heterogeneity in regression coefficients and sampling variances across areas, enhancing both robustness and flexibility. To overcome the computational challenges of the existing algorithm, we propose an efficient method that significantly reduces computation time and improves scalability for large datasets. Additionally, we introduce a novel approach for generating area-specific poverty estimates for out-of-sample areas, yielding less synthetic estimates compared to existing methods. Through design-based simulation studies, we demonstrate that the proposed method outperforms existing approaches in terms of relative bias and relative root mean squared prediction error. Finally, we apply the proposed method to household survey data from the 2002 Albania Living Standards Measurement Survey to estimate poverty indicators for 374 municipalities, incorporating auxiliary information from the 2001 census.

In Chapter 4, we advance the theory of parametric bootstrap in constructing highly efficient empirical best (EB) prediction intervals of small area means. In the context of an area level model where the random effects follow a non-normal known distribution except possibly for unknown hyperparameters, we analytically show that the order of coverage error of empirical best linear (EBL) prediction interval is $O(m^{-3/2})$ even if we relax the normality of the random effects by

the existence of pivot for a suitably standardized random effects when hyperpameters are known. Recognizing the challenge of showing existence of a pivot, we develop a simple moment-based method to claim non-existence of pivot. In absence of a pivot, we find that the order of coverage error of the parametric bootstrap EBL prediction interval is $O(m^{-1})$, which is always positive under certain conditions. In general, we analytically show for the first time that the coverage problem can be corrected by adopting a suitably devised double parametric bootstrap. Our Monte Carlo simulations show that our proposed single bootstrap method performs reasonably well when compared to rival methods.

In Chapter 5, we give a summary of this dissertation and give directions for future research.

Chapter 2: Effects of model misspecification on small area estimators

2.1 Introduction

The effectiveness of any model-based method largely depends on the availability of good auxiliary data. Therefore, considerable attention should be given to selecting auxiliary variables that serve as good predictors of the study variables. Observed best prediction (OBP), proposed by Jiang et al. (2011), aims to reduce the impact of such misspecification. In the methodology of OBP, two different strategies are employed for the area level model and the NER model. For the area-level model, the derivation of OBP involves two models: a working model that can incorporate auxiliary variables, which may be misspecified, and a broader model with no or very weak assumptions, representing the true but unknown underlying process. The working model generates the best predictor (BP) for the small area mean, though this BP may no longer be valid when the assumed model is misspecified. The broader model is used to derive a criterion, such as an objective function, for estimating the parameters of the assumed model. This criterion is model-independent, making OBP more robust to model misspecification.

When the assumed model is the NER model in equation (1.10), Jiang et al. (2011) instead uses the design-based MSPE to derive a criterion for estimating model parameters, while the BP remains derived under the assumed model. The advantage of incorporating the design-based MSPE is that its expectation relies solely on sampling from a finite population, without the need

to specify the model structure. As a result, this approach may provide more robust predictions against potential model misspecifications. Overall, for both area level and NER model, the OBP only finds a way to estimate the parameters involved in the BP, namely β and variance components, to minimize the potential damage in case the model is misspecified.

In the context of area level models, Jiang et al. (2011) used both theoretical and empirical studies to demonstrate that OBP outperforms EBLUP in terms of design-based MPSE when the underlying linear mixed model is misspecified. Subsequently, Jiang et al. (2015) considered OBP for the NER model, where both the mean function and variance components are misspecified. Their simulations indicated that OBP may perform significantly better than EBLUP in terms of both overall and area-specific design-based MSPE. However, our simulation studies indicate that OBP using unit-level auxiliary variables (OBP-UNIT) does not outperform EBLUP unless the number of areas m is extremely large, which is unusual in small area estimation. We found that using area-level auxiliary variables, such as in the unit-context or area-level models, one can improve the performance of OBP.

In this chapter, we investigate the effects of misspecified mean function and variance components on the predictive performances of existing small area estimators. We also provide both analytical and numerical evidence explaining why OBP-UNIT may underperform in terms of design-based MSPE, and how the use of area-level auxiliary variables can enhance its effectiveness. The rest of the chapter is organized as follows. In Section 2.2, we provide further details on OBP methodology and the reasoning why the OBP with area-level auxiliary variables outperforms the OBP with unit level auxiliary variables. In Section 2.3, we present numerical studies and the evaluation results in terms of both overall and area-specific design-based MSPE. Section 2.5 offers conclusions and practical guidelines for implementing OBP in the presence of model

misspecification.

2.2 Analytical comparison between OBP with unit-level auxiliary variables and OBP with area-level auxiliary variables

In this section, we provide analytical evidence demonstrating why OBP using area-level auxiliary information outperforms OBP using unit-level auxiliary information. Let $\boldsymbol{\psi} = (\boldsymbol{\beta}', \sigma_v^2, \sigma_e^2)$ represent the vector of model parameters of the NER model (1.10), and let $\tilde{\boldsymbol{\theta}}(\boldsymbol{\psi}) = [\tilde{\theta}_i(\boldsymbol{\psi})]_{1\leq i\leq m}$ denote the vector of BPs as expressed in (1.11) for the true small area means $\boldsymbol{\theta} = [\theta_i]_{1\leq i\leq m}$, in which the θ_i's are finite population means $\bar{Y}_i = 1/N_i \sum_{j=1}^{N_i} y_{ij}$.

Then, the design-based MSPE of $\tilde{\boldsymbol{\theta}}(\boldsymbol{\psi})$ is defined as

$$\text{MSPE}(\tilde{\boldsymbol{\theta}}(\boldsymbol{\psi})) = \text{E}(|\tilde{\boldsymbol{\theta}}(\boldsymbol{\psi}) - \boldsymbol{\theta}|^2) = \sum_{i=1}^{m} \text{E}(\tilde{\theta}_i(\boldsymbol{\psi}) - \theta_i)^2, \tag{2.1}$$

where the expectation is with respect to the sample design. Following Jiang et al. (2011) and Jiang et al. (2015), the MSPE in (2.1) has an alternative expression, which is a key idea of the OBP. Namely, it leads to the fundamental equation of the OBP,

$$\text{MSPE}(\tilde{\boldsymbol{\theta}}(\boldsymbol{\psi})) = \text{E}\{Q(\boldsymbol{\psi})\}, \tag{2.2}$$

and $Q(\cdot)$ is called the observed MSPE function. Under the NER model, the Q function can be expressed as:

$$
\begin{aligned}
Q &= \boldsymbol{\beta}'(\bar{\boldsymbol{X}} - \boldsymbol{G}\bar{\boldsymbol{x}})'(\bar{\boldsymbol{X}} - \boldsymbol{G}\bar{\boldsymbol{x}})\boldsymbol{\beta} - 2\bar{\boldsymbol{y}}'\{(\boldsymbol{I}_m - 2\boldsymbol{G})\bar{\boldsymbol{X}} + \boldsymbol{G}^2\bar{\boldsymbol{x}}\}\boldsymbol{\beta} \\
&\quad + \bar{\boldsymbol{y}}'\boldsymbol{G}^2\bar{\boldsymbol{y}} + \mathbf{1}_m'(\boldsymbol{I}_m - 2\boldsymbol{G})\hat{\boldsymbol{\mu}}^2,
\end{aligned}
$$

where $\bar{\boldsymbol{X}} = (\bar{\boldsymbol{X}}_i')_{1\le i\le m}$, $\bar{\boldsymbol{x}} = (\bar{\boldsymbol{x}}_i')_{1\le i\le m}$, $\bar{\boldsymbol{x}}_i$ being the vector of unweighted sample means of the auxiliary variables for area i, $\bar{\boldsymbol{y}} = (\bar{y}_i)_{1\le i\le m}$, $\boldsymbol{G} = \text{diag}\{n_i/N_i + (1-n_i/N_i)n_i\sigma_v^2/(n_i\sigma_v^2+\sigma_e^2), 1 \le i \le m\}$, and $\hat{\boldsymbol{\mu}}^2 = (\hat{\mu}_i^2)_{1\le i\le m}$ is the design-unbiased estimator of $(\bar{Y}_i^2)_{1\le i\le m}$ given by

$$
\hat{\mu}_i^2 = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{N_i - 1}{N_i(n_i-1)}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2.
$$

Thus, the best predictive estimators (BPE) of $\boldsymbol{\psi}$, denoted as $\hat{\boldsymbol{\psi}}$, is obtained by minimizing $Q(\boldsymbol{\psi})$. Note that the OBP can be motivated as EBP when the BPE of model parameters are used in place of standard model parameter estimators, such as the maximum likelihood estimator (MLE).

To understand why OBP performs optimally when area-level auxiliary variables are used, it is instructive to consider the case when all the unit-level model parameters $\boldsymbol{\psi}$ are known. A simplification Q as a function of $\bar{\boldsymbol{x}}$ is given by:

$$
\begin{aligned}
Q &= \boldsymbol{\beta}'\{[(\boldsymbol{I}_m - \boldsymbol{G})\bar{\boldsymbol{X}} - \boldsymbol{G}(\bar{\boldsymbol{x}} - \bar{\boldsymbol{X}})]'[(\boldsymbol{I}_m - \boldsymbol{G})\bar{\boldsymbol{X}} - \boldsymbol{G}(\bar{\boldsymbol{x}} - \bar{\boldsymbol{X}})]\}\boldsymbol{\beta} \\
&\quad -2\bar{\boldsymbol{y}}'\{(\boldsymbol{I}_m - \boldsymbol{G})^2\bar{\boldsymbol{X}} + \boldsymbol{G}^2(\bar{\boldsymbol{x}} - \bar{\boldsymbol{X}})\}\boldsymbol{\beta} + \cdots \\
&= -2\boldsymbol{\beta}'\bar{\boldsymbol{X}}'(\boldsymbol{I}_m - \boldsymbol{G})\boldsymbol{G}(\bar{\boldsymbol{x}} - \bar{\boldsymbol{X}})\boldsymbol{\beta} + \boldsymbol{\beta}'(\bar{\boldsymbol{x}} - \bar{\boldsymbol{X}})'\boldsymbol{G}^2(\bar{\boldsymbol{x}} - \bar{\boldsymbol{X}})\boldsymbol{\beta} - 2\bar{\boldsymbol{y}}'\boldsymbol{G}^2(\bar{\boldsymbol{x}} - \bar{\boldsymbol{X}})\boldsymbol{\beta} + \cdots \\
&= Q^*(\bar{\boldsymbol{x}}) + \cdots
\end{aligned}
$$

where the omitted terms in $\cdots$ do not involve $\bar{\boldsymbol{x}}$.

Using the facts that $\mathrm{E}(\bar{\boldsymbol{x}} - \bar{\boldsymbol{X}}) = 0$ and $\mathrm{E}\{\bar{\boldsymbol{y}}'\boldsymbol{G}^2(\bar{\boldsymbol{x}} - \bar{\boldsymbol{X}})\boldsymbol{\beta}\} = \mathrm{cov}(\bar{\boldsymbol{y}}'\boldsymbol{G}^2, \boldsymbol{\beta}'\bar{\boldsymbol{x}}')$, we get

$$\mathrm{E}Q \;=\; \mathrm{E}\{\boldsymbol{\beta}'(\bar{\boldsymbol{x}} - \bar{\boldsymbol{X}})'\boldsymbol{G}^2(\bar{\boldsymbol{x}} - \bar{\boldsymbol{X}})\boldsymbol{\beta}\} + \cdots \tag{2.3}$$

where $\cdots$ are the terms involving only finite population parameters and known model parameters. Since $\boldsymbol{G}^2$ is positive definite, the design-based MSPE, $\mathrm{E}Q$, is minimized at $\bar{\boldsymbol{x}} - \bar{\boldsymbol{X}} = 0$, for any fixed $\boldsymbol{\beta}$. This result indicates that incorporating area-level auxiliary variables leads to optimal OBP performance.

The more detailed derivations of OBP procedure for both area-level and unit-level models can be found in the supplementary materials of Jiang et al. (2011).

2.3 Simulations

In this section, we conduct simulation studies to examine the impact of model misspecification on small area estimators. The simulation settings follow those in Jiang et al. (2015). For simplicity, we consider a case of a single auxiliary variable that is assumed to be linearly associated with the response variable y_{ij} through the following model:

$$y_{ij} = \beta_1 x_{ij} + v_i + e_{ij} \quad i = 1, \ldots, m, j = 1, \ldots, N_i, \tag{2.4}$$

where x_{ij}'s are known values of an auxiliary variable for the jth unit of the ith area; β_1 is an unknown regression coefficient; v_i, e_{ij} are the same as in (1.10). We assume that x_{ij}'s are not all the same in an area.

In many applications, when good unit-level auxiliary variables are unavailable, an alternative model, which is a special case of the NER model and is often referred to as the unit-context model (Newhouse et al., 2022), is used

$$y_{ij} = \bar{\boldsymbol{X}}_i'\boldsymbol{\beta} + v_i + e_{ij}, \quad i = 1, \cdots, m, \quad j = 1, \cdots, N_i, \tag{2.5}$$

where $\bar{\boldsymbol{X}}_i = N_i^{-1} \sum_{j=1}^{N_i} \boldsymbol{x}_{ij}$ is a vector of known population means of auxiliary variables for area i. The same assumptions as the NER model (1.10) on the random effects v_i and sampling errors e_{ij} apply. This corresponds to the case discussed in Section 2.2, where area-level auxiliary information is used for modeling unit-level response variables.

In this simulation context, (2.5) becomes:

$$y_{ij} = \beta_1 \bar{X}_i + v_i + e_{ij} \quad i = 1, \ldots, m, j = 1, \ldots, N_i. \tag{2.6}$$

Since the unit-context model is a special case of unit-level models, OBP can be directly derived by replacing $\boldsymbol{x}_{ij}$ by $\bar{\boldsymbol{X}}_i$.

For comparison, we also consider an area-level model, given by:

$$\bar{y}_i = \beta_1 \bar{X}_i + v_i + \bar{e}_i \quad i = 1, \ldots, m. \tag{2.7}$$

where $\bar{y}_i$ is the unweighted sample mean; v_i's are area-specific random effect; $\bar{e}_i$'s are sampling errors. It is assumed that v_i's and $\bar{e}_i$'s are independent with $v_i \overset{iid}{\sim} N(0, \sigma_v^2)$, $\bar{e}_i \overset{iid}{\sim} N(0, D_i)$, where

D_i is estimated as $\hat{D}_i = s^2/n_i$ with

$$s^2 = \frac{1}{\sum_{i=1}^{m} n_i - 1} \sum_{i=1}^{m} \sum_{i=1}^{n_i} (y_{ij} - \bar{y})^2.$$

For the simulations set up, we draw simple random sampling without replacement (SRSWOR) samples from the population of each small areas and consider the following estimators of the small area means throughout the rest of this section:

(A) Direct estimator (sample mean),

(B) OBP under the assumed unit-context model (2.6) (OBP-UC),

(C) OBP under the basic Fay-Herriot (area-level) model (2.7) (OBP-FH),

(D) OBP under the assumed unit-level model(2.4) (OBP-UNIT),

(E) EBLUP under the assumed unit-level model (2.4) (EBLUP-UNIT),

(F) EBLUP under the assumed unit-context model (2.6) (EBLUP-UC).

To introduce model misspecifications, we first generate y for the finite population from the following superpopulation heteroscedastic nested-error regression model:

$$y_{ij} = b + v_i + e_{ij} \quad i = 1, \ldots, m, j = 1, \ldots, N_i. \tag{2.8}$$

The population and sample sizes are the same for all areas and are fixed at $N_i = 1000$ and $n_i = 4$, respectively. We consider two different values of b: $b = 5, 10$, and three values of the number of small areas: $m = 40, 100$ or 400 with v_i generated from the normal distribution $N(0, 1)$, and

e_{ij} generated from the normal distribution $N(0, \sigma_{ei}^2)$, where σ_{ei}^2 is independently generated from the gamma distribution $\Gamma(3, 0.5)$. In each case, x for the finite population is generated from a log-normal superpopulation distribution with a mean of 1 and a standard deviation of 0.5.

Each scenario is independently simulated $K = 1000$ times. The performance of the estimators (A)-(F), under the above simulation setups, are assessed in terms of both overall and area-specific design-based MSPEs. The area-specific design-based MSPE is defined as $\text{MSPE}(\hat{\bar{Y}}_i) = \text{E}(\hat{\bar{Y}}_i - \bar{Y}_i)^2$, where $\bar{Y}_i = N_i^{-1} \sum_{i=1}^{N_i} y_{ij}$ is the true small area mean, and $\hat{\bar{Y}}_i$ is the predicted value for ith area, either by OBP or EBLUP and E is with respect to the sample design. In Monte Carlo simulations, MSPE for area i and overall MSPE are approximated by:

$$\text{MSPE}_i \approx \frac{1}{K} \sum_{k=1}^{K} (\hat{\bar{Y}}_i^{(k)} - \bar{Y}_i^{(k)})^2, \tag{2.9}$$

$$\text{MSPE} \approx \frac{1}{m} \sum_{i=1}^{m} \left[\frac{1}{K} \sum_{k=1}^{K} (\hat{\bar{Y}}_i^{(k)} - \bar{Y}_i^{(k)})^2 \right], \tag{2.10}$$

respectively, where $\bar{Y}_i^{(k)}$ and $\hat{\bar{Y}}_i^{(k)}$ are the true mean and estimated mean for area i in the k^{th} simulation run, respectively.

Table 2.1: Overall simulated design-based MSPE when the finite population is generated from the heteroscedastic NER model (2.8)

(m, b)	DIRECT	OBP-UC	OBP-FH	OBP-UNIT	EBLUP-UNIT	EBLUP-UC
(40, 5)	1.475	0.689	0.698	1.775	1.474	0.687
(100, 5)	1.488	0.638	0.657	1.409	1.495	0.638
(400, 5)	1.493	0.614	0.637	1.352	1.508	0.613
(40, 10)	1.475	0.696	0.705	3.919	1.522	0.694
(100, 10)	1.488	0.646	0.664	2.364	1.513	0.645
(400, 10)	1.493	0.621	0.644	1.711	1.510	0.621

Table 2.1 reports the simulated overall design-based MSPE for the various simulation conditions and estimators when the finite population is generated based on the true underlying model (2.8). The direct estimator is unaffected by model misspecification as it is not model-based. EBLUP-UNIT behavior is similar to direct because it automatically assigns more weight to the sample mean when the model is weak. While OBP-UNIT was designed to handle model misspecification, it surprisingly performs poorly, even worse than the direct estimator and EBLUP-UNIT when $b = 10$. OBP-UC and EBLUP-UC show similar and better performance than EBLUP-UNIT. The reason why EBLUP-UC also performs well could be that, given the large population size N_i and the fact that the population values come from a common distribution across all areas, the population means $\bar{X}_i$ are roughly equal across areas. As a result, the unit-context model and the area-level model approximate the true model closely. Consequently, predictors using area-level auxiliary variables tend to outperform those using unit-level auxiliary variables. Overall, our findings suggest that in instances of significant model misspecification, OBP-UNIT may not be an effective choice.

As for the area-specific MSPEs, we utilize boxplots to display the distributions of the area-specific design-based MSPEs associated with all the estimators. See Figure 2.1. The boxplot of OBP-UNIT shows much larger median MSPE and variability than those of OBP-UC and OBP-FH and in some cases worse than direct and EBLUP-UNIT. OBP-UC shows slightly larger variability than OBP-FH. It might be because that in contrast to area-level models, unit-context models incorporate the uncertainty resulting from the estimation of model parameters, such as sampling variances.

As discussed in Jiang et al. (2015), the simulation conditions above might be a little extreme in which the assumed models are completely different from the true underlying model.

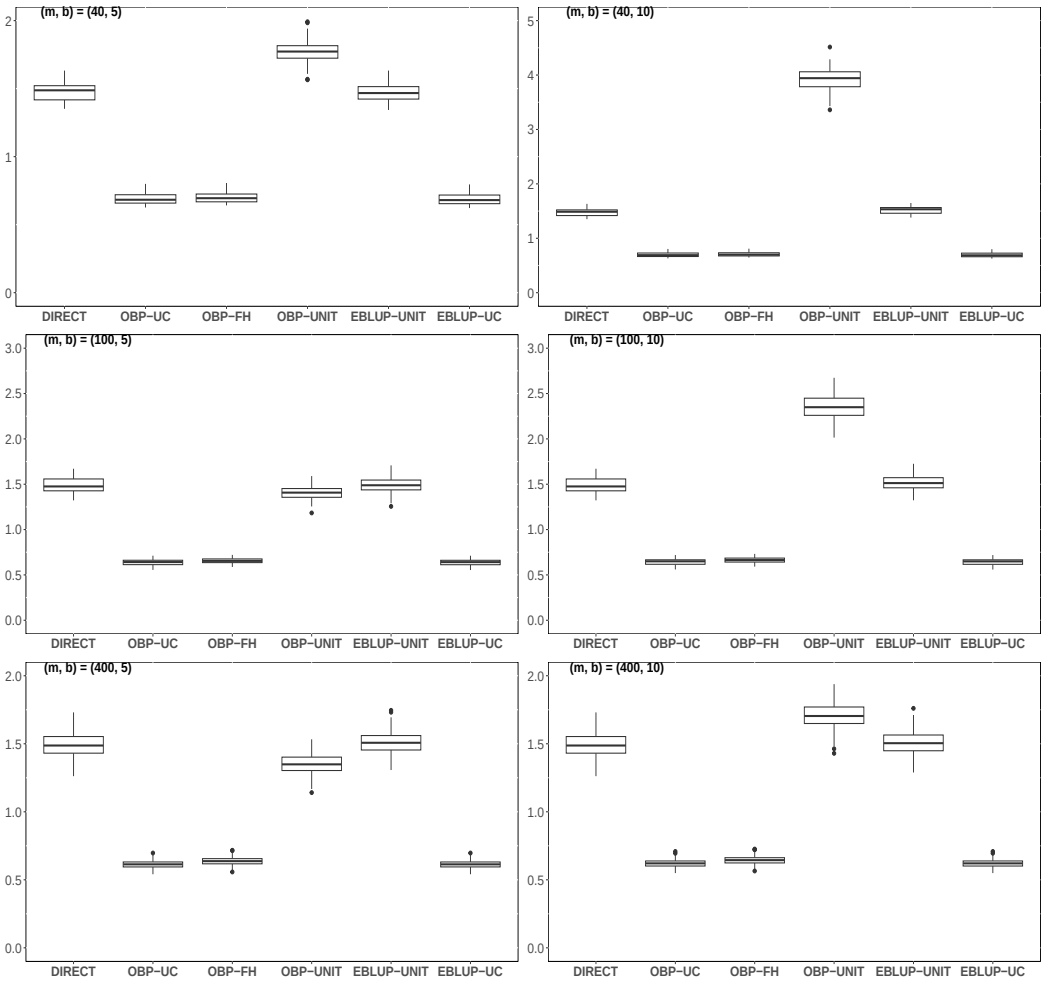


Figure 2.1: Distributions of area-specific simulated design-based MSPEs when the finite population is generated from the heteroscedastic NER model (2.8)

This motivates us to consider some moderate cases where assumed model is partially correct compared to the true model. Keeping the same assumed models (2.4) - (2.7) we generate the finite population for y from the underlying superpopulation model:

$$y_{ij} = b_0 + b_1 x_{ij} + v_i + e_{ij} \quad i = 1, \ldots, m, j = 1, \ldots, N_i \tag{2.11}$$

where $b_0 = 10$, $b_1 = 5$. Note that the slope in (2.11) is not zero and it matches the linear relationship part of the assumed model (2.4). But still, we have slight misspecifications in the

Table 2.2: Overall simulated design-based MSPE when the finite population is generated from the heteroscedastic NER model (2.11)

m	DIRECT	OBP-UC	OBP-FH	OBP-UNIT	EBLUP-UNIT	EBLUP-UC
40	18.243	1.798	1.588	18.230	1.532	1.872
100	18.339	1.341	1.248	7.958	1.514	1.358
400	18.422	1.085	1.059	3.122	1.509	1.087

sense that the true model here also has a nonzero intercept. In addition, e_{ij} is generated from the same normal distribution as described previously, and so we have the issue of heteroscedasticity as well. Three different values of m are considered: $m = 40, 100$ or 400, and v_i is generated from normal distribution $N(0, 1)$.

The results based on $K = 1000$ simulations are displayed in Table 2.2. As expected, in this case EBLUP-UNIT outperforms the direct estimator since the assumed model is closer to the true model and EBLUP can borrow strengths from other related areas. While OBP-UNIT was expected to perform comparably to EBLUP-UNIT in this scenario, our simulation results indicate that OBP-UNIT performs actually worse than EBLUP-UNIT in terms of the simulated MSPE. When $m = 40$, the simulated MSPE for OBP-UNIT is similar to that of the DIRECT estimator, indicating poor performance. In contrast, both OBP-UC and EBLUP-UC show similar results, performing slightly better than EBLUP-UNIT when $m = 100$ and $m = 400$. OBP-FH demonstrates stable performance and performs better than OBP-UC and EBLUP-UC when $m = 40$.

The results in Tables (2.1) and (2.2) indicates that OBP-UNIT may not effectively reduce the impact of model misspecification, compared with EBLUP-UNIT. This surprising outcome prompted further investigation, which we explore in the following Remarks 1 and 2.

2.4 Further numerical studies on OBP

Remark 1: MSPE of OBP-UNIT vs $|\bar{\boldsymbol{X}} - \bar{\boldsymbol{x}}|$

To further investigate the relationship between the performance of OBP-UNIT and the absolute difference between sample means and population means, we conducted a numerical study using the following Nested Error Regression (NER) superpopulation model:

$$y_{ij} = 10 + v_i + e_{ij} \quad i = 1, \ldots, 50,\ j = 1, \ldots, 1000. \tag{2.12}$$

The generative processes for the finite population for x, v_i, and e_{ij} remain consistent with the aforementioned model. Next, we drew SRS samples of size $n_i = 4$ for each small areas.

In a simulation setting, when the number of units within a specific area is relatively small, it is often impossible to have sample mean $\bar{\boldsymbol{x}}$ equal or even close to population mean $\bar{\boldsymbol{X}}$. To explore how deviations of sample means $\bar{\boldsymbol{x}}$ from population means $\bar{\boldsymbol{X}}$ affect the performance of OBP, we introduced a bias term in the sample means. Specifically, for each area, the sample means were adjusted by setting: $\bar{\boldsymbol{x}} = \bar{\boldsymbol{X}} + \text{bias}$, where the bias term varied to take on negative, positive, and zero values.

The design-based MSPE for OBP using unit-level auxiliary variables (MSPE_i) was calculated for each area, and these values were compared with the magnitude of the absolute difference $|\bar{\boldsymbol{X}} - \bar{\boldsymbol{x}}|$. The results were visualized in Figure 2.2, which illustrates that the simulated MSPE_i for OBP-UNIT increases with the difference between $\bar{X}_i$ and $\bar{x}_i$ for area $i = 1$. Notably, when $\bar{X}_i - \bar{x}_i = 0$, corresponding to the unit-context model, simulated MSPE_i is lower compared to

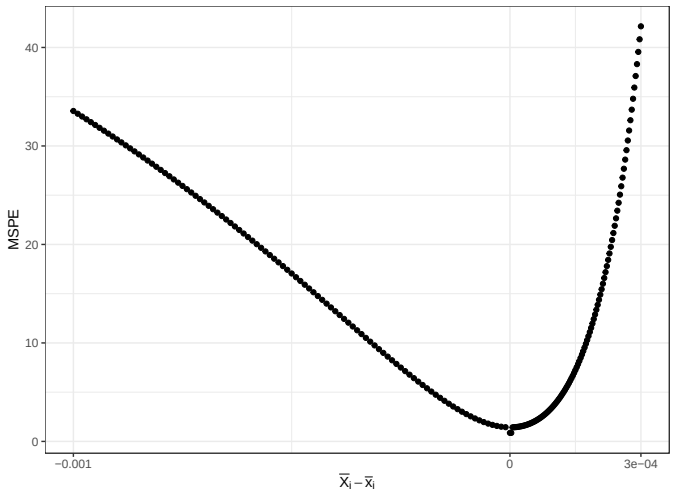


Figure 2.2: Relationship between $|\bar{X}_i - \bar{x}_i|$ and MSPE_i for area $i = 1$

other scenarios with non-zero differences. These results align with our theoretical design-based expectation of Q in section 2.2.

Remark 2: Minimizing $Q(\boldsymbol{\psi})$ vs Minimizing $\text{MSPE}(\tilde{\theta}(\boldsymbol{\psi}))$

Another possible explanation for the underperformance of OBP-UNIT might be that the basic OBP equation, valid for fixed $\boldsymbol{\psi}$, may not hold for random $\boldsymbol{\psi}$, as in the case of BPE or MLE. To investigate this, we begin with the basic OBP equation:

$$\text{E}(|\tilde{\boldsymbol{\theta}}(\boldsymbol{\psi}) - \boldsymbol{\theta}|^2) = \text{E}\{\text{Q}(\boldsymbol{\psi})\}. \tag{2.13}$$

Let $\hat{\boldsymbol{\psi}}$ be the BPE of $\boldsymbol{\psi}$ obtained by minimizing the observed MSPE, and let $\tilde{\boldsymbol{\psi}}$ denote the MLE. The corresponding OBP-UNIT and EBLUP-UNIT are $\tilde{\boldsymbol{\theta}}(\hat{\boldsymbol{\psi}})$ and $\tilde{\boldsymbol{\theta}}(\tilde{\boldsymbol{\psi}})$, respectively. If the equation (2.13) holds for random $\boldsymbol{\psi}$, one would have

$$\text{E}(|\tilde{\boldsymbol{\theta}}(\hat{\boldsymbol{\psi}}) - \boldsymbol{\theta}|^2) = \text{E}\{\text{Q}(\hat{\boldsymbol{\psi}})\} \leq \text{E}\{\text{Q}(\tilde{\boldsymbol{\psi}})\} = \text{E}(|\tilde{\boldsymbol{\theta}}(\tilde{\boldsymbol{\psi}}) - \boldsymbol{\theta}|^2). \tag{2.14}$$

The inequality follows from the definition of the BPE. To test this, we compare the four terms in the inequality using the same simulation settings as in previous analyses. The results are presented in Tables 2.3 and 2.4.

When the assumed model is completely different from the true model, Table 2.3 indicates that $\mathrm{E}(|\tilde{\boldsymbol{\theta}}(\hat{\boldsymbol{\psi}}) - \boldsymbol{\theta}|^2)$, the first term of (2.14), could be much larger than $\mathrm{E}\{\mathrm{Q}(\hat{\boldsymbol{\psi}})\}$, the second term of (2.14), unless m is extremely large. But $\mathrm{E}\{\mathrm{Q}(\tilde{\boldsymbol{\psi}})\}$, third term of (2.14) is approximately equal to $\mathrm{E}(|\tilde{\boldsymbol{\theta}}(\tilde{\boldsymbol{\psi}}) - \boldsymbol{\theta}|^2)$, the third term of (2.14) even for smaller m. When the assumed model is partially correct, Table 2.4 shows neither $\mathrm{E}(|\tilde{\boldsymbol{\theta}}(\hat{\boldsymbol{\psi}}) - \boldsymbol{\theta}|^2) = \mathrm{E}\{\mathrm{Q}(\hat{\boldsymbol{\psi}})\}$ nor $\mathrm{E}\{\mathrm{Q}(\tilde{\boldsymbol{\psi}})\} = \mathrm{E}(|\tilde{\boldsymbol{\theta}}(\tilde{\boldsymbol{\psi}}) - \boldsymbol{\theta}|^2)$ of (2.14) holds, with $\mathrm{E}(|\tilde{\boldsymbol{\theta}}(\hat{\boldsymbol{\psi}}) - \boldsymbol{\theta}|^2)$ being much larger than $\mathrm{EQ}(\hat{\boldsymbol{\psi}})$, especially for smaller m (e.g., 40). Thus our simulation supports the argument that MSPE of OBP-UNIT could be larger than that of MSPE of EBLUP-UNIT unless m is extremely and unreasonably large.

One potential factor contributing to these differences between $\mathrm{E}(|\tilde{\boldsymbol{\theta}}(\hat{\boldsymbol{\psi}}) - \boldsymbol{\theta}|^2)$ and $\mathrm{E}\{\mathrm{Q}(\hat{\boldsymbol{\psi}})\}$ is the numerical stability of $\hat{\boldsymbol{\psi}}$, BPE of $\boldsymbol{\psi}$. Particularly when $m = 40$, nearly 20%, of BPE estimates yielding a zero value for the ratio σ_v^2/σ_e^2. The performance of OBP-UNIT appears to depend on value of m. The performance of OBP-UNIT improves for extremely large m (e.g., 4000 or 40000), where both sides of the inequality of (2.14), i.e., $\mathrm{E}(|\tilde{\boldsymbol{\theta}}(\hat{\boldsymbol{\psi}}) - \boldsymbol{\theta}|^2) = \mathrm{E}\{\mathrm{Q}(\hat{\boldsymbol{\psi}})\}$ and $\mathrm{E}\{\mathrm{Q}(\tilde{\boldsymbol{\psi}})\} = \mathrm{E}(|\tilde{\boldsymbol{\theta}}(\tilde{\boldsymbol{\psi}}) - \boldsymbol{\theta}|^2)$, tend to be approximately hold, and thus OBP-UNIT eventually perform marginally better than EBLUP-UNIT under the model misspecifications. We hypothesize that the BPE of variance components stabilizes as m increases, suggesting that future research should work on improving the BPE of variance components before proceeding the OBP method under unit-level models. Furthermore, it is noteworthy that the computation of ML or REML estimates are routinely done via available software packages, such as `lme4` or

Table 2.3: Comparison among the four terms in (2.14) under Table 2.1 settings

(m, b)	$\text{MSPE}(\tilde{\boldsymbol{\theta}}(\hat{\boldsymbol{\psi}}))$	$\text{E}\{\text{Q}(\hat{\boldsymbol{\psi}})\}$	$\text{E}\{\text{Q}(\tilde{\boldsymbol{\psi}})\}$	$\text{MSPE}(\tilde{\boldsymbol{\theta}}(\tilde{\boldsymbol{\psi}}))$
(40, 5)	1.775	0.448	1.479	1.474
(100, 5)	1.409	0.918	1.493	1.495
(400, 5)	1.352	1.211	1.500	1.508
(1000, 5)	1.384	1.323	1.504	1.513
(2000, 5)	1.414	1.379	1.521	1.522
(4000, 5)	1.414	1.398	1.512	1.515
(40, 10)	3.919	-1.189	1.649	1.522
(100, 10)	2.364	0.493	1.551	1.513
(400, 10)	1.711	1.231	1.508	1.510
(1000, 10)	1.600	1.389	1.530	1.532
(2000, 10)	1.509	1.412	1.500	1.496
(4000, 10)	1.485	1.431	1.493	1.493

`sae`, which are optimized to ensure numerical stability. In contrast, the computation of BPE associated with the OBP is implemented using the code developed by the authors. As such, this implementation may not yet achieve the same level of computational optimization and stability as these established software packages. We have provided an accessible link to the code and encourage readers to review and examine it.

Table 2.4: Comparison among the four terms in (2.14) under Table 2.2 settings

m	$\text{MSPE}(\tilde{\boldsymbol{\theta}}(\hat{\boldsymbol{\psi}}))$	$\text{E}\{\text{Q}(\hat{\boldsymbol{\psi}})\}$	$\text{E}\{\text{Q}(\tilde{\boldsymbol{\psi}})\}$	$\text{MSPE}(\tilde{\boldsymbol{\theta}}(\tilde{\boldsymbol{\psi}}))$
40	18.230	-14.036	3.247	1.532
100	7.958	-4.809	1.586	1.514
400	3.122	-0.584	1.583	1.508
1000	2.173	0.828	1.580	1.569
2000	1.815	1.131	1.546	1.501
4000	1.650	1.267	1.496	1.505
40000	1.488	1.408	1.462	1.503

2.5 Discussion

In this chapter, we investigate the effects of misspecified mean structure and sampling variance in the well-known nested error regression model on EBLUP and OBP that uses (i) unit-level auxiliary variables only and (ii) area-level auxiliary variables only (i.e., unit context model). Through a series of simulations, we demonstrate that under significant model misspecifications, the OBP procedure for the area-level model and the unit context model outperforms the corresponding unit-level model that relies solely on unit-specific auxiliary variables.

We propose two potential reasons for the underperformance of OBP-UNIT: first, the difference between the sample and population means of the auxiliary variables may negatively affect the performance of OBP-UNIT; second, the numerical stability of the BPE for variance components using unit-level auxiliary variables could be a contributing factor. Our results suggest that utilizing the OBP procedure with area-level auxiliary variables is a promising alternative, particularly when challenges such as the lack of census information and model misspecification is a concern.

Furthermore, when choosing between the unit-context model and the area-level model, we recommend the area-level model as it may offer greater stability. In the area-level model, the variance of sampling errors can be estimated using smoothing methods and is less dependent on the assumed model which may be misspecified. Finally, we recommend further theoretical and empirical research to explore and better understand these unexpected outcomes in the OBP for nested error models that utilize only unit-level auxiliary variables.

Chapter 3: Empirical Best Prediction of Poverty Indicators via Nested Error Regression with High Dimensional Parameters

3.1 Introduction

The United Nations emphasizes the priority of eradicating all forms of poverty by making it the first of its 17 Sustainable Development Goals. Measuring poverty levels across regions is essential for designing effective poverty reduction strategies and ensuring equitable allocation of resources. Accurate information about people's living conditions at a regional level is a basic instrument for targeting policies and programs to reduce poverty. However, progress toward this goal is often slowed by a consistent lack of data on important economic indicators, especially for small population groups and specific geographic areas in developing countries. One major reason for the shortage of disaggregated poverty data is that large scale sample surveys are typically designed to produce reliable estimates of various characteristics of interest for large geographic areas or large population subgroups, rather than for smaller, more granular domains (Chatterjee et al., 2008).

As discussed in Chapter 1, direct (or design-based) estimates can become problematic when sample sizes in subgroups of interest are small (or even zero). In such cases, model-based methods offer a more reliable approach for generating small area estimates. These methods

"borrow strength" across areas by linking models that incorporate auxiliary information from sources like censuses or administrative registers. Beyond the small area context, many measures of poverty and inequality are nonlinear functions of the small area distribution of the economic variable underpinning poverty, usually income or a surrogate such as expenditure. This makes many small area estimation methods, typically developed for estimating linear characteristics such as small area means, inapplicable. In this Chapter, we develop a framework for small area poverty estimation, focusing on a class of poverty measures known as the FGT poverty measures, as discussed in Section 1.1, with the relevant formulas provided in equations (1.2) and (1.3).

The small area estimation (SAE) literature for poverty measures was revolutionized by the pioneering work of Elbers et al. (2003), which introduced a unit-level approach (ELL) for poverty mapping. This method has become widely adopted for producing fine-grained estimates of FGT measures. The ELL approach assumes a two-level model with a common regression relationship between the response and auxiliary variables while allowing for heteroskedastic unit-level errors. The unit-level variance components are estimated using a proposed *alpha* model. Simulated census data are then generated to estimate poverty measures. It is important to note that the ELL approach is a synthetic method that does not rely on a conditional distribution. For further details, we refer readers to their paper.

Another popular approach for poverty estimation introduced by Molina and Rao (2010) is the empirical best prediction (EBP) method through Monte Carlo simulations. This method assumes that the transformed population welfare variables is characterized by a homoskedastic nested error regression (NER) model, as proposed by Battese et al. (1988). Specifically, it assumes that the regression coefficients and variance components are the same across all small areas, with normally distributed errors. Model parameters are estimated using the maximum

likelihood (ML) method or the restricted maximum likelihood (REML) method, and EBPs of poverty indicators are then approximated through random draws from the conditional distribution of the out-of-sample units, given the sample data. Molina et al. (2014) introduced a Bayesian version of this approach, which is particularly useful for quantifying the uncertainty of the estimates.

The effectiveness of such EBP method tends to decline when the number of small areas to be combined becomes large. This limitation arises because the assumption of identical regression coefficients and/or variance components across all small areas, inherent in these models, may not hold universally, leading to potential issues with model misspecification. To address this, random area-specific regression coefficient models (Hobza and Morales, 2013; Prasad and Rao, 1990a; Rao and Molina, 2015) and random area-specific sampling variance models (Arora et al., 1997; Kubokawa et al., 2016; Otto and Bell, 1995; Sugasawa and Kubokawa, 2017) have been suggested in the literature. Such modeling, though useful in some applications, needs more nontrivial assumptions on the joint distribution for a large number of random effects. On the other hand, fixed effects assumptions on the area specific regression coefficients and sampling variances generally lead to unstable estimates of these fixed effects due to small area specific sample sizes (Jiang and Nguyen, 2012).

Lahiri and Salvati (2023) introduced a novel approach in the SAE literature, addressing several challenges. Specifically, this approach assumes fixed effects for both area-specific regression coefficients and sampling variances. But they estimate these coefficients using area-specific (with respect to a tuning parameter) estimating equations applied to data from all areas, and then use appropriately constructed residuals for estimation of variance components. When the tuning parameters are known, they haven shown that the model parameters can be consistently estimated.

This approach, termed as the Nested Error Regression Model with High-Dimensional Parameters (NERHDP), not only offers a more flexible SAE modeling framework but also provides a reliable solution to the challenges of high-dimensional parameter estimation in small area analysis. However, their study focused solely on empirical best predictors (EBPs) for small area means, which are linear functions of the data.

In this Chapter, we extend the NERHDP model to obtain EBPs of poverty indicators, as particular complex and nonlinear FGT measures. Unlike the approach in Molina and Rao (2010), our proposed EBPs account for heterogeneity across small areas in both regression coefficients and sampling variances. Such heterogeneity is common in real-world applications due to variations in socio-economic conditions, data quality, and sampling designs among areas. By extending NERHDP to accommodate FGT poverty measures, we enable the generation of more precise and reliable estimates that are sensitive to area-specific characteristics. This innovation not only enhances the flexibility and applicability of the modeling but also provides a robust methodological framework for addressing the complexities inherent in small area poverty estimation.

Additionally, we address the computational challenges associated with the algorithm proposed by Lahiri and Salvati (2023), which involves iterative convergence and can be computationally intensive. We introduce a more efficient algorithm that significantly reduces computation time, producing estimates within seconds. This improvement ensures the proposed model is not only methodologically robust but also practical for large-scale applications.

Another key feature of our extension is its ability to handle out-of-sample areas, which is particularly important when analyzing a large number of areas. In Molina and Rao (2010), the EBPs become purely synthetic estimates for non-sampled areas due to the absence of observations. In this study, we introduce a new method that enables the estimation of area-specific model

parameters even for non-sampled areas. This approach allows for the estimation of poverty measures that are more tailored to specific areas and so less synthetic. While these estimates remain synthetic, they better capture area-specific heterogeneity compared to those in Molina and Rao (2010). Finally, we propose a parametric bootstrap method for quantifying the uncertainty associated with the estimated poverty measures.

We conduct design-based simulation studies using the 2021 Italian dataset from the European Survey on Income and Living Conditions as the fixed population, which exhibits potential heterogeneity in both regression coefficients and sampling variances, to evaluate the performance of our proposed method. We compare it with existing approaches, including the EBP method of Molina and Rao (2010) (MR), the synthetic method of Elbers et al. (2003) (ELL), and the direct estimator. The simulation results demonstrate that our method has better performance than the existing approaches in estimating headcount ratios and poverty gaps for both in-sample and out-of-sample areas, based on relative bias (RB) and relative root mean squared prediction error (RRMSPE).

We demonstrate the benefits of implementing the proposed approach on data from the 2002 Living Standards Measurement Survey (LSMS) in Albania to estimate headcount ratios and poverty gaps for the 374 municipalities, using auxiliary information from the 2001 Census. Of these municipalities, 161 are out-of-sample, meaning they were not included in the original probability sample used for estimation. Consequently, their characteristics were not directly observed in the survey, and any estimates for these areas rely entirely on model-based predictions.

The structure of the Chapter is as follows. Section 3.2 provides a brief review of the empirical best prediction (EBP) methodology for FGT measures. Section 3.3 introduces the 2002 Living Standards Measurement Survey (LSMS) in Albania, describes the estimation problem, and outlines the available auxiliary information. Section 3.4 introduces the best prediction approach

under a nested error linear regression model with high-dimensional parameters, along with the associated new parameter estimation algorithms. In Section 3.5, we describe the parametric bootstrap method used to estimate measures of uncertainty. Section 3.6 presents the results of design-based simulation studies, comparing the performance of the proposed method with existing approaches in terms of bias and variability. In Section 3.7, we apply the new methodology to LSMS data to estimate headcount ratios (HCRs) and poverty gaps (PGs) at the municipality level. Finally, Section 3.8 concludes the Chapter by summarizing the key findings and suggesting directions for future research.

3.2 A review of Empirical Best Prediction for FGT measures

Like in Molina and Rao (2010), we assume that a random sample of size $n < N$ is drawn from the finite population using a specified sampling design. Let Ω, $s \subset \Omega$, and $r = \Omega - s$ denote the set of indices for all population units, the set of sampled units, and the set of indices for the unsampled units with size $N - n$, respectively. For area i, we have the following associated notations: Ω_i, s_i, and n_i. Note that if area i is not sampled, then $n_i = 0$.

Following the notations from Molina and Rao (2010), we consider a random vector, $\mathbf{Y} = (Y_1, \cdots, Y_N)'$, denoting the values of a random variable associated with the N units of a finite population. Let $\mathbf{y}_s$ be the sub-vector of $\mathbf{y}$ corresponding to sample elements s and $\mathbf{y}_r$ be the sub-vector of out-of sample elements r. After reordering the units of the population, we can write $\mathbf{y} = (\mathbf{y}_s', \mathbf{y}_r')'$. The goal is to predict the value of a real-valued function $\delta = h(\mathbf{y})$ of the random vector $\mathbf{y}$ using the sample data $\mathbf{y}_s$. The best predictor (BP) of δ, which minimizing the mean

squared error (MSE), is given by the conditional expectation

$$\hat{\delta}^B = \mathrm{E}_{\mathbf{y}_r}(\delta|\mathbf{y}_s), \tag{3.1}$$

where the expectation is taken with respect to the conditional distribution of $\mathbf{y_r}$ given the observed sample data $\mathbf{y}_s$.

Suppose there is a one-to-one transformation $Y_{ij} = T(E_{ij})$ of the welfare variables, E_{ij}. The vector $\mathbf{y}$ contains the transformed variables Y_{ij} for all population units, following a normal distribution with mean $\boldsymbol{\mu}$ and variance $\mathbf{V}$. Using this transformation, the random variable $F_{\alpha ij}$ defined in (1.3) can be rewritten as:

$$F_{\alpha ij} = \left(\frac{z - T^{-1}(Y_{ij})}{z}\right)^{\alpha} I(T^{-1}(Y_{ij}) < z) =: h_\alpha(Y_{ij}),\ j = 1, \cdots, N_i.$$

From (3.1), the BP of $\delta = F_{\alpha i}$ is given by:

$$\hat{F}^B_{\alpha i} = \mathrm{E}_{\mathbf{y}_r}(F_{\alpha i}|\mathbf{y}_s) = \frac{1}{N_i}\left\{\sum_{j\in s_i} F_{\alpha ij} + \sum_{j\in r_i} \hat{F}^B_{\alpha ij}\right\} \tag{3.2}$$

where s_i and r_i denote the sets of sample and out-of-sample units belonging to area i, respectively, and $\hat{F}^B_{\alpha ij}$ is the BP of $F_{\alpha ij} = h_\alpha(Y_{ij})$. The BP of $F_{\alpha ij}$ for $j \in r_i$ is given by:

$$\hat{F}^B_{\alpha ij} = \mathrm{E}_{\mathbf{y}_r}[h_\alpha(Y_{ij})|\mathbf{y}_s] = \int h_\alpha(t) f_{Y_{ij}}(t|\mathbf{y}_s)dt, \tag{3.3}$$

where $f_{Y_{ij}}(t|\mathbf{y}_s)$ denotes the conditional density of Y_{ij} given the observed sample data $\mathbf{y}_s$.

Due to the complexity of $h_\alpha(t)$, it may not be feasible to explicitly calculate the expectation

in (3.3). Assuming that $\mathbf{y} = (\mathbf{y}_s', \mathbf{y}_r')'$ follows a multivariate normal distribution with mean vector $\boldsymbol{\mu} = (\boldsymbol{\mu}_s', \boldsymbol{\mu}_r')'$ and a covariance matrix partitioned as:

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_s & \mathbf{V}_{sr} \\ \mathbf{V}_{rs} & \mathbf{V}_r \end{pmatrix},$$

where $\mathbf{V}_s = \text{var}(\mathbf{y}_s)$, $\mathbf{V}_r = \text{var}(\mathbf{y}_r)$, and $\mathbf{V}_{sr} = \mathbf{V}_{rs}' = \text{cov}(\mathbf{y}_s, \mathbf{y}_r)$. In the absence of sample selection bias, the conditional distribution of $\mathbf{y}_r$ given $\mathbf{y}_s$ can be expressed as:

$$\mathbf{y}_r|\mathbf{y}_s \sim N(\boldsymbol{\mu}_{r|s}, \mathbf{V}_{r|s}), \tag{3.4}$$

where the conditional mean $\boldsymbol{\mu}_{r|s}$ and conditional variance $\mathbf{V}_{r|s}$ are given by:

$$\boldsymbol{\mu}_{r|s} = \boldsymbol{\mu}_r + \mathbf{V}_{rs}\mathbf{V}_s^{-1}(\mathbf{y}_s - \boldsymbol{\mu}_s) \text{ and } \mathbf{V}_{r|s} = \mathbf{V}_r - \mathbf{V}_{rs}\mathbf{V}_s^{-1}\mathbf{V}_{sr}. \tag{3.5}$$

Molina and Rao (2010) proposed using an empirical approximation of the conditional expectation in (3.3) through Monte Carlo simulations. Specifically, a large number K of vectors $\mathbf{y}_r$ are generated from the conditional distribution in (3.4). Let $Y_{ij}^{(k)}$ denote the kth simulated value of the out-of-sample observation Y_{ij} for $j \in r_i$, where $k = 1, \cdots, K$. The Monte Carlo approximation to the BP for Y_{ij} for $j \in r_i$ is then expressed as:

$$\hat{F}_{\alpha ij}^B = \text{E}_{\mathbf{y}_r}[h_\alpha(Y_{ij})|\mathbf{y}_s] \approx \frac{1}{K}\sum_{k=1}^{K} h_\alpha(Y_{ij}^{(k)}),\ j \in r_i. \tag{3.6}$$

Since the mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{V}$ typically depend on an unknown parameter

vector $\boldsymbol{\theta}$, the conditional density $f_{Y_{ij}}(t|\mathbf{y}_s)$ also depends on $\boldsymbol{\theta}$. Substituting $\boldsymbol{\theta}$ with its estimate $\hat{\boldsymbol{\theta}}$, we generate simulated values $Y_{ij}^{(k)}$ from the estimated density $f_{Y_{ij}}(t|\mathbf{y}_s, \hat{\boldsymbol{\theta}})$. The resulting empirical best predictor (EBP) of the poverty measure $F_{\alpha i}$ is:

$$\hat{F}_{\alpha i}^{EB} = \frac{1}{N_i}\left[\sum_{j\in s_i} F_{\alpha ij} + \sum_{j\in r_i} \hat{F}_{\alpha ij}^{EB}\right].$$

Note that the EBP method from Molina and Rao (2010) requires linking survey and census households to identify sample indices, which may not be feasible in real-world applications. Often, the survey sample is not a subset of the census. To address this limitation, we follow the approach proposed by Rodas et al. (2021) and apply Monte Carlo simulations to all population units rather than restricting it to non-sampled units. Thus, the EBP can be expressed as:

$$\hat{F}_{\alpha i}^{EB} = \frac{1}{N_i}\sum_{j}^{N_i} \hat{F}_{\alpha ij}^{EB}. \tag{3.7}$$

For consistency, and without causing confusion, we use (3.7) as the EBP for the poverty measures throughout this Chapter. When the sampling fraction n_i/N_i is negligible, which is the most cases in practice, the EBP (3.7) is practically equivalent to the original formulation proposed by Molina and Rao (2010) and further refined by Rodas et al. (2021) and Molina et al. (2022).

3.3 The LSMS and Census data

The primary dataset used in this study is the 2002 Albania Living Standards Measurement Survey (LSMS), conducted by INSTAT in the Spring of 2002. This survey provides comprehensive insights into the living conditions of the Albanian population, capturing both income and non-

income dimensions of poverty. The LSMS sample consists of $3,591$ households, covering a broad range of socio-economic indicators critical for poverty assessment in Albania.

The sample design follows a two-stage cluster sampling strategy, with 450 primary sampling units (PSUs) selected from the 2001 pre-census enumeration areas (EAs). Within each PSU, 8 households were randomly chosen. The sampling frame was stratified into four main regions to ensure representativeness across Tirana, other urban areas, rural areas, and the three major agro-ecological/economic zones (Coastal, Central, and Mountain). The final dataset comprises of:

- $1,000$ households from the Coastal regions,
- $1,000$ households from the Mountain regions,
- 991 households from the Central region,
- 600 households from Tirana.

In the sample, a total of $1,640$ households are rural, and $1,951$ are urban.

The 2002 LSMS data includes observations from 213 of the 374 municipalities in Albania, while the remaining 161 municipalities are entirely out-of-sample, meaning no households were directly surveyed in these areas. Consequently, poverty estimates for these out-of-sample municipalities rely solely on model-based predictions, introducing additional uncertainty and potential extrapolation bias if the characteristics of the out-of-sample areas differ significantly from those of the sampled ones.

Among the observed municipalities, sample sizes vary significantly, ranging from just 6 households in municipalities such as Koder Thu and Hajmel to 600 households in Tirana. The

mean sample size is 16.8, with quartiles at 8 (25th percentile), 8 (50th percentile), and 16 (75th percentile). This indicates that several municipalities have very small sample sizes, which hinders the reliability of direct estimates.

For sampled municipalities, we compute direct estimates of FGT poverty measures, using the weighted average estimator:

$$\hat{F}^{w}_{\alpha i} = \frac{1}{\hat{N}_i} \sum_{j \in s_i} w_{ij} F_{\alpha ij},$$

where w_{ij} represents the survey weight for individual j in municipality i, and $\hat{N}_i = \sum_{j \in s_i} w_{ij}$ is a design-unbiased estimator of the population size N_i. The weighted direct estimates and their variances $\widehat{\text{Var}}(\hat{F}^{w}_{\alpha i})$ are computed using the *direct* function from the `emdi R` package.

Figure 3.1 illustrates the percent coefficient of variation (CV) of direct estimates for HCR and PG, emphasizing the increasing uncertainty in estimates for municipalities with smaller sample sizes. Statistical agencies set thresholds for the reliability of estimates based on their coefficient of variation (CV). For instance, Statistics Canada considers an estimate to be publishable if its CV is below 16.6%, whereas estimates with a CV exceeding 33.3% are deemed unpublishable (Statistics Canada, 2015). Similarly, the UK Office for National Statistics (ONS) sets a threshold of 20% for publication suitability (ONS, 2020). If we set the reliability threshold at 20%, direct survey estimates of the 2002 LSMS dataset would be publishable for only 9 out of the 213 municipalities for HCR and just 2 for PG. This finding underscores the limitations of direct estimation methods in small domains, where high variability often renders estimates unreliable.

The application of SAE methods based on the EBP significantly increases the number of municipalities for which reliable estimates can be obtained. This approach requires fitting an appropriate model to the survey data. Estimated model parameters for fixed and random

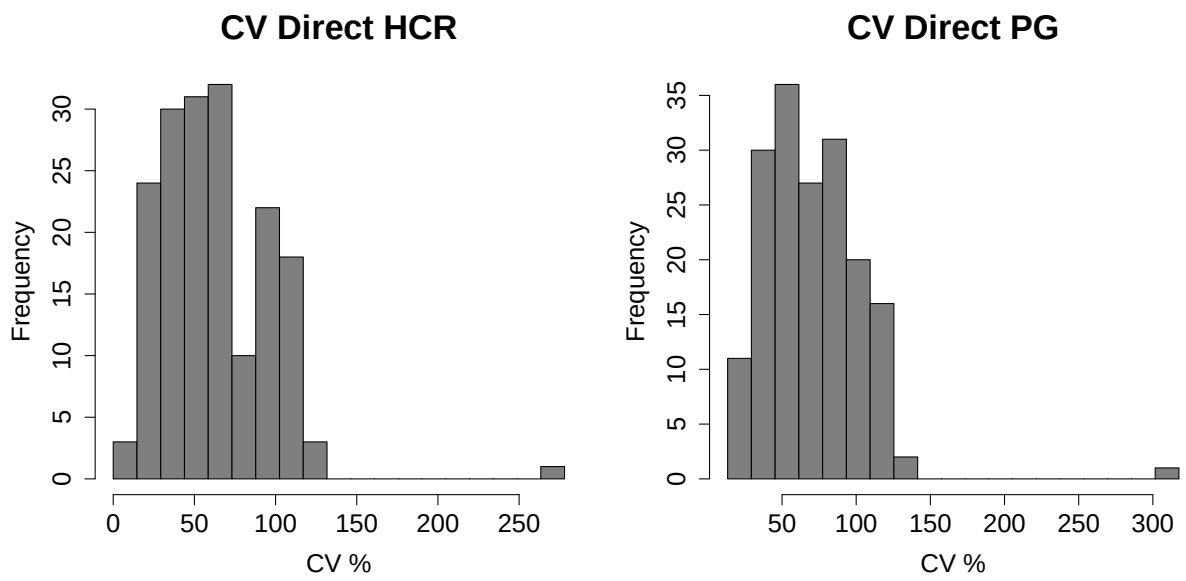


Figure 3.1: Histograms of Percent coefficient of variation of direct estimates for headcount ratios (HCR) and poverty gaps (PG) for all sampled municipalities using 2002 LSMS data.

effects are then combined with known population information for each municipality to predict its levels of HCR and PG. The selection of covariates is based on prior poverty assessment studies in Albania, ensuring consistency with previous methodologies. We employ the same household-level covariates as Betti et al. (2003) and Tzavidis et al. (2008), utilizing information available from both the survey and Census. Specifically, the following variables are considered: household size (a strong predictor of poverty), dwelling facilities (e.g., presence of TV, satellite dish, refrigerator, air conditioning, personal computer), ownership variables (e.g., homeownership, land ownership, car ownership).

To better understand the distribution of these key covariates, Table 3.1 presents the proportions for binary variables and the means and medians for continuous variables from both the LSMS and Census data. This provides insights into the variability of the covariates and their potential predictive power for poverty estimation. Examining these statistics helps assess the representativeness of the sample and the suitability of the selected covariates for small area estimation.

The descriptive statistics highlight a strong consistency between LSMS and Census data across most covariates, reinforcing the validity of using both sources for small area estimation. Household size exhibits nearly identical distributions, with the same median and very similar mean values, suggesting that the sample adequately represents population characteristics. Similarly, ownership variables such as homeownership and land ownership display minimal differences between the two sources, indicating a high degree of reliability in their measurement. Although minor discrepancies arise in asset ownership, such as car ownership and household appliances, these differences are within an expected range and may be attributed to reporting variations between the survey and the Census. Notably, the proportion of households with a television is slightly higher in LSMS than in the Census (0.95 vs. 0.90), while similar patterns are observed for refrigerator and satellite dish ownership. These variations could reflect differences in reporting behavior or coverage between the two data sources. Overall, the alignment of key covariates between the two datasets provides confidence in their joint use for poverty estimation and underscores the representativeness of the survey data in capturing relevant socioeconomic conditions.

Table 3.1: Descriptive statistics of covariates from LSMS and Census data.

Variable	LSMS Data			Census Data		
	Median	Mean/Prop	Std.err.	Median	Mean/Prop	Std.err.
Household size	4	4.28	1.76	4	4.22	1.75
homeownership		0.93	0.26		0.93	0.25
land ownership		0.50	0.50		0.50	0.50
car ownership		0.10	0.30		0.08	0.27
presence of TV		0.95	0.21		0.90	0.30
satellite dish		0.21	0.41		0.23	0.42
refrigerator		0.83	0.37		0.74	0.44
air conditioning		0.02	0.13		0.01	0.11
personal computer		0.02	0.14		0.01	0.12

3.4 Nested error linear regression model with high dimensional parameters

Unlike the traditional NER model used in Molina and Rao (2010), this study extends the NERHDP model for small area poverty estimation, which accounts for heterogeneity across small areas in both regression coefficients and sampling variances. Specifically, let Y_{ij} and $\mathbf{x}_{ij}$ denote the values of the study variable and a $p \times 1$ vector of known auxiliary variables for jth individual of the ith small area. The NERHDP model can be expressed as follows:

$$Y_{ij} = \beta_0 + \mathbf{x}'_{ij}\boldsymbol{\beta}_i + \gamma_i + \epsilon_{ij}, \quad j = 1, \cdots, N_i; \quad i = 1, \cdots, m, \tag{3.8}$$

where the area effects $\gamma_i \overset{\text{iid}}{\sim} N(0, \sigma^2_\gamma)$ and the errors $\epsilon_{ij} \sim N(0, \sigma^2_{\epsilon i})$ are independent. To represent the model at the area level, we define the following vectors and matrices for areas $i = 1, \cdots, m$:

$$\mathbf{y}_i = \underset{1 \le j \le N_i}{\text{col}}(Y_{ij}), \ \boldsymbol{\epsilon}_i = \underset{1 \le j \le N_i}{\text{col}}(\epsilon_{ij}), \ \mathbf{X}_i = \underset{1 \le j \le N_i}{\text{col}}(\mathbf{x}'_{ij}).$$

Under this model, the response vectors $\mathbf{y}_i$, $i = 1, \cdots, m$, are independent with $\mathbf{y}_i \sim N(\boldsymbol{\mu}_i, \mathbf{V}_i)$, where:

$$\boldsymbol{\mu}_i = \beta_0 \mathbf{1}_{N_i} + \mathbf{X}_i\boldsymbol{\beta}_i \ \text{ and } \ \mathbf{V}_i = \sigma^2_\gamma \mathbf{1}_{N_i}\mathbf{1}'_{N_i} + \sigma^2_{\epsilon i}\mathbf{I}_{N_i}.$$

Here, $\mathbf{1}_{N_i}$ is an $N_i \times 1$ vector of ones, and $\mathbf{I}_{N_i}$ is the $N_i \times N_i$ identity matrix. Following with Molina and Rao (2010), for $i = 1, \cdots, m$, we can decompose $\mathbf{y}_i$ into sample and out-of-sample elements $\mathbf{y}_i = (\mathbf{y}'_{is}, \mathbf{y}'_{ir})'$, and the corresponding decomposition on $\mathbf{X}_i$, $\boldsymbol{\mu}_i$ and $\mathbf{V}_i$. When the

sample size $n_i > 0$, the distribution of $\mathbf{y}_{ir}$ given the sample data is

$$\mathbf{y}_{ir}|\mathbf{y}_s \sim N(\boldsymbol{\mu}_{ir|s}, \mathbf{V}_{ir|s}), \tag{3.9}$$

where

$$\boldsymbol{\mu}_{ir|s} = \beta_0 \mathbf{1}_{N_i-n_i} + \mathbf{X}_{ir}\boldsymbol{\beta}_i + \sigma_\gamma^2 \mathbf{1}_{N_i-n_i}\mathbf{1}'_{n_i}\mathbf{V}_{is}^{-1}(\mathbf{y}_{is} - \beta_0\mathbf{1}_{n_i} - \mathbf{X}_{is}\boldsymbol{\beta}_i) \tag{3.10}$$

$$\mathbf{V}_{ir|s} = \sigma_\gamma^2(1-B_i)\mathbf{1}_{N_i-n_i}\mathbf{1}'_{N_i-n_i} + \sigma_{\epsilon i}^2\mathbf{I}_{N_i-n_i}, \tag{3.11}$$

where $\mathbf{V}_{is} = \sigma_\gamma^2 1_{n_i}1'_{n_i} + \sigma_{\epsilon i}^2\mathbf{I}_{n_i}$ and $B_i = \sigma_\gamma^2(\sigma_\gamma^2 + \sigma_{\epsilon i}^2/n_i)^{-1}$.

To avoid the computational burden of simulating multivariate normal vectors $\mathbf{y}_{ir}$ of sizes $N_i - n_i,\ i = 1, \cdots, m$, the conditional distribution in (3.9) can be reformulated using the following model:

$$\mathbf{y}_{ir} = \boldsymbol{\mu}_{ir|s} + u_i\mathbf{1}_{N_i-n_i} + \boldsymbol{e}_{ir}, \tag{3.12}$$

where u_i represents new random effects, and $\boldsymbol{\epsilon}_{ir}$ denotes the error terms. These components are independent and satisfy the following distributions:

$$u_i \sim N(0, \sigma_\gamma^2(1-B_i)), \quad i = 1, \cdots, m, \quad \text{and} \quad \boldsymbol{e}_{ir} \sim N(\mathbf{0}_{N_i-n_i}, \sigma_{\epsilon i}^2\mathbf{I}_{N_i-n_i}).$$

This reformulation simplifies the simulation process while maintaining the statistical properties of the original model.

As mentioned before, in practice the model parameters $\boldsymbol{\phi}_i = (\beta_0, \boldsymbol{\beta}'_i, \sigma_\gamma^2, \sigma_{\epsilon i}^2)$ for $i =$

$1, \cdots, m$, need to be replaced by suitable estimators $\hat{\boldsymbol{\phi}}_i = (\hat{\beta}_0, \hat{\boldsymbol{\beta}}_i', \hat{\sigma}_\gamma^2, \hat{\sigma}_{\epsilon i}^2)$. Then the variables Y_{ij} are generated from the corresponding estimated normal distributions. Given the practical challenges in directly linking the survey sample to the population and under the assumption that the sampling fraction is negligible, for in-sample areas, the elements $\{Y_{ij}, j = 1, \cdots, N_i\}$ are generated from

$$Y_{ij} = \hat{\beta}_0 + \mathbf{x}_{ij}'\hat{\boldsymbol{\beta}}_i + \hat{\sigma}_\gamma^2 \mathbf{1}_{n_i}' \hat{\mathbf{V}}_{is}^{-1}(\mathbf{y}_{is} - \hat{\beta}_0 \mathbf{1}_{n_i} - \mathbf{X}_{is}\hat{\boldsymbol{\beta}}_i) + u_i + e_{ij},$$

where $u_i \overset{\text{iid}}{\sim} N(0, \hat{\sigma}_\gamma^2(1 - \hat{B}_i))$, and $e_{ij} \overset{\text{iid}}{\sim} N(0, \hat{\sigma}_{\epsilon i}^2)$. If an area i is not sampled, then $\{Y_{ij}, j = 1, \cdots, N_i\}$ are generated by bootstrap from

$$Y_{ij} = \hat{\beta}_0 + \mathbf{x}_{ij}'\hat{\boldsymbol{\beta}}_i + \gamma_i^* + \epsilon_{ij}^*,$$

where $\gamma_i^* \overset{\text{iid}}{\sim} N(0, \hat{\sigma}_\gamma^2)$ and $\epsilon_{ij}^* \overset{\text{iid}}{\sim} N(0, \hat{\sigma}_{\epsilon i}^2)$.

Repeat the above generation process K times. Then use the formulas in (3.6) and (3.7) obtain the EBP approximation of $F_{\alpha i}$:

$$\hat{F}_{\alpha i}^{EB} \approx \frac{1}{K}\sum_{k=1}^{K}\left\{\frac{1}{N_i}\sum_{j}^{N_i} h_\alpha(Y_{ij}^{(k)})\right\},$$

for all areas of interest.

3.4.1 Estimation of the high dimensional parameters

In this study, we introduce a data-driven method for estimating the model parameters $\boldsymbol{\phi}_i = (\beta_0, \boldsymbol{\beta}_i', \sigma_\gamma^2, \sigma_{\epsilon i}^2)$ for $i = 1, \cdots, m$, which effectively leverages data across areas to provide consistent estimators for the area-specific parameters when the tuning parameters are known. Moreover, our approach offers a significant improvement over Lahiri and Salvati (2023) by drastically reducing computational time while maintaining comparable accuracy and performance. These enhancements make our method both more efficient and practical for the analysis of large-scale datasets. The algorithm is detailed in the following steps.

Step 1 Define $r_{l;i} = q_i^{-1/2}(\mathbf{y}_{ls} - a_{0i}\mathbf{1}_{n_l} - \mathbf{X}_{ls}\boldsymbol{\beta}_i)$, where q_i is a known scale parameter, $\mathbf{y}_{ls}$ is a $n_l \times 1$ vector of the response variable and $\mathbf{X}_{ls}$ denotes the matrix of dimension $n_l \times p$ individual-level covariates of the sampled units in area l. Obtain the area-specific regression coefficients estimates $\{(\hat{a}_{0i}, \hat{\boldsymbol{\beta}}_i), i = 1, \cdots, m\}$ via solving the following system of equations:

$$\sum_{l=1}^{m} \mathbf{X}_{ls(p+1)}' q_i^{1/2} \psi_i(r_{l;i}) = 0, \tag{3.13}$$

where $\mathbf{X}_{ls(p+1)}$ is the matrix of dimension $n_l \times (p+1)$ including the intercept term and covariates of the sampled units in area l, and $\psi_i(r_{l;i})$ is a $n_l \times 1$ vector obtained from the vector of $r_{l;i}$ with its jth element, say $r_{lj;i}$ replaced by $\psi_i(r_{lj;i})$, a chosen known function of $r_{lj;i}$.

Remark 1 In this section, we assume that the function ψ_i is fully specified. For example, consider the form:

$$\psi_i(r) = 2\psi(r)[\tau_i I(r > 0) + (1 - \tau_i) I(r \leq 0)], -\infty < r < \infty, \tag{3.14}$$

where $\psi(r)$ is a known monotone non-decreasing function satisfying $\psi(-\infty) < \psi(0) < \psi(\infty)$, and $\tau_i \in T = (0, 1)$ is a known tuning parameter. Notably, setting $\tau_i = 0.5$ with the identify function for ψ would lead to the standard least square estimator of the regression coefficient vector. The tuning parameter τ_i introduces small area-specific effects into the estimation of the regression coefficients $\boldsymbol{\beta}_i$ and the sampling variances $\sigma^2_{\epsilon i}$, enhancing the flexibility and adaptability of the model. Following Chambers and Tzavidis (2006), we assume that $\psi(\cdot)$ is the Huber influence function, with a tuning constant $c = 1.345$, ensuring robust performance in the presence of outliers.

Step 2 Define $\tilde{\mathbf{r}}_{l;i} = q_i^{-1/2}(\mathbf{y}_{ls} - \hat{a}_{0i}\mathbf{1}_{n_l} - \mathbf{X}_{ls}\hat{\boldsymbol{\beta}}_i)$ and $\tilde{\mathbf{V}}_{l;i} = \sigma^2_\gamma \mathbf{1}_{n_l}\mathbf{1}'_{n_l} + \sigma^2_{\epsilon i}\mathbf{I}_{n_l}$. Obtain $\{\hat{\sigma}^2_{\epsilon i}, i = 1, \cdots, m\}$ as a solution of the following system of estimation equations:

$$\sum_{l=1}^{m} \left[\psi_i(\tilde{\mathbf{r}}_{l;i})' \, \tilde{\mathbf{V}}^{-1}_{l;i} \, \frac{\partial \mathbf{V}_{i;i}}{\partial \sigma^2_{\epsilon i}} \, \tilde{\mathbf{V}}^{-1}_{l;i} \, \psi_i(\tilde{\boldsymbol{r}}_{l;i}) \; - \mathrm{tr}(\tilde{\mathbf{V}}^{-1}_{l;i} \, \frac{\partial \mathbf{V}_{i;i}}{\partial \sigma^2_{\epsilon i}}) \right] = 0 \tag{3.15}$$

where $i = 1, \cdots, m$, and $\mathbf{V}_{i;i} = \sigma^2_\gamma \mathbf{1}_{n_i}\mathbf{1}'_{n_i} + \sigma^2_{\epsilon i}\mathbf{I}_{n_i}$.

Step 3 Define $\bar{\mathbf{r}}^\star = \{\bar{r}^\star_i; i \in s\}'$, where $\bar{r}^\star_i = n_i^{-1} \sum_{j=1}^{n_i} r^\star_{ij}$ and $r^\star_{ij} = Y_{ij} - \mathbf{x}'_{ij}\hat{\boldsymbol{\beta}}_i$ with $\hat{\boldsymbol{\beta}}_i$ from Step 1. We estimate β_0 by taking the grand residual mean: $\hat{\beta}_0 = 1/n \sum_{i\in s} \sum_{j=1}^{n_i} r^\star_{ij}$. Define $v^\star_i = \sigma^2_\gamma + \hat{d}_i$, where $\hat{d}_i = \hat{\sigma}^2_{\epsilon i}/n_i$ for $i \in s$. Obtain $\hat{\sigma}^2_\gamma$ as a solution of the following estimating equation:

$$\sum_{i\in s} \left[\psi^2\left((v^\star_i)^{-1/2} \bar{r}^\star_i \right) (v^\star_i)^{-1} - (v^\star_i)^{-1} w^\star \right] = 0 \tag{3.16}$$

where $w^\star = \mathrm{E}[\psi^2(u)]$ with u following a standard normal distribution.

3.4.2 Definition and estimation of tuning parameter τ_i

Let the subscript U denotes a finite population indicators. Define the following finite population parameters:

$$\begin{aligned} Y_{ij;\tau} &= \alpha_{0\tau;U} + \mathbf{x}'_{ij}\boldsymbol{\beta}_{\tau;U}; \\ \tau_{ij} &= \text{argmin}_{\tau\in T}(Y_{ij;\tau} - Y_{ij})^2, \end{aligned}$$

for $i = 1, \cdots, m$; $j = 1, \cdots, N_i$. Note that $Y_{ij;\tau}$ and τ_{ij} are unknown even for sampled units since $\alpha_{0\tau;U}$ and $\boldsymbol{\beta}_{\tau;U}$ are unknown finite population parameters. We are interested in estimating the following finite population parameter: $\tau_i = 1/N_i \sum_{j=1}^{N_i} \tau_{ij}$.

For the sampled areas, we use the same data-driven method as Lahiri and Salvati (2023) to estimate the tuning parameters, τ_i with $i \in s$. Specifically, for a fine grid $\tau \in T$, we obtain a collection of fitted values for the entire sample:

$$\hat{Y}_{ij;\tau} = \hat{\alpha}_{0\tau} + \mathbf{x}'_{ij}\hat{\boldsymbol{\beta}}_{\tau}, \quad i \in s; \quad j = 1, \cdots, n_i,$$

using the standard quantile or M-quantile methods, where $\hat{\alpha}_{0\tau}$ and $\hat{\boldsymbol{\beta}}_{\tau}$ are estimated intercept and regression coefficients, respectively. For each observation Y_{ij} in the sample, we identify the fitted line that minimizes the prediction error, defined as the difference between Y_{ij} and the predicted value from the fitted regression at $\mathbf{x}_{ij}$. Let $\hat{\tau}_{ij}$ denote the value of τ in the grid corresponding to this best-fitting line. The variability of $\hat{\tau}_{ij}$ reflects unit-level variability. Provided there are sample observations in area i, and a non-informative sampling method has been used to obtain them, one

can consider an estimate of the area-i-specific tuning parameter is the sample average of $\hat{\tau}_{ij}$ for that area: $\bar{\hat{\tau}}_i = n_i^{-1} \sum_{j \in s_i} \hat{\tau}_{ij}$. Lahiri and Salvati (2023) also introduced an empirical linear best (ELB) predictor for τ_i denoted as $\hat{\tau}_i^{\text{ELB}}$, which enhances the performance of $\bar{\hat{\tau}}_i$ by leveraging data from all areas. For a detailed explanation, we refer readers to their paper.

For out-of-sample areas with no observations, we propose the following unmatched model to leverage information from observed areas, enabling the estimation of area-specific tuning parameters for completely unobserved regions:

$$\begin{aligned} &\text{E}(\bar{\hat{\tau}}_i|\tau_i) = \tau_i, \quad \text{Var}(\bar{\hat{\tau}}_i|\tau_i) = \Delta_i; \\ &\text{E}[\text{logit}(\tau_i)] = \bar{\mathbf{Z}}_i'\boldsymbol{\eta}, \end{aligned}$$

where $i = 1, \ldots, m$. The variance components Δ_i are typically assumed to be known; however, in practice, they need to be estimated. In this study, we employ a smoothing technique to estimate Δ_i

$$\hat{\Delta}_i = \frac{1}{n_i} \sum_{i \in s} \sum_{j=1}^{n_i} \frac{(\hat{\tau}_{ij} - \bar{\hat{\tau}}_i)^2}{n - m_s}, \quad i \in s,$$

where m_s is the number of sampled areas. Here, $\bar{\mathbf{Z}}_i$ represents the population mean of relevant auxiliary variables, and $\boldsymbol{\eta}$ is the unknown vector of coefficients, estimated by $\hat{\eta}$ that minimizes the following objective function, $\mathbf{Q}(\boldsymbol{\eta})$:

$$\mathbf{Q}(\boldsymbol{\eta}) = \sum_{i \in s} \frac{1}{\hat{\Delta}_i} \left[\bar{\hat{\tau}}_i - \frac{\exp\left(\bar{\mathbf{Z}}_i'\boldsymbol{\eta}\right)}{1 + \exp\left(\bar{\mathbf{Z}}_i'\boldsymbol{\eta}\right)} \right]^2.$$

The estimated tuning parameters for the out-of-sample areas are then obtained as follows:

$$\hat{\tau}_i = \frac{\exp\left(\bar{\mathbf{Z}}_i'\hat{\boldsymbol{\eta}}\right)}{1+\exp\left(\bar{\mathbf{Z}}_i'\hat{\boldsymbol{\eta}}\right)}, \quad i \notin s.$$

Then, the area-specific model parameters, $\boldsymbol{\phi}_i = (\beta_0, \boldsymbol{\beta}_i', \sigma_\gamma^2, \sigma_{\epsilon i}^2)$ for $i = 1, \cdots, m$, can be estimated using the algorithm outlined in Section 3.4.1, with the tuning parameters τ_i replaced by their estimates, $\hat{\tau}_i$, for all small areas.

3.5 Uncertainty measures

The assessment of uncertainty of poverty indicators is crucial to analyze the quality of estimates. The MSPE of $\hat{F}_{\alpha i}^{EB}$ is a conventional measure fulfilling this goal, which is given by

$$\text{MSPE}(\hat{F}_{\alpha i}^{EB}) = E_{\boldsymbol{\phi}}(\hat{F}_{\alpha i}^{EB} - F_{\alpha i})^2, \tag{3.17}$$

where $E_{\boldsymbol{\phi}}$ denotes expectation with respect to the super-population model. However, the analytical approximations to the MSPE are difficult to derive when involving complex parameters such as the FGT poverty measures.

In this section, we propose a parametric bootstrap method that can be applied to produce reasonable estimators of various uncertainty measures, not necessarily MSPE. Our bootstrap scheme builds on the method introduced by Lahiri and Salvati (2023), which demonstrated that for known τ_i, the parametric bootstrap estimators of uncertainty measures converge in probability to the true corresponding uncertainty measures as m tends to infinity, under certain regularity conditions. Our approach adapts this framework to the context of poverty mapping. The steps of

the proposed bootstrap are as follows:

Step 1. Fit model (3.8) to sample data and obtain estimators $\hat{\boldsymbol{\phi}}_i = (\hat{\beta}_0, \hat{\boldsymbol{\beta}}_i', \hat{\sigma}_\gamma^2, \hat{\sigma}_{\epsilon i}^2)$ for $i = 1, \cdots, m$, using the proposed method in section 3.4;

Step 2. Given $\hat{\boldsymbol{\phi}}_i$ from Step 1, generate B parametric bootstrap populations using the following model:

$$Y_{ij}^{*(b)} = \hat{\beta}_0 + \mathbf{x}_{ij}'\hat{\boldsymbol{\beta}}_i + u_i^{*(b)} + e_{ij}^{*(b)}, \tag{3.18}$$

where $u_i^{*(b)} \sim N(0, \hat{\sigma}_\gamma^2)$ and $e_{ij}^{*(b)} \sim N(0, \hat{\sigma}_{\epsilon i}^2)$ are all independently distributed, $i = 1, \cdots, m$, $j = 1, \cdots, N_i$.

Step 3. For each population, calculate bootstrap population parameters $F_{\alpha i}^{*(b)} = N_i^{-1} \sum_{j=1}^{N_i} F_{\alpha ij}^{*(b)}$, where $F_{\alpha ij}^{*(b)} = h_\alpha(Y_{ij}^{*(b)})$, $b = 1, \cdots, B$.

Step 4. Using the same $\hat{\boldsymbol{\phi}}_i$ from Step 1, generate bootstrap samples with the same size as the original sample. Specifically, for each bootstrap population b, we have the following bootstrap sample:

$$Y_{ij;s}^{*(b)} = \hat{\beta}_0 + \mathbf{x}_{ij;s}'\hat{\boldsymbol{\beta}}_i + u_{i;s}^{*(b)} + e_{ij;s}^{*(b)}, i \in s, j = 1, \cdots, n_i, \tag{3.19}$$

where $\{u_{i;s}^{*(b)}; i \in s\}$ are selected using the same area indices as in the original sample from $\{u_i^{*(b)}; i = 1, \cdots, m\}$, which are obtained in Step 2. Here $\{e_{ij;s}^{*(b)}; i \in s, j = 1, \cdots, n_i\}$ are independently generated from $N(0, \hat{\sigma}_{\epsilon i}^2)$. These errors are not subsets of the population

errors generated in Step 2 but are instead independently drawn from the same distributions.

Step 5. Fit model (3.8) to bootstrap samples and calculate the bootstrap EBPs, $\hat{F}_{\alpha i}^{EB*(b)}$, $b = 1, \cdots, B$, as described in Sections 3.2 and 3.4.

Step 6. A Monto Carlo approximation to the theoretical bootstrap estimator $\text{MSPE}_*(\hat{F}_{\alpha i}^{EB*(b)}) = E_{\boldsymbol{\phi}^*}(\hat{F}_{\alpha i}^{EB*(b)} - F_{\alpha i}^{*(b)})^2$ of $\hat{F}_{\alpha i}^{EB}$ is calculated as

$$\text{mspe}_*(\hat{F}_{\alpha i}^{EB}) = \frac{1}{B}\sum_{b=1}^{B}(\hat{F}_{\alpha i}^{EB*(b)} - F_{\alpha i}^{*(b)})^2. \tag{3.20}$$

The estimator $\text{mspe}_*(\hat{F}_{\alpha i}^{EB})$ is used to estimate $\text{MSPE}(\hat{F}_{\alpha i}^{EB})$ given in (3.17).

3.6 Design-based Monte Carlo simulations

In this section, we perform design-based simulation experiments to assess the performance of various estimators for poverty measures across repeated samples drawn from a fixed population. Specifically, we utilize the 2021 Italian dataset from the European Survey on Income and Living Conditions (EUSILC), comprising $29,798$ observations. The dataset is treated as the fixed population, with Italy's 107 provinces serving as the areas of interest. This setup allows us to compute the true area-level headcount ratios and poverty gaps, serving as a benchmark for evaluation:

$$F_{\alpha i} = \frac{1}{N_i}\sum_{j=1}^{N_i} F_{\alpha ij}, \quad i = 1, \cdots, 107,$$

where

$$F_{\alpha ij} = \left(\frac{z - E_{ij}}{z} \right)^{\alpha} I(E_{ij} < z),\ j = 1, \cdots, N_i.$$

Here, E_{ij} denotes the equalized annual net income for each individual, and z represents the fixed poverty line, which is roughly equal to 0.6 times the median of the welfare variables E_{dj} for the above fixed population.

Following Molina and Rao (2010), the welfare variable E_{ij} is transformed to ensure positivity by adding a fixed constant c, and then taking the logarithm. Specifically, the response variable is defined as $Y_{ij} = \log(E_{ij}+c)$. The auxiliary variables used in the model comprise both continuous and categorical types. The continuous variables include age, house size in square meters (`MQ`), household size (`hsize`), and total years of work experience (`Years_work`). Binary indicators consist of working status (`Work_Status`) and gender, while categorical variables include six levels of income and living conditions (`cond_ilo`) and marital status (`Marital_status`). Additionally, life satisfaction (`life_satisf`) is treated as a scale variable.

To explore the data, we calculate the sample variance of the response variable to understand its dispersion within each province. Additionally, we fit a regression model for the response variable using the auxiliary variables as predictors, separately for each province. This approach allows us to examine the relationships between the response and auxiliary variables while accounting for provincial differences in patterns and variability. Figure 3.2 shows boxplots of the distribution of residuals by province and the distribution of regression coefficients. These results indicate that the regression coefficients vary significantly across provinces, and within-area variation may also differ between provinces. Consequently, the assumptions of the traditional NER model may not

hold for this dataset.

In the simulation experiments, we consider the following estimators of the FGT measures for each province:

(i) Direct estimator (DIRECT),

(ii) The EBP based on the traditional NER model from Molina and Rao (2010) (MR),

(iii) The proposed EBP based on the NERHDP model (CLS),

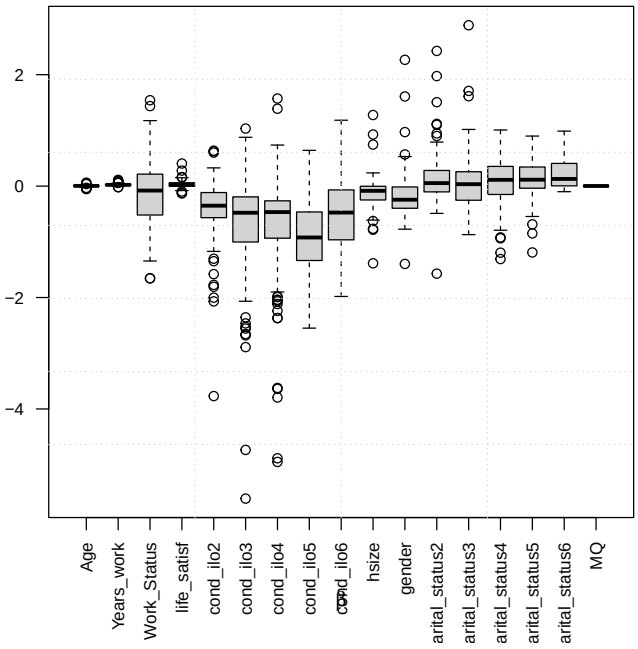
(iv) The ELL estimator using the algorithm from Elbers et al. (2003) (ELL).

For the EBP methods, we use $K = 100$ for the Monte Carlo approximation, and for the ELL method, 100 simulated censuses are generated.

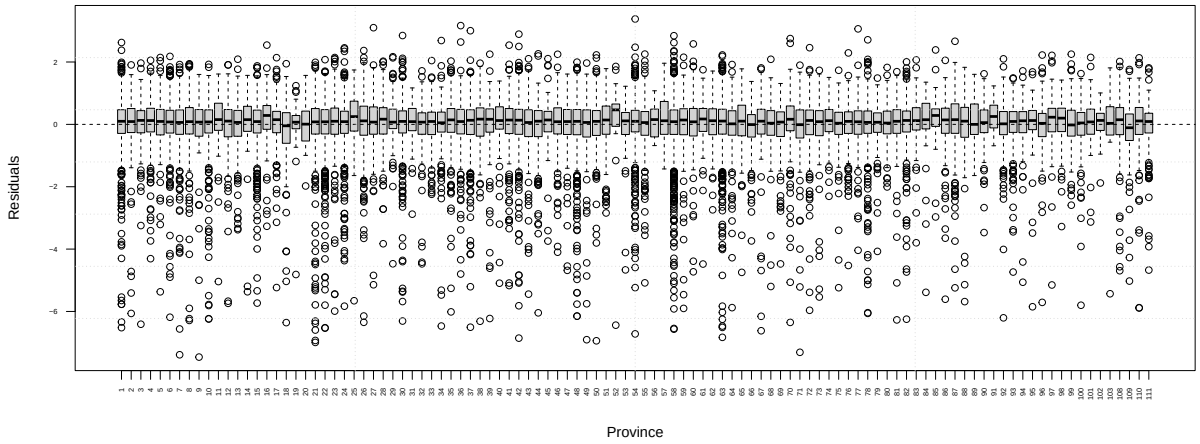
Two experimental settings are considered: (1) all 107 provinces in the fixed population are sampled; (2) only 90 provinces are sampled, leaving 17 provinces with no observations. For each setting, 1000 samples are drawn using simple random sampling (SRS) without replacement within each sampled area, with the sample size set at $0.1 \times N_i$. The performance of the above estimators is assessed using the following two criteria:

(a) Relative Bias (RB): The average RB over all small areas. For a given area, RB is defined as the ratio of the average difference between the estimated and true small-area FGT measures to the true small-area FGT measure, averaged over simulations.

(b) Relative Root Mean Squared Prediction Error (RRMSPE): For a given area, RRMSPE is the ratio of the square root of the average squared difference between the estimated and true small-area FGT measures to the true measure, averaged over simulations.



(a) Distributions of estimated regression coefficients.



(b) Distributions of the residuals by province.

Figure 3.2: Distributions of estimated regression coefficients (a) and residuals (b) obtained by fitting a regression model separately for each province.

(c) Efficiency (EFF) measured as the ratio of the RRMSPE of each estimator/predictor to the RRMSPE of the corresponding the baseline estimator. DIRECT and MR serve as the baseline estimators in experiments 1 and 2, respectively.

First, the computational efficiency of the new parameter estimation algorithm is evaluated, comparing it to the existing algorithm proposed by Lahiri and Salvati (2023). All computations are carried out on a `13th Gen Intel(R) Core(TM) i7-13700F 2.10 GHz` processor. For this comparison, five samples are generated using SRS without replacement within each province, with the sample size set at $0.02 \times N_i$, resulting in a fixed total sample size of 616. Both algorithms are applied to these five samples, and their computation times are recorded. The algorithm proposed by Lahiri and Salvati (2023) requires an average computation time of 1197.59 seconds, while the new algorithm completes the same task in just 2.17 seconds on average. This demonstrates a substantial improvement, as our algorithm requires only about 0.2% of the computation time compared to the existing algorithm.

3.6.1 Experiment 1: All provinces are sampled

In this simulation experiment, we assume that all 107 provinces are included in the samples ($m_s = m = 107$). For each province, samples are repeatedly drawn with a size equal to $0.1 \times N_i$. The total sample size is 2979 and the province-specific sample sizes range from 2 to 180. In this scenario, all four estimators are available for all provinces.

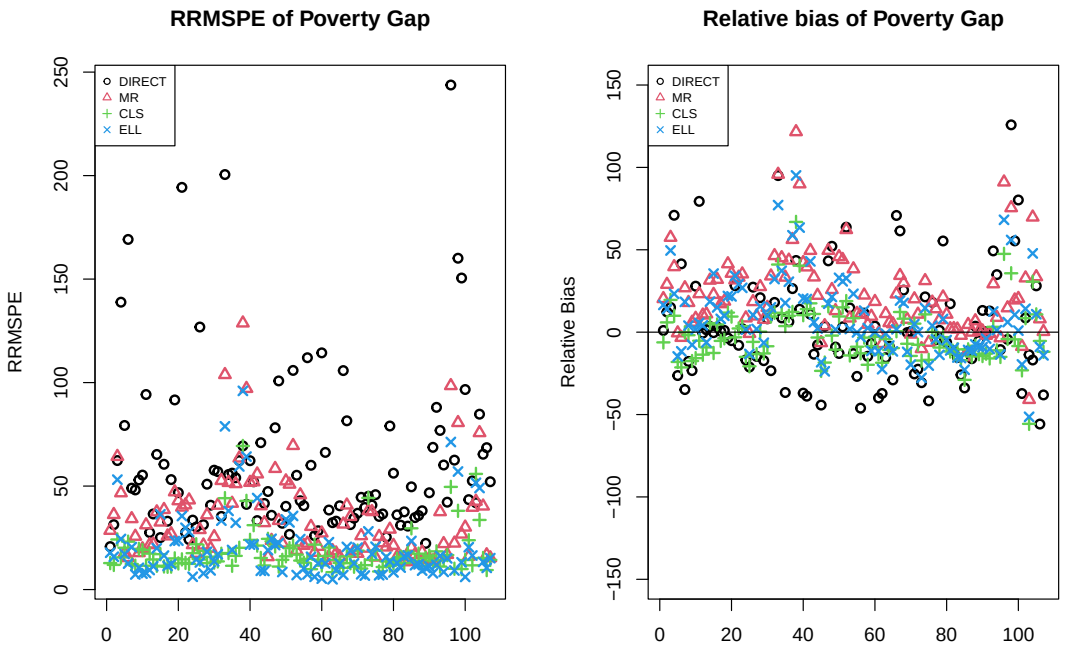
Table 3.2 provides a summary of the overall performance of the four estimators, evaluated based on average RB, average RRMSPE, and EFF, with the DIRECT estimator serving as the baseline for comparison. CLS consistently outperforms MR in terms of both bias and RRMSPE,

likely due to the limitations of the MR method, which assumes identical regression coefficients and variance components across the population—assumptions that may not be valid in this context. Additionally, ELL shows better performance than MR, in contrast to the findings in Molina and Rao (2010). This difference may stem from our use of the algorithm from Elbers et al. (2003), which incorporates heteroscedasticity at the individual level, rather than the simplified version outlined in Molina and Rao (2010). For the poverty gap indicator, CLS outperforms in both relative bias and RRMSPE, while its performance is comparable to ELL for the headcount ratio.

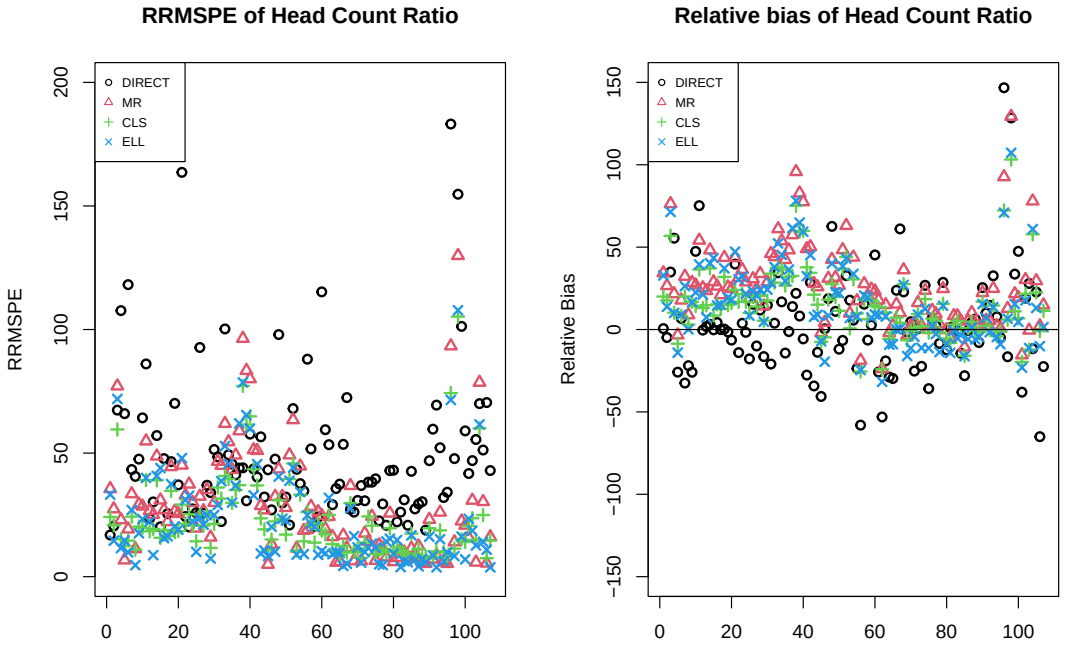
Figure 3.3 illustrates the province-specific performance of the estimators. For the poverty gap indicator, Figure 3.3a shows that CLS outperforms MR in terms of both RRMSPE and relative bias for most provinces, with MR exhibiting a tendency toward positive bias. Similarly, Figure 3.3b highlights the performance of the four methods for estimating the headcount ratio, demonstrating that CLS achieves smaller RRMSPE and absolute bias than MR in the majority of provinces.

Table 3.2: Performance of estimators/predictors of small area FGT measures when all areas are sampled; Average is over 107 areas; numbers in parenthesis are the values of the efficiency over DIRECT in terms of RRMSPE.

Predictor	RRMSPE % (EFF)		Relative Bias %	
	Headcount Ratio	Poverty Gap	Headcount Ratio	Poverty Gap
DIRECT	48.71 (1.00)	60.43 (1.00)	4.16	5.14
MR	28.73 (0.59)	33.39 (0.55)	25.32	21.71
CLS	22.94 (0.47)	17.81 (0.29)	15.34	-2.03
ELL	22.48 (0.46)	20.25 (0.34)	15.61	8.18



(a) The RRMSPE and Relative Bias of the four estimators of poverty gap for each area.



(b) The RRMSPE and Relative Bias of the four estimators of headcount ratio for each area.

Figure 3.3: The RRMSPE and Relative Bias of the four estimators of FGT measures, when all provinces are sampled.

3.6.2 Experiment 2: 17 out-of-sample provinces

In the second experiment, we assess the performance of EBP methods and the ELL method in estimating FGT measures for out-of-sample provinces, where the direct estimator is unavailable due to the absence of observations. Specifically, 90 provinces are randomly selected, and this selection is fixed to draw simple random samples (SRS) repeatedly from these provinces. Consequently, 17 provinces are consistently excluded from the sampling process, ensuring they have no observations across all simulations.

Table 3.3 summarizes the overall performance of the three estimators for the out-of-sample provinces. The results align with those observed in Experiment 1. Overall, CLS outperforms MR in terms of both RRMSPE and relative bias for both FGT measures. CLS also shows better performance than ELL for the poverty gap measure, while their performance is comparable for the headcount ratio.

Figure 3.4a illustrates the province-specific performance for the 17 out-of-sample provinces, demonstrating that the CLS model consistently outperforms both MR and ELL for the poverty gap measure in terms of RRMSPE and relative bias. Notably, CLS achieves significantly smaller RRMSPE than MR for the majority of out-of-sample areas. This superior performance can be

Table 3.3: Performance of estimators/predictors of small area FGT measures when some areas are not sampled; Average is over 17 out-of-sample areas; numbers in parenthesis are the values of the efficiency over MR in terms of RRMSPE.

Predictor	RRMSPE % (EFF)		Relative Bias %	
	Headcount Ratio	Poverty Gap	Headcount Ratio	Poverty Gap
MR	35.36 (1.00)	41.87 (1.00)	31.50	29.78
CLS	27.49 (0.78)	16.69 (0.40)	18.83	-0.93
ELL	26.04 (0.74)	20.55 (0.49)	19.17	10.46

explained by the flexibility of the CLS model, which allows regression coefficients and sampling variances to vary across areas, enabling it to better capture region-specific heterogeneity.

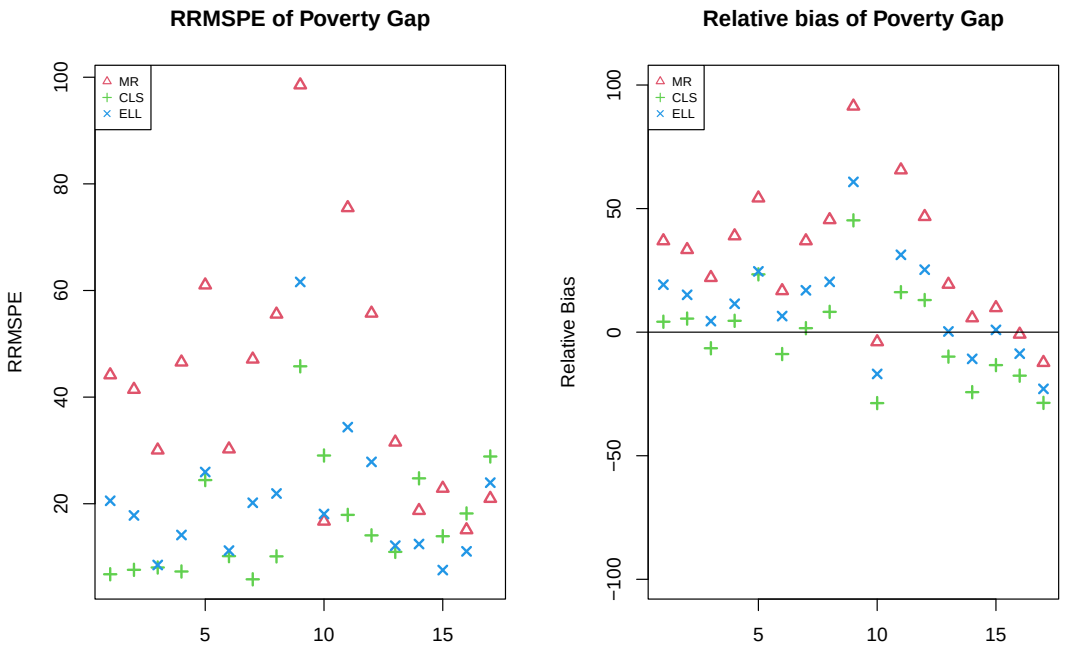
For the headcount ratio, Figure 3.4b shows that CLS also performs better than MR, although the margin of improvement is smaller compared to the poverty gap measure. These results highlight the capability of the CLS model to adapt to area-level differences, offering enhanced precision and reduced bias in poverty estimates.

3.7 Application: poverty mapping for Albania

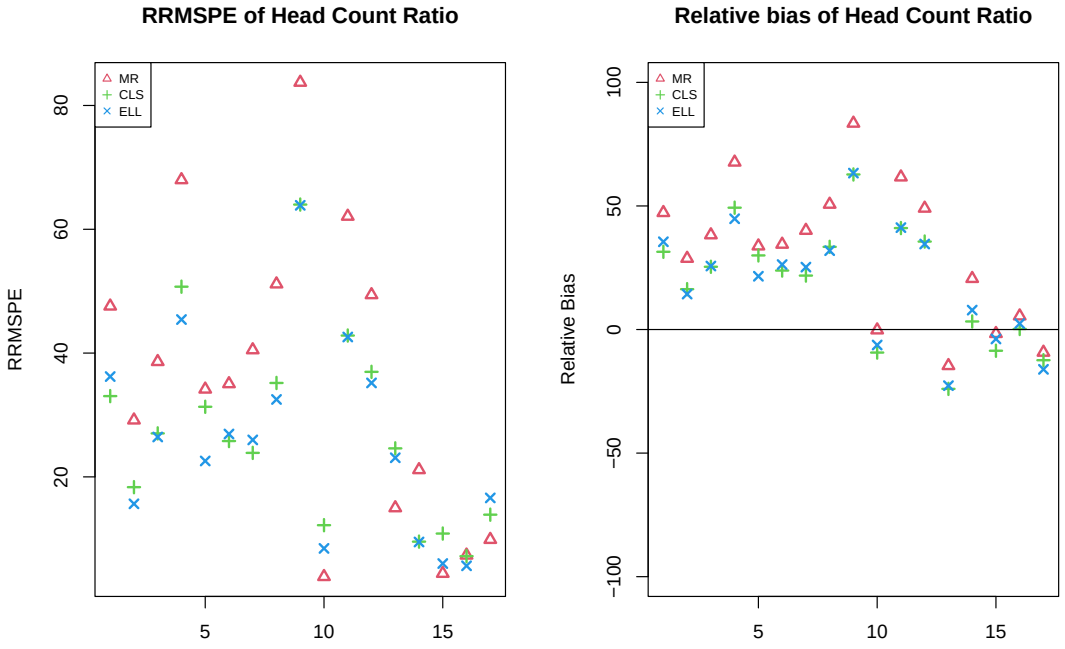
Poverty mapping is a crucial tool for understanding the spatial distribution of poverty and guiding the implementation of poverty alleviation programs. In Albania, where the economy is predominantly rural and income is difficult to measure accurately, poverty is assessed using a consumption-based measure (Betti et al., 2003). In this section, we use 2002 Living Standards Measurement Survey (LSMS) data to estimate poverty indicators for 374 municipalities in Albania. To supplement this information, we incorporate auxiliary data from the 2001 Census, which covers 726,895 households across Albania. The descriptive statistics of these datasets have been presented in Section 3.3, providing an overview of their key characteristics and ensuring their suitability for small area estimation."

Design-based simulation results in Section 3.6 indicate that the CLS method is a promising approach due to: (i) significantly lower bias and RRMSPE compared to the MR method and (ii) greater computational efficiency relative to the existing algorithm proposed in Lahiri and Salvati (2023), improving feasibility for large-scale datasets.

To estimate poverty indicators for Albanian municipalities, we consider household consumption



(a) The RRMSPE and Relative Bias of the EBP estimators of poverty gap for out of sample areas.



(b) The RRMSPE and Relative Bias of the EBP estimators of headcount ratio for out of sample areas.

Figure 3.4: The RRMSPE and Relative Bias of the EBP estimators of FGT measures for out of sample areas.

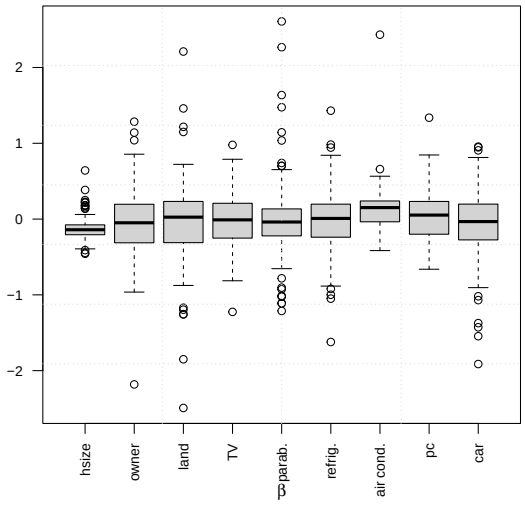
expenditure as the primary welfare variable, denoted as E_{ij}. Following Section 3.6, we apply a shifted logarithm transformation to obtain Y_{ij} as the final response variable in the NERHDP model. The covariates, including household size, dwelling facilities, and ownership variables, are based on prior poverty assessment studies in Albania (Betti et al., 2003; Tzavidis et al., 2008) and have been presented in Section 3.3. The poverty line is set at 4,891 Leks per month for all districts, based on Betti et al. (2003).

For each small area, we estimate the sample variance and fit a regression model where the response variable is the shifted logarithm transformed household consumption expenditure, with covariates previously specified from the LSMS data. Initial data exploration suggests substantial heterogeneity in both within-area variation and regression coefficients across municipalities. Figure 3.5 presents box plots of the distributions of estimated regression coefficients (top panel) and residual distributions by municipality (bottom panel). These results indicate that assuming identical regression coefficients and/or variance components in the MR method is unrealistic for this dataset, emphasizing the need for a more flexible modeling approach that accounts for area-specific variations.

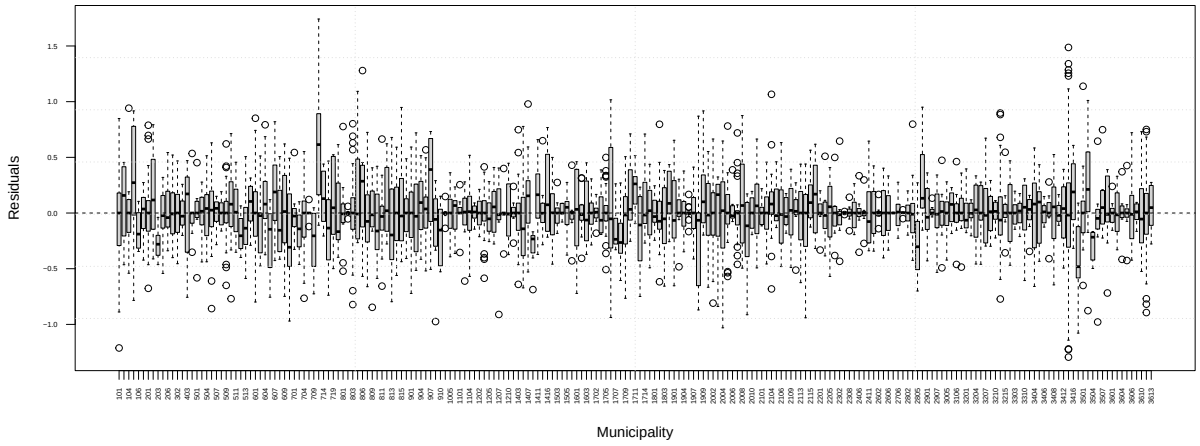
We apply the CLS method to estimate empirical best predictors (EBPs) of headcount ratios and poverty gaps by municipality, denoted as $\hat{F}_{0i}^{CLS}$ and $\hat{F}_{1i}^{CLS}$, respectively. The MSPEs of these FGT measures are estimated using the parametric bootstrap estimator $\text{mspe}_*(\hat{F}_{\alpha i}^{CLS})$ with $B = 100$ replicates. We then compute the associated coefficients of variation (CVs) using:

$$\text{CV}(\hat{F}_{\alpha i}^{CLS}) = \{\text{mspe}_*(\hat{F}_{\alpha i}^{CLS})\}^{1/2} / \hat{F}_{\alpha i}^{CLS}.$$

Following Molina and Rao (2010), we also compute direct estimators of the FGT measures



(a) Distributions of estimated regression coefficients.



(b) Distributions of the residuals by municipality.

Figure 3.5: Distributions of estimated regression coefficients (a) and residuals (b) obtained by fitting a regression model separately for each municipality.

$\hat{F}^w_{\alpha i}$ and their variance estimates $\widehat{\text{Var}}(\hat{F}^w_{\alpha i})$ for sampled municipalities using weighted sample means (see Section 3.3).

To assess the quality of our CLS estimator, we use a set of diagnostic tools based on the requirement that model-based small area estimates should align with the corresponding unbiased

direct estimates, albeit more precise (Marino et al., 2019). Correlation coefficients between CLS and direct estimates for the headcount ratio and poverty gap are 0.796 and 0.802, respectively, suggesting strong agreement between the two approaches. We further evaluate coherence using the goodness-of-fit diagnostic (Brown et al., 2001):

$$W = \sum_{i}^{m_s} \frac{(\hat{F}^{w}_{\alpha i} - \hat{F}^{CLS}_{\alpha i})^2}{\widehat{\text{Var}}(\hat{F}^{w}_{\alpha i}) + \text{mspe}_*(\hat{F}^{CLS}_{\alpha i})}.$$

With $W = 119.59$ for headcount ratio and $W = 175.75$ for poverty gap, both are below the 95th percentile of a χ^2 distribution with 212 degrees of freedom ($\chi^2_{212,0.95} = 177.39$), indicating that CLS estimates are coherent with direct estimates.

To assess precision gains, Table 3.4 presents summary statistics for CLS and direct estimates across municipalities. The results indicate that CLS estimators consistently exhibit lower CVs than direct estimators, even in municipalities with varying sample sizes. Moreover, for some municipalities where the estimated headcount ratio and poverty gap are zero, the CVs of direct estimates are unavailable, highlighting a key limitation of the direct estimation method.

We further compare the empirical cumulative density functions (ECDFs) of the CV of the

Table 3.4: Population size, sample size, direct and CLS estimates of headcount ratios (%) and poverty gaps (%), and the associated CVs (%) of direct and CLS estimators for the Albania municipalities with sample size closest to minimum, first quartile, median, third quartile, and maximum.

			Headcount Ratio				Poverty Gap			
Municipality	N_i	n_i	$\hat{F}^{w}_{0i}$	$\hat{F}^{CLS}_{0i}$	CV($\hat{F}^{w}_{0i}$)	CV($\hat{F}^{CLS}_{0i}$)	$\hat{F}^{w}_{1i}$	$\hat{F}^{CLS}_{1i}$	CV($\hat{F}^{w}_{1i}$)	CV($\hat{F}^{CLS}_{1i}$)
Hajmel	1111	6	33.33	23.82	56.96	27.01	1.52	5.92	70.56	31.31
Orenje	1419	16	12.50	20.69	66.19	15.06	2.00	4.97	66.03	17.44
B.Curri	1488	40	20.00	24.54	32.82	8.57	5.65	5.51	43.01	9.56
Gramsh	2503	64	23.55	17.97	24.48	0.69	4.75	3.64	38.93	21.90
Tirane	89764	600	12.43	11.72	10.91	4.46	2.49	2.31	14.71	4.72

CLS estimator and the direct estimator. Figure 3.6 presents CV (%) values from all sampled municipalities and shows that the ECDFs for the CLS estimators, for both the headcount ratio and the poverty gap, consistently dominate those of the direct estimators, further highlighting the improved precision of CLS estimators. Additionally, we find that about 78% and 89% of Albanian municipalities have CV values associated with the direct estimates of the headcount ratio and poverty gap, respectively, exceeding the conventional 33% reliability threshold in small area estimation (Marino et al., 2019). In contrast, these percentages reduce to 28% and 34% when applying the proposed CLS approach.

We generate poverty maps for Albania using the direct estimation method, as shown in Figure 3.7. The gray areas represent regions where poverty estimates are unavailable due to insufficient sample sizes, underscoring a key limitation of the direct estimation approach in small-area analysis. In contrast, Figure 3.8 presents poverty maps generated using our proposed CLS method, which provides more comprehensive and reliable estimates across all municipalities.

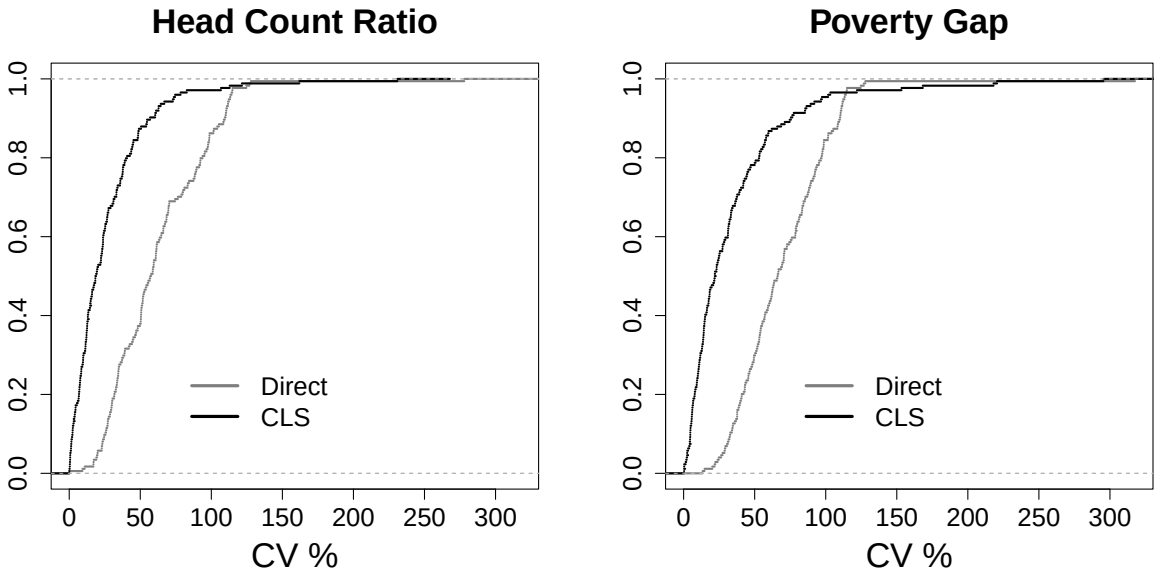


Figure 3.6: CVs empirical cumulative density functions for the CLS and direct estimator.

These maps reveal that several municipalities in the Bulqize district—such as Ostren, Gjorice, Shupenze, and Bulqize — exhibit the highest headcount ratio and poverty gap values, significantly exceeding those observed in other municipalities. Similarly, municipalities like Gjoaaj in the Peqin district, Katund I in the Durres district, and Arren in the Kukes district also show relatively high poverty levels, highlighting the need for targeted policy interventions to improve living conditions in these areas. In contrast, municipalities in the southern regions display substantially lower poverty levels, reflecting better economic conditions and higher living standards.

These results are consistent with the findings of Betti et al. (2003), who used the ELL method, and those of Tzavidis et al. (2008), further validating the robustness of our approach in capturing the spatial distribution of poverty in Albania.

The CLS method offers a significant advantage over direct estimation by producing estimates for all municipalities, thereby addressing data gaps observed in the direct method. This improvement

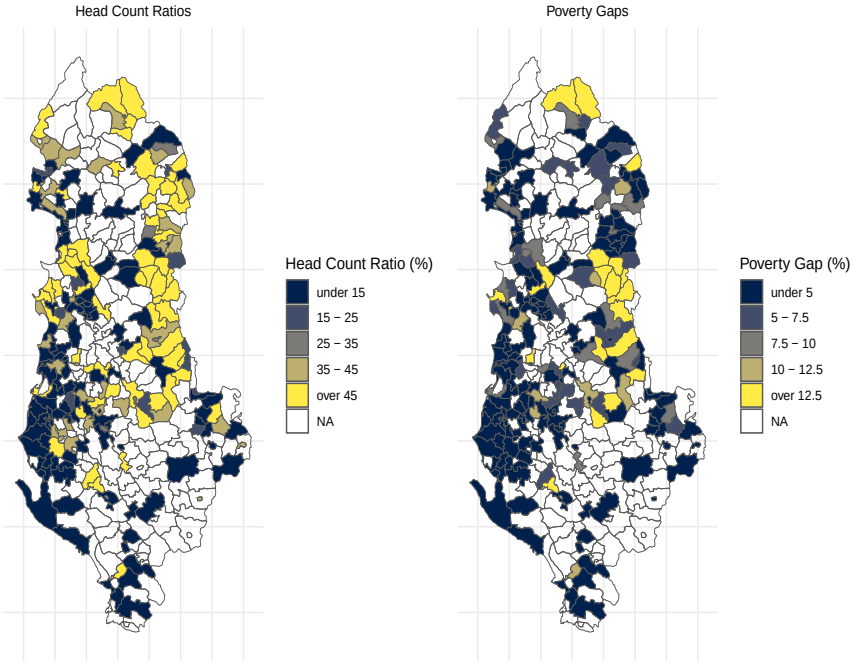


Figure 3.7: Municipality-level direct estimates of headcount ratios and poverty gaps in Albania.

ensures a more complete understanding of poverty distribution, which is crucial for effective policy planning.

The identification of high-poverty areas, particularly in the northern and central regions, highlights the need for geographically targeted poverty alleviation programs. In contrast, the relatively lower poverty levels in the south suggest regional disparities that should be accounted for in national development strategies.

The consistency of CLS estimates across municipalities, even in those with limited survey data, reinforces the reliability of this approach in small area poverty estimation. This robustness makes it a valuable tool for informing social protection policies and resource allocation.

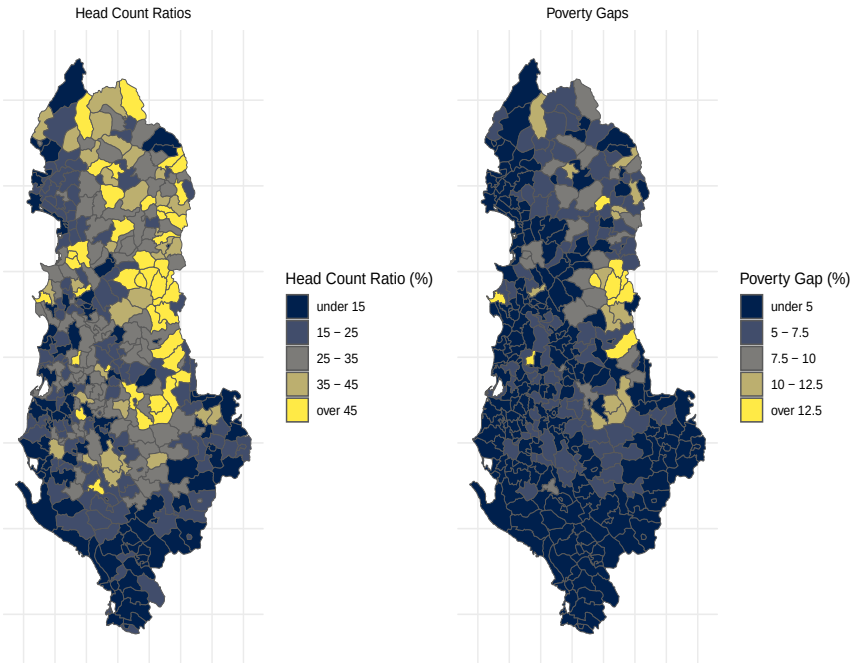


Figure 3.8: Municipality-level CLS estimates of headcount ratios and poverty gaps in Albania.

3.8 Discussion

The proposed model bridges the gap between existing approaches that either assume uniformity in model parameters across areas, potentially leading to misspecification, or rely on random effect specifications that demand strong distributional assumptions. By leveraging area-specific estimating equations and appropriately constructed residuals, we offer a framework for deriving empirical best prediction of complex poverty measures that captures variability across areas without the instability associated with fixed effects in small samples. On the computational side, we introduce a novel, computational efficient, and data-driven procedure for estimating model parameters, which is critical for practical applications. Additionally, we develop a new method to estimate tuning parameters for out-of-sample areas, enabling the derivation of both area-specific regression coefficients and sampling variances, even for out of sample areas.

Our simulation studies highlight the advantages of the proposed CLS method, which consistently outperforms the MR method in terms of both bias and RRMSE, particularly when the assumptions of constant regression coefficients and sampling variances across areas are violated. These results underscore the effectiveness of our approach in handling heterogeneous and complex data structures. It is important to note that estimates for out-of-sample areas, whether derived from CLS or MR, remain synthetic due to the absence of direct observations. Future research could focus on reducing the synthetic nature of these estimates by generating pseudo-random effect estimates or incorporating other information, such as spatial correlations among areas.

The field of SAE is also evolving with the integration of non-traditional data sources, such as satellite imagery, mobile phone data, and geospatial information, alongside the use of machine learning methods to address model misspecification. For instance, Jean et al. (2016) and others

illustrate how combining machine learning algorithms with SAE techniques can enhance poverty estimates, particularly in data-scarce regions. These advancements offer promising avenues for refining the estimation of poverty measures in contexts where traditional survey data alone are insufficient.

Chapter 4: Impact of existence and nonexistence of pivot on the coverage of empirical best linear prediction intervals for small areas

4.1 Introduction

In the area-level model discussed in Section (1.2.1), normality is typically assumed for both the sampling errors and random effects. The assumption of normality for the sampling error e_i is often justified by the Central Limit Theorem (Rao and Molina, 2015). However, recent research has modified the normality assumption for random effects, proposing alternative distributions to improve the robustness of small area mean predictors, particularly in the presence of outliers. Motivated by these considerations, this Chapter explores the following general area-level model: For $i = 1, \cdots, m$,

Level 1 (Sampling model): $y_i \,|\, \theta_i \overset{\text{ind}}{\sim} \mathcal{N}(\theta_i, D_i)$;

Level 2 (Linking model): $\theta_i \overset{\text{ind}}{\sim} \mathcal{G}(\boldsymbol{x_i'}\boldsymbol{\beta}, A, \boldsymbol{\phi})$,

where $\mathcal{G}$ is a fully parametric distribution, not necessarily normal, with mean $\mathrm{E}(\theta_i) = \boldsymbol{x_i'}\boldsymbol{\beta}$, variance $\mathrm{Var}(\theta_i) = A \geq 0$, and any additional parameters $\boldsymbol{\phi}$.

To improve robustness against outliers, various extensions of the traditional normality assumption for $\mathcal{G}$ have been explored. Datta and Lahiri (1995) considered modeling $\mathcal{G}$ as a scale mixture of normal distributions. Later, Bell and Huang (2006) and Xie et al. (2007)

introduced a t-distribution as a special case of the scale mixture of normals at level 2, with mean $\boldsymbol{x}_i'\boldsymbol{\beta}$, variance A, and degrees of freedom ϕ. Fabrizi and Trivisano (2010) proposed two alternative robust area-level models: one assuming an exponential power distribution for θ_i with mean $\boldsymbol{x}_i'\boldsymbol{\beta}$, variance A, and shape parameter ϕ; and another assuming a skewed exponential power distribution characterized by mean $\boldsymbol{x}_i'\boldsymbol{\beta}$, variance A, shape parameter ψ, and skewness parameter λ, where the parameter vector is given by $\boldsymbol{\phi} = (\psi, \lambda)'$. These modifications offer greater flexibility in modeling random effects, leading to more reliable small area estimates in the presence of non-normality or extreme observations.

Similar to (1.6), the above two-level model can be rewritten as the following simple linear mixed model

$$y_i = \theta_i + e_i = \boldsymbol{x}_i'\boldsymbol{\beta} + u_i + e_i,\ i = 1, \cdots, m, \tag{4.1}$$

where the random effect $u_i \overset{\text{iid}}{\sim} \mathcal{G}(0, A, \boldsymbol{\phi})$ and the sampling error $e_i \overset{\text{ind}}{\sim} \mathcal{N}(0, D_i)$ are independent. When $\mathcal{G}$ is different from normal distribution, the best predictor (BP) of θ_i, $\tilde{\theta}_i^{\text{BP}} = \text{E}(\theta_i|y_i)$ may not have a closed form. Instead, the best linear predictor (BLP) of θ_i always has the explicit form as below:

$$\tilde{\theta}_i^{\text{BLP}} = (1 - B_i)y_i + B_i\boldsymbol{x}_i'\boldsymbol{\beta}, \tag{4.2}$$

where $B_i = D_i/(A + D_i)$, and the BLP minimizes the mean squared prediction error (MSPE) among all linear predictors of θ_i. The variance of the prediction error $(\theta_i - \tilde{\theta}_i^{\text{BLP}})$ is denoted by $g_{1i}(A) \coloneqq \text{Var}(\theta_i - \tilde{\theta}_i^{\text{BLP}}) = AD_i/(A + D_i)$. If A is known, one can obtain the standard weighted least squares estimator of $\boldsymbol{\beta}$, denoting as $\tilde{\boldsymbol{\beta}}(A)$. Replacing $\boldsymbol{\beta}$ in (4.2) by $\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(A)$, a best linear unbiased prediction (BLUP) estimator of θ_i is given by

$$\tilde{\theta}_i^{\text{BLUP}} = (1 - B_i)y_i + B_i \boldsymbol{x_i'}\tilde{\boldsymbol{\beta}}, \tag{4.3}$$

which does not require normality assumptions typically used in small area estimation models. In practice, it is common that both $\boldsymbol{\beta}$ and A are unknown and need to be estimated based on data. After plugging in their estimators, an empirical best linear unbiased predictor (EBLUP) of θ_i is given by:

$$\hat{\theta}_i^{\text{EBLUP}} = (1 - \hat{B}_i)y_i + \hat{B}_i \boldsymbol{x_i'}\hat{\boldsymbol{\beta}}, \tag{4.4}$$

where $\hat{B}_i = D_i/(\hat{A} + D_i)$ and $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\hat{A})$, and $\hat{A}$ is a consistent estimator of A for large m. After plugging in the estimator $\hat{A}$, we have $g_{1i}(\hat{A}) = \hat{A}D_i/(\hat{A} + D_i)$. To simplify the notation, throughout the remainder of this Chapter, we will use g_{1i} and $\hat{g}_{1i}$ to denote $g_{1i}(A)$ and $g_{1i}(\hat{A})$, respectively.

Point prediction of small area parameters using EBLUP and the associated MSPE estimation have been extensively studied; see Jiang and Nguyen (2007) and Rao and Molina (2015) for a comprehensive review. However, research on interval estimation in small area studies has primarily focused on specific cases of area-level models, particularly the linear mixed normal model (Jiang and Lahiri, 2006). In this Chapter, we address prediction interval approximations for small area means θ_i under the more general area-level model (4.1).

A prediction interval of θ_i, denoted by I_i, is called a $100(1-\alpha)\%$ interval for θ_i if $\mathrm{P}(\theta_i \in I_i|\boldsymbol{\beta}, A, \boldsymbol{\phi}) = 1 - \alpha$, for any fixed $\boldsymbol{\beta} \in \mathcal{R}^p$, $A \in \mathcal{R}^+$ and $\boldsymbol{\phi} \in \Theta_{\boldsymbol{\phi}}$, the parameter space of $\boldsymbol{\phi}$. The probability P is with respect to the area-level model. There are several options for constructing interval estimates for θ_i. The prediction interval based on only the Level 1 model for the observed data is given by $y_i \pm z_{\alpha/2}\sqrt{D_i}$, where $z_{\alpha/2}$ is the $100(1-\alpha/2)$th standard normal

percentile. While the coverage probability for this direct interval is $1-\alpha$, it is not efficient when D_i is large as in the case of small area estimation. Hall and Maiti (2006) considered the interval based the regression synthetic estimator: $(\boldsymbol{x_i'}\hat{\boldsymbol{\beta}} + b_{\hat{\alpha}_l^s}\sqrt{\hat{A}}, \boldsymbol{x_i'}\hat{\boldsymbol{\beta}} + b_{\hat{\alpha}_u^s}\sqrt{\hat{A}})$, where $b_{\hat{\alpha}_l^s}$ and $b_{\hat{\alpha}_u^s}$ are obtained using a parametric bootstrap method (described in detail in a later section). However, this interval is synthetic in the sense that it is constructed using synthetic regression estimator of θ_i and its associated uncertainty measure, which are not area specific in the outcome variable and hence may cause larger length of the confidence interval.

It is of importance to combine both levels of the model in the interval estimation. A effective approach is to use empirical best methodology. We call an interval empirical best (EB) prediction interval if it is based on empirical best predictor of θ_i. For a special case of the area-level model that θ_i follows a normal distribution, Cox (1975) initiated the exact empirical Bayes interval: $\hat{\theta}_i^{\text{EB}} \pm z_{\alpha/2}\sqrt{\hat{g}_{1i}}$, where $\hat{\theta}_i^{\text{EB}}$ is the empirical Bayes estimator of θ_i. Although the Cox interval always has smaller length than that of the direct interval, its coverage error is of the order $O(m^{-1})$, not accurate enough in most small area applications. Yoshimori and Lahiri (2014) improved the cox-type empirical Bayes interval by using a carefully devised adjusted residual maximum likelihood (ARML) estimator of A. Their interval has a coverage error of order $O(m^{-3/2})$. Additionally, they analytically showed that their interval always produces shorter length than the corresponding direct interval. However, the properties of both the ARML estimator of A and the associated interval have not yet been explored for cases involving non-normally distributed random effects.

A function of the data and parameters is called to be a pivot if its distribution does not depend on any unknown quantity (Hall, 2013; Shao, 2008). When we have a linear mixed normal model on (4.1) , $(\theta_i - \tilde{\theta}_i^{\text{BLP}})/\sqrt{g_{1i}}$ is a standard normal pivot. The traditional method of interval

estimation for θ_i is of the form $\text{EBLUP} \pm z_{\alpha/2}\sqrt{\text{mspe}}$, where mspe is an estimate of the true mean squared prediction error (MSPE) of the EBLUP. Unfortunately, $(\theta_i - \hat{\theta}_i^{\text{EBLUP}})/\sqrt{\hat{g}_{1i}}$ is not a pivot and this traditional approach produces too short or too long intervals. The coverage error of such interval is of the order $O(m^{-1})$, not accurate enough in most small area applications. Recognizing that $(\theta_i - \hat{\theta}_i^{\text{EBLUP}})/\sqrt{\hat{g}_{1i}}$ does not follow a standard normal distribution, Chatterjee et al. (2008) and Li and Lahiri (2010) developed a parametric bootstrap method to approximate its distribution and obtained a EB prediction interval for θ_i in linear mixed normal models. They showed such interval has the coverage error of the order $O(m^{-3/2})$. However, the property remains unknown for non-normal distributed random effects.

In this Chapter, one main aim is to bring out the virtues of pivoting or rescaling, which can decrease the dependence of our statistics on unknown parameters and yield improved prediction interval approximation for small areas under the general model (4.1). Analogous to Chatterjee et al. (2008), we propose parametric bootstrap methods to approximate the distribution of a suitably centered and scaled EBLUP under the general model (4.1), and apply it to construct a $100(1-\alpha)\%$ prediction interval estimation of θ_i. Here, we define an interval based on the EBLUP of θ_i as an Empirical Best Linear (EBL) prediction interval. Specifically, we introduce these two key quantities: one is the centered and scaled EBLUP following the F_{1i} distribution, which can be expressed as below:

$$H_i(\hat{\boldsymbol{\beta}}, \hat{A}) := (\theta_i - \hat{\theta}_i^{\text{EBLUP}})/\sqrt{\hat{g}_{1i}} \sim F_{1i},$$

and the other is based on BLP which has the F_{2i} distribution and can be expressed as

$$H_i(\boldsymbol{\beta}, A) := (\theta_i - \tilde{\theta}_i^{\text{BLP}})/\sqrt{g_{1i}} \sim F_{2i}.$$

If F_{2i} does not depend on any unknown parameters, $H_i(\boldsymbol{\beta}, A)$ can be referred as a pivot; otherwise, $H_i(\boldsymbol{\beta}, A)$ is not a pivot and we rewrite F_{2i} as $F_{2i}(\boldsymbol{\nu})$, where $\boldsymbol{\nu} = (\nu_1, \cdots, \nu_s)'$ is the unknown parameter vector determining the distribution of $H_i(\boldsymbol{\beta}, A)$. We define q_α as the α-level quantile of the distribution $F_{2i}(\boldsymbol{\nu})$, if $\mathrm{P}(H_i(\boldsymbol{\beta}, A) \leq q_\alpha | \, \boldsymbol{\nu}) = \alpha$ for any fixed $\boldsymbol{\nu}$.

In Section 4.2, we introduce a list of notation and regularity conditions used throughout the Chapter. In Section 4.3, we propose a single parametric bootstrap EBL prediction interval for small area means in the context of an area-level model, where the random effects follow a general (non-normal) known distribution, except possibly for unknown hyperparameters. We analytically demonstrate that the coverage error order of the EBL prediction interval remains the same as in Chatterjee et al. (2008), even when relaxing the normality of random effects through the existence of a pivot for suitably standardized random effects when hyperparameters are known. However, when a pivot does not exist, the EBL prediction interval fails to achieve the desired coverage error order, i.e. $O(m^{-3/2})$. Surprisingly, we find that the order $O(m^{-1})$ term is always positive under certain conditions, indicating potential overcoverage in the current single parametric bootstrap EBL prediction interval. Recognizing the challenge of showing existence of a pivot, we develop a simple moment-based method to claim non-existence of pivot. In Section 4.4, we propose a double parametric bootstrap method under a general area level model and, for the first time, analytically show that this approach can correct the coverage problem. In Section 4.5, we compare our proposed EBL prediction interval methods with the direct method, other traditional approaches, and the parametric bootstrap prediction interval proposed by Hall and

Maiti (2006).

4.2 A list of notations and regularity conditions

We use the following notations throughout the Chapter:

$\boldsymbol{Y} = (y_1, \cdots, y_m)'$, a $m \times 1$ column vector of direct estimates;

$\boldsymbol{X'} = (\boldsymbol{x_1}, \cdots, \boldsymbol{x_m})$, a $p \times m$ known matrix of rank p;

$\boldsymbol{\Sigma} = \text{diag}(A + D_1, \cdots, A + D_m)$, a $m \times m$ diagonal matrix;

$\tilde{\boldsymbol{\beta}} = (\boldsymbol{X'\Sigma^{-1}X})^{-1}\boldsymbol{X'\Sigma^{-1}Y}$, weighted least square estimator of $\boldsymbol{\beta}$ with known A;

$\boldsymbol{P} = \boldsymbol{\Sigma^{-1}} - \boldsymbol{\Sigma^{-1}X}(\boldsymbol{X'\Sigma^{-1}X})^{-1}\boldsymbol{X'\Sigma^{-1}}$;

$\partial F_{2i}(x; \tilde{\boldsymbol{\nu}})/\partial \boldsymbol{\nu} = \partial F_{2i}(x; \boldsymbol{\nu})/\partial \boldsymbol{\nu}|_{\boldsymbol{\nu}=\tilde{\boldsymbol{\nu}}}$, derivative with respect to $\boldsymbol{\nu}$ evaluated at $\tilde{\boldsymbol{\nu}}$.

We assume following regularity conditions for proving various results presented in this Chapter:

R1: The rank of $\boldsymbol{X}$, rank$(\boldsymbol{X}) = p$, is bounded for a large m;

R2: $0 < \inf_{i \geq 1} D_i \leq \sup_{i \geq 1} D_i < \infty$, $A \in (0, \infty)$ and the true $\boldsymbol{\phi} \in \Theta^0_{\boldsymbol{\phi}}$, the interior of $\Theta_{\boldsymbol{\phi}}$;

R3: $\sup_{i \geq 1} h_{ii} \equiv \boldsymbol{x_i'(X'X)^{-1}x_i} = O(m^{-1})$;

R4: $\text{E}|u_i|^{8+\delta} < \infty$ for $0 < \delta < 1$;

R5: The distribution function $F_{2i}(\cdot)$ is three times continuously differentiable with respect to $(\cdot)$, and third derivative is uniformly bounded. When $H_i(\boldsymbol{\beta}, A)$ is not a pivot, $F_{2i}(\cdot; \boldsymbol{\nu})$ is

three times continuously differentiable with respect to the parameter vector $\boldsymbol{\nu}$, and its third derivative is uniformly bounded;

R6: When having a non-pivot $H_i(\boldsymbol{\beta}, A)$, we assume the estimator $\hat{\boldsymbol{\nu}}$ satisfies:

$$\begin{aligned} \mathrm{E}\big\{(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu})'\partial F_{2i}(x;\boldsymbol{\nu})/\partial\hat{\boldsymbol{\nu}}\big\} &= O(m^{-1}), \\ \mathrm{E}\Big\{(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu})'\left[\partial^2 F_{2i}(x;\boldsymbol{\nu})/\partial\hat{\boldsymbol{\nu}}\partial\hat{\boldsymbol{\nu}}'\right](\hat{\boldsymbol{\nu}}-\boldsymbol{\nu})\Big\} &= O(m^{-1}), \\ \mathrm{E}\|(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu})\|^3 = \sum_{i,j,k}\mathrm{E}[(\hat{\nu}_i-\nu_i)(\hat{\nu}_j-\nu_j)(\hat{\nu}_k-\nu_k)] &= o(m^{-1}), \end{aligned}$$

where the expectation is taken at the true $\boldsymbol{\nu}$. For a special case that $\boldsymbol{\nu} = A$, using the method of moment from Prasad and Rao (1990b) to estimate A, Lahiri and Rao (1995) showed that $\mathrm{E}(\hat{A}_{\mathrm{PR}} - A) = o(m^{-1})$. Moreover, they provided the Lemma C.1 that under the regularity conditions R1 - R5, $\mathrm{E}|\hat{A}_{\mathrm{PR}} - A|^{2q} = O(m^{-q})$ for any q satisfying $0 < q \leq 2 + \delta'$ with $0 < \delta' < \frac{1}{4}\delta$.

4.3 Single parametric bootstrap

For the remainder of this Chapter, without further explicit mention, we will use $\tilde{\theta}_i$ to represent the BLP of θ_i, and $\hat{\theta}_i$ to denote the EBLUP of θ_i. The single parametric bootstrap method has been widely studied for its simplicity in obtaining prediction intervals directly from the bootstrap histogram. Ideally, a prediction interval of θ_i can be constructed based on the distribution of $(\theta_i - \hat{\theta}_i)/\sqrt{\hat{g}_{1i}}$, although this distribution function F_{1i} is complex and difficult to approximate analytically. In this Chapter, we firstly follow the procedure introduced by Chatterjee et al. (2008) and Li and Lahiri (2010) to provide a bootstrap approximation of F_{1i}

by using a parametric bootstrap method. The implementation is straightforward, following these steps:

1. Conditionally on the data $(\boldsymbol{X}, \boldsymbol{Y})$, draw θ_i^* for $i = 1, \cdots, m$, independently from the distribution $\mathcal{G}(\boldsymbol{x_i'}\hat{\boldsymbol{\beta}}, \hat{A}, \hat{\boldsymbol{\phi}})$;

2. Given θ_i^*, draw y_i^* from the distribution $\mathcal{N}(\theta_i^*, D_i)$;

3. Construct the bootstrap estimators $\hat{\boldsymbol{\beta}}^*$, $\hat{A}^*$ using the data $(\boldsymbol{X}, \boldsymbol{Y^*})$ and then obtain $\hat{\theta}_i^* = (1 - \hat{B}_i^*)y_i + \hat{B}_i^* \boldsymbol{x_i'}\hat{\boldsymbol{\beta}}^*$ and $\hat{g}_{1i}^* = \hat{A}^* D_i/(\hat{A}^* + D_i)$;

4. Calibrate on α using the bootstrap method. Let $(q_{\alpha_l}, q_{\alpha_u}) = (q_{\hat{\alpha}_l^s}, q_{\hat{\alpha}_u^s})$ be the solution of the following equations:

$$\begin{aligned} \mathrm{P}^*(H_i(\hat{\boldsymbol{\beta}}^*, \hat{A}^*) \leq q_{\alpha_u} \,|\, \hat{\boldsymbol{\beta}}, \hat{A}, \hat{\boldsymbol{\phi}}) &= 1 - \alpha/2 \\ \mathrm{P}^*(H_i(\hat{\boldsymbol{\beta}}^*, \hat{A}^*) \leq q_{\alpha_l} \,|\, \hat{\boldsymbol{\beta}}, \hat{A}, \hat{\boldsymbol{\phi}}) &= \alpha/2, \end{aligned} \tag{4.5}$$

   where $H_i(\hat{\boldsymbol{\beta}}^*, \hat{A}^*) = (\theta_i^* - \hat{\theta}_i^*)/\sqrt{\hat{g}_{1i}^*}$;

5. The single bootstrap calibrated prediction interval is constructed by:

$$\hat{I}_i^s = \left(\hat{\theta}_i + q_{\hat{\alpha}_l^s}\sqrt{\hat{g}_{1i}}\,,\, \hat{\theta}_i + q_{\hat{\alpha}_u^s}\sqrt{\hat{g}_{1i}}\right).$$

One of our main results from the algorithm above is that we relax the normality of the random effects by the existence of a pivot $H_i(\boldsymbol{\beta}, A)$, the prediction interval $\hat{I}_i^s$ obtained above still has a high degree of coverage accuracy. That is, it brings the coverage error down to $O(m^{-3/2})$.

Theorem 4.1 *Under regularity conditions, for a preassigned* $\alpha \in (0,1)$ *and arbitrary* $i = 1, \cdots, m$*, when* $H_i(\boldsymbol{\beta}, A)$ *is a pivot, we have*

$$\mathrm{P}(\theta_i \in \hat{I}_i^s) = \mathrm{P}\left(\hat{\theta}_i + q_{\hat{\alpha}_l^s}\sqrt{\hat{g}_{1i}} \le \theta_i \le \hat{\theta}_i + q_{\hat{\alpha}_u^s}\sqrt{\hat{g}_{1i}}\right) = 1 - \alpha + O(m^{-3/2}), \tag{4.6}$$

where $q_{\hat{\alpha}_l^s}$ *and* $q_{\hat{\alpha}_u^s}$ *are determined via the single parametric bootstrap procedure described above.*

Proof of Theorem 4.1: Under regularity conditions, with a pivot $H_i(\boldsymbol{\beta}, A)$, we have

$$\begin{aligned} F_{1i}(q_\alpha) &= \mathrm{P}((\theta_i - \hat{\theta}_i)/\sqrt{\hat{g}_{1i}} \le q_\alpha) \\ &= E\left[\mathrm{P}\left((\theta_i - \tilde{\theta}_i)/\sqrt{g_{1i}} \le q_\alpha + (\hat{\theta}_i - \tilde{\theta}_i + q_\alpha(\sqrt{\hat{g}_{1i}} - \sqrt{g_{1i}}))/\sqrt{g_{1i}}\right)\right] \\ &= E\left[F_{2i}(q_\alpha + Q(q_\alpha + \boldsymbol{Y}))\right] \\ &= F_{2i}(q_\alpha) + m^{-1}\gamma(\boldsymbol{\beta}, A; q_\alpha) + O(m^{-3/2}). \end{aligned} \tag{4.7}$$

where $Q(q_\alpha + \boldsymbol{Y}) = (\hat{\theta}_i - \tilde{\theta}_i + q_\alpha(\sqrt{\hat{g}_{1i}} - \sqrt{g_{1i}}))/\sqrt{g_{1i}}$, and γ is a smooth function of order $O(1)$. The last equation is derived from similar results of the Chapter 4 of Li (2007). See details of its derivation in Appendix A.

Then, from equations (4.5) and (4.7), we have

$$\begin{aligned} \mathrm{E}\left[\mathrm{P}^*(H_i(\hat{\boldsymbol{\beta}}^*, \hat{A}^*) \le q_{\hat{\alpha}_u^s} \,|\, \hat{\boldsymbol{\beta}}, \hat{A}, \hat{\boldsymbol{\phi}})\right] &= \mathrm{E}\left[F_{2i}(q_{\hat{\alpha}_u^s}) + m^{-1}\gamma(\hat{\boldsymbol{\beta}}, \hat{A}; q_{\hat{\alpha}_u^s})\right] + O(m^{-3/2}) \\ &= 1 - \alpha/2, \\ \mathrm{E}\left[\mathrm{P}^*(H_i(\hat{\boldsymbol{\beta}}^*, \hat{A}^*) \le q_{\hat{\alpha}_l^s} \,|\, \hat{\boldsymbol{\beta}}, \hat{A}, \hat{\boldsymbol{\phi}})\right] &= \mathrm{E}\left[F_{2i}(q_{\hat{\alpha}_l^s}) + m^{-1}\gamma(\hat{\boldsymbol{\beta}}, \hat{A}; q_{\hat{\alpha}_l^s})\right] + O(m^{-3/2}) \\ &= \alpha/2. \end{aligned}$$

Therefore, we have

$$\mathrm{E}\left[F_{2i}(q_{\hat{\alpha}_u^s})\right] = 1 - \alpha/2 - \mathrm{E}\left[m^{-1}\gamma(\hat{\boldsymbol{\beta}}, \hat{A}; q_{\hat{\alpha}_u^s})\right] + O(m^{-3/2}),$$

and

$$\mathrm{E}\left[F_{2i}(q_{\hat{\alpha}_l^s})\right] = \alpha/2 - \mathrm{E}\left[m^{-1}\gamma(\hat{\boldsymbol{\beta}}, \hat{A}; q_{\hat{\alpha}_l^s})\right] + O(m^{-3/2}).$$

Finally, we have

$$\begin{aligned}
& \mathrm{P}\left(\hat{\theta}_i + q_{\hat{\alpha}_l^s}\sqrt{\hat{g}_{1i}} \le \theta_i \le \hat{\theta}_i + q_{\hat{\alpha}_u^s}\sqrt{\hat{g}_{1i}}\right) \\
= \; & \mathrm{P}\left(H_i(\hat{\boldsymbol{\beta}}, \hat{A}) \le q_{\hat{\alpha}_u^s}\right) - \mathrm{P}\left(H_i(\hat{\boldsymbol{\beta}}, \hat{A}) \le q_{\hat{\alpha}_l^s}\right) \\
= \; & \mathrm{E}\left[F_{2i}(q_{\hat{\alpha}_u^s}) + m^{-1}\gamma(\boldsymbol{\beta}, A; q_{\hat{\alpha}_u^s})\right] - \mathrm{E}\left[F_{2i}(q_{\hat{\alpha}_l^s}) + m^{-1}\gamma(\boldsymbol{\beta}, A; q_{\hat{\alpha}_l^s})\right] + O(m^{-3/2}) \\
= \; & 1 - \alpha/2 + \mathrm{E}\left[m^{-1}\gamma(\boldsymbol{\beta}, A; q_{\hat{\alpha}_u^s}) - m^{-1}\gamma(\hat{\boldsymbol{\beta}}, \hat{A}; q_{\hat{\alpha}_u^s})\right] - \alpha/2 \\
& + \mathrm{E}\left[m^{-1}\gamma(\hat{\boldsymbol{\beta}}, \hat{A}; q_{\hat{\alpha}_l^s}) - m^{-1}\gamma(\boldsymbol{\beta}, A; q_{\hat{\alpha}_l^s})\right] + O(m^{-3/2}) \\
= \; & 1 - \alpha + O(m^{-3/2}).
\end{aligned}$$

An example of Theorem 4.1 is that when $\mathcal{G}(\boldsymbol{x_i'}\boldsymbol{\beta}, A, \boldsymbol{\phi})$ is a normal distribution, $\mathcal{N}(\boldsymbol{x_i'}\boldsymbol{\beta}, A)$, using Theorem 3.2 of Chatterjee et al. (2008), we have

$$\mathrm{P}(\theta_i \in \hat{I}_i^s) = 1 - \alpha + O(m^{-3/2}). \tag{4.8}$$

Proposition 4.1 *When $H_i(\boldsymbol{\beta}, A)$ is not a pivot, we have*

$$\mathrm{P}(\theta_i \in \hat{I}_i^s) = 1 - \alpha + O(m^{-1}), \tag{4.9}$$

where $\hat{I}_i^s$ is obtained from the single parametric bootstrap procedure described above.

Proof of Proposition 4.1: Consider the case where $H(\boldsymbol{\beta}, A)$ is not a pivot, meaning that F_{2i} depends some unknown true parameters, $\boldsymbol{\nu}$. Here, we rewrite the distribution function of $H(\boldsymbol{\beta}, A)$ as $F_{2i}(\boldsymbol{\nu})$, and Taylor expand $F_{2i}(x; \boldsymbol{\nu})$ around $\boldsymbol{\nu} = \boldsymbol{\nu}_1$, obtaining,

$$F_{2i}(\boldsymbol{\nu}) = F_{2i}(\boldsymbol{\nu_1}) + (\boldsymbol{\nu} - \boldsymbol{\nu_1})' \frac{\partial F_{2i}(\boldsymbol{\nu_1})}{\partial \boldsymbol{\nu}} + \frac{1}{2}(\boldsymbol{\nu} - \boldsymbol{\nu_1})' \frac{\partial^2 F_{2i}(\boldsymbol{\nu_1})}{\partial \boldsymbol{\nu} \partial \boldsymbol{\nu'}} (\boldsymbol{\nu} - \boldsymbol{\nu_1}) + \mathbf{R}(\boldsymbol{\nu}), \tag{4.10}$$

where

$$\mathbf{R}(\boldsymbol{\nu}) = \frac{1}{6} \sum_{i,j,k} \frac{\partial^3 F_{2i}}{\partial \boldsymbol{\nu_i} \partial \boldsymbol{\nu_j} \partial \boldsymbol{\nu_k}} \bigg|_{\boldsymbol{\nu}^*} (\boldsymbol{\nu_i} - \boldsymbol{\nu_{1i}})(\boldsymbol{\nu_j} - \boldsymbol{\nu_{1j}})(\boldsymbol{\nu_k} - \boldsymbol{\nu_{1k}})$$

with $\boldsymbol{\nu}^*$ lying between $\boldsymbol{\nu}$ and $\boldsymbol{\nu_1}$.

Under the regularity conditions, we have,

$$\mathrm{E}\left[(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu})' \frac{\partial F_{2i}(\boldsymbol{\nu})}{\partial \hat{\boldsymbol{\nu}}}\right] + \mathrm{E}\left[\frac{1}{2}(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu})' \frac{\partial^2 F_{2i}(\boldsymbol{\nu})}{\partial \hat{\boldsymbol{\nu}} \partial \hat{\boldsymbol{\nu}}'} (\hat{\boldsymbol{\nu}} - \boldsymbol{\nu})\right] = m^{-1} c(\boldsymbol{\nu}), \tag{4.11}$$

where c denotes a smooth function of order $O(1)$, and

$$\mathrm{E}[\mathbf{R}(\hat{\boldsymbol{\nu}})] = \frac{1}{6} \sum_{i,j,k} \mathrm{E}\left[\frac{\partial^3 F_{2i}(\boldsymbol{\nu^*})}{\partial \hat{\boldsymbol{\nu}}_{\boldsymbol{i}} \partial \hat{\boldsymbol{\nu}}_{\boldsymbol{j}} \partial \hat{\boldsymbol{\nu}}_{\boldsymbol{k}}} (\hat{\boldsymbol{\nu}}_{\boldsymbol{i}} - \boldsymbol{\nu_i})(\hat{\boldsymbol{\nu}}_{\boldsymbol{j}} - \boldsymbol{\nu_j})(\hat{\boldsymbol{\nu}}_{\boldsymbol{k}} - \boldsymbol{\nu_k})\right] = o(m^{-1}) \tag{4.12}$$

with $\boldsymbol{\nu}^*$ lying between $\hat{\boldsymbol{\nu}}$ and $\boldsymbol{\nu}$.

Thus, using equations (4.5) and (4.7), with a non-pivot $H_i(\boldsymbol{\beta}, A)$, we have

$$\begin{aligned}
\mathrm{E}\left[\mathrm{P}^*(H_i(\hat{\boldsymbol{\beta}}^*, \hat{A}^*) \le q_{\hat{\alpha}_u^s} \,|\, \hat{\boldsymbol{\beta}}, \hat{A}, \hat{\boldsymbol{\phi}})\right] &= \mathrm{E}\left[F_{2i}(q_{\hat{\alpha}_u^s}; \hat{\boldsymbol{\nu}}) + m^{-1}\gamma(\hat{\boldsymbol{\beta}}, \hat{A}; q_{\hat{\alpha}_u^s})\right] + o(m^{-1}) \\
&= \mathrm{E}\left[F_{2i}(q_{\hat{\alpha}_u^s}; \boldsymbol{\nu}) + m^{-1}c(\boldsymbol{\nu}; q_{\hat{\alpha}_u^s}) + m^{-1}\gamma(\hat{\boldsymbol{\beta}}, \hat{A}; q_{\hat{\alpha}_u^s})\right] + o(m^{-1}) \\
&= 1 - \alpha/2,
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{E}\left[\mathrm{P}^*(H_i(\hat{\boldsymbol{\beta}}^*, \hat{A}^*) \le q_{\hat{\alpha}_l^s} \,|\, \hat{\boldsymbol{\beta}}, \hat{A}, \hat{\boldsymbol{\phi}})\right] &= \mathrm{E}\left[F_{2i}(q_{\hat{\alpha}_l^s}; \hat{\boldsymbol{\nu}}) + m^{-1}\gamma(\hat{\boldsymbol{\beta}}, \hat{A}; q_{\hat{\alpha}_l^s})\right] + o(m^{-1}) \\
&= \mathrm{E}\left[F_{2i}(q_{\hat{\alpha}_l^s}; \boldsymbol{\nu}) + m^{-1}c(\boldsymbol{\nu}; q_{\hat{\alpha}_l^s}) + m^{-1}\gamma(\hat{\boldsymbol{\beta}}, \hat{A}; q_{\hat{\alpha}_l^s})\right] + o(m^{-1}) \\
&= \alpha/2.
\end{aligned}$$

Finally, we obtain,

$$\begin{aligned}
&\mathrm{P}\left(\hat{\theta}_i + q_{\hat{\alpha}_l^s}\sqrt{\hat{g}_{1i}} \le \theta_i \le \hat{\theta}_i + q_{\hat{\alpha}_u^s}\sqrt{\hat{g}_{1i}}\right) \\
= \;& \mathrm{E}\left[F_{2i}(q_{\hat{\alpha}_u^s}; \boldsymbol{\nu}) + m^{-1}\gamma(\boldsymbol{\beta}, A; q_{\hat{\alpha}_u^s})\right] - \mathrm{E}\left[F_{2i}(q_{\hat{\alpha}_l^s}; \boldsymbol{\nu}) + m^{-1}\gamma(\boldsymbol{\beta}, A; q_{\hat{\alpha}_l^s})\right] + o(m^{-1}) \\
= \;& \mathrm{E}\left[1 - \alpha/2 - m^{-1}c(\boldsymbol{\nu}; q_{\hat{\alpha}_u^s}) - m^{-1}\gamma(\hat{\boldsymbol{\beta}}, \hat{A}; q_{\hat{\alpha}_u^s}) + m^{-1}\gamma(\boldsymbol{\beta}, A; q_{\hat{\alpha}_u^s})\right] \\
& -\mathrm{E}\left[\alpha/2 - m^{-1}c(\boldsymbol{\nu}; q_{\hat{\alpha}_l^s}) - m^{-1}\gamma(\hat{\boldsymbol{\beta}}, \hat{A}; q_{\hat{\alpha}_l^s}) + m^{-1}\gamma(\boldsymbol{\beta}, A; q_{\hat{\alpha}_l^s})\right] + o(m^{-1}) \\
= \;& 1 - \alpha - \mathrm{E}\left[m^{-1}c(\boldsymbol{\nu}; q_{\hat{\alpha}_u^s}) - m^{-1}c(\boldsymbol{\nu}; q_{\hat{\alpha}_l^s})\right] + o(m^{-1}) \\
= \;& 1 - \alpha + O(m^{-1}). \qquad (4.13)
\end{aligned}$$

Proposition 4.2 *Suppose that:*

(i) The random effects u_i *are symmetrically distributed;*

(ii) $\partial F_{2i}(x;\boldsymbol{\nu})/\partial\nu_i > 0,\ 1 \le i \le s$ *and* $\lambda_{\max}(\partial^2 F_{2i}(x;\boldsymbol{\nu})/\partial\boldsymbol{\nu}\partial\boldsymbol{\nu}') < 0$*, where* $\lambda_{\max}$ *means the largest eigenvalue.*

This condition is satisfied for some continuous distributions. For instance, when u_i *follows a logistic or* t *distribution with known degrees of freedom. The only unknown parameter of* $F_{2i}(x;\boldsymbol{\nu})$ *is the variance* A*. As Remark 4.1 indicates that the kurtosis of* $H_i(\boldsymbol{\beta}, A)$ *is a decreasing function of* A*, it is not difficult to show that* $\partial F_{2i}(x;A)/\partial A > 0$ *and* $\partial^2 F_{2i}(x;A)/\partial A^2 < 0$*;*

(iii) The estimators of the unknown parameters, $\hat{\boldsymbol{\nu}} = (\hat{\nu}_1, \cdots, \hat{\nu}_s)$*, are either second-order unbiased or negatively biased, that is* $E(\hat{\nu}_i - \nu_i) = b_i + o(m^{-1})$ *with* $b_i \le 0$*, for* $i = 1, \cdots, s$*.*

Under these conditions, the prediction interval (4.9) *has an overcoverage property. More specifically, we can rewrite* (4.9) *as below:*

$$\mathrm{P}(\theta_i \in \hat{I}_i^s) = 1 - \alpha + T_1 + o(m^{-1}), \tag{4.14}$$

where T_1 *is of the order* $O(m^{-1})$ *with* $T_1 > 0$*.*

Proof of Proposition 4.2: From the third equation of (4.13) and (4.11), we have

$$\begin{aligned}
T_1 &= -\mathrm{E}\left[m^{-1}c(\boldsymbol{\nu};q_{\hat{\alpha}_u^s}) - m^{-1}c(\boldsymbol{\nu};q_{\hat{\alpha}_l^s})\right] \\
&= -\mathrm{E}\left[(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu})'\Big\{\frac{\partial F_{2i}(q_{\hat{\alpha}_u^s};\boldsymbol{\nu})}{\partial\hat{\boldsymbol{\nu}}} - \frac{\partial F_{2i}(q_{\hat{\alpha}_l^s};\boldsymbol{\nu})}{\partial\hat{\boldsymbol{\nu}}}\Big\}\right. \\
&\quad \left. +\frac{1}{2}(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu})'\Big\{\frac{\partial^2 F_{2i}(q_{\hat{\alpha}_u^s};\boldsymbol{\nu})}{\partial\hat{\boldsymbol{\nu}}\partial\hat{\boldsymbol{\nu}}'} - \frac{\partial^2 F_{2i}(q_{\hat{\alpha}_l^s};\boldsymbol{\nu})}{\partial\hat{\boldsymbol{\nu}}\partial\hat{\boldsymbol{\nu}}'}\Big\}(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu})\right] \\
&= -\mathrm{E}\left[(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu})'\frac{\partial}{\partial\hat{\boldsymbol{\nu}}}\Big\{F_{2i}(q_{\alpha_u};\boldsymbol{\nu}) + \frac{\partial F_{2i}(q_{\alpha_u};\boldsymbol{\nu})}{\partial\hat{\boldsymbol{\nu}}'}(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu})\Big\}\right. \\
&\quad \left. -(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu})'\frac{\partial}{\partial\hat{\boldsymbol{\nu}}}\Big\{F_{2i}(q_{\alpha_l};\boldsymbol{\nu}) + \frac{\partial F_{2i}(q_{\alpha_l};\boldsymbol{\nu})}{\partial\hat{\boldsymbol{\nu}}'}(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu})\Big\}\right] \\
&\quad -\mathrm{E}\left[\frac{1}{2}(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu})'\Big\{\frac{\partial^2 F_{2i}(q_{\alpha_u};\boldsymbol{\nu})}{\partial\hat{\boldsymbol{\nu}}\partial\hat{\boldsymbol{\nu}}'} - \frac{\partial^2 F_{2i}(q_{\alpha_l};\boldsymbol{\nu})}{\partial\hat{\boldsymbol{\nu}}\partial\hat{\boldsymbol{\nu}}'}\Big\}(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu})\right] + o(m^{-1}) \\
&= -\mathrm{E}\left[(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu})'\Big\{\frac{\partial F_{2i}(q_{\alpha_u};\boldsymbol{\nu})}{\partial\hat{\boldsymbol{\nu}}} - \frac{\partial F_{2i}(q_{\alpha_l};\boldsymbol{\nu})}{\partial\hat{\boldsymbol{\nu}}}\Big\}\right] \\
&\quad -\frac{3}{2}\mathrm{E}\left[(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu})'\Big\{\frac{\partial^2 F_{2i}(q_{\alpha_u};\boldsymbol{\nu})}{\partial\hat{\boldsymbol{\nu}}\partial\hat{\boldsymbol{\nu}}'} - \frac{\partial^2 F_{2i}(q_{\alpha_l};\boldsymbol{\nu})}{\partial\hat{\boldsymbol{\nu}}\partial\hat{\boldsymbol{\nu}}'}\Big\}(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu})\right] + o(m^{-1}),
\end{aligned} \tag{4.15}$$

where q_u and q_l are unknown but fixed quantites.

In view of the above, we can write

$$T_1 = T_{11} + T_{12} + o(m^{-1}), \tag{4.16}$$

where

$$\begin{aligned}
T_{11} &= -\mathrm{E}\left[(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu})'\Big\{\frac{\partial F_{2i}(q_{\alpha_u};\boldsymbol{\nu})}{\partial\hat{\boldsymbol{\nu}}} - \frac{\partial F_{2i}(q_{\alpha_l};\boldsymbol{\nu})}{\partial\hat{\boldsymbol{\nu}}}\Big\}\right], \\
T_{12} &= -\frac{3}{2}\mathrm{E}\left[(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu})'\Big\{\frac{\partial^2 F_{2i}(q_{\alpha_u};\boldsymbol{\nu})}{\partial\hat{\boldsymbol{\nu}}\partial\hat{\boldsymbol{\nu}}'} - \frac{\partial^2 F_{2i}(q_{\alpha_l};\boldsymbol{\nu})}{\partial\hat{\boldsymbol{\nu}}\partial\hat{\boldsymbol{\nu}}'}\Big\}(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu})\right].
\end{aligned}$$

With the assumption of symmetrically distributed random effects u_i, it is easy to show that

$F_{2i}(q_{\alpha_u};\boldsymbol{\nu}) = 1 - F_{2i}(q_{\alpha_l};\boldsymbol{\nu})$. Hence, we have

$$\begin{aligned} T_{11} &= -2\mathrm{E}\left[(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu})'\frac{\partial F_{2i}(q_{\alpha_u};\boldsymbol{\nu})}{\partial\hat{\boldsymbol{\nu}}}\right], \\ T_{12} &= -3\mathrm{E}\left[(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu})'\frac{\partial^2 F_{2i}(q_{\alpha_u};\boldsymbol{\nu})}{\partial\hat{\boldsymbol{\nu}}\partial\hat{\boldsymbol{\nu}}'}(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu})\right]. \end{aligned}$$

Given the assumptions (i)-(iii), we have

$$T_{11} + T_{12} = O(m^{-1}) > 0,$$

under the regularity conditions. Thus, $T_1 > 0$.

The proposition indicates that under the regularity conditions, the prediction intervals conducted by the proposed single parametric bootstrap can produce higher coverage than the nominal coverage with a non pivot $H_i(\boldsymbol{\beta}, A)$ up to the order of $O(m^{-1})$, which could be beneficial for practitioners without considering other properties of prediction intervals.

Remark 4.1 *When $\mathcal{G}(\boldsymbol{x}_i'\boldsymbol{\beta}, A, \boldsymbol{\phi})$ is not a normal distribution, it is challenging to obtain the explicit form of the distribution of $H_i(\boldsymbol{\beta}, A)$. Consequently, it is difficult to verify whether $H_i(\boldsymbol{\beta}, A)$ is a pivot. Note the fact that for $H_i(\boldsymbol{\beta}, A)$ to be a pivot, its moments should not depend on any unknown parameters. Based on that, we develop a simple moment-based method to claim non-existence of pivot. Under the symmetry assumption of u_i, the odd moments of $H_i(\boldsymbol{\beta}, A)$ are zero if they exist, and the second moment equals to 1 because it is standardized. To verify if $H_i(\boldsymbol{\beta}, A)$ is a pivot, we calculate its fourth moment as follows:*

$$
\begin{aligned}
& E\left[(\theta_i - \tilde{\theta}_i)/\sqrt{g_{1i}}\right]^4 \\
= \; & \frac{1}{g_{1i}^2} E\left[\boldsymbol{x_i'\beta} + u_i - (1-B_i)y_i - B_i\boldsymbol{x_i'\beta}\right]^4 \\
= \; & \frac{1}{g_{1i}^2} E\left[B_i u_i - (1-B_i)e_i\right]^4 \\
= \; & \frac{1}{g_{1i}^2} E\left[B_i^4 u_i^4 - 4B_i^3 u_i^3(1-B_i)e_i + 6B_i^2 u_i^2(1-B_i)^2 e_i^2\right. \\
& \left. -4B_i u_i(1-B_i)^3 e_i^3 + (1-B_i)^4 e_i^4\right] \\
= \; & \frac{1}{g_{1i}^2} E\left[B_i^4 u_i^4 + 6B_i^2 u_i^2(1-B_i)^2 e_i^2 + (1-B_i)^4 e_i^4\right] \\
= \; & 3 + \frac{D_i^2}{(A+D_i)^2}\left[\frac{E(u_i^4)}{A^2} - 3\right], \qquad (4.17)
\end{aligned}
$$

where $[E(u_i^4)/A^2 - 3]$ *is the excess kurtosis of* u_i*. Note that when* u_i *is normally distributed,* $[E(u_i^4)/A^2 - 3]$ *is zero. When the distribution of* u_i *is other than normal, such as t, double exponential, and logistic,* $[E(u_i^4)/A^2 - 3]$ *is a constant other than zero, indicating that the fourth moment of* $H_i(\boldsymbol{\beta}, A)$ *depends on the unknown parameter* A *and thus* $H_i(\boldsymbol{\beta}, A)$ *is not a pivot. Moreover, in these cases, the fourth moment of* $H_i(\boldsymbol{\beta}, A)$ *is a decreasing function of* A*.*

4.4 Double parametric bootstrap

Hall and Maiti (2006) considered parametric bootstrap methods to approximate the distribution of $(\theta_i - \boldsymbol{x_i'}\hat{\boldsymbol{\beta}})/\sqrt{\hat{A}}$. Then, the prediction interval can be constructed as $(\boldsymbol{x_i'}\hat{\boldsymbol{\beta}} + b_{\hat{\alpha}_l^s}\sqrt{\hat{A}}, \boldsymbol{x_i'}\hat{\boldsymbol{\beta}} + b_{\hat{\alpha}_u^s}\sqrt{\hat{A}})$, where $b_{\hat{\alpha}_l^s}$ and $b_{\hat{\alpha}_u^s}$ are obtained from single bootstrap approximation based on their algorithm. Their prediction interval is based on a synthetic model or the regression model, which does not permit approximation of the conditional distribution of θ_i given the data $\boldsymbol{Y}$. As a

consequence, it is likely to underweight the area specific data. When the level 2 distribution $\mathcal{G}$ determined only by $\boldsymbol{x}_i'\boldsymbol{\beta}$ and A, it is easy to know that the quantity, $(\theta_i - \boldsymbol{x}_i'\boldsymbol{\beta})/\sqrt{A}$, is a pivot. As Hall and Maiti (2006) stated, their prediction interval is effective as $\mathrm{P}(\boldsymbol{x}_i'\hat{\boldsymbol{\beta}} + b_{\hat{\alpha}_l^s}\sqrt{\hat{A}} \leq \theta_i \leq \boldsymbol{x}_i'\hat{\boldsymbol{\beta}} + b_{\hat{\alpha}_u^s}\sqrt{\hat{A}}) = 1 - \alpha + O(m^{-2})$. However, when additional parameters involved into the distribution of random effects, $(\theta_i - \boldsymbol{x}_i'\boldsymbol{\beta})/\sqrt{A}$ might not a pivot and we may lose the effectiveness.

Proposition 4.3 *When considering a general distribution $\mathcal{G}(\boldsymbol{x}_i'\boldsymbol{\beta}, A, \boldsymbol{\phi})$, such as a t distribution with unknown degree of freedom, we have*

$$\mathrm{P}(\boldsymbol{x}_i'\hat{\boldsymbol{\beta}} + b_{\hat{\alpha}_l^s}\sqrt{\hat{A}} \leq \theta_i \leq \boldsymbol{x}_i'\hat{\boldsymbol{\beta}} + b_{\hat{\alpha}_u^s}\sqrt{\hat{A}}) = 1 - \alpha + O(m^{-1}), \tag{4.18}$$

where $b_{\hat{\alpha}_l^s}$ and $b_{\hat{\alpha}_u^s}$ are obtained from single bootstrap approximation based on Hall and Maiti (2006) algorithm.

Proof of Proposition 4.3 Let $M_i(\boldsymbol{\beta}, A) = (\theta_i - f_i(\boldsymbol{\beta}))/\sqrt{A} \sim J_{2i}(\boldsymbol{\delta})$, and $M_i(\hat{\boldsymbol{\beta}}, \hat{A}) = (\theta_i - f_i(\hat{\boldsymbol{\beta}}))/\sqrt{\hat{A}} \sim J_{1i}(\boldsymbol{\delta})$, where $\boldsymbol{\delta}$ is a vector of unknown parameters. We have

$$\begin{aligned}
& \mathrm{E}\left[\mathrm{P}^*(M_i(\hat{\boldsymbol{\beta}}^*, \hat{A}^*) \leq b_{\hat{\alpha}_u^s} \,\middle|\, \hat{\boldsymbol{\beta}}, \hat{A}, \hat{\boldsymbol{\phi}})\right] \\
= \; & \mathrm{E}\left[J_{2i}(b_{\hat{\alpha}_u^s}; \hat{\boldsymbol{\delta}}) + m^{-1}c_1(\hat{\boldsymbol{\beta}}, \hat{A}; b_{\hat{\alpha}_u^s})\right] + o(m^{-1}) \\
= \; & \mathrm{E}\left[J_{2i}(b_{\hat{\alpha}_u^s}; \boldsymbol{\delta}) + m^{-1}c_2(b_{\hat{\alpha}_u^s}; \boldsymbol{\delta}) + m^{-1}c_1(\hat{\boldsymbol{\beta}}, \hat{A}; b_{\hat{\alpha}_u^s})\right] + o(m^{-1}) \\
= \; & 1 - \alpha/2,
\end{aligned} \tag{4.19}$$

and

$$\begin{aligned}
& \mathrm{E}\left[\mathrm{P}^*(M_i(\hat{\boldsymbol{\beta}}^*, \hat{A}^*) \le b_{\hat{\alpha}_l^s} \,\middle|\, \hat{\boldsymbol{\beta}}, \hat{A}, \hat{\boldsymbol{\phi}})\right] \\
= \; & \mathrm{E}\left[J_{2i}(b_{\hat{\alpha}_l^s}; \hat{\boldsymbol{\delta}}) + m^{-1} c_1(\hat{\boldsymbol{\beta}}, \hat{A}; b_{\hat{\alpha}_l^s})\right] + o(m^{-1}) \\
= \; & \mathrm{E}\left[J_{2i}(b_{\hat{\alpha}_l^s}; \boldsymbol{\delta}) + m^{-1} c_2(\boldsymbol{\delta}; b_{\hat{\alpha}_l^s}) + m^{-1} c_1(\hat{\boldsymbol{\beta}}, \hat{A}; b_{\hat{\alpha}_l^s})\right] + o(m^{-1}) \\
= \; & \alpha/2, \qquad (4.20)
\end{aligned}$$

where c_1 and c_2 are smooth functions of order $O(1)$.

Eventually, we have

$$\begin{aligned}
& \mathrm{P}(f_i(\hat{\boldsymbol{\beta}}) + b_{\hat{\alpha}_l^s} \hat{A}^{1/2} \le \theta_i \le f_i(\hat{\boldsymbol{\beta}}) + b_{\hat{\alpha}_u^s} \hat{A}^{1/2}) \\
= \; & \mathrm{E}\left[J_{2i}(b_{\hat{\alpha}_u^s}; \boldsymbol{\delta}) + m^{-1} c_1(\boldsymbol{\beta}, A; b_{\hat{\alpha}_u^s})\right] - \mathrm{E}\left[J_{2i}(b_{\hat{\alpha}_l^s}; \boldsymbol{\delta}) + m^{-1} c_1(\boldsymbol{\beta}, A; b_{\hat{\alpha}_l^s})\right] \\
= \; & \mathrm{E}\left[1 - \alpha/2 - m^{-1} c_2(\boldsymbol{\delta}; q_{\hat{\alpha}_u^s}) - m^{-1} c_1(\hat{\boldsymbol{\beta}}, \hat{A}; b_{\hat{\alpha}_u^s}) + m^{-1} c_1(\boldsymbol{\beta}, A; b_{\hat{\alpha}_u^s})\right] \\
& -\mathrm{E}\left[\alpha/2 - m^{-1} c_2(\boldsymbol{\delta}; b_{\hat{\alpha}_l^s}) - m^{-1} c_1(\hat{\boldsymbol{\beta}}, \hat{A}; b_{\hat{\alpha}_l^s}) + m^{-1} c_1(\boldsymbol{\beta}, A; b_{\hat{\alpha}_l^s})\right] + o(m^{-1}) \\
= \; & 1 - \alpha - \mathrm{E}\left[m^{-1} c_2(\boldsymbol{\delta}; b_{\hat{\alpha}_u^s}) - m^{-1} c_2(\boldsymbol{\delta}; b_{\hat{\alpha}_l^s})\right] + o(m^{-1}) \\
= \; & 1 - \alpha + O(m^{-1}). \qquad (4.21)
\end{aligned}$$

Hall and Maiti (2006) proposed a double-bootstrap method to calibrate $(\hat{\alpha}_l^s, \hat{\alpha}_u^s)$ from single parametric bootstraps, which can achieve a high degree of coverage accuracy $O(m^{-3})$. However, their calibration approach can overcorrect and produce a calibrated $\hat{\alpha}$ greater than 1, which makes the implementation infeasible.

While the order $O(m^{-1})$ term T_1 in (4.14) is theoretically positive under certain conditions,

it remains unclear whether this positiveness holds when u_i is asymmetrically distributed. Unlike Hall and Maiti (2006), we introduce a new double bootstrap method, which does not require a pivot or symmetric u_i. This method reduces the coverage error to $o(m^{-1})$, even when u_i is asymmetrically distributed. Our double bootstrap approach is based on the algorithm from Shi (1992), where the double bootstrap is proposed to obtain accurate and efficient confidence interval for parameters of interest in both nonparametric and univariate parametric distribution settings. Later on, McCullough and Vinod (1998) discussed the theory of the double bootstrap both with and without pivoting, and provided the implementations in some nonlinear production functions. In this Chapter, we develop the double bootstrap in the context of our mixed effect model and apply it to obtain the EBL prediction intervals of θ_i. The framework of our double parametric bootstrap is as below:

1. ***First-stage bootstrap***

   1.1 Conditionally on the data $(\boldsymbol{X}, \boldsymbol{Y})$, draw θ_i^*, $i = 1, \cdots, m$, from the distribution $\mathcal{G}(\boldsymbol{x_i'}\hat{\boldsymbol{\beta}}, \hat{A}, \hat{\boldsymbol{\phi}})$;

   1.2 Given θ_i^*, draw y_i^* from the distribution $\mathcal{N}(\theta_i^*, D_i)$;

   1.3 Compute $\hat{\boldsymbol{\beta}}^*$, $\hat{A}^*$, $\hat{\boldsymbol{\phi}}^*$ from the data $(\boldsymbol{X}, \boldsymbol{Y^*})$, and obtain $\hat{\theta}_i^*$ and $\hat{g}_{1i}^*$;

2. ***Second-stage bootstrap***

   2.1 Given $\hat{\boldsymbol{\beta}}^*$, $\hat{A}^*$, $\hat{\boldsymbol{\phi}}^*$, draw θ_i^{**} from the distribution $\mathcal{G}(\boldsymbol{x_i'}\hat{\boldsymbol{\beta}}^*, \hat{A}^*, \hat{\boldsymbol{\phi}}^*)$;

   2.2 Given θ_i^{**}, draw y_i^{**} from the distribution $\mathcal{N}(\theta_i^{**}, D_i)$;

   2.3 Compute $\hat{\boldsymbol{\beta}}^{**}$, A^{**} from the data $(\boldsymbol{X}, \boldsymbol{Y^{**}})$, also obtain $\hat{\theta}_i^{**}$ and $\hat{g}_{1i}^{**}$;

2.4 Consider to obtain a $(1-\alpha)$-level, two-sided, equal-tailed prediction interval. Define

$$\hat{G}^*(z) \equiv \mathrm{P}^{**}(H_i(\hat{\boldsymbol{\beta}}^{**}, \hat{A}^{**}) \leq z \,|\, \hat{\boldsymbol{\beta}}^*, \hat{A}^*, \hat{\phi}^*) \tag{4.22}$$

For seeking the upper limit, we firstly solve the following system of equations in order to obtain $\hat{\alpha}_u$ such that

$$\begin{cases} \mathrm{P}^*(H_i(\hat{\boldsymbol{\beta}}^*, \hat{A}^*) \leq q_{\hat{\alpha}_u} \,|\, \hat{\boldsymbol{\beta}}, \hat{A}, \hat{\boldsymbol{\phi}}) = 1 - \alpha/2 \\ \mathrm{P}^{**}(H_i(\hat{\boldsymbol{\beta}}^{**}, \hat{A}^{**}) \leq q_{\hat{\alpha}_u} \,|\, \hat{\boldsymbol{\beta}}^*, \hat{A}^*, \hat{\boldsymbol{\phi}}^*) = \hat{\alpha}_u. \end{cases}, \tag{4.23}$$

Using the definition (4.22), we have $q_{\hat{\alpha}_u} = \hat{G}^{*-1}(\hat{\alpha}_u)$. Then, the above system of equations is equivalent to

$$\mathrm{P}^* \left(H_i(\hat{\boldsymbol{\beta}}^*, \hat{A}^*) \leq \hat{G}^{*-1}(\hat{\alpha}_u) \,|\, \hat{\boldsymbol{\beta}}, \hat{A}, \hat{\boldsymbol{\phi}} \right) = 1 - \alpha/2. \tag{4.24}$$

Rewriting (4.24) gives

$$\mathrm{P}^* \left[\mathrm{P}^{**} \left(H_i(\hat{\boldsymbol{\beta}}^{**}, \hat{A}^{**}) \leq H_i(\hat{\boldsymbol{\beta}}^*, \hat{A}^*) \,|\, \hat{\boldsymbol{\beta}}^*, \hat{A}^*, \hat{\boldsymbol{\phi}}^* \right) \leq \hat{\alpha}_u \,|\, \hat{\boldsymbol{\beta}}, \hat{A}, \hat{\boldsymbol{\phi}} \right] = 1 - \alpha/2. \tag{4.25}$$

Note that the inner probability, $\mathrm{P}^{**}(H_i(\hat{\boldsymbol{\beta}}^{**}, \hat{A}^{**}) \leq H_i(\hat{\boldsymbol{\beta}}^*, \hat{A}^*) \,|\, \hat{\boldsymbol{\beta}}^*, \hat{A}^*, \hat{\boldsymbol{\phi}}^*)$, is a function of the first-stage bootstrap resample. More specifically, on the jth first-stage bootstrap sample, after all K second-stage bootstrap operations are completed, let Z_j

be the proportion of times that $H_i(\hat{\boldsymbol{\beta}}^{**}, \hat{A}^{**}) \leq H_i(\hat{\boldsymbol{\beta}}^{*}, \hat{A}^{*})$, i.e.,

$$Z_j = \mathrm{P}^{**}\left(H_i(\hat{\boldsymbol{\beta}}^{**}, \hat{A}^{**}) \leq H_i(\hat{\boldsymbol{\beta}}^{*}, \hat{A}^{*}) \,|\, \hat{\boldsymbol{\beta}}^{*}, \hat{A}^{*}, \hat{\boldsymbol{\phi}}^{*}\right) \approx \#(H_i(\hat{\boldsymbol{\beta}}^{**}, \hat{A}^{**}) \leq H_i(\hat{\boldsymbol{\beta}}^{*}, \hat{A}^{*}))/K$$

We will use Z_j to adjust the first-stage intervals. After all bootstrapping operations are complete, we have estimates Z_j, $j = 1, 2, \cdots, J$, where J is the number of first-stage bootstrap operations. Sort Z_j and choose the $1 - \alpha/2$ quantile as the percentile point $\hat{\alpha}_u$ for defining the double-bootstrap upper limit.

2.5 After completing the step 2.4, we have all estimates $H_i^{(j)}(\hat{\boldsymbol{\beta}}^{*}, \hat{A}^{*})$, $j = 1, \cdots, J$ and $\hat{\alpha}_u$. Choose $q_{\alpha_u} = q_{\hat{\alpha}_u^d}$ such that

$$\mathrm{P}^{*}(H_i(\hat{\boldsymbol{\beta}}^{*}, \hat{A}^{*}) \leq q_{\alpha_u} \,|\, \hat{\boldsymbol{\beta}}, \hat{A}, \hat{\boldsymbol{\phi}}) = \hat{\alpha}_u. \tag{4.26}$$

2.6 Take $q_{\hat{\alpha}_u^d}$ as the upper limit of the double bootstrap prediction interval. Similar operations determine the lower limit, $q_{\hat{\alpha}_l^d}$. Finally, construct the prediction interval of θ_i as

$$\hat{I}_i^d = \left(\hat{\theta}_i + q_{\hat{\alpha}_l^d}\sqrt{\hat{g}_{1i}}\,,\ \hat{\theta}_i + q_{\hat{\alpha}_u^d}\sqrt{\hat{g}_{1i}}\right).$$

The algorithm above shows that the single parametric bootstrap $\hat{I}_i^s$ is calibrated by the second-stage bootstrap. In general, such a calibration improves the coverage accuracy to $o(m^{-1})$, even when the pivot does not exist. One more advantage of our double bootstrap algorithm is that it avoids the problem of over correction and so make it more practical than that of Hall and Maiti (2006).

Theorem 4.2 *Under regularity conditions, for a preassigned* $\alpha \in (0,1)$ *and arbitrary* $i = 1, \cdots, m$*, we have*

$$\mathrm{P}\left(\hat{\theta}_i + q_{\hat{\alpha}_l^d}\sqrt{\hat{g}_{1i}} \le \theta_i \le \hat{\theta}_i + q_{\hat{\alpha}_u^d}\sqrt{\hat{g}_{1i}}\right) = 1 - \alpha + o(m^{-1}), \tag{4.27}$$

where $q_{\hat{\alpha}_l^d}$ *and* $q_{\hat{\alpha}_u^d}$ *are obtained from the double parametric bootstrap procedure described above.*

Proof of Theorem 4.2: With equations (4.26) and (4.7), for the upper limit, we can obtain

$$\begin{aligned}
&\mathrm{E}\left[\hat{\alpha}_u\right] = \mathrm{E}\left[\mathrm{P}^*(H_i(\hat{\boldsymbol{\beta}}^*, \hat{A}^*) \le q_{\hat{\alpha}_u^d} \,|\, \hat{\boldsymbol{\beta}}, \hat{A}, \hat{\boldsymbol{\phi}})\right] \\
= \;& \mathrm{E}\left[F_{1i}(q_{\hat{\alpha}_u^d}; \hat{\boldsymbol{\nu}})\right] \\
= \;& \mathrm{E}\left[F_{2i}(q_{\hat{\alpha}_u^d}; \hat{\boldsymbol{\nu}}) + m^{-1}\gamma(\hat{\boldsymbol{\beta}}, \hat{A}; q_{\hat{\alpha}_u^d})\right] + O(m^{-3/2}) \\
= \;& \mathrm{E}\left[F_{2i}(q_{\hat{\alpha}_u^d}; \boldsymbol{\nu}) + m^{-1}c(\boldsymbol{\nu}; q_{\hat{\alpha}_u^d}) + m^{-1}\gamma(\hat{\boldsymbol{\beta}}, \hat{A}; q_{\hat{\alpha}_u^d})\right] + o(m^{-1}).
\end{aligned} \tag{4.28}$$

Replacing α_u by $\hat{\alpha}_u$ in the system of equations (4.23), we obtain

$$\begin{aligned}
&\mathrm{E}\left[1 - \alpha/2 - \hat{\alpha}_u\right] \\
= \;& \mathrm{E}\left[F_{1i}(q_{\hat{\alpha}_u}; \hat{\boldsymbol{\nu}}) - F_{1i}(q_{\hat{\alpha}_u}; \hat{\boldsymbol{\nu}}^*)\right] \\
= \;& \mathrm{E}\left[F_{2i}(q_{\hat{\alpha}_u}; \hat{\boldsymbol{\nu}}) - F_{2i}(q_{\hat{\alpha}_u}; \hat{\boldsymbol{\nu}}^*) + m^{-1}\gamma(\hat{\boldsymbol{\beta}}, \hat{A}; q_{\hat{\alpha}_u}) - m^{-1}\gamma(\hat{\boldsymbol{\beta}}^*, \hat{A}^*; q_{\hat{\alpha}_u})\right] + O(m^{-3/2}) \\
= \;& \mathrm{E}\left[F_{2i}(q_{\hat{\alpha}_u}; \hat{\boldsymbol{\nu}}) - F_{2i}(q_{\hat{\alpha}_u}; \hat{\boldsymbol{\nu}}^*)\right] + O(m^{-3/2}) \\
= \;& \mathrm{E}\left[F_{2i}(q_{\hat{\alpha}_u}; \hat{\boldsymbol{\nu}}) - F_{2i}(q_{\hat{\alpha}_u}; \hat{\boldsymbol{\nu}}) - m^{-1}c(\hat{\boldsymbol{\nu}}; q_{\hat{\alpha}_u})\right] + o(m^{-1}) \\
= \;& \mathrm{E}\left[-m^{-1}c(\hat{\boldsymbol{\nu}}; q_{\hat{\alpha}_u})\right] + o(m^{-1}).
\end{aligned} \tag{4.29}$$

Therefore,

$$\mathrm{E}\left[\hat{\alpha}_u\right] = 1 - \alpha/2 + \mathrm{E}\left[m^{-1}c(\hat{\boldsymbol{\nu}}; q_{\hat{\alpha}_u})\right] + o(m^{-1}). \tag{4.30}$$

For the upper limit, using the last equation of (4.28) and (4.30), we have

$$\begin{aligned}
& \mathrm{P}(\theta_i \le \hat{\theta}_i + q_{\hat{\alpha}_u^d}\sqrt{\hat{g}_{1i}}) \\
= \; & \mathrm{E}\left[F_{1i}(q_{\hat{\alpha}_u^d})\right] \\
= \; & \mathrm{E}\left[F_{2i}(q_{\hat{\alpha}_u^d}; \boldsymbol{\nu}) + m^{-1}\gamma(\boldsymbol{\beta}, A; q_{\hat{\alpha}_u^d})\right] + O(m^{-3/2}) \\
= \; & \mathrm{E}\left[1 - \alpha/2 + m^{-1}c(\hat{\boldsymbol{\nu}}; q_{\hat{\alpha}_u^d}) - m^{-1}c(\boldsymbol{\nu}; q_{\hat{\alpha}_u^d})\right. \\
& \left. -m^{-1}\gamma(\hat{\boldsymbol{\beta}}, \hat{A}; q_{\hat{\alpha}_u^d}) + m^{-1}\gamma(\boldsymbol{\beta}, A; q_{\hat{\alpha}_u^d})\right] + o(m^{-1}) \\
= \; & 1 - \alpha/2 + o(m^{-1}).
\end{aligned} \tag{4.31}$$

Similarly, for the lower limit, we have

$$\begin{aligned}
\mathrm{P}(\theta_i \le \hat{\theta}_i + q_{\hat{\alpha}_l^d}\sqrt{\hat{g}_{1i}}) &= \mathrm{E}\left[F_{1i}(q_{\hat{\alpha}_l^d})\right] \\
&= \alpha/2 + o(m^{-1}).
\end{aligned} \tag{4.32}$$

Finally, obtain

$$\begin{aligned}
& \mathrm{P}(\hat{\theta}_i + q_{\hat{\alpha}_l^d}\sqrt{\hat{g}_{1i}} \le \theta_i \le \hat{\theta}_i + q_{\hat{\alpha}_u^d}\sqrt{\hat{g}_{1i}}) \\
= \; & \mathrm{E}\left[F_{1i}(q_{\hat{\alpha}_u^d}) - F_{1i}(q_{\hat{\alpha}_l^d})\right] \\
= \; & 1 - \alpha/2 - \alpha/2 + o(m^{-1}) \\
= \; & 1 - \alpha + o(m^{-1}).
\end{aligned}$$

4.5 Monte Carlo Simulations

In this section, we compare the performance of the proposed parametric bootstrap with their competitors where available, using Monte Carlo simulation studies. To maintain comparability with existing studies, we adopt part of the simulation framework of Chatterjee et al. (2008). We consider an area-level model with $\boldsymbol{x}_i'\boldsymbol{\beta} = 0$, and five groups of small areas. Within each group, the sampling variances D_i's remain the same. Two patterns for the D_i's are considered: (i) $(4.0, 0.6, 0.5, 0.4, 0.2)$; (ii) $(8.0, 1.2, 1.0, 0.8, 0.4)$. For pattern (i), we take $A = 1$; for pattern (ii), we take $A = 2$ doubling the variances of pattern (i) while preserving the $B_i = D_i/(A+D_i)$ ratios. To examine the effect of m, we consider $m = 15$ and 50. With the increase of m, all methods improve and get closer to one another, supporting our asymptotic theory. Since we obtained virtually identical results under these two patterns, for the full study we confined attention to the pattern (i) and the results for pattern (ii) are provided in the Appendix B.

The Prasad-Rao method-of-moments (Prasad and Rao, 1990b), and the Fay-Herriot method of estimating the variance component A are considered. For the Fay-Herriot estimator of A, we employ the method of scoring to solve the estimating equation, which has showed to be more stable (Datta et al., 2005) than the Newton-Raphson method originally used in Fay and Herriot (1979). The estimation equation of the Fay-Herriot estimator is

$$f(A) = \sum_{i=1}^{m} \frac{(y_i - \boldsymbol{x}_i'\tilde{\boldsymbol{\beta}})^2}{A + D_i} = 0.$$

Here, $f(A)$ is a decreasing function with respect to A, and the expectation of its first derivative $\mathrm{E}(f'(A)) = -\mathrm{tr}(\mathrm{P}) < 0$. To improve computational efficiency, we implemented a slight

modification to the algorithm. Specifically, the revised Fay-Herriot algorithm begins by calculating $f(A)$ at the initial point $A_0 = 0$, as in Fay and Herriot (1979). If $f(A_0) < 0$, we truncate the estimator to $\hat{A}_{\text{FH}} = 0.01$. Otherwise, the iterative process continues to search for a positive solution, with $\hat{A}_{\text{FH}} = 0.01$ also applied if no positive solution is found. This revised Fay-Herriot algorithm further enhances computational efficiency compared to the original method.

4.5.1 Simulations on symmetric cases

First, we consider the scenario that u_i is symmetrically distributed. Specifically, we assume $\mathcal{G}$ is a t distribution with 9 degrees of freedom. In this setting, we compare coverage probabilities and average lengths of the following seven different prediction intervals of θ_i:

- Two prediction intervals based on the proposed single parametric bootstrap method with two different variance estimators: $\hat{A}_{\text{FH}}$ and $\hat{A}_{\text{PR}}$, denoted as **SB.FH** and **SB.PR**, respectively;
- Two prediction intervals based on the single parametric bootstrap methods proposed by Hall and Maiti (2006), using the same variance estimators, denoted as **HM.FH** and **HM.PR**;
- Two traditional prediction intervals of the form $\hat{\theta}_i \pm z_{\alpha/2}\sqrt{\text{mspe}(\hat{\theta}_i)}$, where the variance estimator $\hat{A}_{\text{FH}}$ is used along with the MSPE estimator of Datta et al. (2005), denoted as **FH**, and the variance estimator $\hat{A}_{\text{PR}}$ is used with the Prasad-Rao MSPE estimator (Prasad and Rao, 1990b), denoted as **PR**;
- The direct confidence interval (**DIRECT**) given by $y_i \pm z_{\alpha/2}\sqrt{D_i}$.

Each reported result is based on 1000 simulation runs. For all cases, we consider single bootstrap sample of size 400 and three different nominal coverages: $\alpha \in (80\%, 90\%, 95\%)$.

We report the percentages of negative estimates for A under t_9 distribution in Table 4.1. For $m = 15$, the Prasad-Rao method-of-moment approach yields as high as about 13% negative estimates of A under pattern (i) of D_i. Fay-Herroit method produces significantly fewer negative estimates than the Prasad-Rao method-of-moment approach across all scenarios.

Table 4.1: Percentages of negative estimates of A ($\hat{A}^*$) for different estimation methods and different patterns of D_i under t_9.

	Pattern (i)		Pattern (ii)	
m	FH	PR	FH	PR
15	1.6 (6.4)	13.1 (25.0)	1.1 (5.7)	11.8 (23.2)
50	0 (0.07)	1.2(6.6)	0 (0.07)	1.2 (6.1)

Table 4.2 presents coverage probabilities and average lengths for each prediction interval method with $m = 50$ and the t_9 distribution under the pattern (i) of D_i. The **SB.PR** and **HM.PR** prediction intervals consistently over-cover. **SB.PR** prediction intervals have shorter lengths than **HM.PR**, especially for groups 2-5 with smaller sampling variances D_i. The **SB.FH** and **HM.FH** prediction intervals perform well regarding coverage errors. Specifically, their coverage probabilities are very close to all three nominal coverages. The **SB.FH** method produces the shortest average lengths among these four single parametric bootstrap methods. The **FH** intervals tend to undercover for G1 group when $\alpha = 90\%$ and 95%. The **PR** intervals have the undercoverage issue for G1 group across all three nominal coverages.

Table 4.3 reports the results for $m = 15$. As illustrated in Table 4.1, the Prasad–Rao method produces an extremely high percentage of zero estimates for a small $m = 15$. Thus, it is not surprising to note that the **SB.PR** intervals have severe undercoverage problem when $\alpha = 90\%$ and 95%. The high percentage of negative estimates of A might also contribute to the similar undercoverage problem of **HM.PR** intervals at $\alpha = 90\%$ and 95% as well as the large average

Table 4.2: Coverage probabilities (average length) of different intervals under t_9 distribution with m = 50 and Pattern (i)

	SB.FH	HM.FH	SB.PR	HM.PR	FH	PR	DIRECT
1) $\alpha = 80\%$							
G1	79.94 (2.31)	80.30 (2.53)	83.70 (2.59)	84.45 (2.84)	79.89 (2.30)	78.66 (2.31)	79.43 (5.13)
G2	80.15 (1.59)	80.31 (2.49)	83.63 (1.79)	84.54 (2.81)	80.38 (1.59)	81.55 (1.67)	79.85 (1.99)
G3	79.71 (1.50)	79.53 (2.48)	83.12 (1.69)	83.75 (2.80)	79.86 (1.50)	80.88 (1.58)	79.99 (1.81)
G4	80.47 (1.39)	80.03 (2.48)	84.06 (1.57)	83.95 (2.80)	80.49 (1.39)	82.06 (1.47)	80.29 (1.62)
G5	79.65 (1.05)	80.48 (2.46)	83.32 (1.20)	84.89 (2.78)	79.99 (1.06)	81.80 (1.14)	80.32 (1.15)
2) $\alpha = 90\%$							
G1	89.88 (3.05)	90.35 (3.39)	94.13 (3.79)	94.25 (4.21)	88.90 (2.96)	87.82 (2.97)	89.42 (6.58)
G2	90.17 (2.05)	90.37 (3.30)	93.25 (2.59)	94.27 (4.14)	89.89 (2.04)	90.58 (2.14)	90.19 (2.55)
G3	90.06 (1.94)	89.34 (3.30)	93.38 (2.44)	93.76 (4.14)	90.15 (1.92)	90.78 (2.03)	90.11 (2.33)
G4	90.07 (1.79)	89.50 (3.29)	93.41 (2.26)	94.24 (4.13)	90.12 (1.78)	91.19 (1.89)	90.24 (2.08)
G5	90.29 (1.36)	90.28 (3.26)	93.33 (1.72)	94.20 (4.09)	90.26 (1.36)	91.56 (1.47)	90.32 (1.47)
3) $\alpha = 95\%$							
G1	95.13 (3.75)	95.04 (4.22)	97.30 (5.31)	97.42 (5.91)	93.57 (3.52)	92.65 (3.53)	94.77 (7.84)
G2	95.11 (2.47)	95.35 (4.07)	96.99 (3.61)	97.50 (5.79)	94.87 (2.43)	95.52 (2.55)	95.38 (3.04)
G3	94.96 (2.32)	94.45 (4.06)	97.05 (3.39)	97.42 (5.77)	94.70 (2.29)	95.51 (2.41)	95.00 (2.77)
G4	95.06 (2.14)	95.03 (4.04)	96.86 (3.14)	97.58 (5.75)	95.12 (2.12)	95.86 (2.25)	95.07 (2.48)
G5	95.31 (1.62)	94.98 (3.99)	96.87 (2.38)	97.42 (5.69)	95.28 (1.62)	96.01 (1.75)	95.17 (1.75)

lengths of both **SB.PR** and **HM.PR** intervals. The **SB.FH** and **HM.FH** intervals uniformly tend to overcover. Still, **SB.FH** intervals have the shortest average lengths among the four types of parametric bootstrap intervals. The **FH** intervals significantly undercover for group G1 at three normial coverages and have slight undercoverage for groups G2 and G3, when $\alpha = 90\%$ and 95%. The **PR** intervals undercover for group G1 and switch to overcover for rest of the groups at all three nominal coverages.

4.5.2 Further simulations on asymmetric cases

While some of our theoretical results for single bootstrap are based on the symmetry assumption, we also use simulation studies to assess our proposed parametric bootstrap method under asymmetric conditions. Specifically, for the same models and parameter choices as above, we change $\mathcal{G}$ to be a shifted exponential distribution (SE). Besides the seven prediction intervals

Table 4.3: Coverage probabilities (average length) of different intervals under t_9 distribution with m = 15 and Pattern (i)

	SB.FH	HM.FH	SB.PR	HM.PR	FH	PR	DIRECT
1) $\alpha = 80\%$							
G1	82.43 (2.68)	82.50 (2.94)	81.50 (3.37)	82.30 (3.74)	75.07 (2.31)	76.23 (2.45)	81.37 (5.13)
G2	82.80 (1.78)	83.50 (2.78)	80.77 (2.25)	82.13 (3.60)	79.43 (1.63)	88.97 (2.19)	80.03 (1.99)
G3	81.40 (1.67)	81.37 (2.76)	80.57 (2.12)	82.27 (3.58)	79.07 (1.54)	88.20 (2.15)	80.40 (1.81)
G4	84.07 (1.55)	84.17 (2.74)	81.50 (1.95)	82.97 (3.54)	81.40 (1.43)	90.47 (2.1)	80.67 (1.62)
G5	81.80 (1.17)	83.17 (2.66)	79.33 (1.47)	82.23 (3.42)	80.73 (1.11)	88.50 (1.98)	80.00 (1.15)
2) $\alpha = 90\%$							
G1	93.07 (3.89)	92.87 (4.33)	87.00 (5.77)	87.30 (6.45)	83.97 (2.97)	85.80 (3.14)	90.33 (6.58)
G2	93.00 (2.51)	93.10 (3.96)	86.30 (3.78)	87.27 (6.03)	89.37 (2.09)	95.70 (2.81)	90.87 (2.55)
G3	92.63 (2.35)	92.50 (3.91)	86.50 (3.57)	87.77 (6.01)	88.37 (1.98)	95.27 (2.76)	90.20 (2.33)
G4	93.90 (2.16)	93.83 (3.86)	87.53 (3.28)	87.87 (5.94)	90.17 (1.84)	96.67 (2.70)	90.93 (2.08)
G5	91.83 (1.60)	92.87 (3.67)	86.03 (2.42)	87.53 (5.62)	90.03 (1.43)	94.73 (2.54)	89.60 (1.47)
3) $\alpha = 95\%$							
G1	97.13 (5.46)	97.07 (6.11)	88.77 (8.91)	88.80 (9.98)	88.73 (3.54)	90.73 (3.75)	94.73 (7.84)
G2	96.33 (3.41)	96.83 (5.36)	88.43 (5.73)	88.87 (9.17)	93.70 (2.49)	97.87 (3.35)	95.70 (3.04)
G3	96.60 (3.18)	96.83 (5.26)	89.20 (5.39)	89.50 (9.10)	93.33 (2.36)	97.90 (3.29)	94.77 (2.77)
G4	97.23 (2.90)	97.37 (5.15)	89.87 (4.93)	89.90 (8.95)	95.33 (2.19)	98.83 (3.22)	95.63 (2.48)
G5	96.30 (2.11)	96.93 (4.77)	88.93 (3.59)	89.30 (8.38)	95.13 (1.70)	97.80 (3.03)	94.57 (1.75)

mentioned before, we also include our proposed double parametric bootstrap intervals in this subsection and we expect they might improve the potential coverage error under asymmetric conditions. Below, the double bootstrap method based on the Fay-Herriot variance estimator is denoted by **DB.FH**, and the double bootstrap method based on the Prasad-Rao variance estimator is indicated by **DB.PR**. We keep bootstrap sample of size 400 in the first stage and apply two different sizes $B_2 \in (50, 100)$ in the second stage. Since for $B_2 = 50$ and $B_2 = 100$, we obtained very similar results. Moreover, the two patterns of D_i gave us similar results. In the following, we provide detailed discussions only for $B_2 = 100$ under the pattern (i), but the results for $B_2 = 50$ are provided in the Appendix B.

Table 4.4 shows the percentages of negative estimates for A under the SE distribution. When $m = 15$, the Prasad-Rao method-of-moments approach yields very high percentages of negative estimates at both the first and second stages of bootstrapping. Although the Fay-

Herriot method results in fewer negative estimates than the Prasad-Rao approach, the occurrence of negative estimates remains notable, especially when the number of small areas is small (e.g., $m = 15$).

Table 4.4: Percentages of negative estimates of A ($\hat{A}^*$)[$\hat{A}^{**}$] for different estimation methods and different patterns of D_i under SE distribution.

	Pattern (i)		Pattern (ii)	
m	FH	PR	FH	PR
15	3.6 (10.72) [16.32]	16.7 (27.98) [31.95]	3.9 (10.93) [16.46]	18.1 (28.49) [31.97]
50	0 (0.37) [1.44]	1.80 (7.89) [12.82]	0 (0.44) [1.56]	2.80 (8.89) [13.68]

Table 4.5 displays the coverage probabilities and average lengths for each prediction interval under $m = 50$. The parametric bootstrap methods based on Prasad-Rao estimators of A, **SB.PR**, **HM.PR** and **DB.PR**, consistently tend to over-cover at $\alpha = 80\%$ and 90%. **DB.PR** intervals bring the coverage probabilities closer to the nominal coverages but have larger lengths than single bootstrap **SB.PR** intervals. Even with $m = 50$, the relatively high percentage of zero estimates of A in the second bootstrap stage showed in Table 4.4, might contribute to the increment on interval length of **DB.PR**. When applying the Fay-Herriot method in estimating A, our proposed single parametric bootstrap intervals **SB.FH** already showed good performance, **DB.FH** has little or no effect. The **FH** intervals perform well, except overcoverage at $\alpha = 80\%$. **PR** intervals have undercoverage problem for group G1 at $\alpha = 95\%$. Overall, the **SB.FH** and **FH** intervals perform the best in terms of both coverage probabilities and average lengths in this setting.

With $m = 15$, similar to the cases under t_9 distribution, the Prasad-Rao variance estimator produces very high percentages of zero estimates of A, especially at the bootstrap stages. See Table 4.4. Thus, it is not abnormal to have the results that the three parametric bootstrap intervals based on $\hat{A}_{PR}$ have very low coverage probabilities when the nominal coverage $\alpha = 90\%$ or

Table 4.5: Coverage probabilities (average length) of different intervals under shifted exponential distribution with m = 50, B_2 = 100 and Pattern (i)

	SB.FH	HM.FH	DB.FH	SB.PR	HM.PR	DB.PR	FH	PR	DIRECT
1) $\alpha = 80\%$									
G1	80.29 (2.20)	80.75 (2.35)	79.58 (2.22)	84.50 (2.50)	87.12 (2.68)	83.16 (2.86)	83.45 (2.28)	81.94 (2.30)	79.90 (5.13)
G2	79.99 (1.55)	81.01 (2.31)	79.42 (1.56)	83.66 (1.77)	87.17 (2.65)	82.99 (2.07)	80.56 (1.57)	82.99 (1.67)	79.94 (1.99)
G3	80.06 (1.47)	81.27 (2.30)	79.62 (1.47)	83.41 (1.68)	87.39 (2.65)	83.09 (1.97)	80.38 (1.48)	82.61 (1.59)	79.86 (1.81)
G4	80.60 (1.36)	80.68 (2.30)	80.20 (1.37)	83.71 (1.56)	87.00 (2.64)	82.96 (1.83)	81.05 (1.37)	83.78 (1.48)	80.15 (1.62)
G5	80.81 (1.05)	80.52 (2.29)	80.34 (1.05)	83.79 (1.20)	87.13 (2.63)	83.23 (1.4)	80.95 (1.05)	83.54 (1.17)	80.69 (1.15)
2) $\alpha = 90\%$									
G1	90.46 (3.01)	91.81 (3.26)	89.56 (3.03)	93.49 (3.79)	94.70 (4.16)	92.15 (5.53)	90.98 (2.92)	90.23 (2.95)	90.01 (6.58)
G2	90.22 (2.04)	91.74 (3.16)	89.53 (2.05)	93.01 (2.61)	94.91 (4.10)	92.09 (4.05)	90.14 (2.02)	92.22 (2.14)	89.81 (2.55)
G3	90.00 (1.93)	91.93 (3.15)	89.48 (1.93)	92.51 (2.46)	94.89 (4.10)	91.69 (3.85)	90.20 (1.90)	91.69 (2.03)	89.80 (2.33)
G4	90.48 (1.78)	91.74 (3.14)	89.8 (1.79)	92.72 (2.28)	94.65 (4.09)	92.09 (3.59)	90.47 (1.76)	92.06 (1.90)	90.17 (2.08)
G5	90.69 (1.36)	91.34 (3.10)	90.24 (1.37)	92.97 (1.74)	94.64 (4.04)	92.23 (2.76)	90.71 (1.35)	92.35 (1.50)	90.29 (1.47)
3) $\alpha = 95\%$									
G1	95.43 (3.83)	96.68 (4.18)	94.37 (3.90)	96.45 (5.44)	96.80 (6.02)	95.73 (9.56)	94.27 (3.48)	93.58 (3.52)	95.04 (7.84)
G2	95.32 (2.51)	96.69 (3.98)	94.77 (2.58)	96.37 (3.69)	96.91 (5.90)	95.88 (6.82)	94.91 (2.40)	96.04 (2.55)	95.07 (3.04)
G3	94.87 (2.36)	96.54 (3.96)	94.35 (2.43)	96.34 (3.49)	97.19 (5.89)	95.72 (6.48)	94.72 (2.27)	95.51 (2.42)	94.79 (2.77)
G4	95.41 (2.17)	96.36 (3.93)	94.87 (2.23)	96.35 (3.22)	97.01 (5.86)	95.87 (6.06)	95.13 (2.10)	95.92 (2.27)	95.17 (2.48)
G5	95.34 (1.64)	96.27 (3.86)	95.16 (1.69)	96.34 (2.43)	96.99 (5.79)	95.92 (4.60)	95.31 (1.61)	96.30 (1.78)	95.44 (1.75)

95%. See Table 4.6 for the cases $m = 15$. The other three parametric bootstrap intervals based on $\hat{A}_{FH}$ overcover when nominal covarage is 80% or 90%. When $\alpha = 95\%$, **SB.FH** intervals have good performance as well as **DB.FH** in terms of coverage error, while **SB.FH** intervals have shorter average lengths than those of **HM.FH** and **DB.FH**. The **FH** intervals suffer from severe undercoverage for group G1 at all nominal coverages as well as **PR** intervals at $\alpha = 95\%$. With $\alpha = 80\%$ and 90%, **PR** intervals have longer average lengths than those of **SB.FH** intervals for groups G2 - G5.

Table 4.6: Coverage probabilities (average length) of different intervals under shifted exponential distribution with m = 15, B_2 = 100 and Pattern (i)

	SB.FH	HM.FH	DB.FH	SB.PR	HM.PR	DB.PR	FH	PR	DIRECT
1) $\alpha = 80\%$									
G1	84.63 (2.66)	86.00 (2.90)	83.63 (2.68)	79.77 (3.31)	81.33 (3.69)	80.60 (4.89)	77.17 (2.23)	80.43 (2.43)	79.80 (5.13)
G2	82.40 (1.78)	85.50 (2.74)	81.63 (1.80)	79.50 (2.20)	81.33 (3.56)	79.57 (3.26)	79.43 (1.59)	88.60 (2.21)	79.23 (1.99)
G3	82.77 (1.68)	85.87 (2.72)	81.97 (1.70)	80.17 (2.08)	81.53 (3.55)	80.40 (3.07)	80.13 (1.51)	89.27 (2.18)	78.70 (1.81)
G4	84.13 (1.55)	84.90 (2.70)	82.97 (1.57)	79.63 (1.92)	80.83 (3.51)	79.63 (2.80)	81.83 (1.41)	90.9 (2.15)	82.03 (1.62)
G5	82.80 (1.18)	84.23 (2.61)	81.70 (1.18)	79.7 (1.44)	80.67 (3.38)	80.00 (2.00)	81.40 (1.12)	88.93 (2.07)	80.60 (1.15)
2) $\alpha = 90\%$									
G1	93.77 (4.08)	94.33 (4.54)	93.13 (4.83)	85.83 (5.84)	86.10 (6.59)	86.07 (13.91)	86.13 (2.86)	89.37 (3.12)	90.43 (6.58)
G2	92.70 (2.59)	94.00 (4.06)	91.53 (2.93)	86.00 (3.78)	86.33 (6.19)	85.97 (8.66)	89.07 (2.04)	95.13 (2.84)	89.67 (2.55)
G3	92.83 (2.43)	93.97 (4.01)	91.67 (2.69)	86.80 (3.55)	86.80 (6.14)	86.73 (8.16)	89.50 (1.94)	95.77 (2.80)	90.20 (2.33)
G4	93.20 (2.23)	93.83 (3.94)	92.00 (2.48)	85.93 (3.26)	86.37 (6.07)	85.50 (7.39)	91.43 (1.81)	96.2 (2.76)	91.60 (2.08)
G5	92.17 (1.65)	93.10 (3.71)	91.00 (1.75)	86.53 (2.39)	86.57 (5.74)	86.33 (5.20)	91.30 (1.43)	95.23 (2.66)	90.23 (1.47)
3) $\alpha = 95\%$									
G1	96.60 (6.02)	96.77 (6.72)	96.50 (10.31)	87.97 (9.34)	88.07 (10.58)	88.40 (22.55)	90.33 (3.41)	93.63 (3.71)	95.07 (7.84)
G2	96.17 (3.64)	96.67 (5.67)	95.43 (5.73)	88.80 (5.86)	88.80 (9.64)	88.93 (13.23)	93.73 (2.43)	97.57 (3.38)	95.00 (3.04)
G3	96.07 (3.38)	96.30 (5.53)	95.57 (5.13)	89.13 (5.48)	89.00 (9.53)	89.80 (12.69)	93.97 (2.31)	98.00 (3.34)	95.67 (2.77)
G4	95.93 (3.09)	96.63 (5.41)	95.40 (4.59)	88.53 (5.03)	88.30 (9.42)	88.67 (11.26)	95.33 (2.16)	98.03 (3.28)	96.00 (2.48)
G5	96.17 (2.22)	96.40 (4.95)	95.47 (2.99)	88.83 (3.63)	88.63 (8.76)	89.07 (8.17)	95.77 (1.71)	97.77 (3.16)	94.90 (1.75)

4.6 Discussion

In this study, we put forward parametric bootstrap approaches to construct prediction intervals in the contexts of small area estimation under a general (non-normal) area-level model. Our simulation results show that the proposed single bootstrap method with the Fay-Herriot method of variance estimator performs well for all cases. Moreover, it is more efficient in terms of average lengths than the existing parametric bootstrap method proposed by Hall and Maiti (2006).

When the number of area is small, there is a high likelihood to obtain negative estimates of the variance A when applying Prasad-Rao method. Throughout our simulation studies, we notice that the likelihood of obtaining negative estimates of variance might affect the performance of the parametric bootstrap methods in terms of both coverage probabilities and average lengths. One might consider a better variance estimator, such as the Fay-Herriot estimator we used, or a sensible truncation of negative estimates. In this study, we arbitrarily truncate the negative estimates of the variance A at 0.01 as similar to Datta et al. (2005). To this end, in the future we will consider an extension of adjusted maximum likelihood estimators of A such as the ones considered by Li and Lahiri (2010) or Hirose and Lahiri (2018) to the model proposed in this Chapter.

There is an another issue that even though the double bootstrap calibration can bring the coverage accuracy to $o(m^{-1})$ regardless the existence of pivot, our simulations suggest that it is not always beneficial to attempt boosting the theoretical coverage probability via double bootstrap, disregarding other properties of the interval. Specifically, variability of calibrated intervals are greater than uncalibrated ones, minimum length property is almost never preserved, and the results are quite dependent on the parameters and fixed constants of the problem, such

as estimation of the variance components A in this study. For instance, Table 4.5 shows that the proposed single bootstrap intervals already produced good performance and thus the calibration using double bootstrap has little or no effect, where $m = 50$. When m is relatively small, i.e., $m = 15$, double bootstrap improves the coverage probability marginally but it produces much larger interval length than the corresponding single bootstrap method.

Chapter 5: Future direction

Model misspecification remains a persistent challenge in small area estimation. In Chapter 2, we investigated various methods for estimating model parameters and incorporating auxiliary information to improve the robustness of small area inference when the assumed model may be misspecified. However, our approach remains within the framework of linear mixed effect models. While these models are effective for structured survey data, their predictive accuracy is highly sensitive to assumption validity (Jiang and Rao, 2020).

Future research should explore semiparametric and nonparametric methods, such as machine learning approaches, to enhance estimation accuracy by integrating multiple data sources. Although machine learning techniques offer flexibility and require fewer parametric assumptions, most supervised learning methods do not account for the dependencies inherent in complex survey data (Mendez, 2008). Addressing this limitation will be crucial for leveraging machine learning in small area estimation.

Looking ahead, my research will focus on integrating machine learning techniques with mixed effects models. Specifically, I plan to use machine learning methods to model the fixed component, as they can better capture complex relationships between the response and auxiliary variables than traditional linear models (Krennmair and Schmid, 2022). To account for residual dependencies, I will incorporate random effects, thereby enhancing the overall predictive accuracy

of the model.

Additionally, these combined techniques can improve data integration from diverse sources, such as satellite imagery and other geospatial data. As discussed in Chapter 3, such data have shown significant potential for small area poverty estimation. By leveraging machine learning within the mixed effects framework, my research aims to develop more robust and flexible approaches for small area estimation.

Another key challenge in applying machine learning methods is the quantification of uncertainty. In future work, I plan to refine uncertainty measurement techniques by incorporating semi-parametric and nonparametric methods. Additionally, I aim to explore double bootstrap techniques to mitigate common issues of uncertainty underestimation. Advancing these methodologies will contribute to more robust statistical inference in small area estimation and related fields.

Beyond uncertainty quantification, my research will also focus on developing innovative parameter estimation methods, which are crucial for reliable statistical analysis. As demonstrated in the simulation results of Chapter 4, standard variance component estimation methods, such as the method of moments, can sometimes produce negative estimates for the variances of random effects—values that should be strictly positive. This issue is particularly concerning because it can cause the EBLUP to collapse into a basic regression estimator, leading to excessive shrinkage and reduced predictive performance. To address this, I plan to investigate adjusted estimation methods that enforce positive and consistent variance estimates across a wide range of scenarios, particularly for non-normal random effects.

Appendix A: Further details of proofs

Let $f_{2i}(\cdot)$ denote the probability density function of $H_i(\boldsymbol{\beta}, A)$. Let f'_{2i} and f''_{2i} be the first and second derivative of $f_{2i}(\cdot)$. Then (4.7) can be expressed as:

$$\begin{aligned} & F_{1i}(q_\alpha) = \mathrm{P}((\theta_i - \hat{\theta}_i)/\sqrt{\hat{g}_{1i}} \leq q_\alpha) \\ = \; & E\left[F_{2i}(q_\alpha + Q(q_\alpha + \boldsymbol{Y}))\right] \\ = \; & F_{2i}(q_\alpha) + f_{2i}(q_\alpha)\mathrm{E}Q + \frac{1}{2}f'_{2i}(q_\alpha)\mathrm{E}Q^2 + \frac{1}{2}\mathrm{E}\left[\int_{q_\alpha}^{q_\alpha+Q} f''_{2i}(t)(q_\alpha + Q - t)^2 dt\right], \end{aligned} \tag{A.1}$$

where $Q = Q(q_\alpha + \boldsymbol{Y})$. Note that for $t \in (q_\alpha, q_\alpha + Q)$, we have $0 \leq |q_\alpha + Q - t| \leq |Q|$. Assuming that the third derivative is uniformly bounded, we have

$$\mathrm{E}\int_{q_\alpha}^{q_\alpha+Q} f''_{2i}(t)(q_\alpha + Q - t)^2 dt \leq C\mathrm{E}|Q|^3 \tag{A.2}$$

for some constant $C > 0$. Then, to show the last equation of (4.7), it reduces to the evaluation of $\mathrm{E}Q$, $\mathrm{E}Q^2$ and $\mathrm{E}|Q|^3$. In particular, if $\mathrm{E}Q = O(m^{-1})$, $\mathrm{E}Q^2 = O(m^{-1})$ and $\mathrm{E}Q^4 = O(m^{-2})$, then it follows that

$$\mathrm{E}|Q|^3 \leq (\mathrm{E}Q^4)^{3/4} = O(m^{-3/2}), \tag{A.3}$$

by the Lyapunov inequality so that

$$F_{1i}(q_\alpha) = F_{2i}(q_\alpha) + m^{-1}\gamma(\boldsymbol{\beta}, A; q_\alpha) + O(m^{-3/2}).$$

First, note that

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X'}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X'}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{Y} = \boldsymbol{\beta} + (\boldsymbol{X'}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X'}\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{u}+\boldsymbol{e}).$$

We now simplify the expression for $Q = Q(q_\alpha + \boldsymbol{Y})$:

$$\begin{aligned}
\hat{\theta}_i - \tilde{\theta}_i &= \frac{\hat{A}}{\hat{A}+D_i}y_i + \frac{D_i}{\hat{A}+D_i}\boldsymbol{x}_i'\hat{\boldsymbol{\beta}} - \frac{A}{A+D_i}y_i - \frac{D_i}{A+D_i}\boldsymbol{x}_i'\boldsymbol{\beta} \\
&= \frac{D_i}{A+D_i}\boldsymbol{x}_i'(\boldsymbol{X'}\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X'}\boldsymbol{\Sigma}^{-1}(\boldsymbol{u}+\boldsymbol{e}) \qquad \text{(A.4)} \\
&\quad + \frac{D_i}{\hat{A}+D_i}\boldsymbol{x}_i'(\boldsymbol{X'}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X'}\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{u}+\boldsymbol{e}) - \frac{D_i}{A+D_i}\boldsymbol{x}_i'(\boldsymbol{X'}\Sigma^{-1}\boldsymbol{X})^{-1}\boldsymbol{X'}\Sigma^{-1}(\boldsymbol{u}+\boldsymbol{e}) \\
&\quad + (\frac{\hat{A}}{\hat{A}+D_i} - \frac{A}{A+D_i})(u_i + e_i).
\end{aligned}$$

In view of the above, similar to Li (2007) and Saegusa et al. (2022), we can write

$$Q(q_\alpha, \boldsymbol{Y}) = Q_1 + Q_2(\boldsymbol{Y}) + Q_3(\boldsymbol{Y}) + Q_4(\boldsymbol{Y}).$$

where

$$
\begin{aligned}
Q_1 &= \frac{1}{\sqrt{g_{1i}}}\frac{D_i}{A+D_i}\boldsymbol{x}_i'(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{u}+\boldsymbol{e}),\\
Q_2(\boldsymbol{Y}) &= \frac{1}{\sqrt{g_{1i}}}\left[\frac{D_i}{\hat{A}+D_i}\boldsymbol{x}_i'(\boldsymbol{X}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{u}+\boldsymbol{e})\right.\\
&\quad \left.-\frac{D_i}{A+D_i}\boldsymbol{x}_i'(\boldsymbol{X}'\Sigma^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\Sigma^{-1}(\boldsymbol{u}+\boldsymbol{e})\right],\\
Q_3(\boldsymbol{Y}) &= \frac{1}{\sqrt{g_{1i}}}\left[(\frac{\hat{A}}{\hat{A}+D_i}-\frac{A}{A+D_i})(u_i+e_i)\right],\\
Q_4(\boldsymbol{Y}) &= q_\alpha(\sqrt{\hat{g}_{1i}}-\sqrt{g_{1i}})/\sqrt{g_{1i}}.
\end{aligned}
$$

Therefore, in order to obtain the last equation of (A.1), we need to show that $\mathrm{E}Q_i = O(m^{-1})$, $\mathrm{E}Q_i^2 = O(m^{-1})$ and $\mathrm{E}Q_i^4 = O(m^{-2})$, $i = 1, \cdots, 4$. Firstly, it is easy to show that $\mathrm{E}Q_1 = 0$. Using the regularity condition that $\sup_{i\geq 1} h_{ii} \equiv \boldsymbol{x_i'(X'X)^{-1}x_i} = O(m^{-1})$, we have

$$
\begin{aligned}
\mathrm{E}Q_1^2 &= \frac{D_i}{A(A+D_i)}\boldsymbol{x}_i'(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\mathrm{E}(\boldsymbol{u}+\boldsymbol{e})(\boldsymbol{u}+\boldsymbol{e})'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{x}_i\\
&= \frac{D_i}{A(A+D_i)}h_{ii} \qquad (A.5)\\
&= O(m^{-1}).
\end{aligned}
$$

Under the assumption $h_{ii} = O(m^{-1})$, it is easily to see that

$$
\begin{aligned}
\tilde{h}_{ii} &= \boldsymbol{x_i'(X\Sigma^{-1}X)^{-1}X'\Sigma^{-2}(X\Sigma^{-1}X)^{-1}x_i}\\
&= \boldsymbol{d_i'P_XP_X'd_i}\\
&= O(m^{-1})
\end{aligned}
$$

where $\boldsymbol{d_i}$ is a column vector with the i^{th} element unity and the rest zeros, $P_X = \boldsymbol{X(X\Sigma^{-1}X)^{-1}X'\Sigma^{-1}}$. An application of the Cauchy–Schwartz inequality yields

$$\begin{aligned}
& \mathrm{E}Q_1^4 = \frac{D_i^2}{A^2(A+D_i)^2}\mathrm{E}(\boldsymbol{x}_i'(\boldsymbol{X'\Sigma^{-1}X})^{-1}\boldsymbol{X'\Sigma^{-1}}(\boldsymbol{u}+\boldsymbol{e}))^4 \\
= & \frac{D_i^2}{A^2(A+D_i)^2}\mathrm{E}(\boldsymbol{d_i'P_XP_X}(\boldsymbol{u}+\boldsymbol{e}))^4 \\
& \leq \frac{D_i^2}{A^2(A+D_i)^2}(\boldsymbol{d_i'P_XP_X'd_i})^2\{\mathrm{E}[(\boldsymbol{u}+\boldsymbol{e})'\boldsymbol{P_X'P_X}(\boldsymbol{u}+\boldsymbol{e})]\}^2 \\
= & Cm^{-2}\{\mathrm{tr}(\boldsymbol{X(X'\Sigma^{-1}X)^{-1}X})\}^2 \\
= & O(m^{-2}). \qquad \text{(A.6)}
\end{aligned}$$

To evaluate the fourth moment of rest of the terms in Q, we need to evaluate the moment of $(\hat{A}+D_i)^{-1}-(A+D_i)^{-1}$ and $\hat{\boldsymbol{\Sigma}}^{-1}-\boldsymbol{\Sigma}^{-1}$. To see this, for example, we can break Q_2 down in terms of more tractable variables and remainder terms:

$$\begin{aligned}
Q_2(\boldsymbol{Y}) = & \frac{1}{\sqrt{g_{1i}}}\left[\frac{D_i}{\hat{A}+D_i}\boldsymbol{x}_i'(\boldsymbol{X'\hat{\Sigma}^{-1}X})^{-1}\boldsymbol{X'\hat{\Sigma}^{-1}}(\boldsymbol{u}+\boldsymbol{e}) - \frac{D_i}{A+D_i}\boldsymbol{x}_i'(\boldsymbol{X'}\Sigma^{-1}\boldsymbol{X})^{-1}\boldsymbol{X'\Sigma^{-1}}(\boldsymbol{u}+\boldsymbol{e})\right] \\
= & \frac{1}{\sqrt{g_{1i}}}\Bigg[\left(\frac{D_i}{\hat{A}+D_i}-\frac{D_i}{A+D_i}\right)\boldsymbol{x}_i'\left((\boldsymbol{X'\hat{\Sigma}^{-1}X})^{-1}-(\boldsymbol{X'\Sigma^{-1}X})^{-1}\right)\boldsymbol{X'}\left(\hat{\boldsymbol{\Sigma}}^{-1}-\boldsymbol{\Sigma}^{-1}\right)(\boldsymbol{u}+\boldsymbol{e}) \\
& +\left(\frac{D_i}{\hat{A}+D_i}-\frac{D_i}{A+D_i}\right)\boldsymbol{x}_i'\left((\boldsymbol{X'\hat{\Sigma}^{-1}X})^{-1}-(\boldsymbol{X'\Sigma^{-1}X})^{-1}\right)\boldsymbol{X'\Sigma^{-1}}(\boldsymbol{u}+\boldsymbol{e}) \\
& +\left(\frac{D_i}{\hat{A}+D_i}-\frac{D_i}{A+D_i}\right)\boldsymbol{x}_i'(\boldsymbol{X'\Sigma^{-1}X})^{-1}\boldsymbol{X'}\left(\hat{\boldsymbol{\Sigma}}^{-1}-\boldsymbol{\Sigma}^{-1}\right)(\boldsymbol{u}+\boldsymbol{e}) \qquad \text{(A.7)} \\
& +\left(\frac{D_i}{\hat{A}+D_i}-\frac{D_i}{A+D_i}\right)\boldsymbol{x}_i'(\boldsymbol{X'\Sigma^{-1}X})^{-1}\boldsymbol{X'\Sigma^{-1}}(\boldsymbol{u}+\boldsymbol{e}) \\
& +\frac{D_i}{A+D_i}\boldsymbol{x}_i'\left((\boldsymbol{X'\hat{\Sigma}^{-1}X})^{-1}-(\boldsymbol{X'\Sigma^{-1}X})^{-1}\right)\boldsymbol{X'}\left(\hat{\boldsymbol{\Sigma}}^{-1}-\boldsymbol{\Sigma}^{-1}\right)(\boldsymbol{u}+\boldsymbol{e}) \\
& +\frac{D_i}{A+D_i}\boldsymbol{x}_i'\left((\boldsymbol{X'\hat{\Sigma}^{-1}X})^{-1}-(\boldsymbol{X'\Sigma^{-1}X})^{-1}\right)\boldsymbol{X'\Sigma^{-1}}(\boldsymbol{u}+\boldsymbol{e}) \\
& +\frac{D_i}{A+D_i}\boldsymbol{x}_i'(\boldsymbol{X'\Sigma^{-1}X})^{-1}\boldsymbol{X'}\left(\hat{\boldsymbol{\Sigma}}^{-1}-\boldsymbol{\Sigma}^{-1}\right)(\boldsymbol{u}+\boldsymbol{e})\Bigg].
\end{aligned}$$

Using the Taylor series expansion, we have

$$\begin{aligned}
&\frac{D_i}{\hat{A}+D_i} - \frac{D_i}{A+D_i} \\
= \;& -(\hat{A}-A)\frac{D_i}{(A+D_i)^2} + (\hat{A}-A)^2\frac{D_i}{(A+D_i)^3} - (\hat{A}-A)^3\frac{D_i}{(A^*+D_i)^4},
\end{aligned}$$

$$\begin{aligned}
&\hat{\boldsymbol{\Sigma}}^{-1} - \boldsymbol{\Sigma}^{-1} \\
= \;& -(\hat{A}-A)\text{diag}\Big\{\frac{1}{(A+D_i)^2}\Big\}_{i=1}^m + (\hat{A}-A)^2\text{diag}\Big\{\frac{1}{(A+D_i)^3}\Big\}_{i=1}^m \\
& -(\hat{A}-A)^3\text{diag}\Big\{\frac{1}{(A^*+D_i)^4}\Big\}_{i=1}^m,
\end{aligned}$$

and for Q_3, we have

$$\begin{aligned}
&\frac{\hat{A}}{\hat{A}+D_i} - \frac{A}{A+D_i} \\
= \;& (\hat{A}-A)\frac{D_i}{(A+D_i)^2} - (\hat{A}-A)^2\frac{D_i}{(A+D_i)^3} + (\hat{A}-A)^3\frac{D_i}{(A^*+D_i)^4},
\end{aligned}$$

where A^* lies between A and $\hat{A}$.

For the last term Q_4, letting $W = g_{1i}^{-1}(\hat{g}_{1i} - g_{1i})$, we have

$$\begin{aligned} Q_4(q_\alpha, \boldsymbol{Y}) &= q_\alpha(\sqrt{\hat{g}_{1i}} - \sqrt{g_{1i}})/\sqrt{g_{1i}} \\ &= q_\alpha(\sqrt{\hat{g}_{1i}}/\sqrt{g_{1i}} - 1) \\ &= q_\alpha\left[\{g_{1i}^{-1}(\hat{g}_{1i} - g_{1i}) + 1\}^{1/2} - 1\right] \\ &= q_\alpha\left[\{W + 1\}^{1/2} - 1\right] \\ &= q_\alpha\left[W/2 - W^2/8 + r_n\right], \end{aligned} \tag{A.8}$$

where the remainder being $r_n = O(W^3)$. Then, apply Taylor expansion on W:

$$\begin{aligned} W &= g_{1i}^{-1}(\hat{g}_{1i} - g_{1i}) \\ &= g_{1i}^{-1}\left(\frac{\hat{A}D_i}{\hat{A} + D_i} - \frac{AD_i}{A + D_i}\right) \\ &= g_{1i}^{-1}\left((\hat{A} - A)\frac{D_i^2}{(A + D_i)^2} - (\hat{A} - A)^2\frac{D_i^2}{(A + D_i)^3} + (\hat{A} - A)^3\frac{D_i^2}{(A^* + D_i)^4}\right). \end{aligned} \tag{A.9}$$

As pointed out in Chatterjee et al. (2008) and Saegusa et al. (2022), this evaluation of $(\hat{A}+D_i)^{-1}-(A+D_i)^{-1}$ and $\hat{\boldsymbol{\Sigma}}^{-1}-\boldsymbol{\Sigma}^{-1}$ involves numerous elementary calculations. In the end, both fourth moments reduce to the fourth moment of $\hat{A} - A$. If using the Prasad-Rao method of moment estimator $\hat{A}_{PR}$, Lahiri and Rao (1995) showed that $\mathrm{E}(\hat{A}_{\mathrm{PR}} - A) = o(m^{-1})$ and $\mathrm{E}|\hat{A}_{\mathrm{PR}} - A|^{2q} = O(m^{-q})$ for any q satisfying $0 < q \leq 2 + \delta'$ with $0 < \delta' < \frac{1}{4}\delta$ under regularity conditions.

Therefore, the above arguments support the final expression:

$$F_{1i}(q_\alpha) = F_{2i}(q_\alpha) + m^{-1}\gamma(\boldsymbol{\beta}, A; q_\alpha) + O(m^{-3/2}).$$

Appendix B: More simulation results for prediction intervals

Table B.1: Coverage probabilities (average length) of different intervals under t_9 distribution with m = 50 and Pattern (ii)

	SB.FH	HM.FH	SB.PR	HM.PR	FH	PR	DIRECT
1) $\alpha = 80\%$							
G1	79.94 (3.26)	80.30 (3.58)	84.56 (3.76)	85.44 (4.12)	79.89 (3.26)	78.65 (3.27)	79.43 (7.25)
G2	80.15 (2.25)	80.31 (3.52)	84.51 (2.61)	85.40 (4.07)	80.38 (2.25)	81.56 (2.36)	79.85 (2.81)
G3	79.71 (2.12)	79.53 (3.51)	83.85 (2.47)	84.58 (4.07)	79.86 (2.12)	80.88 (2.23)	79.99 (2.56)
G4	80.47 (1.96)	80.05 (3.51)	84.72 (2.29)	84.72 (4.06)	80.49 (1.96)	82.06 (2.08)	80.29 (2.29)
G5	79.65 (1.49)	80.48 (3.48)	84.04 (1.76)	85.73 (4.03)	79.99 (1.49)	81.8 (1.62)	80.32 (1.62)
2) $\alpha = 90\%$							
G1	89.89 (4.31)	90.38 (4.80)	94.62 (5.73)	94.68 (6.33)	88.90 (4.18)	87.79 (4.19)	89.42 (9.30)
G2	90.18 (2.91)	90.37 (4.67)	93.44 (3.96)	94.56 (6.24)	89.89 (2.89)	90.58 (3.03)	90.19 (3.60)
G3	90.06 (2.74)	89.34 (4.66)	93.68 (3.74)	94.17 (6.23)	90.15 (2.72)	90.78 (2.87)	90.11 (3.29)
G4	90.09 (2.53)	89.5 (4.65)	93.69 (3.46)	94.65 (6.21)	90.12 (2.52)	91.18 (2.67)	90.24 (2.94)
G5	90.29 (1.92)	90.28 (4.60)	93.56 (2.65)	94.73 (6.15)	90.26 (1.92)	91.56 (2.07)	90.32 (2.08)
3) $\alpha = 95\%$							
G1	95.13 (5.30)	95.04 (5.97)	97.45 (8.42)	97.61 (9.31)	93.57 (4.98)	92.61 (5.00)	94.77 (11.09)
G2	95.12 (3.49)	95.37 (5.76)	97.15 (5.82)	97.62 (9.13)	94.87 (3.44)	95.51 (3.60)	95.38 (4.29)
G3	94.97 (3.29)	94.46 (5.74)	97.22 (5.48)	97.60 (9.08)	94.7 (3.24)	95.51 (3.42)	95 (3.92)
G4	95.07 (3.03)	95.03 (5.71)	96.93 (5.08)	97.68 (9.04)	95.12 (3.00)	95.86 (3.18)	95.07 (3.51)
G5	95.31 (2.29)	95.01 (5.64)	96.99 (3.88)	97.52 (8.94)	95.28 (2.29)	96.02 (2.47)	95.17 (2.48)

Table B.2: Coverage probabilities (average length) of different intervals under t_9 distribution with m = 15 and Pattern (ii)

	SB.FH	HM.FH	SB.PR	HM.PR	FH	PR	DIRECT
1) $\alpha = 80\%$							
G1	84.30 (3.89)	84.53 (4.26)	82.80 (5.42)	84.1 (5.98)	76.07 (3.30)	76.50 (3.48)	79.97 (7.25)
G2	83.43 (2.59)	83.77 (4.03)	81.33 (3.65)	83.77 (5.74)	79.57 (2.32)	87.10 (3.09)	79.17 (2.81)
G3	84.20 (2.43)	84.53 (3.99)	82.73 (3.42)	84.57 (5.67)	80.00 (2.19)	88.30 (3.03)	80.03 (2.56)
G4	83.07 (2.25)	84.50 (3.97)	81.87 (3.17)	83.67 (5.65)	79.77 (2.04)	89.33 (2.97)	80.57 (2.29)
G5	82.53 (1.69)	84.87 (3.87)	80.73 (2.38)	84 (5.42)	79.17 (1.57)	87.20 (2.82)	78.13 (1.62)
2) $\alpha = 90\%$							
G1	94.50 (5.78)	94.97 (6.43)	87.57 (10.07)	87.8 (11.18)	85.23 (4.23)	85.83 (4.47)	89.5 (9.3)
G2	93.73 (3.77)	94.00 (5.88)	87.57 (6.69)	87.87 (10.47)	88.80 (2.98)	94.87 (3.96)	90.03 (3.60)
G3	93.83 (3.52)	95.10 (5.80)	87.60 (6.26)	88.33 (10.32)	88.93 (2.81)	95.37 (3.89)	89.47 (3.29)
G4	93.87 (3.23)	94.50 (5.72)	87.60 (5.81)	88.53 (10.25)	89.83 (2.61)	95.50 (3.81)	89.83 (2.94)
G5	91.73 (2.39)	94.47 (5.42)	87.17 (4.28)	88.7 (9.66)	89.17 (2.02)	93.73 (3.62)	88.5 (2.08)
3) $\alpha = 95\%$							
G1	98.10 (8.48)	97.9 (9.45)	89.27 (16.44)	89.37 (18.35)	90.4 (5.04)	91.17 (5.33)	94.40 (11.09)
G2	97.73 (5.36)	97.73 (8.3)	89.33 (10.69)	89.43 (16.85)	93.8 (3.55)	98.50 (4.72)	94.77 (4.29)
G3	97.50 (4.97)	97.73 (8.1)	89.33 (9.96)	89.50 (16.57)	93.67 (3.35)	98.27 (4.64)	94.53 (3.92)
G4	97.23 (4.55)	97.80 (7.92)	89.67 (9.20)	89.87 (16.39)	94.2 (3.11)	97.93 (4.54)	94.90 (3.51)
G5	96.5 (3.29)	97.4 (7.30)	89.93 (6.70)	90.53 (15.28)	94.57 (2.40)	97.13 (4.32)	93.40 (2.48)

Table B.3: Coverage probabilities (average length) of different intervals under shifted exponential distribution with m = 50, B_2 = 50 and Pattern (i)

	SB.FH	HM.FH	DB.FH	SB.PR	HM.PR	DB.PR	FH	PR	DIRECT
1) $\alpha = 80\%$									
G1	80.41 (2.22)	81.98 (2.36)	80.46 (2.27)	84.82 (2.51)	87.88 (2.69)	84.59 (3.08)	83.69 (2.29)	82.36 (2.32)	79.51 (5.13)
G2	79.91 (1.56)	80.96 (2.33)	80.36 (1.59)	83.76 (1.77)	87.22 (2.66)	84.04 (2.26)	80.64 (1.58)	82.97 (1.67)	80.08 (1.99)
G3	79.99 (1.48)	81.47 (2.32)	80.06 (1.50)	83.42 (1.67)	87.41 (2.65)	83.80 (2.06)	80.62 (1.49)	82.90 (1.59)	79.56 (1.81)
G4	80.32 (1.37)	81.26 (2.32)	80.89 (1.4)	83.69 (1.55)	87.52 (2.65)	83.83 (1.96)	80.67 (1.38)	83.41 (1.48)	80.17 (1.62)
G5	80.10 (1.05)	81.69 (2.30)	80.41 (1.07)	83.22 (1.20)	87.29 (2.63)	83.32 (1.51)	80.14 (1.05)	82.80 (1.17)	80.30 (1.15)
2) $\alpha = 90\%$									
G1	90.86 (3.03)	92.24 (3.28)	90.70 (3.16)	93.66 (3.78)	95.04 (4.14)	93.44 (6.27)	91.23 (2.94)	90.27 (2.97)	89.62 (6.58)
G2	90.53 (2.05)	91.43 (3.19)	90.85 (2.13)	92.97 (2.59)	94.46 (4.09)	93.16 (4.54)	90.63 (2.02)	92.15 (2.15)	89.99 (2.55)
G3	89.86 (1.93)	91.99 (3.18)	90.27 (2.00)	93.28 (2.44)	95.21 (4.08)	92.79 (4.27)	89.95 (1.91)	91.67 (2.04)	89.71 (2.33)
G4	90.34 (1.79)	91.91 (3.17)	90.64 (1.85)	92.80 (2.26)	94.41 (4.07)	92.7 (4.08)	90.20 (1.77)	91.98 (1.90)	89.74 (2.08)
G5	90.28 (1.36)	92.04 (3.13)	90.80 (1.41)	92.83 (1.72)	94.90 (4.03)	92.48 (3.21)	90.47 (1.35)	92.10 (1.50)	90.20 (1.47)
3) $\alpha = 95\%$									
G1	95.90 (3.85)	96.88 (4.21)	95.72 (4.42)	96.67 (5.39)	97.25 (5.96)	96.93 (11.17)	94.28 (3.51)	93.54 (3.54)	94.54 (7.84)
G2	95.51 (2.51)	96.10 (4.00)	95.72 (2.82)	96.56 (3.64)	96.98 (5.84)	96.83 (7.92)	94.98 (2.41)	95.36 (2.56)	94.81 (3.04)
G3	94.99 (2.36)	96.76 (3.98)	95.51 (2.65)	96.43 (3.42)	97.28 (5.82)	96.70 (7.51)	94.85 (2.27)	95.94 (2.43)	94.68 (2.77)
G4	95.34 (2.17)	96.54 (3.96)	95.64 (2.43)	96.40 (3.17)	96.99 (5.82)	96.30 (7.07)	94.98 (2.11)	95.91 (2.27)	94.84 (2.48)
G5	95.51 (1.64)	96.57 (3.89)	95.76 (1.83)	96.41 (2.40)	97.08 (5.74)	96.5 (5.53)	95.13 (1.61)	96.10 (1.78)	95.21 (1.75)

Table B.4: Coverage probabilities (average length) of different intervals under shifted exponential distribution with m = 15, B_2 = 50 and Pattern (i)

	SB.FH	HM.FH	DB.FH	SB.PR	HM.PR	DB.PR	FH	PR	DIRECT
1) $\alpha = 80\%$									
G1	83.90(2.66)	85.60 (2.90)	84.43 (2.77)	79.27 (3.29)	80.23 (3.67)	80.27 (5.36)	76.83 (2.23)	80.90 (2.43)	79.83 (5.13)
G2	83.00 (1.78)	85.90 (2.74)	82.90 (1.85)	79.73 (2.19)	80.67 (3.54)	80.20 (3.49)	79.77 (1.59)	89.60 (2.22)	79.40 (1.99)
G3	83.40 (1.68)	85.47 (2.71)	83.23 (1.74)	79.67 (2.06)	80.77 (3.52)	80.10 (3.28)	80.57 (1.51)	89.90 (2.19)	79.90 (1.81)
G4	83.23 (1.55)	85.70 (2.69)	82.93 (1.60)	78.90 (1.90)	80.77 (3.48)	78.97 (2.98)	81.30 (1.41)	90.30 (2.15)	79.40 (1.62)
G5	83.17 (1.18)	85.00 (2.60)	82.73 (1.21)	80.40 (1.43)	81.03 (3.36)	80.50 (2.08)	83.20 (1.12)	89.50 (2.08)	80.77 (1.15)
2) $\alpha = 90\%$									
G1	93.30 (4.08)	93.93 (4.54)	93.40 (5.54)	85.07 (5.83)	85.27 (6.58)	85.53 (15.6)	85.53 (2.87)	89.27 (3.12)	90.10 (6.58)
G2	92.87 (2.59)	93.83 (4.06)	92.63 (3.31)	85.77 (3.75)	86.07 (6.14)	86.43 (9.76)	89.83 (2.04)	95.43 (2.84)	90.33 (2.55)
G3	92.87 (2.42)	93.50 (3.99)	92.60 (3.04)	85.33 (3.53)	85.67 (6.12)	85.90 (9.26)	89.83 (1.94)	95.83 (2.81)	89.67 (2.33)
G4	92.77 (2.22)	94.53 (3.92)	92.87 (2.7)	85.70 (3.23)	86.37 (6.02)	86.07 (8.26)	90.43 (1.81)	95.57 (2.76)	89.53 (2.08)
G5	92.23 (1.65)	93.40 (3.71)	92.00 (1.88)	86.87 (2.38)	87.03 (5.73)	87.13 (5.79)	91.47 (1.43)	95.30 (2.67)	90.6 (1.47)
3) $\alpha = 95\%$									
G1	96.33 (5.99)	96.57 (6.68)	96.70 (12.50)	87.33 (9.30)	87.43 (10.53)	88.53 (24.21)	89.53 (3.41)	92.73 (3.71)	94.93 (7.84)
G2	96.17 (3.64)	96.6 (5.65)	96.10 (7.03)	87.97 (5.82)	87.90 (9.59)	89.23 (14.30)	93.93 (2.44)	97.77 (3.39)	95.33 (3.04)
G3	95.93 (3.38)	96.13 (5.53)	96.30 (6.24)	87.70 (5.46)	87.70 (9.51)	88.87 (13.53)	94.07 (2.31)	97.9 (3.34)	94.97 (2.77)
G4	96.47 (3.08)	97.07 (5.39)	96.53 (5.67)	87.93 (5.01)	88.2 (9.39)	89.40 (12.21)	95.03 (2.16)	98.13 (3.29)	94.63 (2.48)
G5	96.07 (2.22)	96.50 (4.95)	96.23 (3.69)	89.17 (3.60)	88.90 (8.72)	90.27 (8.69)	96.00 (1.71)	98.00 (3.19)	95.07 (1.75)

Bibliography

Arora, V., P. Lahiri, and K. Mukherjee (1997). Empirical bayes estimation of finite population means from complex surveys. *Journal of the American Statistical Association 92*(440), 1555–1562.

Battese, G. E., R. M. Harter, and W. A. Fuller (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association 83*(401), 28–36.

Bell, W. R. and E. T. Huang (2006). Using the t-distribution to deal with outliers in small area estimation. In *Proceedings of statistics Canada symposium*.

Betti, G., L. Neri, and F. Bellini (2003). Poverty and inequality mapping in albania: final report.

Brown, G., R. Chambers, P. Heady, and D. Heasman (2001). Evaluation of small area estimation methods—an application to unemployment estimates from the uk lfs. In *Proceedings of statistics Canada symposium*, Volume 2001, pp. 1–10. Statistics Canada.

Chambers, R. and N. Tzavidis (2006). M-quantile models for small area estimation. *Biometrika 93*(2), 255–268.

Chatterjee, S., P. Lahiri, and H. Li (2008). Parametric bootstrap approximation to the distribution of EBLUP and related prediction intervals in linear mixed models. *The Annals of Statistics 36*(3), 1221 – 1245.

Cox, D. (1975). Prediction intervals and empirical bayes confidence intervals. *Journal of Applied Probability 12*(S1), 47–55.

Datta, G. S. and P. Lahiri (1995). Robust hierarchical bayes estimation of small area characteristics in the presence of covariates and outliers. *Journal of Multivariate Analysis 54*(2), 310–328.

Datta, G. S., J. Rao, and D. D. Smith (2005). On measuring the variability of small area estimators under a basic area level model. *Biometrika 92*(1), 183–196.

Edochie, I., D. Newhouse, N. Tzavidis, T. Schmid, E. Foster, A. L. Hernandez, A. Ouedraogo, A. Sanoh, and A. Savadogo (2024). Small area estimation of poverty in four west african countries by integrating survey and geospatial data. *Journal of Official Statistics*, 0282423X241284890.

Elbers, C., J. Lanjouw, and P. Lanjouw (2003). Micro-level estimation of poverty and inequality. *Econometrica 71*(1), 355–364.

Fabrizi, E. and C. Trivisano (2010). Robust linear mixed models for small area estimation. *Journal of Statistical Planning and Inference 140*(2), 433–443.

Fay, R. E. and R. A. Herriot (1979). Estimates of income for small places: an application of james-stein procedures to census data. *Journal of the American Statistical Association 74*(366a), 269–277.

Foster, J., J. Greer, and E. Thorbecke (1984). A class of decomposable poverty measures. *Econometrica: journal of the econometric society*, 761–766.

Fuller, W. A. (2011). *Sampling statistics*. John Wiley & Sons.

Ghosh, M. and P. Lahiri (1987). Robust empirical bayes estimation of means from stratified samples. *Journal of the American Statistical Association 82*(400), 1153–1162.

Ha, N. S., P. Lahiri, and V. Parsons (2014). Methods and results for small area estimation using smoking data from the 2008 national health interview survey. *Statistics in Medicine 33*(22), 3932–3945.

Hall, P. (2013). *The bootstrap and Edgeworth expansion*. Springer Science & Business Media.

Hall, P. and T. Maiti (2006). On parametric bootstrap methods for small area prediction. *Journal of the Royal Statistical Society Series B: Statistical Methodology 68*(2), 221–238.

Hawala, S. and P. Lahiri (2018). Variance modeling for domains. *Statistics and Applications 16*(1), 399–409.

Hirose, M. Y. and P. Lahiri (2018). Estimating variance of random effects to solve multiple problems simultaneously. *The Annals of Statistics 46*(4), 1721–1741.

Hobza, T. and D. Morales (2013). Small area estimation under random regression coefficient models. *Journal of Statistical Computation and Simulation 83 (11)*, 2160–2177.

Jean, N., M. Burke, M. Xie, W. Davis, D. Lobell, and S. Ermon (2016). Micro-level estimation of poverty and inequality. *Science 353*(6301), 790–794.

Jiang, J. and P. Lahiri (2006). Mixed model prediction and small area estimation. *Test 15*, 1–96.

Jiang, J. and T. Nguyen (2007). *Linear and generalized linear mixed models and their applications*, Volume 1. Springer.

Jiang, J. and T. Nguyen (2012). Small area estimation via heteroscedastic nested-error regression. *Canadian Journal of Statistics 40*, 588–603.

Jiang, J., T. Nguyen, and J. S. Rao (2011). Best predictive small area estimation. *Journal of the American Statistical Association 106*(494), 732–745.

Jiang, J., T. Nguyen, and J. S. Rao (2015). Observed best prediction via nested-error regression with potentially misspecified mean and variance. *Survey Methodology 41*(1), 37–56.

Jiang, J. and J. S. Rao (2020). Robust small area estimation: An overview. *Annual review of statistics and its application 7*, 337–360.

Kalton, G. (1983). Models in the practice of survey sampling. *International Statistical Review/Revue Internationale de Statistique*, 175–188.

Kalton, G. and C. F. Citro (2000). *Small-area Income and Poverty Estimates: Priorities for 2000 and beyond*. National Academies Press.

Krennmair, P. and T. Schmid (2022). Flexible domain prediction using mixed effects random forests. *Journal of the Royal Statistical Society Series C: Applied Statistics 71*(5), 1865–1894.

Kubokawa, T., S. Sugasawa, M. Ghosh, and S. Chaudhuri (2016). Prediction in heteroscedastic nested error regression models with random dispersions. *Statistica Sinica 26*, 465–492.

Lahiri, P. and K. Mukherjee (2007). Hierarchical bayes estimation of small area means under generalized linear models and design consistency. *Annals of Statistics 35*, 724–737.

Lahiri, P. and J. Rao (1995). Robust estimation of mean squared error of small area estimators. *Journal of the American Statistical Association 90*(430), 758–766.

Lahiri, P. and N. Salvati (2023). A nested error regression model with high-dimensional parameter for small area estimation. *Journal of the Royal Statistical Society Series B: Statistical Methodology 85*(2), 212–239.

Li, H. (2007). *Small area estimation: An empirical best linear unbiased prediction approach*. University of Maryland, College Park.

Li, H. and P. Lahiri (2010). An adjusted maximum likelihood method for solving small area estimation problems. *Journal of multivariate analysis 101*(4), 882–892.

Marino, M. F., M. G. Ranalli, N. Salvati, and M. Alfo (2019). Semiparametric empirical best prediction for small area estimation of unemployment indicators. *The Annals of Applied Statistics 13*(2), 1166–1197.

McCullough, B. and H. Vinod (1998). Implementing the double bootstrap. *Computational Economics 12*, 79–95.

Mendez, G. (2008). *Tree-based methods to model dependent data*. Arizona State University.

Molina, I., P. Corral, and M. Nguyen (2022). Estimation of poverty and inequality in small areas: review and discussion. *Test 31*(4), 1143–1166.

Molina, I., B. Nandram, and J. N. K. Rao (2014). Small area estimation of general parameters with application to poverty indicators: A hierarchical Bayes approach. *The Annals of Applied Statistics 8*(2), 852 – 885.

Molina, I. and J. N. Rao (2010). Small area estimation of poverty indicators. *Canadian Journal of statistics 38*(3), 369–385.

Newhouse, D. L., J. D. Merfeld, A. Ramakrishnan, T. Swartz, and P. Lahiri (2022). Small area estimation of monetary poverty in mexico using satellite imagery and machine learning. *Available at SSRN 4235976*.

Otto, M. C. and W. R. Bell (1995). Sampling error modelling of poverty and income statistics for states. In *American Statistical Association, Proceedings of the Section on Government Statistics*, pp. 160–165.

Prasad, N. G. N. and J. N. K. Rao (1990a). The estimation of the mean squared error of small area estimators. *Journal of the American Statistical Association 85 (409)*, 163–171.

Prasad, N. N. and J. N. Rao (1990b). The estimation of the mean squared error of small-area estimators. *Journal of the American statistical association 85*(409), 163–171.

Rao, J. N. and I. Molina (2015). *Small area estimation*. John Wiley & Sons.

Rodas, P. C., I. Molina, and M. Nguyen (2021). Pull your small area estimates up by the bootstraps. *Journal of Statistical Computation and Simulation 91*(16), 3304–3357.

Saegusa, T., S. Sugasawa, and P. Lahiri (2022). Parametric bootstrap confidence intervals for the multivariate fay–herriot model. *Journal of Survey Statistics and Methodology 10*(1), 115–130.

Shao, J. (2008). *Mathematical statistics*. Springer Science & Business Media.

Shi, S. G. (1992). Accurate and efficient double-bootstrap confidence limit method. *Computational statistics & data analysis 13*(1), 21–32.

Stukel, D. M. and J. Rao (1999). On small-area estimation under two-fold nested error regression models. *Journal of statistical planning and inference 78*(1-2), 131–147.

Sugasawa, S. and T. Kubokawa (2017). Heteroscedastic nested error regression models with variance functions. *Statistica Sinica 27*(3), 1101–1123.

Torabi, M. and J. Rao (2014). On small area estimation under a sub-area level model. *Journal of Multivariate Analysis 127*, 36–55.

Tzavidis, N., N. Salvati, M. Pratesi, and R. Chambers (2008). M-quantile models with application to poverty mapping. *Statistical Methods and Applications 17*, 393–411.

Xie, D., T. E. Raghunathan, and J. M. Lepkowski (2007). Estimation of the proportion of overweight individuals in small areas—a robust extension of the fay–herriot model. *Statistics in Medicine 26*(13), 2699–2715.

Yoshimori, M. and P. Lahiri (2014). A second-order efficient empirical Bayes confidence interval. *The Annals of Statistics 42*(4), 1233 – 1261.

You, Y. and J. Rao (2002). A pseudo-empirical best linear unbiased prediction approach to small area estimation using survey weights. *Canadian Journal of Statistics 30*(3), 431–439.