On Small Area Estimation Problems with Measurement Errors and Clustering

by

Elaheh Torkashvand

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfilment of the requirements of the degree of

Doctor of Philosophy

Department of Statistics
University of Manitoba
Winnipeg

Abstract

In this dissertation, we develop some new statistical methodologies for solving small area estimation problems with measurement errors. The first problem we consider is to predict small area means for the unit-level regression model with the functional measurement error in the area-specific covariate. To this end, we obtain the James-Stein (JS) estimate of the true area-specific covariate. Consequently, we construct the pseudo Bayes (PB) and pseudo empirical Bayes (PEB) predictors of small area means and estimate the mean squared prediction error (MSPE) associated with each predictor using the jackknife method. We apply our methodology to a cross sectional study in New Zealand to predict diastolic blood pressure using the cholesterol level. The second problem we study is to modify the point estimation of the true area-specific covariate under the unit-level regression model with the functional measurement error in the area-specific covariate such that the histogram of the predictors of the small area means gets closer to the true histogram of the small area means. We propose the constrained Bayes (CB) estimate of the true area-specific covariate. We show the superiority of the CB estimate of the true area-specific covariate over its counterpart based on the maximum likelihood (ML) estimate in terms of the Bayes risk. We also show that the PB predictor of the small area mean based on the CB estimate of the true area-specific covariate dominates its

counterpart based on the ML estimate of the true area-specific covariate in terms of the Bayes risk. Further, we compare the performance of different predictors of the small area means using measures such as sensitivity, specificity, positive predictive value, negative predictive value, and MSPE. Similar to the first problem, we analyze the cross sectional study in New Zealand to predict diastolic blood pressure using the cholesterol level. We believe that using the PEB and pseudo hierarchical Bayes predictors of small area means based on the constrained empirical Bayes (CEB) and constrained hierarchical Bayes (CHB) estimates of the true area-specific covariate offers higher precision in recognizing socio-economic groups which are in danger of the prehypertension. Clustering the small areas to understand the behaviour of the random effects better and accordingly, to predict the small area means is the final problem we address. We consider the Fay-Herriot model for this problem. In the first step, we design a statistical test to evaluate the assumption of the equality of the variance components in different clusters. In the case of the rejection of the null hypothesis of the equality of the variance components, we implement a modified version of Tukey's method to combine some clusters. Subsequently, we calculate the MSPE to evaluate the effect of the clustering on the precision of the predictors of the small area means. We apply our methodology to the National Health and Nutrition Examination Survey (NHANES) for 2011-2012 to predict the waist circumference based on the Body Mass Index (BMI).

Acknowledgments

First and foremost I want to thank my supervisor Dr. Mohammad Jafari Jozani. It has been an honour to be his PhD student. He has taught me how to enjoy being in Academia and interacting with students. I appreciate all his contributions of time, ideas, and funding to make my Ph.D program fruitful and promising. During tough time of my PhD program, his encouragement and motivation made me continue my research. Thank you!

I also thank my co-supervisor, Dr. Mahmoud Torabi, for his constructive comments during my PhD pursuit. His applied view towards the research helped me to strengthen my computational skills. Our discussions always elevated my understanding of the small area estimation. I appreciate them all.

Examiner committee with their constructive comments improved the quality of this research thesis. I thank Dr. Malay Ghosh, Dr. Liqun Wang, and Dr. Rashid Ahmed for their time and effort.

I would also like to thank my professors, Dr. Brad Johnson, Dr. Theodore Koulis, Dr. Alexandre Leblanc, Dr. Liqun Wang, and Dr. Xikui Wang, fellow students, specially Dr. Mohammad Nourmohammadi, Dr. Armin Hatefi, and Rojiar Haddadian, and department staff for the friendly atmosphere I experienced during my PhD program. I express my appreciation to Carrie Paquette, my mentor during teaching times, to my officemate, Jenna Tichon, for her kindness, and to Dave Gabrielson for his technical support.

Finally, financial support by University of Manitoba Graduate Fellowship (UMGF), Manitoba Graduate Scholarship (MGS), Faculty of Science Studentship, and Natural Sciences and Engineering Research Council of Canada (NSERC) are acknowledged.

To my father, Valiollah:

for all his encouragement;

my mother, Mahnaz:

for her patience;

and

angels of my life:

Fereshteh and Hossein.

Publications

- Torkashvand E, Jafari Jozani M, Torabi M. (2015) Pseudo-Empirical Bayes Estimation of Small Area Means Based on the James-Stein Estimation in Linear Regression Models with Functional Measurement Errors. Canadian Journal of Statistics 43: 265–287

- Torkashvand E., Jafari Jozani M., and Torabi M. (2016) The Constrained Bayes Estimation in Small Area Models with Functional Measurement Error. TEST: 1–21.

- Torkashvand E., Jafari Jozani M., and Torabi M. (2016+) Clustering in Small Area Estimation with Area-Level Linear Mixed Models. Under Revision.

Contents

Contents **v**

List of Tables **ix**

List of Figures **xiii**

1 Introduction **1**

1.1 Small Area Estimation 2

1.2 Measurement Error 7

1.3 Clustering 12

1.4 Thesis Overview 16

2 PEB Predictor Using the JS Estimation **21**

2.1 A New Pseudo-Bayes Predictor 23

2.2 Pseudo-EB Predictor 27

2.3 The Jackknife Estimation of MSPE($\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$) 38

2.4 Empirical Results 40

2.4.1 Efficiency of PB Estimators 41

2.4.2 Simulation Study 43

2.5 An Application 46

3 PEB Predictor Using the CB Estimation **55**

3.1 CB Estimate of the True Area-Specific Covariate 58

3.2 Performance measures 68

3.3 PEB and PHB Predictors of Small Area Means 71

3.4 An Application 73

3.5 Simulation Study 77

4 Clustering in Small Area Estimation **87**

4.1 Clustering in Area-Level Small Area Model 89

4.2 Testing the Equality of Variance Components 91

4.3 EBLUP and its MSPE Estimation 101

4.4 Real Data Analysis 104

4.5 Simulation Studies 108

4.5.1 Evaluating the test statistic and the effect of the complete and combined clustering on EMSPE and mspe 110

4.5.2 Assessing the performance of the estimator of D_i 119

5 Summary and Future Work **125**

Bibliography **135**

List of Tables

2.1 Numerical values of the MSPE of $\widehat{\gamma}_{i\mathrm{GS}}^{\mathrm{PB}}$, $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}}$ and $\widehat{\gamma}_{i\mathrm{DRT}}^{\mathrm{PB}}$ and the empirical MSPE of $\widehat{\gamma}_{i\mathrm{GS}}^{\mathrm{PEB}}$, $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$, $\widehat{\gamma}_{i\mathrm{DRT}}^{\mathrm{PEB}}$ and $\widehat{\gamma}_{i\mathrm{NAI}}^{\mathrm{PEB}}$ 51

2.2 Numerical values of the weighted MSPE of $\widehat{\gamma}_{\mathrm{GS}}^{\mathrm{PB}}$, $\widehat{\gamma}_{JS}^{\mathrm{PB}}$, $\widehat{\gamma}_{\mathrm{DRT}}^{\mathrm{PB}}$, and $\widehat{\gamma}_{NAI}^{\mathrm{PB}}$ for different weight schemes 52

2.3 Numerical values of the weighted EMSPE of $\widehat{\gamma}_{\mathrm{GS}}^{\mathrm{PEB}}$, $\widehat{\gamma}_{JS}^{\mathrm{PEB}}$, $\widehat{\gamma}_{\mathrm{DRT}}^{\mathrm{PEB}}$, and $\widehat{\gamma}_{NAI}^{\mathrm{PEB}}$ for different weight schemes 52

2.4 The estimated values of the parameters and their corresponding biases and MSEs 52

2.5 The values of the components of the EMSPE of $\widehat{\gamma}_{\mathrm{GS}}^{\mathrm{PEB}}$, $\widehat{\gamma}_{JS}^{\mathrm{PEB}}$, $\widehat{\gamma}_{\mathrm{DRT}}^{\mathrm{PEB}}$, and $\widehat{\gamma}_{NAI}^{\mathrm{PEB}}$ 53

2.6 Percent relative bias of the jackknife estimators of MSPE and the bias of different PEB estimators of the γ_i's 54

3.1 The MSE of the model parameters using GS and HB methods 78

3.2 Numerical values of the MSPE of $\hat{\gamma}_{iB}^{PB}$, $\hat{\gamma}_{iCB}^{PB}$ and $\hat{\gamma}_{iML}^{PB}$ and the EMSPE of $\hat{\gamma}_{iEB}^{PEB}$, $\hat{\gamma}_{iCEB}^{PEB}$, $\hat{\gamma}_{iHB}^{PEB}$, $\hat{\gamma}_{iCHB}^{PEB}$, and $\hat{\gamma}_{iML}^{PEB}$ 86

4.1 The comparison between the complete, combined, and simple clustering in terms of the quantiles of their corresponding mspe 108

4.2 MSE of the REML and MMM estimates for different variance components 112

4.3 MSE of the REML and MMM estimates for similar variance components in some clusters 113

4.4 The power of the test for different values of $\sigma_{u_l}^2$'s and D_j's . . . 114

4.5 The effect of removing clusters on the power of the test for the case of $\sigma_u^2 = (1, 1, 2, 4, 2.5, 4, 3, 3, 5, 6)$ 115

4.6 The percentage of times the null hypothesis is rejected by mistake115

4.7 Comparison of EMSPE of predictors of small area means using different approaches based on their deciles: (a) $\frac{\text{EMSPE}_s}{\text{EMSPE}_c}$, (b) $\frac{\text{EMSPE}_s}{\text{EMSPE}_{cb}}$, (c) $\frac{\text{EMSPE}_{cb}}{\text{EMSPE}_c}$ when $\sigma_u^2 = (1, 8, 25,\ 5, 64, 10, 49, 36, 40, 13)$ and $D_{jl} \sim \text{Uniform}(0.25, 0.5)$. For columns (a) and (b), values larger than 1 indicate the complete and combined clustering approaches estimate the small area mean with a higher precision in comparison with the simple approach, respectively. Values larger than 1 in column (c) indicate the complete clustering approach performs better than the combined clustering approach in estimating the small area mean. 117

4.8 The summary statistics for the RB of the estimator of MSPE of small area means using different approaches for $\sigma_u^2 = (1, 8, 25, 5, 64, 10, 49, 36, 40, 13)$: (a) the complete clustering approach, (b) the combined clustering approach, and (c) the simple approach . . 118

4.9 Comparison of EMSPE of predictors of small area means using different approaches based on their deciles: (a) $\frac{\text{EMSPE}_s}{\text{EMSPE}_c}$, (b) $\frac{\text{EMSPE}_s}{\text{EMSPE}_{cb}}$, (c) $\frac{\text{EMSPE}_{cb}}{\text{EMSPE}_c}$ when $\sigma_u^2 = (1, 8, 25,\ 25, 64, 25, 49, 40, 40, 13)$ and $D_{jl} \sim \text{Uniform}(0.25, 0.5)$. For columns (a) and (b), values larger than 1 indicate the complete and combined clustering approaches estimate the small area mean with a higher precision in comparison with the simple approach, respectively. Values larger than 1 in column (c) indicate the complete clustering approach performs better than the combined clustering approach in estimating the small area mean. 122

4.10 The summary statistics for the RB of the estimator of MSPE of small area means using different approaches for $\sigma_u^2 = (1, 8, 25, 25, 64, 25, 49, 40, 40, 13)$: (a) the complete clustering approach, (b) the combined clustering approach, and (c) the simple approach . . 123

List of Figures

1.1 (a) Number of clusters using K-means clustering, and (b) Silhouette average for k=2 clusters 16

1.2 The dendrogram of data using the hierarchical clustering method 17

2.1 EMSPE and bias of $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$ (dashed line), $\widehat{\gamma}_{IDRT}^{\mathrm{PEB}}$ (solid line) and $\widehat{\gamma}_{i\mathrm{NAI}}^{\mathrm{PEB}}$ (dot line) for the different areas 47

2.2 Diastolic Blood Pressure versus Cholesterol 49

2.3 The pseudo-empirical Bayes predictors of small area means. . . 50

2.4 The MSPE estimation of pseudo-empirical Bayes predictors of small area means. 50

3.1 Average Cholesterol Level Estimates Using Different Approaches 75

3.2 The estimated values of the small area mean. 75

3.3 The estimated values of the measures of performance of different areas. 76

3.4 Exact versus asymptotic values of ν using hierarchical and empirical Bayes estimates of the model parameters 79

3.5 The density of the estimated values of the true area-specific covariate using different approaches (solid line) vs the density of the true area-specific covariate (dashed line) 81

3.6 The simulated values of the measures of performance of different areas. 82

3.7 The histograms of the predicted small area means for the first nine areas 83

4.1 The effect of number of clusters on within groups sums of squares in the k-means clustering 105

4.2 Histogram of the mspe's: (a) the complete clustering method, (b) the combined clustering method, (c) the simple method . . 109

4.3 Histogram of predicted small area means: (a) the complete clustering method, (b) the combined clustering method, (c) the simple method 109

4.4 Histogram of the neglected terms in estimation of $\sigma^2_{e_i}$'s given in (4.5.2). 121

4.5 (a) histogram of the true D_i's vs. (b) histogram of the estimates of D_i's 121

Chapter 1

Introduction

Small area estimation is a statistical methodology that has been given considerable attention in recent years. A sample is obtained from the population. A short while later, the interest is to estimate parameters in subpopulations. The traditional area-specific direct predictors tend to have inadequate precision because of the small sample sizes corresponding to each small area (subpopulation). For instance, in health services research, policy decisions about implementing programs or projects in the small areas are made using predictors of underlying parameters. In the context of small area estimation necessary approaches are developed to estimate parameters of interest in subpopulations and survey researchers work to provide more reliable predictions for small areas. In model-based small area estimation, the model is used to borrow information from other small areas to estimate parameters of interest in each small area using the response variable and corresponding covariates. There are situations

where we cannot observe the exact value of the covariate of interest due to device malfunction, human error, or cost of the operation. This unwanted error will reduce the precision of estimated values in small areas, so, the prediction of the small area mean or total will be less reliable. In order to address this problem, the measurement error model is used. Measurement error is an important statistical approach which takes into account the uncontrollable error to give more precise estimates of parameters, and eventually the predictors of small area mean and total. There are different types of measurement errors. In this thesis, we focus on the linear functional measurement error and leave the discussion regarding the structural measurement error as future research areas. The other statistical technique we use to obtain more accurate estimates of small area parameters is clustering. We have shown clustering increases the efficiency of small area estimation because of the homogeneity between elements (small areas) of the same cluster. In this chapter, we provide an overview of small area estimation, measurement error, and clustering. At the end, we present an overview of the thesis.

1.1. Small Area Estimation

Sample surveys are commonly conducted to provide reliable predictors for the finite population parameters such as totals or means. Methods to derive such "direct" predictors are based on total sample size. In recent years, there has been increasing demand to get reliable predictors for sub-populations (small

areas), such as counties or gender-age groups, due to their growing use in formulating policies and programs, allocation of government funds, regional planning, marketing decisions at local level, and other uses. Design-based and generalized regression estimators are two common approaches to derive direct estimation of small area parameters. In this chapter, we give a short review of the design-based approach as a direct method to estimate small area parameters.

Suppose V shows the survey population consisting of N distinct elements identified through the labels $j = 1, \ldots, N$. Further, assume there is no measurement error in the characteristic of interest, y_j for $j = 1, \ldots, N$. The parameters of interest are either population total $Y = \sum_{j=1}^{N} y_j$ or the population mean $\gamma = \frac{Y}{N}$. A sampling design is used to draw a sample s from V with probability $p(s)$. Design variables such as stratum indicator variables and size measures of clusters can be taken into account to construct the sample selection probability $p(s)$. Assume there is a complete response, i.e., all sample units corresponding to characteristic of interest, y_j for $j \in s$, can be observed. An estimator $\hat{Y}$ is said to be design-unbiased (or p-unbiased) if

$$\mathrm{E}_p(\hat{Y}) = \sum_{s} p(s)\hat{Y}_s = Y$$

where s belongs to the overall possible samples under the specified sampling design. The design variance of $\hat{Y}$ is $V_p(\hat{Y}) = \mathrm{E}_p(\hat{Y} - \mathrm{E}_p(\hat{Y}))^2$. Accordingly, $v(\hat{Y}) = s^2(\hat{Y})$ estimates $V_p(\hat{Y})$. The sampling variance $v(\hat{Y})$ is $p-$unbiased

for $V_p(\hat{Y})$ if $\mathrm{E}_p(v(\hat{Y})) = V_p(\hat{Y})$. An estimator $\hat{Y}$ is called design-consistent (or $p-$consistent) if $\hat{Y}$ converges in probability to Y. Rao and Molina (2015) presented a sufficient condition for the design consistency as $\frac{1}{N}(\mathrm{E}_p(\hat{Y}) - Y)$ and $\frac{1}{N^2}V_p(\hat{Y})$ tends to zero when the sample size increases.

The problem with the "direct" estimators of small area parameters is their unacceptable large standard deviation for small areas with few or no sample units. In order to increase the effective sample size and thus decrease the standard error of the corresponding estimators, traditional implicit models including synthetic estimators (Gonzalez, 1973), composite estimators (Schaible, 1978), and the James-Stein estimators (Purcell and Kish, 1979) have been proposed. Explicit model based estimators (Jiang and Lahiri, 2006; Jiang, 2010; Rao and Molina, 2015) have also been suggested to borrow strength from related areas by introducing random effects into the study. In the context of linear mixed models, such small area models may be classified into two broad types: (i) Area-level models that relate small area direct estimates to area-specific covariates; such models are used if unit-level data are not available (Fay and Herriot, 1979). (ii) Unit-level models that relate the unit values of a study variable to associated unit-level covariates with known area-specific covariates means (Battese et al., 1988). A comprehensive account of model-based small area estimation under area-level and unit-level models is given by Rao and Molina (2015).

The commonly used area-level regression model is given by Fay and Herriot (1979)

$$y_i = \mathbf{X}_i\beta + u_i + e_i, \quad i = 1, \ldots, m, \tag{1.1.1}$$

where y_i is the variable of interest, $\mathbf{X}_i = (1, X_{i1}, X_{i2}, \ldots, X_{ip})$ is the vector of covariates, $\beta = (\beta_0, \beta_1, \ldots, \beta_p)'$ is the vector of regression coefficients, m, u_i's, and e_i's are the number of areas, the area-level random effects, and the random errors, respectively. Also, assume that u_i are independently and identically distributed (i.i.d) random variables with $N(0, \sigma_u^2)$, and e_i's are independently and normally distributed with $e_i \sim N(0, D_i)$, where D_i's are all known and σ_u^2 is unknown. We assume that there is no sample selection bias and the sampling design is not informative (Rao and Molina, 2015).

On the other hand, the following nested error linear regression model describes a unit-level model (Battese et al., 1988)

$$y_{ij} = \mathbf{X}_{ij}\beta + u_i + e_{ij}, \quad i = 1, \ldots, m, j = 1, \ldots, N_i, \tag{1.1.2}$$

where N_i is the known population size of the ith area (i=1,...,m), y_{ij} is the value of the study variable associated with the jth unit in the ith area and $\mathbf{X}_{ij} = (1, X_{ij1}, X_{ij2}, \ldots, X_{ijp})$. Further, the random errors e_{ij} and the area-level random effects u_i are assumed to be mutually independent with $e_{ij} \overset{i.i.d}{\sim} N(0, \sigma_e^2)$ and $u_i \overset{i.i.d}{\sim} N(0, \sigma_u^2)$ where σ_u^2 and σ_e^2 are unknown.

Models (1.1.1) and (1.1.2) can be summarized as

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}u + \mathbf{e}, \tag{1.1.3}$$

where $\mathbf{y}$ is the $n \times 1$ vector of sample observations, $\mathbf{X} = (\mathbf{X}_1', \ldots, \mathbf{X}_m')'$ where $\mathbf{X}_i' = (X_{i1}, X_{i2}, \ldots, X_{in_i})'$ with $X_{ij} = (1, X_{ij1}, X_{ij2}, \ldots, X_{ijp})$ for $j = 1, \ldots, n_i$, is a known $n \times p$ matrix, $\mathbf{Z}$ is a known $n \times h$ matrix, and $n = \sum_{i=1}^{m} n_i$. We assume both $\mathbf{X}$ and $\mathbf{Z}$ are full rank. Further, it is assumed that $\mathbf{u}$ and $\mathbf{e}$ are independently normally distributed with means $\mathbf{0}$ and variance matrices $\mathbf{G}$ and $\mathbf{R}$, respectively. $\mathbf{G}$ and $\mathbf{R}$ are functions of q unknown parameters $\delta = (\delta_1, \delta_2, \ldots, \delta_q)$. The variance-covariance matrix of $\mathbf{Y}$ is given by $\mathrm{Var}(\mathbf{y}) = \mathbf{V} = \mathbf{R} + \mathbf{Z}\mathbf{G}\mathbf{Z}'$.

Consider $\mu = \mathbf{l}'\beta + \mathbf{m}'\mathbf{u}$. The interest is to estimate μ for specified vectors $\mathbf{l}$ and $\mathbf{m}$. To this end, we need to find a linear estimator of μ of the form $\hat{\mu} = \mathbf{a}'\mathbf{Y} + \mathbf{b}$ for real valued vectors $\mathbf{a}$ and $\mathbf{b}$ such that the model is unbiased under the squared error loss function, $L(\hat{\mu}, \mu) = (\hat{\mu} - \mu)^2$ i.e. $\mathrm{E}(\mu) = \mathrm{E}(\hat{\mu})$. The best linear unbiased predictor (BLUP) of μ in terms of minimum mean squared prediction error (MSPE) is given by

$$\begin{aligned} \tilde{\mu} &= t(\delta, \mathbf{Y}) \\ &= \mathbf{l}'\tilde{\beta} + \mathbf{m}'\tilde{\mathbf{u}} \\ &= \mathbf{l}'\tilde{\beta} + \mathbf{m}'\mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\beta}), \end{aligned} \tag{1.1.4}$$

where $\tilde{\beta} = \tilde{\beta}(\delta) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$ (Henderson, 1950). Accordingly, MSPE of BLUP of μ as a measure of precision is given by

$$MSPE(t(\delta, \mathbf{Y})) = \mathrm{E}(t(\delta, \mathbf{Y}) - \mu)^2.$$

Consequently, the empirical best linear unbiased predictor (EBLUP) of μ is obtained by substituting the estimate of δ in (1.1.4). There are approaches such as restricted maximum likelihood (REML) and maximum likelihood (ML) to estimate δ. We refer the reader for a comprehensive explanation on the estimation of δ to Rao and Molina (2015). The MSPE of the EBLUP is defined as

$$\text{MSPE}(t(\hat{\delta}, \mathbf{Y})) = \text{E}(t(\hat{\delta}, \mathbf{Y}) - \mu)^2.$$

Prasad and Rao (1990) obtained the second order unbiased approximation to the MSPE of the EBLUP under the model (1.1.3). The second order unbiased approximation of MSPE is dependent on model parameters. The estimated value of this approximation is also presented in their work. We refer the reader for a comprehensive explanation on the estimation of $\text{MSPE}(t(\hat{\delta}, \mathbf{Y}))$ to Prasad and Rao (1990).

1.2. Measurement Error

Measurement error occurs when it is not possible to observe the true variable of interest. There are many circumstances where covariates are subject to measurement errors. For example, consider the problem of predicting the diastolic blood pressure using cholesterol level as the covariate. The observed value of the cholesterol level is different from the true one because the patient might have fatty food before the blood test more than their usual consumption level or vice versa. The other factor that affects the result is device malfunction

in the laboratory. Thus, it seems likely that the measured covariate is influenced by some error.

In order to use the measurement error model, it is important to determine whether we model the distribution of the observed variable given the true one or vice versa. The first case is referred to as the classical measurement error while the second one is the Berkson model. In this thesis, the focus is on the classical measurement error.

For quantitative variables, the additive measurement error model is commonly used. We have

$$X = x + \eta$$

where η is a random variable such that $\mathrm{E}(X|x) = x$. This means X, the observed variable, is an unbiased estimator of x, the true one. In the nonadditive measurement error, $\mathrm{E}(X|x) \neq x$ which implies X is a biased observation of x (Buonaccorsi, 2010).

In a pioneering paper, Ghosh and Sinha (2007), henceforth abbreviated GS, proposed a nested error linear regression model with an area-level covariate subject to the functional measurement error.

Following them, we consider the nested error linear regression population model as

$$y_{ij} = \beta_0 + \beta_1 x_i + u_i + e_{ij}, \quad i = 1, \ldots, m, j = 1, \ldots, N_i, \tag{1.2.1}$$

with

$$X_{ij} = x_i + \eta_{ij}, \quad i = 1, \ldots, m, j = 1, \ldots, N_i, \tag{1.2.2}$$

where N_i is the known population size of the ith area (i=1,...,m), y_{ij} is the value of the study variable associated with the jth unit in the ith area and x_i is the unknown true area-specific covariate associated with y_{ij}. Further, the random errors e_{ij}, measurement errors η_{ij} and the area-level random effects u_i are assumed to be mutually independent with $e_{ij} \overset{i.i.d}{\sim} N(0, \sigma_e^2)$, $\eta_{ij} \overset{i.i.d}{\sim} N(0, \sigma_\eta^2)$ and $u_i \overset{i.i.d}{\sim} N(0, \sigma_u^2)$. The observable covariate in the study is X_{ij} while the true area-specific covariate is x_i, $x_{i1} = x_{i2} = \cdots = x_{iN_i}$, for $i = 1, \ldots, m$. In (1.2.2), the distribution of the observed covariate is modelled given the true area-specific covariate. This indicates we deal with classical measurement error. The functional measurement error is considered for the observed covariate in (1.2.2) where x_i's are fixed and unknown true area-specific covariate (Fuller, 2009). When there is no measurement error in covariate (i.e., $\sigma_\eta^2 = 0$) we have $x_i = X_{ij}$, and (1.2.1) and (1.2.2) reduce to the unit level regression model

$$y_{ij} = b_0 + b_1 X_{ij} + u_i + e_{ij}, \quad i = 1, \ldots, m, j = 1, \ldots, N_i.$$

proposed by Battese et al. (1988). Assuming that there is no sample selection bias (i.e. the sampling scheme is not informative), a sample of size n_i is selected from the i'th area and the sample data, without loss of generality, is denoted by $(\mathbf{y}, \mathbf{X}) = \{(y_{ij}, X_{ij}),\ i = 1, \ldots, m,\ j = 1, \ldots, n_i\}$. Also, it is assumed that the covariate is only observed for the units in the sample.

In this thesis, we consider the simple case of one covariate to predict the small area mean. As an example, we estimate the diastolic blood pressure using the cholesterol level. Definitely, there are other factors that influence the

response variable, however, in this thesis, we only consider one covariate in the model. The random variable, u_i for $i = 1, \ldots, m$, shows the role of unattended covariates in the prediction of the small area mean.

However, the covariate which is studied to predict the small area mean is subject to measurement error. In our example, the level of cholesterol shown on a blood test is not an accurate value because it is subject to some types of error. There might be some device error included in the reading of cholesterol. Also, the value shown on the blood test is highly dependent on the type of food the patient took before test. This justifies the measurement error in the covariate conceptually. Technically, we use the classical linear measurement error shown in model (1.2.2). In the application, we use many variables (gender, age, BMI, ethnic, and smoking status) to make domains (small areas). This suggests that we have homogeneous sample units in each small area and their corresponding covariates are very close to each other which explains the assumption of equal fixed area-specific covariate. Also, it is reasonable to consider any departure from the true area-specific covariate mainly as the result of measurement error rather than sampling error. Note that even if the assumption of the equality of the true covariate units does not hold, due to the clean grouping, it is expected that the difference between the true covariates is negligible in an small area. In this case, we consider x_i to be the mean of the covariate obtained from all small area units. Furthermore, the difference between observed covariates in the small area is mainly because of the measurement error.

GS obtained the Bayes predictor of the small area mean, $\gamma_i = \frac{1}{N_i}\sum_{j=1}^{N_i} y_{ij}$, as

$$\widehat{\gamma}_i^B = \widehat{\gamma}_i^B(x_i, \phi) = (1 - f_i B_i)\bar{y}_i + f_i B_i(\beta_0 + \beta_1 x_i), \tag{1.2.3}$$

where $\phi = (\beta_0, \beta_1, \sigma_u^2, \sigma_e^2)$, $f_i = 1 - \frac{n_i}{N_i}$ is the finite population correction factor, and $B_i = \frac{\sigma_e^2}{\sigma_e^2 + n_i\sigma_u^2}$, for $i = 1, \ldots, m$. However, (1.2.3) depends on x_i's, the true values of the small area-specific covariate that are unknown.

To estimate the unknown x_i's, GS proposed a pseudo-Bayes (PB) predictor of γ_i by replacing x_i with its moment estimator, $\bar{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$. Datta et al. (2010), henceforth abbreviated DRT, suggested a new PB predictor of γ_i by using the maximum likelihood (ML) estimates of x_i's, i.e.

$$\tilde{x}_i(\psi) = \bar{X}_i + \frac{\beta_1\sigma_\eta^2(\bar{y}_i - \beta_0 - \beta_1\bar{X}_i)}{\sigma_e^2 + n_i\sigma_u^2 + b_1^2\sigma_\eta^2}, \tag{1.2.4}$$

where $\psi = (\phi, \sigma_\eta^2)$. The PB predictor proposed by DRT accounts for the variance of the measurement error model (i.e. σ_η^2) unlike the PB predictor of small area means proposed by GS.

GS and DRT also obtained a pseudo-empirical Bayes (PEB) predictor of γ_i by using the method of moments estimators of ψ proposed by GS as

$$\widehat{\sigma}_\eta^2 = \mathrm{MSW}_x, \qquad \widehat{\sigma}_e^2 = \mathrm{MSW}_y,$$

$$\widehat{b}_1 = \frac{\mathrm{MSB}_x}{\mathrm{MSB}_x - \mathrm{MSW}_x}\tilde{b}_1, \qquad \widehat{b}_0 = \bar{y} - \widehat{b}_1\bar{X},$$

and

$$\widehat{\sigma}_u^2 = \max\left\{0, \{\mathrm{MSB}_y - \mathrm{MSW}_y - \widehat{b}_1^2(\mathrm{MSB}_x - \mathrm{MSW}_x)\}\frac{m-1}{g_m}\right\},$$

where

$$\text{MSB}_y = \frac{1}{m-1}\sum_{i=1}^{m} n_i(\bar{y}_i - \bar{y})^2, \qquad \text{MSW}_y = \frac{1}{n-m}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2,$$
$$\text{MSB}_x = \frac{1}{m-1}\sum_{i=1}^{m} n_i(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})^2, \qquad \text{MSW}_x = \frac{1}{n-m}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(X_{ij} - \bar{\mathbf{X}}_i)^2,$$
$$g_m = n - \sum_{i=1}^{m} n_i^2/n, \qquad \bar{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{m} n_i\bar{\mathbf{X}}_i,$$
$$\bar{y} = \frac{1}{n}\sum_{i=1}^{m} n_i\bar{y}_i, \qquad \text{and } \tilde{b}_1 = \frac{\sum_{i=1}^{m} n_i\bar{y}_i(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})}{(m-1)\text{MSB}_x}.$$

Accordingly, GS and DRT obtained the MSPE estimation of their proposed PEB predictors using the jackknife method.

1.3. Clustering

It is a common practice in mixed-effects models to assume homogenous random effects. However, there are many applications where such an assumption does not hold. This is simply because in different areas the lurking variables affecting the response variable could be different, hence the random effects do not need to be homogenous. Thus, in order to provide more realistic prediction of small area means, one needs to work with non-homogenous random effects. Recently, there has been some research in this direction. For example, Maiti et al. (2011) considered the idea of clustering based on consequences of differences between the covariate where they studied the effect of the poverty on the educational performance of students in different school districts. Small areas (school districts) were clustered based on their poverty status and different regression models were assumed for different clusters. Random effects of school districts

belonging to the same cluster considered to follow a Normal distribution with a different variance component due to the effect of socio-economic status of families and areas on the students' education. Another example where clustering can be used in the small area estimation study is evaluating the effect of the body mass index (BMI) on the waste circumference for different groups based on age, sex, poverty, education and ethnic. We expect the small areas that are similar in terms of the Euclidean distance between their covariates show similar pattern in terms of the random effects. Obviously, the trend of accumulating body fat for underweight and obese people is different. So, one can easily consider different variance components for different clusters. The ultimate goal is to predict the small area mean with a higher precision by taking into account this inherent difference between areas. There are different approaches to cluster small areas such as hierarchical, $K-$means, and Silhouette clustering (Kaufman and Rousseeuw, 2009).

In the context of small area estimation, all small areas are considered as a single cluster. This produces the greatest amount of impairment to the available information regarding to the behaviour of the random effects. On the other hand, in the hierarchical clustering, initially there exist m clusters which means that we treat each small area as a separate cluster. The aim is to reduce the number of clusters to k clusters such that the least impairment of the available information in m clusters is produced. To this end, we use the Euclidean distance between the covariate of (1.1.1), $\|\mathbf{X}_i - \mathbf{X}_j\|_2 = \sqrt{\sum_{t=1}^{p}(X_{it} - X_{jt})^2}$, as a measure of similarity of small areas. Stepwise, the number of clusters are

reduced. The most two similar small areas in terms of Euclidean distance are put in a single cluster which reduces the number of clusters to $m-1$. This is achieved by $\frac{m(m-1)}{2}$ comparisons between the small areas. In the next step, we evaluate all $\frac{(m-1)(m-2)}{2}$ combination of elements to find the optimal union of the small areas in order to have $m-2$ clusters. The procedure continues until we get k clusters (Ward Jr, 1963).

The problem occurs when we want to obtain the dissimilarity between two clusters where either of them might have more than one element. The concept of dissimilarity is extended between two clusters using the notion of linkage. There are different types of linkage: complete, average, single, and centroid. We use the complete linkage where the maximum Euclidean distance between elements inside different clusters is used as the measure of dissimilarity. Using statistical softwares, the dendrogram of the observations is available at once to view the clusterings obtained for each possible number of clusters (James et al., 2013).

K-means clustering is another method we can use for the purpose of clustering. We choose the number of clusters in the beginning. The aim is to minimize the mean of the squared Euclidean distance within the clusters, $\sum_{l=1}^{k} \frac{1}{|C_l|} \sum_{i,j \in C_l; i \neq j} ||\mathbf{X}_i - \mathbf{X}_j||_2^2 = \sum_{l=1}^{k} \frac{1}{|C_l|} \sum_{i,j \in C_l; i \neq j} \sum_{t=1}^{p} (X_{it} - X_{jt})^2$ where C_l and $|C_l|$ are the l'th cluster and their number of elements, respectively, for $l = 1, \ldots, k$. In order to do K-means clustering, we first randomly assign each observation

to a cluster. The centroid of each cluster is found as follows

$$\bar{X}_l = (\bar{X}_{l1}, \ldots, \bar{X}_{lp}),$$

where $\bar{X}_{lt} = \frac{1}{|C_l|} \sum_{i \in C_l} X_{lti}$, for $t = 1, \ldots, p$. In the next step, we rearrange each observation such that the squared Euclidean distance between the observation and the centroid of the cluster is minimum. In case of any change, the new centroid of each cluster is found accordingly and the procedure is repeated until there is no change in clusters.

In this thesis, we use the Silhouette clustering to obtain the required number of clusters. Rousseeuw (1987) introduced this concept as a measure of how tightly the data is grouped. In order to obtain the appropriate number of clusters for the data, Silhouette plots and averages can be used. The idea in the Silhouette method is to find the average distance of each element from other elements in the same cluster, $a(i)$, for $i \in C_l$, $l = 1, \ldots, k$, and compare it to the minimum of the average of distance from other clusters, $b(i)$, for $i \in C_l$, $l = 1, \ldots, k$. The cluster with the minimum $b(i)$ is called neighbouring cluster. A silhouette for a data point is defined as follows

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}; \quad \text{for } i \in C_l,\ l = 1, \ldots, k. \tag{1.3.1}$$

where we have $-1 \leq s(i) \leq 1$, for $i \in C_l$, $l = 1, \ldots, k$. For an inappropriate choice of k, the average of $s(i)$'s will get closer to -1.

Example 1. *We are interested to cluster the following data*

1.1 0.89 0.09 2.04 1.7 4.5 1.85 0.43 0.56 0.01 2.93

First, the required number of clusters is determined using the Silhouette method and K-means clustering. As Figure 1.1 shows, $k = 2$ is an optimal number of clusters.

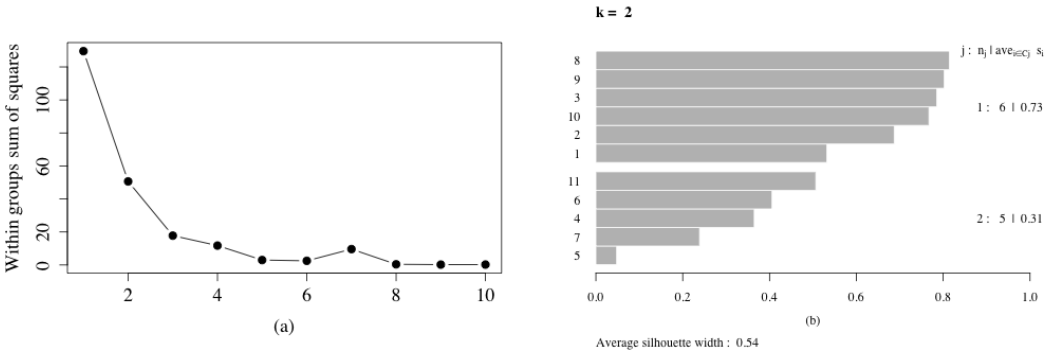

Figure 1.1: (a) Number of clusters using K-means clustering, and (b) Silhouette average for k=2 clusters

`kmeans` command in `R` is used for the purpose of clustering. K-means clustering gives $C_1 = \{1.1, 0.89, 0.09, 0.43, 0.56, 0.01\}$ *and* $C_2 = \{2.04, 1.7, 4.5, 1.85, 2.93\}$. *Figure 1.2 shows the hierarchical clustering for this data using `hclust` in `R`. For* $k = 2$*, the hierarchical clustering results in the same clustering as K-means.*

There are also available packages in `R` for the purpose of clustering.

1.4. Thesis Overview

The outline of the thesis is as follows. In chapter 2, we consider the unit-level nested error linear regression model which is commonly used in small area

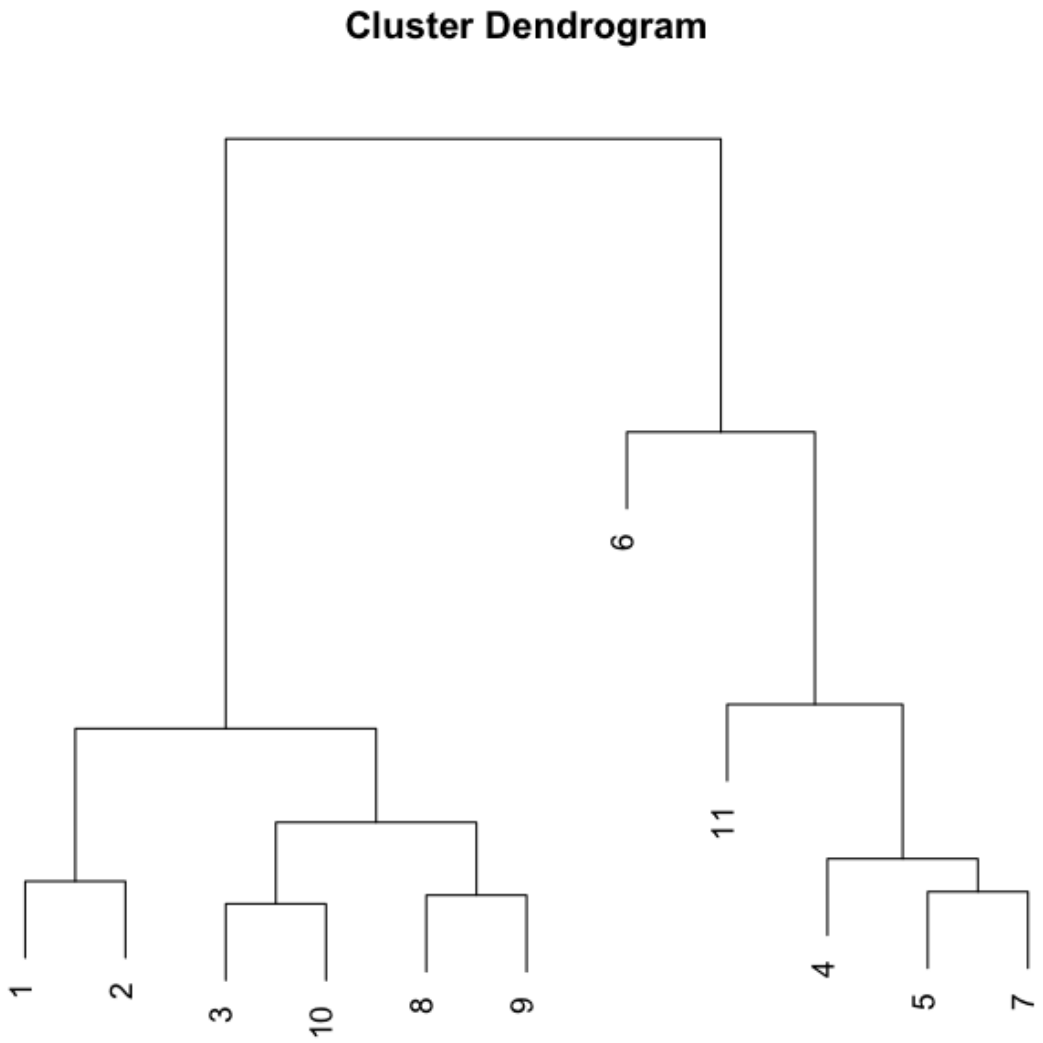

Figure 1.2: The dendrogram of data using the hierarchical clustering method

estimation. We study the case where the covariate in the model is assumed to have functional measurement error. We find the James-Stein (JS) estimator of the true covariate subject to the functional measurement error. We obtain

a new PEB predictor of small area means based on the JS estimator. Then, we show that the new PEB predictor is asymptotically optimal. The weighted and unweighted jackknife estimators of the MSPE of the new PEB predictor are also derived. Simulation studies are conducted to evaluate the performance of the proposed approach. We observe that the PEB predictor based on the JS estimator performs better than those based on the ML method and the method of moments. Finally, we apply the proposed methodology to a cross sectional study in New Zealand to predict the diastolic blood pressure using the cholesterol level.

In chapter 3, we propose to use constrained Bayes (CB) method to estimate the true covariate in order to build the small area PB predictor to study the unit-level nested error linear regression model with the functional measurement error in the covariate. We also provide some measures of performance such as sensitivity, specificity, and positive/negative predictive values for the CB predictor. We estimate the model parameters using the method of moments and Bayesian approach to get corresponding empirical and hierarchical Bayes predictors. We used the same dataset as Chapter 2 to assess our methodology. Finally, the performance of our proposed approach is evaluated through a simulation study.

Clustering small areas based on the Euclidean distance between their corresponding covariates is proposed in chapter 4 to get smaller MSPE for the predicted values of small area means using the area-level linear mixed models. We use the hierarchical clustering technique to cluster small areas

based on the Euclidean distance between their corresponding covariates and construct different groups of small areas where inside each group, areas are homogeneous and areas from different groups are non-homogenous. We start with an assumption that the random effect in each cluster follows a Normal distribution with a different variance component to give more flexibility to the behaviour of the random effects. We propose a statistical test for an area-level data to investigate the homogeneity of variance components among clusters. Then, we obtain the EBLUP of the small area means by taking into account the difference between variance components in different clusters. The performance of our proposed statistical test as well as the effect of the clustering on the MSPE of the small area means is studied using simulation studies. We also obtain a second-order approximation to the MSPE of small area means and derive an estimator of MSPE that is second-order unbiased. The results show that the MSPE of small area means can be improved when the variance components are different. The improvement in the MSPE is significant when the difference between variance components is considerable. Finally, the proposed methodology is applied to the National Health and Nutrition Examination Survey (NHANES) for 2011-2012. The summary of the thesis and future research areas are presented in chapter 5.

Chapter 2

Pseudo Empirical Bayes Predictors of Small Area Means Using the James-Stein Estimation

Without loss of generality, assume the sample units from the i'th small area are the first n_i units. The following gives the sample part of the population models (1.2.1) and (1.2.2)

$$y_{ij} = \beta_0 + \beta_1 x_i + u_i + e_{ij}, \quad i = 1, \ldots, m, j = 1, \ldots, n_i, \tag{2.0.1}$$

with

$$X_{ij} = x_i + \eta_{ij}, \quad i = 1, \ldots, m, j = 1, \ldots, n_i, \tag{2.0.2}$$

where n_i is the known sample size of the ith area (i=1,...,m), y_{ij} is the value of the study variable associated with the jth unit in the ith area and x_i is the unknown true area-specific covariate associated with y_{ij}. Further, the random errors e_{ij}, measurement errors η_{ij} and the area-level random effects u_i are assumed to be mutually independent with $e_{ij} \overset{i.i.d}{\sim} N(0, \sigma_e^2)$, $\eta_{ij} \overset{i.i.d}{\sim} N(0, \sigma_\eta^2)$ and $u_i \overset{i.i.d}{\sim} N(0, \sigma_u^2)$. The functional measurement error is considered for the

observed covariate in (2.0.2) where x_i's are fixed and unknown true area-specific covariate (Fuller, 2009).

In this chapter, we propose to estimate the unknown x_i's of models (2.0.1) and (2.0.2) by the James-Stein (JS) estimator as a natural competitor to the ML and the method of moments estimators. The JS estimator can be used to estimate the mean vector of a multivariate normal distribution which is more efficient than the ML estimate in terms of the sum of the weighted mean squared error,

$$\mathrm{E}\left[\sum_{i=1}^{m} W_i(\widehat{\gamma}_i - \gamma_i)^2\right],$$

where $m \geq 3$, $\widehat{\gamma}_i$ is the estimated value of γ_i, and W_i's are variance stabilizing constants (Efron and Morris, 1972). The JS estimator can be obtained as an empirical Bayes estimator, and it is robust against misspecifying the prior distribution (Lehmann and Casella, 1998).

Whittemore (1989) previously proposed the JS estimator for the linear regression model, $y_i = b_0 + b_1 x_i + e_i$, where $e_i \overset{i.i.d.}{\sim} N(0, \sigma_e^2)$, $i = 1, \ldots, m$, with the measurement errors in the covariate. Following Efron and Morris (1972), one can temporarily use the prior distribution $N(\mu, \tau^2)$, on the unknown x_i's in the measurement error model $X_i = x_i + \eta_i$ with $\eta_i \overset{i.i.d.}{\sim} N(0, \sigma_\eta^2)$ and known σ_η^2, to obtain the JS estimator, as an empirical Bayes estimator, as follows

$$\widehat{x}_i = \widehat{B}\bar{X} + (1 - \widehat{B})X_i \qquad i = 1, \ldots, m,$$

where $\widehat{B} = (m-3)\sigma_\eta^2 / \sum_{i=1}^{m}(X_i - \bar{X})^2$ and $\bar{X} = m^{-1}\sum_{i=1}^{m} X_i$.

We use a similar idea to construct the JS estimator of the true covariate, x_i, subject to the functional measurement error (see Carroll et al., 2010, Sec. 9.1.3 and Carroll et al., 1999 for more details). To this end, in Section 2.1, we first obtain the JS estimator of the true area-specific covariate, x_i, in the unit level regression model (1.2.1). In Section 2.2, a new pseudo empirical Bayes (PEB) predictor of the small area mean, γ_i, is constructed and its optimality is investigated. In Section 2.3, the jackknife estimators of the mean squared prediction error (MSPE) of the proposed PEB predictor of the small area mean are derived. The performance of the proposed approach is evaluated through simulation studies in Section 2.4. Finally, we apply our results to a cross-sectional study from the New Zealand population in Section 2.5.

2.1. A New Pseudo-Bayes Predictor

Following models (2.0.1) and (2.0.2), let

$$\bar{y}_i = b_0 + b_1 x_i + u_i + \bar{e}_i, \tag{2.1.1}$$

and

$$\bar{X}_i = x_i + \bar{\eta}_i, \tag{2.1.2}$$

where $\bar{e}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} e_{ij}$ and $\bar{\eta}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \eta_{ij}$. Define $\bar{Z}_i^*$ as a linear estimator of x_i based on $\bar{X}_i$ and $\bar{y}_i$ as follows

$$\begin{aligned}\bar{Z}_i^* =& \bar{X}_i + h_i(\bar{y}_i - b_0 - b_1\bar{X}_i)\\ =& x_i + h_i u_i + h_i\bar{e}_i + \bar{\eta}_i(1 - h_i b_1), \quad i = 1, \ldots, m.\end{aligned} \tag{2.1.3}$$

Since $\mathrm{E}(u_i) = \mathrm{E}(\bar{e}_i) = \mathrm{E}(\bar{\eta}_i) = 0$, it is easy to show that $\bar{Z}_i^*$ is an unbiased estimator of x_i. Also, one can easily see that

$$Var_{h_i}(\bar{Z}_i^*|x_i) = Var(x_i + h_i u_i + h_i \bar{e}_i + \bar{\eta}_i(1 - h_i b_1)) \tag{2.1.4}$$

$$= h_i^2(\sigma_u^2 + \frac{\sigma_e^2}{n_i}) + \frac{\sigma_\eta^2}{n_i}(1 - h_i b_1)^2, \tag{2.1.5}$$

so,

$$\bar{Z}_i^*|x_i \sim N\left(x_i, h_i^2(\sigma_u^2 + \frac{\sigma_e^2}{n_i}) + \frac{\sigma_\eta^2}{n_i}(1 - h_i b_1)^2\right).$$

We first obtain the best linear estimator of $\bar{Z}_i^*$ by minimizing $\mathrm{var}_{h_i}(\bar{Z}_i^*|x_i)$ with respect to h_i. To this end, by taking the derivative of $\mathrm{var}_{h_i}(\bar{Z}_i^*|x_i)$ with respect to h_i, we have

$$h_i\left(\sigma_u^2 + \frac{\sigma_e^2}{n_i} + b_1^2\frac{\sigma_\eta^2}{n_i}\right) = \frac{b_1\sigma_\eta^2}{n_i},$$

which results in the optimum value of h_i as follows

$$h_i^* = \frac{b_1\sigma_\eta^2}{n_i\sigma_u^2 + \sigma_e^2 + b_1^2\sigma_\eta^2}, \quad i = 1, \ldots, m.$$

One can easily see that $\bar{Z}_i^*$ using h_i^* leads to the ML estimator of x_i defined in (1.2.4).

To get the JS estimator, the method described in Efron and Morris (1972, 1975) is used by temporarily assuming that $x_i \overset{i.i.d.}{\sim} N(\mu, \tau^2)$. Let $\sigma_{0i}^2 = \mathrm{var}_{h_i^*}(\bar{Z}_i^*|x_i)$ and observe that

$$x_i|\bar{Z}_i^* \sim N\left(\frac{\sigma_{0i}^2}{\sigma_{0i}^2 + \tau^2}\mu + \frac{\tau^2}{\sigma_{0i}^2 + \tau^2}\bar{Z}_i{}^*, \frac{\sigma_{0i}^2\tau^2}{\sigma_{0i}^2 + \tau^2}\right), \quad i = 1, \ldots, m.$$

The Bayes estimator of x_i under the quadratic loss function is

$$\mathrm{E}(x_i|\bar{Z}_i^*) = C_i\mu + (1-C_i)\bar{Z}_i^*, \quad i = 1,\ldots,m, \tag{2.1.6}$$

where $C_i = \sigma_{0i}^2/(\sigma_{0i}^2+\tau^2)$. Now, the JS estimator of x_i is given by the empirical Bayes estimate of $\mathrm{E}(x_i|\bar{Z}_i^*)$. To this end, we first need to estimate the unknown parameters τ^2 and μ. There are different methods available in the literature to estimate τ^2. One can estimate τ^2 using the marginal distribution of $\bar{Z}_i{}^* \sim N(\mu, \sigma_{0i}^2+\tau^2)$, $i = 1,\ldots,m$. Efron and Morris (1975) gave an estimate of τ^2 as the solution of

$$\tau^2 = \frac{\sum_{i=1}^m((\bar{Z}_i^*-\mu)^2-\sigma_{0i}^2)\mathrm{I}_i(\tau^2)}{\sum_{i=1}^m \mathrm{I}_i(\tau^2)}, \tag{2.1.7}$$

where $\mathrm{I}_i(\tau^2) = \left\{\mathrm{var}(\bar{Z}_i^*-\mu)^2\right\}^{-1} = \frac{1}{2}(\sigma_{0i}^2+\tau^2)^{-2}$ is the Fisher information of τ^2 contained in $(\bar{Z}_i^*-\mu)^2$. The other method is the MLE of τ^2 using the marginal distribution of $\bar{Z}_i{}^*$. In this research thesis, we follow Efron and Morris (1975) to estimate τ^2.

In (2.1.7), $(\bar{Z}_i^*-\mu)^2$ is used to remove the effect of the non-centrality parameter. We also estimate μ using its ML estimator, given by

$$\bar{Z}^* = \frac{\sum_{i=1}^m \bar{Z}_i^*/(\sigma_{0i}^2+\tau^2)}{\sum_{i=1}^m 1/(\sigma_{0i}^2+\tau^2)}.$$

Hence, the JS estimator of x_i is obtained as follows

$$x_{i\mathrm{JS}} = C_i\bar{Z}^* + (1-C_i)\bar{Z}_i^*, \quad i = 1,\ldots,m. \tag{2.1.8}$$

The new PB predictor of γ_i, following (1.2.3), is then given by

$$\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}} = (1-f_iB_i)\bar{y}_i + f_iB_i(b_0+b_1x_{i\mathrm{JS}}), \quad i = 1,\ldots,m, \tag{2.1.9}$$

where $f_i = 1-\frac{n_i}{N_i}$ and $B_i = \frac{\sigma_e^2}{\sigma_e^2+n_i\sigma_u^2}$, for $i = 1,\ldots,m$.

Remark 1. *In the structural measurement error, we have the assumption that* $x_i \overset{i.i.d}{\sim} N(\mu_x, \sigma_x^2)$*, for* $i = 1, \ldots, m$ *(Carroll et al., 2010). One can easily observe that the JS predictors of* x_i*'s under the structural measurement error (assuming* x_i*'s are random variables) are equal to the JS estimators of* x_i*'s under the functional measurement error (assuming* x_i*'s are unknown fixed values). Note that,* x_i*'s in the structural measurement error are predicted while they are being estimated in the context of the functional measurement error.*

We now derive the MSPE of $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}}$ in (2.1.9) as follows:

$$\begin{aligned}
\mathrm{E}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}} - \gamma_i)^2 &= \mathrm{E}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}} - \widehat{\gamma}_i^B)^2 + \mathrm{E}(\widehat{\gamma}_i^B - \gamma_i)^2 \\
&= (f_i B_i b_1)^2 \left\{ \left(C_i x_i (d_i - 1) + C_i \sum_{j \neq i} x_j d_j \right)^2 + \sigma_{0i}^2 \left(1 + C_i (d_i - 1)\right)^2 \right. \\
&\left. + C_i^2 \sum_{j \neq i} \sigma_{0j}^2 d_j^2 \right\} + f_i^2 \left(\sigma_e^2 \left(\frac{(1 - B_i)^2}{n_i} + \frac{1}{N_i - n_i} \right) + B_i^2 \sigma_u^2 \right) \qquad (2.1.10) \\
&\equiv g_{1i}(\delta), \quad i = 1, \ldots, m,
\end{aligned}$$

where

$$d_i = \frac{1/(\sigma_{0i}^2 + \tau^2)}{\sum_{i=1}^m 1/(\sigma_{0i}^2 + \tau^2)},$$

and $\delta = \left(b_1, \sigma_u^2, \sigma_e^2, \sigma_\eta^2, (x_1, \ldots, x_m)\right)^\top$.

In fact, $\mathrm{MSPE}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}})$ in (2.1.10) is the measure of variability of $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}}$ where the model parameters are known. However, δ is unknown and we need to

estimate δ to obtain a pseudo-empirical Bayes predictor of γ_i and to evaluate its optimality.

Remark 2. *One can observe that the* MSPE *obtained in (2.1.10) depends on* $(x_1, \ldots, x_m)^\top$. *As* $m \to \infty$, *the dimension of the parameter space goes to infinity. However, using (2.1.8), we note that* $(x_1, \ldots, x_m)^\top$ *can be obtained by estimating* $\psi = (\phi, \sigma_\eta^2)$ *and* τ^2 *no matter how large* m *is. Also, following the Bayesian view towards the functional measurement error, and by applying the information provided in the prior distribution* $N(\mu, \tau^2)$ *in (2.1.10), we have*

$$\begin{aligned}\bar{\mathrm{E}}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}} - \gamma_i)^2 &= \mathrm{E}\{\mathrm{E}[(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}} - \gamma_i)^2 | x_1, \ldots, x_m]\} \\ &= (f_i B_i b_1)^2 \left(C_i^2 \tau^2 \left(1 + \sum_{j=1}^{m} d_j^2 - 2d_i \right) + \sigma_{0i}^2 \left(1 + C_i(d_i - 1)\right)^2 \right. \\ &\quad \left. + C_i^2 \sum_{j \neq i} \sigma_{0j}^2 d_j^2 \right) + f_i^2 \left(\sigma_e^2 \left(\frac{(1 - B_i)^2}{n_i} + \frac{1}{N_i - n_i} \right) + B_i^2 \sigma_u^2 \right),\end{aligned}$$

where $\bar{\mathrm{E}}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}} - \gamma_i)^2$ *is obtained using the total law of expectation with* $\bar{\mathrm{E}}$ *being used for double expectations. Thus, the unconditional* MSPE, *denoted by* $\bar{\mathrm{E}}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}} - \gamma_i)^2$ *does not depend on the specific values of* x_i*'s.*

2.2. Pseudo-EB Predictor

In this section, we first obtain a pseudo EB (PEB) predictor, $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$, of γ_i by using the method of moments to estimate the model parameters. We then show that this PEB predictor is asymptotically optimal in the sense that

$\frac{1}{m}\sum_{i=1}^{m} \mathrm{E}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}} - \widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}})^2 \to 0$ as $m \to \infty$. To this end, similar to Ghosh and Sinha (2007), henceforth abbreviated GS, and Datta et al. (2010), henceforth abbreviated DRT, the method of moments estimator of $\psi = (b_0, b_1, \sigma_u^2, \sigma_e^2, \sigma_\eta^2)$, $\widehat{\psi}$, is used.

The PEB predictor of γ_i is then given by

$$\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}} = (1 - f_i\widehat{B}_i)\bar{y}_i + f_i\widehat{B}_i(\widehat{b}_0 + \widehat{b}_1\widehat{x}_{i\mathrm{JS}}), \quad i = 1, \ldots, m,$$

where $\widehat{B}_i$ and $\widehat{x}_{i\mathrm{JS}}$ are estimated values of B_i and $x_{i\mathrm{JS}}$ obtained by replacing consistent estimates of parameters. For asymptotic optimality of the proposed PEB predictor, we need the following assumptions from Ghosh and Sinha (2007)

$$(i)\ \max_{1\le i\le m} n_i \le k_0 < \infty, \quad (ii)\ \frac{1}{m-1}\sum_{i=1}^{m} n_i(x_i - \bar{x})^2 \to c\,(>0)\ \text{as}\, m \to \infty. \quad (2.2.1)$$

Theorem 1 establishes the asymptotic optimality of $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$.

Theorem 1. *Under the assumptions (2.2.1), $\widehat{\gamma}_{iJS}^{PEB}$ is asymptotically optimal in the sense that*

$$m^{-1}\sum_{i=1}^{m} \mathrm{E}(\widehat{\gamma}_{iJS}^{PEB} - \widehat{\gamma}_{iJS}^{PB})^2 \to 0, \text{ as } m \to \infty, \quad (2.2.2)$$

where the expectation is with respect to random variables $\mathbf{X}, \mathbf{Y}$, and x_{iJS}, (for $i = 1, \ldots, m$).

Proof. Straightforward calculation shows that

$$\begin{aligned}
&\frac{1}{m}\sum_{i=1}^{m}\mathrm{E}\left(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}-\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}}\right)^2\\
&=\frac{1}{m}\sum_{i=1}^{m}\mathrm{E}\left(f_i(B_i-\widehat{B}_i)(\bar{y}_i-b_0-b_1x_{i\mathrm{JS}})-f_i\widehat{B}_i(b_0+b_1x_{i\mathrm{JS}})+f_i\widehat{B}_i(\widehat{b}_0+\widehat{b}_1\widehat{x}_{i\mathrm{JS}})\right)^2\\
&=\frac{1}{m}\sum_{i=1}^{m}\mathrm{E}\left(f_i(B_i-\widehat{B}_i)(\bar{y}_i-b_0-b_1x_i-b_1(x_{i\mathrm{JS}}-x_i)\right.\\
&\left.+f_i\widehat{B}_i((\widehat{b}_0-b_0)+(\widehat{b}_1-b_1)\widehat{x}_{i\mathrm{JS}}+b_1(\widehat{x}_{i\mathrm{JS}}-x_{i\mathrm{JS}}))\right)^2\\
&\le\frac{5}{m}\sum_{i=1}^{m}\mathrm{E}\left(f_i^2(B_i-\widehat{B}_i)^2(\bar{y}_i-b_0-b_1x_i)^2+f_i^2(B_i-\widehat{B}_i)^2b_1^2(x_{i\mathrm{JS}}-x_i)^2\right.\\
&\left.+f_i^2\widehat{B}_i^2(\widehat{b}_0-b_0)^2+f_i^2\widehat{B}_i^2(\widehat{b}_1-b_1)^2\widehat{x}_{i\mathrm{JS}}^2+f_i^2\widehat{B}_i^2b_1^2(\widehat{x}_{i\mathrm{JS}}-x_{i\mathrm{JS}})^2\right)\\
&\le\frac{5}{m}\sum_{i=1}^{m}\mathrm{E}\left((B_i-\widehat{B}_i)^2(\bar{y}_i-b_0-b_1x_i)^2+(B_i-\widehat{B}_i)^2b_1^2(x_{i\mathrm{JS}}-x_i)^2\right.\\
&\left.+(\widehat{b}_0-b_0)^2+(\widehat{b}_1-b_1)^2\widehat{x}_{i\mathrm{JS}}^2+b_1^2(\widehat{x}_{i\mathrm{JS}}-x_{i\mathrm{JS}})^2\right). \qquad (2.2.3)
\end{aligned}$$

The first inequality is due to the well-known partial sums moment inequality (e.g., DasGupta, 2008)

$$\mathrm{E}|\sum_{i=1}^{n}X_i|^p\le n^{p-1}\sum_{i=1}^{n}\mathrm{E}|X_i|^p,\quad p>1, \qquad (2.2.4)$$

while the second inequality follows from $f_i\le 1$ and $\widehat{B}_i\le 1$, $i=1,\ldots,m$. Under

the assumption (2.2.1) and the consistency of $\widehat{\sigma}_u^2$ and $\widehat{\sigma}_e^2$, we have

$$\begin{aligned}|B_i - \widehat{B}_i| &= \left| \frac{\sigma_e^2}{\sigma_e^2 + n_i\sigma_u^2} - \frac{\widehat{\sigma}_e^2}{\widehat{\sigma}_e^2 + n_i\widehat{\sigma}_u^2} \right| \\ &= \left| \frac{n_i(\widehat{\sigma}_u^2(\sigma_e - \widehat{\sigma}_e^2) + \widehat{\sigma}_e^2(\widehat{\sigma}_u^2 - \sigma_u^2))}{(\sigma_e^2 + n_i\sigma_u^2)(\widehat{\sigma}_e^2 + n_i\widehat{\sigma}_u^2)} \right| \\ &\leq \frac{k_0}{\sigma_e^2 + \sigma_u^2}(|\widehat{\sigma}_e^2 - \sigma_e^2| + |\widehat{\sigma}_u^2 - \sigma_u^2|) \to 0, \text{ as } m \to \infty. \end{aligned} \tag{2.2.5}$$

Thus

$$\max_{1\leq i\leq m} |B_i - \widehat{B}_i|^2 \to 0, \text{ as } m \to \infty. \tag{2.2.6}$$

In addition,

$$\begin{aligned}\max_{1\leq i\leq m} |B_i - \widehat{B}_i| &\leq \frac{k_0}{\sigma_e^2 + \sigma_u^2}(\widehat{\sigma}_e^2 + \sigma_e^2 + \widehat{\sigma}_u^2 + \sigma_u^2) \\ &\leq k_0 + \frac{k_0}{\sigma_e^2 + \sigma_u^2}(\mathrm{MSB}_y + \mathrm{MSW}_y). \end{aligned} \tag{2.2.7}$$

Furthermore,

$$\begin{aligned} m^{-1}\sum_{i=1}^{m} \mathrm{E}(\bar{y}_i - b_0 - b_1 x_i)^2 &= m^{-1}\sum_{i=1}^{m}(\sigma_u^2 + \frac{\sigma_e^2}{n_i}) \\ &\leq \sigma_u^2 + \sigma_e^2 = O(1). \end{aligned} \tag{2.2.8}$$

Also,

$$\frac{1}{m}\sum_{i=1}^{m} |B_i - \widehat{B}_i|^2(\bar{y}_i - b_0 - b_1 x_i)^2 \leq \max_{1\leq i\leq m} |B_i - \widehat{B}_i|^2 \frac{1}{m}\sum_{i=1}^{m}(\bar{y}_i - b_0 - b_1 x_i)^2. \tag{2.2.9}$$

Now, from (2.2.6) and (2.2.8), we have

$$\frac{1}{m}\sum_{i=1}^{m} |B_i - \widehat{B}_i|^2(\bar{y}_i - b_0 - b_1 x_i)^2 \xrightarrow{p} 0 \quad \text{as } m \to \infty. \tag{2.2.10}$$

To use the Dominated Convergence Theorem (DCT), first note that by using (2.2.9), (2.2.4) and the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
&\mathrm{E}\left[\frac{1}{m}\sum_{i=1}^{m}|B_i-\widehat{B}_i|^2(\bar{y}_i-b_0-b_1x_i)^2\right]\\
&\leq \mathrm{E}\left[\max_{1\leq i\leq m}|B_i-\widehat{B}_i|^2\frac{1}{m}\sum_{i=1}^{m}(\bar{y}_i-b_0-b_1x_i)^2\right]\\
&\leq \sqrt{\mathrm{E}\left(\max_{1\leq i\leq m}|B_i-\widehat{B}_i|^2\right)^2\mathrm{E}\left(\frac{1}{m}\sum_{i=1}^{m}(\bar{y}_i-b_0-b_1x_i)^2\right)^2}\\
&\leq \sqrt{\mathrm{E}\left(\max_{1\leq i\leq m}|B_i-\widehat{B}_i|\right)^4\mathrm{E}\left(\frac{1}{m}\sum_{i=1}^{m}(\bar{y}_i-b_0-b_1x_i)^4\right)}. \qquad (2.2.11)
\end{aligned}$$

Also,

$$\begin{aligned}
\mathrm{E}\left(\frac{1}{m}\sum_{i=1}^{m}(\bar{y}_i-b_0-b_1x_i)^4\right) &= 3\frac{1}{m}\sum_{i=1}^{m}(\sigma_u^2+\frac{\sigma_e^2}{n_i})^2\\
&\leq 3(\sigma_u^2+\sigma_e^2)^2 = O(1). \qquad (2.2.12)
\end{aligned}$$

As (2.2.7) shows, $\mathrm{E}\left(\max_{1\leq i\leq m}|B_i-\widehat{B}_i|\right)^4$ is integrable. Therefore

$$\mathrm{E}\left(\frac{1}{m}\sum_{i=1}^{m}|B_i-\widehat{B}_i|^2(\bar{y}_i-b_0-b_1x_i)^2\right)\to 0, \quad \text{as } m\to\infty.$$

In the next step, we show that

$$\frac{1}{m}\sum_{i=1}^{m}\mathrm{E}\left[(B_i-\widehat{B}_i)^2(x_{i\mathrm{JS}}-x_i)^2\right]\to 0, \quad \text{as } m\to\infty. \qquad (2.2.13)$$

By using the Cauchy-Schwarz inequality and (2.2.7), we first show that

$$\frac{1}{m}\sum_{i=1}^{m}\mathrm{E}(x_{i\mathrm{JS}}-x_i)^4 \qquad (2.2.14)$$

is finite. To this end, since $x_{i\mathrm{JS}}$, $1 \le i \le m$, follows a normal distribution, we have

$$\mathrm{E}(x_{i\mathrm{JS}}^4) < \infty, \quad 1 \le i \le m. \tag{2.2.15}$$

Also, (2.2.14) is finite from (2.2.4). Now using DCT, (2.2.13) follows easily.

To show $\mathrm{E}(\widehat{b}_0 - b_0)^2 \to 0$, as $m \to \infty$, we first write

$$\mathrm{E}(\widehat{b}_0 - b_0)^2 = \mathrm{E}\left[\mathrm{var}(\widehat{b}_0|\mathbf{X}) + (\mathrm{E}(\widehat{b}_0|\mathbf{X}) - b_0)^2\right]. \tag{2.2.16}$$

Note that

$$\begin{aligned}
\mathrm{var}(\widehat{b}_0|\mathbf{X}) &= \mathrm{var}(\bar{y} - \widehat{b}_1\bar{\mathbf{X}}|\mathbf{X}) \\
&= \mathrm{var}(\bar{y}) + \mathrm{var}(\widehat{b}_1|\mathbf{X})\bar{\mathbf{X}}^2 - 2\bar{\mathbf{X}}\mathrm{cov}(\bar{y}, \widehat{b}_1|\mathbf{X}) \\
&= \frac{\sum_{i=1}^m n_i(\sigma_e^2 + n_i\sigma_u^2)}{n^2} + \mathrm{var}(\widehat{b}_1|\mathbf{X})\bar{\mathbf{X}}^2 \\
&\quad - 2\bar{\mathbf{X}}\frac{\mathrm{MSB}_x}{\mathrm{MSB}_x - \mathrm{MSW}_x}\frac{\sum_{i=1}^m n_i^2(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})\mathrm{var}(\bar{y}_i)}{n(m-1)\mathrm{MSB}_x} \\
&\le \frac{\sigma_e^2 + k_0\sigma_u^2}{n} + \mathrm{var}(\widehat{b}_1|\mathbf{X})\bar{\mathbf{X}}^2 - 2\frac{\bar{\mathbf{X}}\sum_{i=1}^m n_i(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})(\sigma_e^2 + n_i\sigma_u^2)}{(\mathrm{MSB}_x - \mathrm{MSW}_x)(m-1)n},
\end{aligned} \tag{2.2.17}$$

where $\mathbf{X}$ is the vector of covariates with measurement error. Ghosh and Sinha (2007) showed that

$$\mathrm{var}(\widehat{b}_1|\mathbf{X}) \le \frac{\sigma_e^2 + k_0\sigma_u^2}{m-1}\frac{\mathrm{MSB}_x}{(\mathrm{MSB}_x - \mathrm{MSW}_x)^2} \tag{2.2.18}$$

By (2.2.17) and (2.2.18), we have $\mathrm{var}(\widehat{b}_0|\mathbf{X}) = O(m^{-1})$. Noting that

$$\mathrm{E}\left(\frac{\sigma_e^2 + k_0\sigma_u^2}{m-1}\frac{\mathrm{MSB}_x}{(\mathrm{MSB}_x - \mathrm{MSW}_x)^2} + \frac{\sigma_e^2 + k_0\sigma_u^2}{n} - 2\frac{\bar{\mathbf{X}}\sum_{i=1}^m n_i(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})(\sigma_e^2 + n_i\sigma_u^2)}{(\mathrm{MSB}_x - \mathrm{MSW}_x)(m-1)n}\right),$$

is finite, we have

$$\mathrm{E}(\mathrm{var}(\widehat{b_0}|\mathbf{X})) \to 0, \quad as\, m \to \infty. \tag{2.2.19}$$

Using orthogonal transformation introduced by Ghosh and Sinha (2007),

$$\begin{aligned}
\mathrm{E}(\mathrm{E}(\widehat{b_0}|\mathbf{X}) - b_0)^2 &= \mathrm{E}(\mathrm{E}(\bar{y}) - \mathrm{E}(\widehat{b_1}|\mathbf{X})\bar{X} - b_0)^2 \\
&= \mathrm{E}\left(b_0 + b_1\bar{x} - b_1\bar{\mathbf{X}}\frac{\sum_{i=1}^m n_i\bar{\mathbf{X}}_i(x_i-\bar{x})}{(m-1)(\mathrm{MSB}_x - \mathrm{MSW}_x)} - b_0\right)^2 \\
&= b_1^2\mathrm{E}\left(\bar{x} - \frac{\sum_{i=1}^m n_i\bar{\mathbf{X}}_i(x_i-\bar{x})}{(m-1)(\mathrm{MSB}_x - \mathrm{MSW}_x)}\bar{\mathbf{X}}\right)^2 \\
&= b_1^2\mathrm{E}\left(\bar{x} - \frac{Z_1Z_2(\sum_{i=1}^m n_i(x_i-\bar{x})^2)^{\frac{1}{2}}}{(m-1)(\mathrm{MSB}_x - \mathrm{MSW}_x)}\sqrt{n}\right)^2,
\end{aligned} \tag{2.2.20}$$

where $(Z_1, Z_2, \ldots, Z_m)^\top = \mathbf{C}(\sqrt{n_1}\bar{\mathbf{X}}_1, \ldots, \sqrt{n_m}\bar{\mathbf{X}}_m)^\top$, and $\mathbf{C}$ is an orthogonal matrix with first two rows given by

$$\left(\sqrt{\frac{n_1}{n}}, \ldots, \sqrt{\frac{n_m}{n}}\right),$$

and

$$\left(\frac{\sqrt{n_1}(x_1-\bar{x})}{\sqrt{\sum n_i(x_i-\bar{x})^2}}, \ldots, \frac{\sqrt{n_m}(x_m-\bar{x})}{\sqrt{\sum n_i(x_i-\bar{x})^2}}\right).$$

Moreover,

$$\bar{x} - \frac{Z_1Z_2(\sum_{i=1}^m n_i(x_i-\bar{x})^2)^{\frac{1}{2}}}{(m-1)(\mathrm{MSB}_x - \mathrm{MSW}_x)}\sqrt{n} \longrightarrow 0,$$

as $m \to \infty$. Under (2.2.1), $\frac{Z_2}{\sqrt{m-1}} \to \sqrt{c}$, $\mathrm{MSB}_x - \mathrm{MSW}_x \to c$ and $\bar{\mathbf{X}} \to \bar{x}$. Similar to Ghosh and Sinha (2007), under the uniform integrability argument,

$\mathrm{E}(\mathrm{E}(\widehat{b_0}|\mathbf{X}) - b_0)^2 \to 0$. Next, we need to show $\mathrm{E}((\widehat{b_1} - b_1)^2 \frac{1}{m}\sum_{i=1}^m \widehat{x}_{i\mathrm{JS}}^2) \to 0$ when $m \to \infty$. Using the Cauchy-Schwarz inequality, we have

$$\mathrm{E}((\widehat{b_1} - b_1)^2 \frac{1}{m}\sum_{i=1}^m \widehat{x}_{i\mathrm{JS}}^2) \le \sqrt{\mathrm{E}(\widehat{b_1} - b_1)^4 \mathrm{E}\left(\frac{1}{m}\sum_{i=1}^m \widehat{x}_{i\mathrm{JS}}^2\right)^2} \tag{2.2.21}$$

$$\le \sqrt{\mathrm{E}(\widehat{b_1} - b_1)^4 \mathrm{E}\left(\frac{1}{m}\sum_{i=1}^m \widehat{x}_{i\mathrm{JS}}^4\right)}, \quad \text{by } C_r \text{ inequality.} \tag{2.2.22}$$

Moreover,

$$(\widehat{b_1} - b_1)^4 = \left(\frac{\mathrm{MSB}_x}{\mathrm{MSB}_x - \mathrm{MSW}_x} \frac{\sum_{i=1}^m n_i \bar{y}_i (\bar{\mathbf{X}}_i - \bar{\mathbf{X}})}{\sum_{i=1}^m n_i (\bar{\mathbf{X}}_i - \bar{\mathbf{X}})^2} - b_1\right)^4$$

$$\le \left(\frac{1}{(m-1)(\mathrm{MSB}_x - \mathrm{MSW}_x)} \sum_{i=1}^m n_i \bar{y}_i (\bar{\mathbf{X}}_i - \bar{\mathbf{X}}) - b_1\right)^4$$

$$\le 8\left(\left(\frac{1}{(m-1)(\mathrm{MSB}_x - \mathrm{MSW}_x)} \sum_{i=1}^m n_i \bar{y}_i (\bar{\mathbf{X}}_i - \bar{\mathbf{X}})\right)^4 + b_1^4\right). \tag{2.2.23}$$

Again, by the Cauchy-Schwarz inequality for integration and (2.2.23) we get

$$\mathrm{E}\left(\frac{1}{(m-1)(\mathrm{MSB}_x - \mathrm{MSW}_x)} \sum_{i=1}^m n_i \bar{y}_i (\bar{\mathbf{X}}_i - \bar{\mathbf{X}})\right)^4$$

$$\le \sqrt{\mathrm{E}\left(\frac{1}{\mathrm{MSB}_x - \mathrm{MSW}_x}\right)^8 \mathrm{E}\left(\frac{\sum_{i=1}^m n_i \bar{y}_i (\bar{\mathbf{X}}_i - \bar{\mathbf{X}})}{(m-1)}\right)^8}. \tag{2.2.24}$$

By (2.2.4) and independency of $\bar{y}_i$ and $(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})$, we have

$$\frac{1}{(m-1)^8}\mathrm{E}\left(\sum_{i=1}^m n_i \bar{y}_i (\bar{\mathbf{X}}_i - \bar{\mathbf{X}})\right)^8 \le \frac{m^7}{(m-1)^8} \sum_{i=1}^m \mathrm{E}\left(n_i^8 \bar{y}_i^8 (\bar{\mathbf{X}}_i - \bar{\mathbf{X}})^8\right)$$

$$\le \frac{k_0^8 m^7}{(m-1)^8} \sum_{i=1}^m \mathrm{E}\left(\bar{y}_i^8\right) \mathrm{E}\left(\bar{\mathbf{X}}_i - \bar{\mathbf{X}}\right)^8 < \infty.$$

The last inequality is due to the fact that $\bar{y}_i$ and $\bar{X}_i - \bar{X}$ follow the normal distribution. Because $\dfrac{1}{(\mathrm{MSB}_x - \mathrm{MSW}_x)^8} \xrightarrow{p} \dfrac{1}{c^8}$, as $m \to \infty$, similar to Ghosh and Sinha (2007),

$$\mathrm{E}\left(\frac{1}{(\mathrm{MSB}_x - \mathrm{MSW}_x)^8}\right) \to \frac{1}{c^8}$$

as $m \to \infty$, under the uniform integrability argument. Therefore, (2.2.24) is finite, and DCT implies $\mathrm{E}(\widehat{b}_1 - b_1)^4 \to 0$. If $\mathrm{E}(\frac{1}{m}\sum_{i=1}^m \widehat{x}_{i\mathrm{JS}})^4 < \infty$, then $\mathrm{E}((\widehat{b}_1 - b_1)^2 \frac{1}{m}\sum_{i=1}^m \widehat{x}_{i\mathrm{JS}}^2) \to 0$. Note that

$$\begin{aligned}\widehat{x}_{i\mathrm{JS}}^4 &= (\widehat{C}_i\widehat{\bar{Z}}^* + (1-\widehat{C}_i)\widehat{\bar{Z}}_i^*)^4 \\ &\le 8(\widehat{C}_i^4\widehat{\bar{Z}}^{*4} + (1-\widehat{C}_i)^4\widehat{\bar{Z}}_i^{*4}), \quad \text{by } C_r \text{ inequality.} \end{aligned} \tag{2.2.25}$$

As $0 \le C_i \le 1$, $i = 1, \ldots, m$, one can easily show that

$$\begin{aligned}\widehat{\bar{Z}}_i^{*4} &\le 8(\bar{X}_i^4 + (\bar{y}_i - (\bar{y} - \widehat{b}_1(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})))^4) \\ &\le 8\bar{X}_i^4 + 216(\bar{y}_i^4 + \bar{y}^4 + \widehat{b}_1^4(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})^4), \quad \text{by } C_r \text{ inequality.} \end{aligned} \tag{2.2.26}$$

Using the Cauchy-Schwarz inequality for the expectation and similar to (2.2.24), $\mathrm{E}(\widehat{b}_1^8) < \infty$. Thus, $\mathrm{E}(\widehat{\bar{Z}}_i^{*4}) < \infty$. In addition, $\mathrm{E}(\widehat{\bar{Z}}_i^{*p}) < \infty$, for $p = 1, 2, 3$. Also, we have

$$\begin{aligned}\widehat{\bar{Z}}^{*4} &= \frac{(\sum_{i=1}^m \widehat{\bar{Z}}_i^*/(\widehat{\sigma}_{0i}^2 + \widehat{\tau}^2))^4}{(\sum_{i=1}^m 1/(\widehat{\sigma}_{0i}^2 + \widehat{\tau}^2))^4} \\ &= \frac{\sum_{i=1}^m\sum_{j=1}^m\sum_{k=1}^m\sum_{l=1}^m \widehat{\omega}_{ijkl}\widehat{\bar{Z}}_i^*\widehat{\bar{Z}}_j^*\widehat{\bar{Z}}_k^*\widehat{\bar{Z}}_l^*}{\sum_{i=1}^m\sum_{j=1}^m\sum_{k=1}^m\sum_{l=1}^m \widehat{\omega}_{ijkl}}, \end{aligned} \tag{2.2.27}$$

where $\widehat{\omega}_{ijkl} = (\widehat{\sigma}_{0i}^2 + \widehat{\tau}^2)^{-1}(\widehat{\sigma}_{0j}^2 + \widehat{\tau}^2)^{-1}(\widehat{\sigma}_{0k}^2 + \widehat{\tau}^2)^{-1}(\widehat{\sigma}_{0l}^2 + \widehat{\tau}^2)^{-1}$. Now, using (2.2.27),

we can easily observe that

$$\begin{aligned}
\mathrm{E}(\widehat{\bar{Z}}^{*4}) &= \mathrm{E}\left(\frac{\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m}\sum_{l=1}^{m}\widehat{\omega}_{ijkl}\widehat{\bar{Z}}_i^{*}\widehat{\bar{Z}}_j^{*}\widehat{\bar{Z}}_k^{*}\widehat{\bar{Z}}_l^{*}}{\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m}\sum_{l=1}^{m}\widehat{\omega}_{ijkl}}\right) \\
&= \mathrm{E}(\max_{1\le i\le m}\widehat{\bar{Z}}_i^{*})^4 \\
&\le \mathrm{E}(\max_{1\le i\le m}\widehat{\bar{Z}}_i^{*4}).
\end{aligned}$$

From (2.2.26), we can simply find a bound for $\widehat{\bar{Z}}_i^{*4}$, $i = 1, \ldots, m$. Therefore, noting that mean of finite numbers is finite and using (2.2.4) and (2.2.25) gives

$$\mathrm{E}\left(\frac{1}{m}\sum_{i=1}^{m}\widehat{x}_{i\mathrm{JS}}\right)^4 < \infty. \tag{2.2.28}$$

To prove the last part of (2.2.3), first, we claim $\widehat{x}_{i\mathrm{JS}} \xrightarrow{p} x_{i\mathrm{JS}}$ as $m \to \infty$. Note that the method of moments estimate of variance components gives consistent estimates of the parameter. Thus, it remains to prove $\widehat{\tau}^2$ is a consistent estimator of τ^2.

As Small and Yang (1999) pointed out, based on Crowder (1986), the estimating equations always have one consistent root so, we need to show that (2.1.7) has a unique solution. Determining the exact number of roots is not possible due to the complicated mathematical form of (2.1.7). To this end, we rely on numerical evaluations of (2.1.7) by plotting it for each run of the simulation. Numerical studies with different dataset show that there exists

only one positive root for this function. Finally, using (2.2.4), we have

$$\mathrm{E}\left(\frac{1}{m}\sum_{i=1}^{m}(\widehat{x}_{i\mathrm{JS}}-x_{i\mathrm{JS}})^2\right)\leq \mathrm{E}\left(\frac{2}{m}\sum_{i=1}^{m}\widehat{x}_{i\mathrm{JS}}^4\right)+\mathrm{E}\left(\frac{2}{m}\sum_{i=1}^{m}x_{i\mathrm{JS}}^4\right).$$

As $x_{i\mathrm{JS}}$'s are normally distributed, $\mathrm{E}(\frac{2}{m}\sum_{i=1}^{m}x_{i\mathrm{JS}}^4)<\infty$. By (2.2.28),

$$\mathrm{E}(\frac{2}{m}\sum_{i=1}^{m}\widehat{x}_{i\mathrm{JS}}^4)<\infty.$$

Now, using DCT the result is obtained and this completes the proof.

Following Ghosh and Sinha (2007) and Datta et al. (2010), we refer to the optimality in the sense of Robbins (1956). To get a measure of the variability of $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$, we calculate $\mathrm{MSPE}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}})$ which is represented as

$$\mathrm{MSPE}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}})=\mathrm{E}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}}-\gamma_i)^2+\mathrm{E}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}-\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}})^2+2\mathrm{E}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}}-\gamma_i)(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}-\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}}),\tag{2.2.29}$$

where $\mathrm{E}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}}-\gamma_i)^2=g_{1i}(\delta)$. Note that $\mathrm{E}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}-\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}})^2+2\mathrm{E}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}}-\gamma_i)(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}-\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}})$ accounts for the variability of the estimation of the model parameters in (2.2.29). In Section 2.3, we show $\mathrm{E}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}}-\gamma_i)(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}-\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}})\to 0$, as $m\to\infty$, for $i=1,\ldots,m$. According to Theorem 1, $\mathrm{E}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}-\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}})^2\to 0$, as $m\to\infty$, for $i=1,\ldots,m$. They imply

$$\mathrm{E}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}-\gamma_i)^2\to\mathrm{E}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}}-\gamma_i)^2,\ \text{as}\ m\to\infty.$$

In Section 2.3, using the jackknife method, we present an estimate of $\mathrm{MSPE}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}})$.

2.3. The Jackknife Estimation of MSPE($\widehat{\gamma}_{i\mathbf{JS}}^{\mathbf{PEB}}$)

To find an estimator of

$$\mathrm{MSPE}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}) = \mathrm{E}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}} - \gamma_i)^2,$$

we follow Jiang et al. (2002), Chen and Lahiri (2002) and Datta et al. (2010) using the weighted and unweighted jackknife methods. Let

$$M_{1i} := \mathrm{E}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}} - \gamma_i)^2 = g_{1i}(\delta),$$

$$M_{2i} := \mathrm{E}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}} - \widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}})^2,$$

and

$$M_{3i} := \mathrm{E}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}} - \gamma_i)(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}} - \widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}}).$$

We have

$$\mathrm{MSPE}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}) = M_{1i} + M_{2i} + 2M_{3i}. \tag{2.3.1}$$

Due to the structure of M_{3i}, it is not easy to find the explicit form of M_{3i}. As it is shown in Theorem 1, $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$ is asymptotically optimal for $\gamma_{i\mathrm{JS}}^{\mathrm{PB}}$ in the sense of Robbins (1956). Using Theorem 1 and the Cauchy-Schwarz inequality, similar to Ghosh and Sinha (2007), we have

$$\begin{aligned}
\mathrm{E}\left(\tfrac{1}{m}\sum_{i=1}^{m}\left|(\gamma_{i\mathrm{JS}}^{\mathrm{PB}} - \gamma_i)(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}} - \gamma_{i\mathrm{JS}}^{\mathrm{PB}})\right|\right) &\le \tfrac{1}{m}\sum_{i=1}^{m}\left[\mathrm{E}^{1/2}(\gamma_{i\mathrm{JS}}^{\mathrm{PB}} - \gamma_i)^2\mathrm{E}^{1/2}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}} - \gamma_{i\mathrm{JS}}^{\mathrm{PB}})^2\right] \\
&\le \max_{1\le i\le m}\mathrm{E}^{1/2}(\gamma_{i\mathrm{JS}}^{\mathrm{PB}} - \gamma_i)^2\left[\tfrac{1}{m}\sum_{i=1}^{m}\mathrm{E}^{1/2}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}} - \gamma_{i\mathrm{JS}}^{\mathrm{PB}})^2\right] \\
&\to 0, \text{as } m \to \infty.
\end{aligned}$$

Knowing the mean of some positive terms goes to zero in limit implies that each term goes to zero in limit. So, we have

$$\mathrm{E}^{1/2}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}} - \gamma_{i\mathrm{JS}}^{\mathrm{PB}})^2 \to 0, \, as \, m \to \infty,$$

for $i = 1, \ldots, m$. Thus, as

$$M_{3i} \leq \mathrm{E}\left|(\gamma_{i\mathrm{JS}}^{\mathrm{PB}} - \gamma_i)(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}} - \gamma_{i\mathrm{JS}}^{\mathrm{PB}})\right| \leq max_{1\leq i\leq m}\mathrm{E}^{1/2}(\gamma_{i\mathrm{JS}}^{\mathrm{PB}} - \gamma_i)^2\mathrm{E}^{1/2}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}} - \gamma_{i\mathrm{JS}}^{\mathrm{PB}})^2,$$

we have $M_{3i} \to 0$, as $m \to \infty$, for $i = 1, \ldots, m$. Therefore, we ignore the cross product term M_{3i}, similar to Ybarra and Lohr (2008) and Datta et al. (2010), and use the approximation

$$\mathrm{MSPE}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}) \approx M_{1i} + M_{2i}, \tag{2.3.2}$$

in the jackknife method. However, we address the magnitude of the ignored term, M_{3i}, in the simulation study (Section 2.4). In particular, we found that (2.3.2) is quite accurate for the proposed predictor $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$.

To find the jackknife estimator of $\mathrm{MSPE}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}})$, denoted by $mspe_J$, let

$$\widehat{M}_{1iJ} = g_{1i}(\widehat{\delta}) - \sum_{l=1}^{m} w_l[g_{1i}(\widehat{\delta}_{-l}) - g_{1i}(\widehat{\delta})],$$

and

$$\widehat{M}_{2iJ} = \sum_{l=1}^{m} w_l(\widehat{\gamma}_{iJS,-l}^{\mathrm{PEB}} - \widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}})^2,$$

where $\widehat{\delta}_{-l}$ is obtained by removing observations of the lth area, $l = 1, \ldots, m$, and finding the model parameters estimate by using the information of other areas. Note that $\widehat{\gamma}_{iJS,-l}^{\mathrm{PEB}}$ is also PEB predictor of γ_i without using the information of lth area. Finally, weight w_l, $l = 1, \ldots, m$, can be either $1 - \bar{\mathbf{X}}_l^{\top}(\sum_{t=1}^{m} \bar{\mathbf{X}}_t\bar{\mathbf{X}}_t^{\top})^{-1}\bar{\mathbf{X}}_l$,

where $\bar{\mathbf{X}}_l = (1, \bar{X}_l)^\top$, or $(m-1)/m$ leading to the weighted (Chen and Lahiri, 2002) or unweighted (Jiang et al., 2002) jackknife estimators of MSPE($\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$), respectively. Hence, the $mspe_J$ is defined in either case as

$$mspe_J = \widehat{M}_{1iJ} + \widehat{M}_{2iJ}, \tag{2.3.3}$$

noting that the $mspe_J$ is a nearly unbiased estimator of MSPE($\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$) if and only if $M_{3i} \approx 0$. Similarly, we use $mspe_{wJ}$ to show the weighted jackknife estimator of MSPE($\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$).

2.4. Empirical Results

In this section, we first provide a numerical study to compare the performance of the MSPE of the proposed pseudo-Bayes predictor $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}}$ with those obtained by GS and DRT. We then conduct a simulation study to evaluate the relative efficiency of the proposed PEB predictor, $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$, compared with the GS estimator, $\widehat{\gamma}_{i\mathrm{GS}}^{\mathrm{PEB}}$, the DRT estimator, $\widehat{\gamma}_{i\mathrm{DRT}}^{\mathrm{PEB}}$ and the naive estimator (henceforth abbreviated NAI), $\widehat{\gamma}_{i\mathrm{NAI}}^{\mathrm{PEB}}$. The NAI estimator is obtained in the absence of the measurement error (Battese et al., 1988). The performance of the proposed jackknife estimator, $mspe_J$, is also evaluated and compared with its competitors based on the GS, DRT and NAI methods.

Following Ghosh and Sinha (2007) and Datta et al. (2010), we assume that the responses y_{ij} for the population units and the observed covariates which contain measurement errors are generated from the model given by

(2.0.1) and (2.0.2) with $\sigma_\eta^2 = 25$, $\sigma_u^2 = 16$, $\sigma_e^2 = 100$, $b_1 = 2$, $b_0 = 100$. The number of units in population is $N = 3950$ spread across 20 areas of sizes N_i given by 50, 250, 50, 100, 200, 150, 50, 150, 100, 150, 100, 50, 300, 350, 400, 200, 250, 300, 350, and 400. The sample sizes (n_i) within areas are taken to be 1, 5, 1, 2, 4, 3, 1, 3, 2, 3, 2, 1, 6, 7, 8, 4, 5, 6, 7, and 8, respectively. We also generate x_i's ranging from a uniform distribution between 191 to 199 given by $\mathbf{x} = (197, 198, 197, 192, 192, 195, 192, 196, 194, 192, 191, 197, 191, 193, 199, 198, 194, 199, 191, 196)$, and treat them fixed through the simulation study. We conduct the simulation study for $R = 5000$ iterations and for each iteration, we generate small area population $\mathbf{Y}_i^{(r)} = (y_{i1}^{(r)}, y_{i2}^{(r)}, \ldots, y_{iN_i}^{(r)})$ and associated simple random samples $(y_{i1}^{(r)}, \ldots, y_{in_i}^{(r)})$ and $(X_{i1}^{(r)}, \ldots, X_{in_i}^{(r)})$, $i = 1, \ldots, m$, $r = 1, \ldots, R$, independently.

2.4.1. Efficiency of PB Estimators

We first evaluate the performance of the proposed PB estimator and compare it with the GS, NAI, and DRT estimators based on their MSPE's.

The MSPE of the GS predictor $\widehat{\bar{\gamma}}_{i\mathrm{GS}}^{\mathrm{PB}}$ can be obtained from the equation (3.5) of GS as follows

$$\mathrm{MSPE}(\widehat{\bar{\gamma}}_{i\mathrm{GS}}^{\mathrm{PB}}) = f_i^2 \left[\sigma_e^2 \left\{ \frac{(1-B_i)^2}{n_i} + \frac{1}{N_i - n_i} \right\} + B_i^2 \sigma_u^2 + \frac{b_1^2 B_i^2 \sigma_\eta^2}{n_i} \right],$$

while the NAI estimator gives

$$\text{MSPE}(\widehat{\gamma}_{i\text{NAI}}^{\text{PB}}) = f_i^2 \left[\sigma_e^2 \left\{ \frac{(1-B_i)^2}{n_i} + \frac{1}{N_i - n_i} \right\} + B_i^2 \sigma_u^2 \right].$$

Similarly, the $\text{MSPE}(\widehat{\gamma}_{i\text{DRT}}^{\text{PB}})$ can be obtained from the equation (2.5) of DRT given by

$$\text{MSPE}(\widehat{\gamma}_{i\text{DRT}}^{\text{PB}}) = \frac{f_i^2 \sigma_e^2 (1-A_i)}{n_i} + \frac{1}{N_i} f_i \sigma_e^2,$$

where $A_i = \sigma_e^2/(\sigma_e^2 + n_i\sigma_u^2 + b_1^2\sigma_\eta^2)$.

Table 2.1 reports the values of $\text{MSPE}(\widehat{\gamma}_i^{\text{PB}})$ based on the JS estimator given by (2.1.10), and the estimators based on the DRT and GS methods. The relative efficiency of $\widehat{\gamma}_{i\text{JS}}^{\text{PB}}$ over $\widehat{\gamma}_{i\text{DRT}}^{\text{PB}}$, defined by $\text{MSPE}(\widehat{\gamma}_{i\text{DRT}}^{\text{PB}})/\text{MSPE}(\widehat{\gamma}_{i\text{JS}}^{\text{PB}})$, ranged from 93.90% to 193.15%, and the relative efficiency of $\widehat{\gamma}_{i\text{JS}}^{\text{PB}}$ over $\widehat{\gamma}_{i\text{GS}}^{\text{PB}}$ ranged from 102.43% to 312.33%. In particular, our results confirm that in small areas with a very small number of samples, the JS estimator is more efficient than the corresponding estimators based on the GS and DRT methods. One can also notice that the NAI method results in the smallest values of $\text{MSPE}(\widehat{\gamma}^{\text{PB}})$, however, as we show in Section 2.4.2, the EMSPE and the relative bias associated with this method are the largest ones among all methods considered in this chapter. This is because the NAI method does not account for the measurement errors in the auxiliary variables (see Section 2.4.2 for more details).

The JS estimator dominates the ML estimator in terms of the sum of the

weighted MSPE, that is,

$$\sum_{i=1}^{m} W_i \mathrm{E}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}} - \gamma_i)^2 \leq \sum_{i=1}^{m} W_i \mathrm{E}(\widehat{\gamma}_{i\mathrm{DRT}}^{\mathrm{PB}} - \gamma_i)^2,$$

with different weighting schemes as presented in Table 2.2 and 2.3. It means that although the MSPE($\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}}$) in some areas are slightly larger than the corresponding MSPE($\widehat{\gamma}_{i\mathrm{DRT}}^{\mathrm{PB}}$) which is based on the ML estimates of x_i's, however, the sum of the weighted MSPE($\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}}$) is smaller than the corresponding sum of the weighted MSPE($\widehat{\gamma}_{i\mathrm{DRT}}^{\mathrm{PB}}$).

As indicated in (2.1.8), the JS estimator employs the information from the data in two ways. First, it combines the information in $\bar{X}_i$ and $\bar{y}_i$ ($i = 1, \ldots, m$), called $\bar{Z}_i^*$. Then, it uses the information of all areas to estimate each area covariate, $x_{i\mathrm{JS}}$ ($i = 1, \ldots, m$). This method improves the GS and the naive estimator in terms of combining the information of $\bar{X}_i$ and $\bar{y}_i$ ($i = 1, \ldots, m$). It also has an advantage over the DRT estimator as it utilizes the information of the whole dataset to estimate the specific area covariate.

2.4.2. Simulation Study

In this section, we conduct a simulation study to evaluate the performance of the new PEB predictor, $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$. To this end, we first compute $\gamma_i^{(r)}$'s, $\gamma_i^{(r)} = \frac{\sum_{j=1}^{N_i} y_{ij}^{(r)}}{N_i}$, based on the small area populations in the rth iteration. Then, we find $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}(r)}$, $\widehat{\gamma}_{i\mathrm{GS}}^{\mathrm{PEB}(r)}$, $\widehat{\gamma}_{i\mathrm{DRT}}^{\mathrm{PEB}(r)}$, and $\widehat{\gamma}_{i\mathrm{NAI}}^{\mathrm{PEB}(r)}$ from the sample units. In order to find the pseudo-

empirical Bayes, model parameters are also estimated using the consistent estimators defined in Section 2.2. Table 2.4 shows the model parameters estimates and their corresponding biases and mean squared errors (MSE). The large value of MSE and bias for the method of moments estimates of model parameters has an important effect on the observed MSPE of small area predictors. Note that introducing estimators of model parameters with smaller MSE will improve the estimation of MSPE of small area predictors. The empirical MSPE of $\widehat{\gamma}_i^{\text{PEB}}$ for different methods ($\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}$, $\widehat{\gamma}_{i\text{GS}}^{\text{PEB}}$, $\widehat{\gamma}_{i\text{DRT}}^{\text{PEB}}$, and $\widehat{\gamma}_{i\text{NAI}}^{\text{PEB}}$) is defined as

$$\text{EMSPE}(\widehat{\gamma}_i^{\text{PEB}}) = \frac{1}{R}\sum_{r=1}^{R}(\widehat{\gamma}_i^{\text{PEB}(r)} - \gamma_i^{(r)})^2.$$

In addition, we decompose $\text{EMSPE}(\widehat{\gamma}_i^{\text{PEB}})$ as

$$\text{EMSPE}(\widehat{\gamma}_i^{\text{PEB}}) = M_{1i} + M_{2i} + 2M_{3i},$$

where

$$M_{1i} = \frac{1}{R}\sum_{r=1}^{R}(\widehat{\gamma}_i^{\text{PB}(r)} - \gamma_i^{(r)})^2, M_{2i} = \frac{1}{R}\sum_{r=1}^{R}(\widehat{\gamma}_i^{\text{PEB}(r)} - \widehat{\gamma}_i^{\text{PB}(r)})^2,$$

and

$$M_{3i} = \frac{1}{R1}\sum_{r=1}^{R}(\widehat{\gamma}_i^{\text{PB}(r)} - \gamma_i^{(r)})(\widehat{\gamma}_i^{\text{PEB}(r)} - \widehat{\gamma}_i^{\text{PB}(r)}).$$

Table 2.1 gives the empirical MSPE of the PEB predictors. One can easily observe that the JS predictor outperforms other methods in most areas. The relative efficiency of the PEB based on the JS estimator over its counterpart based on the ML estimator ranges from 105.06% to 195.78%. Also, the relative

efficiency of the PEB based on the JS estimator over the PEB based on the method of moments ranges from 269.60% to 664.11%. Finally, the relative efficiency of the PEB based on the JS estimator over the NAI estimator ranges from 100.00% to 138.44%. Table 2.3 gives the corresponding values of the weighted EMSPE of PEB predictors for different weight schemes.

As we discussed in Section 2.3, the jackknife estimator of MSPE($\widehat{\gamma}_i^{\text{PEB}}$) is nearly unbiased if $M_{3i} \approx 0$ ($i = 1, \ldots, m$). Table 2.5 gives the decomposition of the EMSPE for the JS, DRT, GS and NAI predictors. The JS approach gives relatively small values of M_{3i} compared to the GS and the NAI predictors, and it is similar to the DRT predictor.

We have also studied the performance of the weighted and unweighted jackknife estimators of MSPE($\widehat{\gamma}_i^{\text{PEB}}$), denoted by $mspe_{wJ}$ and $mspe_J$, of the proposed PEB predictor, $\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}$, and similar estimators for DRT, GS and NAI methods. The relative bias (RB) of MSPE estimators, $mspe$, is given by

$$\text{RB} = \frac{\text{E}(mspe)}{\text{EMSPE}} - 1,$$

where for example to calculate the RB of the weighted jackknife estimation of MSPE($\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}$) for area i we have

$$\text{E}(mspe_i) = \tfrac{1}{R} \sum_{r=1}^{R} mspe_{iw\text{J}}^{(r)}(\widehat{\gamma}_{iw\text{JS}}^{\text{PEB}}).$$

Table 2.6 presents the RB of the weighted and unweighted jackknife estimators of MSPE($\widehat{\gamma}_i^{\text{PEB}}$) for GS, the JS, DRT, and NAI predictors. The

absolute value of the RB of the weighted and unweighted jackknife estimators of MSPE($\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}$) is less than 12% for all areas. On the other hand, the absolute value of the RB of the weighted and unweighted jackknife estimators of MSPE($\widehat{\gamma}_{i\text{DRT}}^{\text{PEB}}$) is less than 15% for all areas while the NAI estimators result in the larger RBs. Our simulation studies show the GS performs poorly in terms of RB. Interestingly, based on our simulation studies, weighted and unweighted jackknife methods perform very similar in terms of the RB.

Table 2.6 also shows the empirical bias of the different PEB predictors. The empircal bias of the PEB predictors based on the JS estimate of the true area covariate is approximately equal to the empirical bias of the NAI estimator. Moreover, DRT and GS have larger biases in predicting the small area means. Figure 2.1 illustrates the EMSPE and bias of the different PEB predictors. Due to the results obtained in Table 2.6, we decided not to report the EMSPE and bias of the PEB predictor of the small area mean based on GS in Figure 2.1.

2.5. An Application

In this section, we evaluate the proposed approach using a cross sectional study in New Zealand to predict the diastolic blood pressure using the cholesterol level. The data frame contains 10529 observations on 58 variables including age, sex, height, weight, and education. In this study we focus on 222 Maori or Asian female observations. This dataset is available as `xs.nz` in the `VGAMdata package` in *R*.

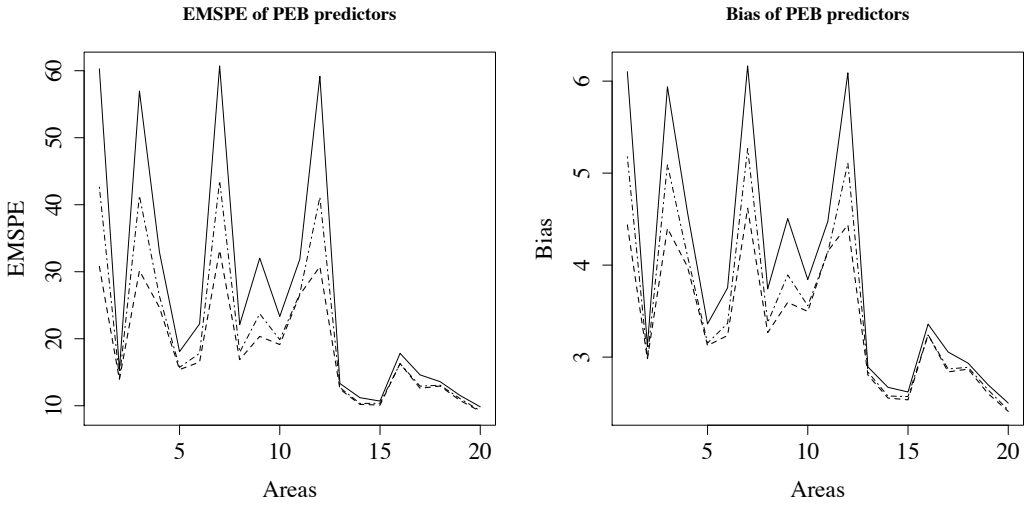

Figure 2.1: EMSPE and bias of $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$ (dashed line), $\widehat{\gamma}_{IDRT}^{\mathrm{PEB}}$ (solid line) and $\widehat{\gamma}_{i\mathrm{NAI}}^{\mathrm{PEB}}$ (dot line) for the different areas

To apply our proposed methodology, we first group the underlying population (i.e., the female) based on the age (16–32, 32–42, 42–52, and 52–88), Body Mass Index (BMI) (12.80–23.53, 23.54–25.86, 25.87–28.68, 28.69–88.43), ethnic (Maori or Asian), and smoking status (0 or 1). This results in 43 small areas (domains) with the number of samples in areas, denoted by n_i, $i = 1, \ldots, 43$, ranges from 0 to 15, with 21 areas having no samples. We assume the sampling error of the covariate in each small area to be negligible in comparison with the measurement error and this seems to be a reasonable assumption due to our refine grouping. Hence, we consider the x_i's as the true mean of the cholesterol level in each area. The idea is to predict the average diastolic blood pressure in each area assuming that the cholesterol level of each individual is measured

with error. To model this measurement error problem, the cholesterol levels of individuals in each area are assumed to be close to the average cholesterol level of the corresponding area with some error. It is worth noting that one can also use the data set to predict the diastolic blood pressure of each individual, however this is not of general interest in the context of the small area estimation and the problem will be studied in a separate work. Figure 2.2 shows the diastolic blood pressure versus the cholesterol level. We use equations (2.0.1) and (2.0.2) to model the data where X_{ij} is the observed value of the cholesterol level and Y_{ij} is the diastolic blood pressure for the j'th person in the i'th small area.

Using the method of moments, we get $\widehat{b}_0 = 24.62$, $\widehat{b}_1 = 9.86$, $\widehat{\sigma}_e^2 = 93.39$, $\widehat{\sigma}_u^2 = 26.07$, and $\widehat{\sigma}_\eta^2 = 0.97$. In this data set X_{ij}'s vary from 2 to 10 while Y_{ij}'s vary from 40 to 100. The estimated values of the diastolic blood pressure means, $\widehat{\gamma}_i$, using different approaches are given in Figure 2.3.

For the small areas with no sample units, we use $\widehat{\gamma}_{i\text{DRT}}^{\text{PEB}} = \widehat{b}_0$ as the DRT estimate of the true area covariate does not exist. The same will be used for $\widehat{\gamma}_{i\text{GS}}^{\text{PEB}}$, while for the JS method, a natural estimator is $\widehat{\gamma}_{i\text{JS}}^{\text{PEB}} = \widehat{b}_0 + \widehat{b}_1\widehat{x}_{i\text{JS}}$. We get $\widehat{\gamma}_{i\text{JS}}^{\text{PEB}} = 74.54$ for areas with no sample units. Finally, Figure 2.4 presents the weighted and unweighted jackknife estimates of MSPE($\widehat{\gamma}_i^{\text{PEB}}$) for different methods.

Based on our result, there are areas with the high diastolic blood pressure. These areas belong to areas with overweight women. The smoking status does

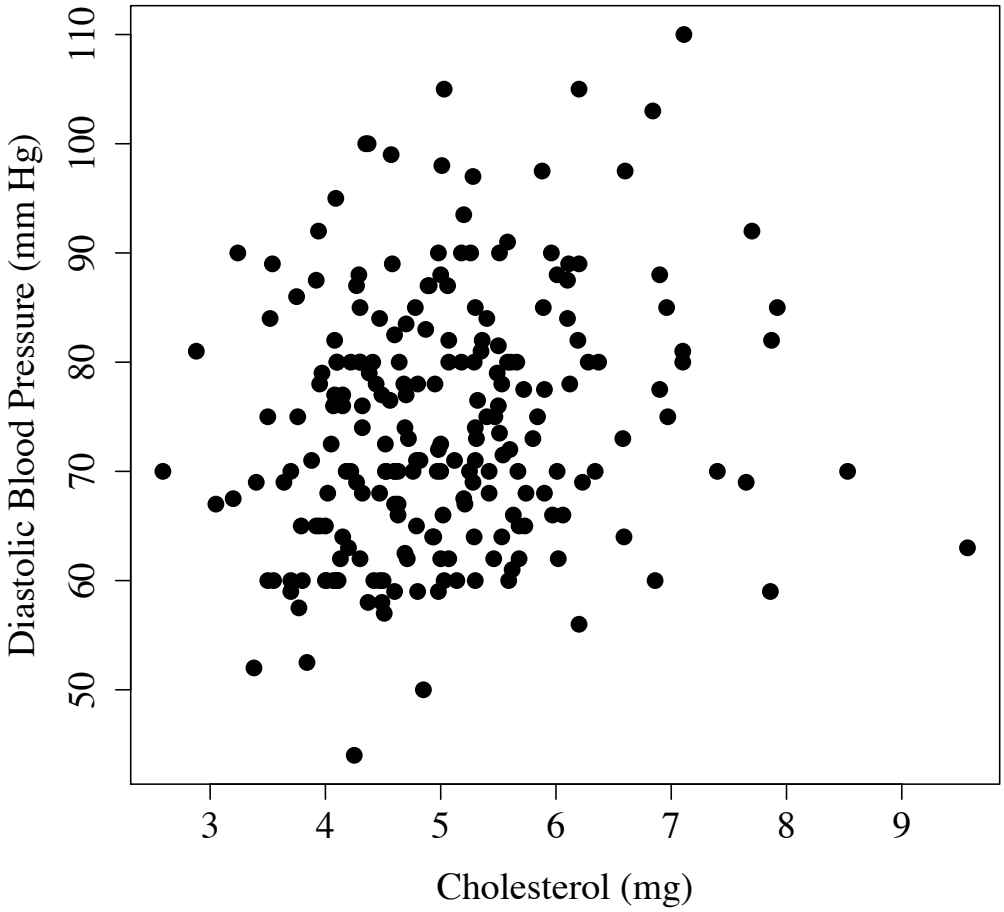

Figure 2.2: Diastolic Blood Pressure versus Cholesterol

not have a significant effect on the diastolic blood pressure, while Age is also an influential factor as older women with larger BMI suffer more from the high diastolic blood pressure.

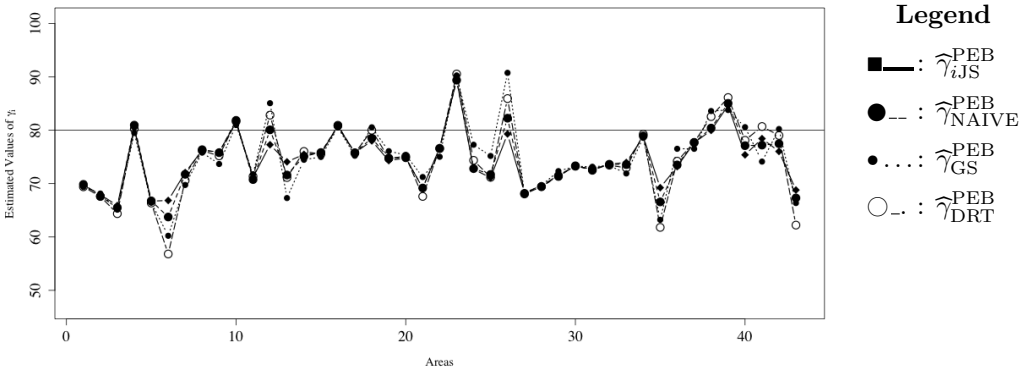

Figure 2.3: The pseudo-empirical Bayes predictors of small area means.

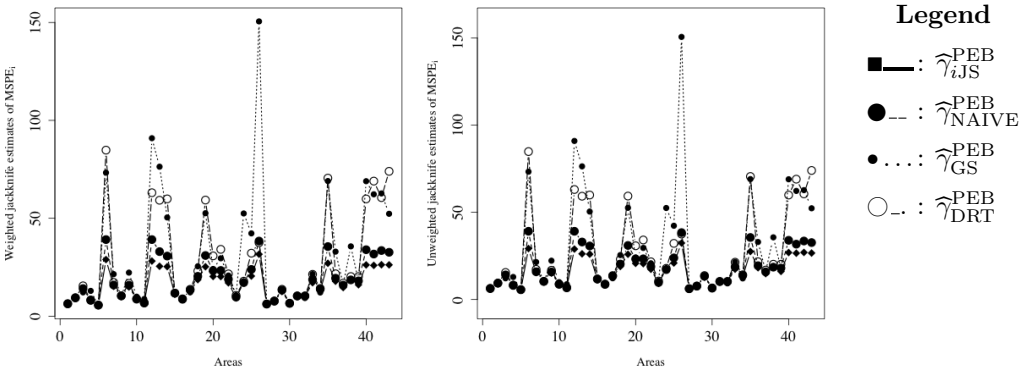

Figure 2.4: The MSPE estimation of pseudo-empirical Bayes predictors of small area means.

Table 2.1: Numerical values of the MSPE of $\widehat{\gamma}_{i\mathrm{GS}}^{\mathrm{PB}}$, $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}}$ and $\widehat{\gamma}_{i\mathrm{DRT}}^{\mathrm{PB}}$ and the empirical MSPE of $\widehat{\gamma}_{i\mathrm{GS}}^{\mathrm{PEB}}$, $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$, $\widehat{\gamma}_{i\mathrm{DRT}}^{\mathrm{PEB}}$ and $\widehat{\gamma}_{i\mathrm{NAI}}^{\mathrm{PEB}}$

		MSPE				EMSPE			
Area	n_i	$\widehat{\gamma}_{i\mathrm{GS}}^{\mathrm{PB}}$	$\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}}$	$\widehat{\gamma}_{i\mathrm{DRT}}^{\mathrm{PB}}$	$\widehat{\gamma}_{i\mathrm{NAI}}^{\mathrm{PB}}$	$\widehat{\gamma}_{i\mathrm{GS}}^{\mathrm{PEB}}$	$\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$	$\widehat{\gamma}_{i\mathrm{DRT}}^{\mathrm{PEB}}$	$\widehat{\gamma}_{i\mathrm{NAI}}^{\mathrm{PEB}}$
1	1	86.58	27.72	53.54	15.21	204.50	30.79	60.29	42.63
2	5	14.86	12.46	12.74	8.93	49.44	13.92	15.27	14.02
3	1	86.58	27.72	53.54	15.21	185.36	30.16	56.97	41.33
4	2	40.18	21.71	28.30	12.62	89.90	24.65	32.89	26.31
5	4	18.79	14.04	15.41	9.86	52.38	15.40	18.07	15.73
6	3	25.65	14.43	19.76	11.04	54.59	16.51	22.22	17.82
7	1	86.58	30.47	53.54	15.21	196.91	33.20	60.72	43.54
8	3	25.65	14.97	19.76	11.04	55.05	16.88	22.10	17.98
9	2	40.18	17.56	28.30	12.62	85.08	20.35	32.04	23.73
10	3	25.65	17.00	19.76	11.04	74.50	19.13	23.32	19.86
11	2	40.18	25.59	28.30	12.62	120.66	26.55	31.87	26.73
12	1	86.58	27.72	53.54	15.21	182.17	30.77	59.18	41.13
13	6	12.33	11.19	10.93	8.17	55.20	12.48	13.32	12.71
14	7	10.58	8.97	9.60	7.53	31.04	10.20	11.20	10.34
15	8	9.29	9.07	8.59	6.98	50.56	10.10	10.70	10.37
16	4	18.79	14.74	15.41	9.86	64.02	16.29	17.82	16.29
17	5	14.86	10.93	12.74	8.93	37.31	12.60	14.60	12.92
18	6	12.33	11.64	10.93	8.17	57.54	12.94	13.59	13.06
19	7	10.58	9.86	9.60	7.53	50.77	10.77	11.52	11.05
20	8	9.29	8.05	8.59	6.98	24.68	9.16	9.84	9.26

Table 2.2: Numerical values of the weighted MSPE of $\widehat{\gamma}_{\mathrm{GS}}^{\mathrm{PB}}$, $\widehat{\gamma}_{JS}^{\mathrm{PB}}$, $\widehat{\gamma}_{\mathrm{DRT}}^{\mathrm{PB}}$, and $\widehat{\gamma}_{NAI}^{\mathrm{PB}}$ for different weight schemes

Weighting Schemes	$\widehat{\gamma}_{\mathrm{GS}}^{\mathrm{PB}}$	$\widehat{\gamma}_{\mathrm{JS}}^{\mathrm{PB}}$	$\widehat{\gamma}_{\mathrm{DRT}}^{\mathrm{PB}}$	$\widehat{\gamma}_{\mathrm{NAI}}^{\mathrm{PB}}$
$W_i = \frac{1/(\sigma_u^2+\frac{\sigma_e^2}{N_i})}{\sum_i 1/(\sigma_u^2+\frac{\sigma_e^2}{N_i})}$	32.76	16.54	23.06	10.64
$W_i = \frac{N_i}{\sum_i N_i}$	19.77	12.92	15.38	9.10
$W_i = \frac{1}{m}$	33.78	16.79	23.64	10.74
$W_i = \frac{1/\sigma_{0i}^2}{\sum_i 1/\sigma_{0i}^2}$	20.96	13.29	16.11	9.27

Table 2.3: Numerical values of the weighted EMSPE of $\widehat{\gamma}_{\mathrm{GS}}^{\mathrm{PEB}}$, $\widehat{\gamma}_{JS}^{\mathrm{PEB}}$, $\widehat{\gamma}_{\mathrm{DRT}}^{\mathrm{PEB}}$, and $\widehat{\gamma}_{NAI}^{\mathrm{PEB}}$ for different weight schemes

Weighting Schemes	$\widehat{\gamma}_{\mathrm{GS}}^{\mathrm{PEB}}$	$\widehat{\gamma}_{\mathrm{JS}}^{\mathrm{PEB}}$	$\widehat{\gamma}_{\mathrm{DRT}}^{\mathrm{PEB}}$	$\widehat{\gamma}_{\mathrm{NAI}}^{\mathrm{PEB}}$
$W_i = \frac{1/(\widehat{\sigma}_u^2+\frac{\widehat{\sigma}_e^2}{N_i})}{\sum_i 1/(\widehat{\sigma}_u^2+\frac{\widehat{\sigma}_e^2}{N_i})}$	83.78	18.33	26.15	20.87
$W_i = \frac{N_i}{\sum_i N_i}$	59.03	14.41	17.87	15.36
$W_i = \frac{1}{m}$	86.08	18.64	26.88	21.34
$W_i = \frac{1/\widehat{\sigma}_{0i}^2}{\sum_i 1/\widehat{\sigma}_{0i}^2}$	63.15	15.17	19.32	16.35

Table 2.4: The estimated values of the parameters and their corresponding biases and MSEs

	b_0	b_1	σ_e^2	σ_u^2	σ_η^2
Parameter Estimate	-356.65	4.34	100.21	13.83	24.92
Bias	564.92	2.90	14.81	14.01	3.70
MSE	290057746.75	7609.13	346.54	283.31	21.77

Table 2.5: The values of the components of the EMSPE of $\widehat{\gamma}_{\text{GS}}^{\text{PEB}}$, $\widehat{\gamma}_{JS}^{\text{PEB}}$, $\widehat{\gamma}_{\text{DRT}}^{\text{PEB}}$, and $\widehat{\gamma}_{NAI}^{\text{PEB}}$

M_{1GS}	M_{2GS}	M_{3GS}	M_{1JS}	M_{2JS}	M_{3JS}	M_{1DRT}	M_{2DRT}	M_{3DRT}	M_{1NAI}	M_{2NAI}	M_{3NAI}
88.58	92.54	11.69	27.94	3.64	-0.39	54.76	6.42	-0.44	88.61	30.43	-38.20
15.16	32.67	0.81	12.67	2.30	-0.52	12.90	2.33	0.02	15.15	7.37	-4.25
87.42	79.37	9.29	28.15	3.25	-0.62	53.25	5.80	-1.04	87.47	29.05	-37.60
40.38	44.51	2.50	22.81	3.31	-0.73	28.89	4.04	-0.02	40.39	17.77	-15.92
18.65	30.72	1.50	13.88	2.30	-0.39	15.31	2.39	0.19	18.63	8.32	-5.61
26.16	22.61	2.91	14.59	1.90	0.01	19.86	2.10	0.13	26.13	10.08	-9.19
89.57	88.73	9.30	31.21	3.72	-0.87	54.78	6.45	-0.26	89.64	31.69	-38.89
25.58	24.01	2.73	14.97	2.07	-0.08	19.63	2.31	0.08	25.60	10.43	-9.02
39.91	36.02	4.58	17.87	2.56	-0.04	28.42	3.23	0.19	39.92	15.48	-15.83
26.24	40.16	4.05	17.41	2.67	-0.47	20.18	2.88	0.13	26.23	11.34	-8.86
39.40	76.02	2.62	25.24	3.84	-1.26	27.36	4.60	-0.04	39.47	18.86	-15.80
85.91	78.02	9.12	28.53	3.58	-0.68	53.07	6.63	-0.26	85.83	30.53	-37.61
12.30	42.94	-0.03	11.18	2.25	-0.48	10.86	2.28	0.09	12.30	6.36	-2.97
10.58	18.52	0.97	8.99	1.35	-0.07	9.58	1.38	0.12	10.58	3.98	-2.11
9.17	42.12	-0.37	9.00	1.97	-0.43	8.50	2.06	0.07	9.17	4.59	-1.70
18.42	42.53	1.53	14.74	2.51	-0.48	15.01	2.61	0.10	18.39	9.29	-5.70
14.55	19.12	1.82	11.08	1.53	-0.01	12.75	1.60	0.13	14.56	5.53	-3.59
12.50	45.59	-0.28	11.81	2.43	-0.65	11.02	2.51	0.03	12.48	6.64	-3.03
10.18	40.34	0.13	9.60	2.03	-0.43	9.17	2.07	0.14	10.18	5.08	-2.10
9.11	13.99	0.79	8.04	1.15	-0.02	8.53	1.17	0.07	9.10	3.14	-1.49

Table 2.6: Percent relative bias of the jackknife estimators of MSPE and the bias of different PEB estimators of the γ_i's

		Relative Bias of $mspe(\widehat{\gamma}_i)$								Bias of $\widehat{\gamma}_i$			
Area	n_i	GS_w	GS_{uw}	JS_w	JS_{uw}	DRT_w	DRT_{uw}	NAI_w	NAI_{uw}	GS	JS	DRT	NAI
1	1	–683.51	-684.50	-7.18	-6.74	7.68	7.16	-12.18	-11.69	9.46	4.44	6.10	5.18
2	5	204.20	204.35	4.01	4.17	7.71	7.78	17.32	17.58	4.63	2.98	3.12	2.98
3	1	-713.26	-713.66	-6.17	-5.74	14.13	13.59	-9.90	-9.46	9.20	4.40	5.94	5.09
4	2	-579.42	-581.18	-5.63	-5.45	6.25	5.94	5.45	5.73	6.62	3.98	4.57	4.10
5	4	-60.77	-62.44	4.28	4.44	7.26	7.18	19.17	19.41	4.98	3.13	3.36	3.15
6	3	-1069.78	-1071.91	6.11	6.38	6.48	6.18	21.59	21.80	5.35	3.23	3.75	3.37
7	1	-478.40	-480.36	-11.48	-11.11	8.12	7.59	-13.37	-12.95	9.54	4.62	6.17	5.27
8	3	-1119.88	-1122.37	5.32	5.58	7.87	7.59	20.87	21.10	5.34	3.27	3.74	3.37
9	2	-772.00	-774.38	5.64	5.95	6.38	5.98	13.66	13.89	6.44	3.59	4.51	3.89
10	3	-308.56	-310.60	-0.81	-0.65	5.88	5.71	12.62	12.85	5.74	3.50	3.84	3.56
11	2	36.61	35.57	-6.33	-6.22	12.94	12.70	6.24	6.57	6.93	4.15	4.47	4.14
12	1	-525.68	-526.03	-6.82	-6.40	10.13	9.59	-9.12	-8.70	9.32	4.44	6.09	5.11
13	6	1572.66	1572.53	6.54	6.70	10.38	10.58	16.70	16.96	4.52	2.80	2.89	2.84
14	7	-207.99	-208.51	7.21	7.46	7.78	7.81	21.40	21.58	3.78	2.56	2.67	2.58
15	8	1846.38	1849.17	11.21	11.48	12.28	12.73	18.63	19.00	4.08	2.54	2.62	2.57
16	4	371.57	371.88	1.62	1.76	10.47	10.44	16.87	17.13	5.04	3.24	3.36	3.24
17	5	-669.14	-671.01	3.91	4.16	5.53	5.40	21.84	22.02	4.30	2.84	3.06	2.87
18	6	1121.81	1123.59	5.78	5.94	9.81	10.09	15.51	15.85	4.62	2.87	2.93	2.89
19	7	1008.16	1007.53	11.37	11.58	13.37	13.58	20.22	20.45	4.26	2.60	2.70	2.64
20	8	-371.06	-372.12	6.66	6.95	7.81	7.89	21.12	21.31	3.48	2.41	2.50	2.43

Chapter 3

Pseudo Empirical Bayes Predictor of Small Area Means Using the Constrained Bayes Estimation

In Chapter 2, a variant of the PEB predictor of the small area mean using the JS estimate of the area-level covariate was proposed. It was also shown that the new PEB predictor outperforms previously proposed predictors using the method of moments (MM) and the ML estimator of the true area-specific covariate in terms of the MSPE and relative bias.

A drawback of the proposed JS estimator in Torkashvand et al. (2015), and essentially any other methods which uses the Bayesian and/or empirical Bayesian methodology to estimate the true area-specific covariate, is that the empirical histogram of these model-based predictors is underdispersed as an estimate of the histogram of the true small area means; noting that the corresponding predictors overshrink the direct estimates to their regression estimates (Lyles et al., 2007; Spjøtvoll and Thomsen, 1987). However, in

practice, there are many situations where the interest lies in producing an ensemble of parameter estimates whose distribution is close to the distribution of area-specific parameters. For example, in a hypertension study (see Section 3.4), one might be interested in identifying small areas whose true mean (diastolic) blood pressures are either below or above certain thresholds to identify groups that are more at risk for having hypertension. Another instance is to identify small areas with average income less than the poverty line or above a specific threshold (Ghosh, 1992) in order to have a better understanding of the socio-economic status of the population.

Louis (1984) proposed constrained (empirical) Bayes estimates of small area means in order to adjust the overshrinkage of small area means towards the prior distribution for the special case of the Fay-Herriot model. Lahiri (1990) gave an exact expression for the overshrinkage of the Bayes estimates of small area means when the underlying distribution of small area means are determined only up to their mean and variance and the conditional distribution of the response variable given the small area mean belongs to an exponential family with a quadratic variance function. Ghosh (1992) found a general expression for the constrained Bayes and showed the superiority of constrained Bayes estimates of small area means over the ML estimates in terms of the Bayes risk for the normal distribution.

Since Louis (1984), constrained Bayes (CB) estimation has been widely used in small area estimation for different purposes. For example, Datta et al. (2011)

used the CB approach for Bayesian benchmarking in small area estimation to provide an overall agreement between model-based area estimates and direct estimates at an aggregate level. Later, Ghosh and Steorts (2013) extended these results to a multi-stage benchmarking scenario. Ha (2013) proposed a general benchmarking method for complex benchmarking questions. Later on, Ha and Lahiri (2014) pointed out potential problems caused by benchmarking and recommended to implement the benchmarking with caution. See Jiang and Lahiri (2006), Ugarte et al. (2009),Kubokawa et al. (2013), Pfeffermann et al. (2014) and references therein on the theory and applications of the CB method in different problems including small area estimation.

Our goal in this chapter is to implement the CB methodology in small area estimation problems with measurement errors when the interest lies in getting a more precise picture of the true structure of small area means in order to make classifications rather than point prediction of individual area mean. To this end, we propose to use the CB method to adjust the estimators of true area-specific covariates of models (2.0.1) and (2.0.2) and consequently construct new PB predictors of small area means. We show that using this approach, one can obtain a more accurate estimate of the underlying histogram of the true area-specific covariate subject to the functional measurement error, and consequently a more precise histogram of small area predictors.

The outline of the chapter is as follows. In Section 3.1, we study the unit-level regression model with the functional measurement error in the area-

specific covariate and construct the CB estimator of the true area-specific covariate. Further, we obtain the PB predictors of area means based on the CB estimate of the true area-specific covariate which dominate the PB predictors of area means based on the ML estimate of the true area-specific covariate. In Sections 3.2 and 3.3, we obtain the constrained empirical Bayes (CEB) and constrained hierarchical Bayes (CHB) predictors of small area means and evaluate the performance of different predictors using some statistical criteria. The performance of our proposed approach is evaluated using a real data (a blood pressure study in New Zealand) and a simulation study in Sections 3.4 and 3.5, respectively.

3.1. Constrained Bayes Estimates of the True Area-Specific Covariate

In Chapter 2, we used the idea of normal prior distribution for the functional measurement error to derive the Bayes estimator of the true area-specific covariate similar to Carroll et al. (2010)[Sec. 9.1.3] and Carroll et al. (1999). Estimators of the true area-specific covariate are proposed under the quadratic loss function, $L(x_i, \lambda(\mathbf{y}, \mathbf{X})) = (x_i - \lambda(\mathbf{y}, \mathbf{X}))^2$, where $\lambda(\mathbf{y}, \mathbf{X})$ is the estimator of the true area-specific covariate, such that some optimal properties are preserved. The Bayes estimator of x_i, $x_{iB} = \mathrm{E}(x_i|\bar{Z}_i^*)$, was obtained in (2.1.6).

Even though the Bayesian method gives the best estimate of x_i's in terms

of the minimum Bayes risk, it overshrinks the Bayes estimates toward the prior mean (μ) in the sense that

$$\mathrm{E}(\sum_{i=1}^{m}(x_i-\bar{x})^2|\bar{\mathbf{Z}}^*) \geq \sum_{i=1}^{m}(x_{iB}(\bar{Z}_i^*)-\bar{x}_B(\bar{\mathbf{Z}}^*))^2, \tag{3.1.1}$$

where $\bar{x}_B(\bar{\mathbf{Z}}^*) = \frac{1}{m}\sum_{i=1}^{m} x_{iB}(\bar{Z}_i^*)$, $\bar{\mathbf{Z}}^* = (\bar{Z}_1^*, \bar{Z}_2^*, \ldots, \bar{Z}_m^*)$, and $\bar{x} = \frac{1}{m}\sum_{i=1}^{m} x_i$. Formula (3.1.1) states that the variance of x_{iB} is less than the posterior variance of x_i's. In other words, as x_{iB}'s are marginally unbiased estimators of μ, the inequality implies that they are more concentrated around μ in comparison with the true posterior distribution of x_i's. The equality holds if and only if all $(x_1-\bar{x}),\ldots,(x_m-\bar{x})$ have degenerate posterior distributions (Ghosh, 1992). This overshrinkage results in the poor performance of the Bayes estimate in tails of the histogram of x_i's (Louis, 1984; Ghosh, 1992; Lyles et al., 2007), and consequently, the resulting PB predictors of small area means based on the Bayes estimates of x_i's will perform poorly (see Section 3.5). Louis (1984) proposed to use the CB estimation method in order to address the problem of overshrinkage of the standard Bayes estimator toward the prior mean, when the underlying distribution is a multivariate normal distribution. This was later extended by Ghosh (1992) to other distributions. Following Ghosh (1992), the CB estimator of x_i is obtained by minimizing the posterior risk under the sum of squared error loss function

$$\mathrm{E}\left[\sum_{i=1}^{m}(x_i-t_i)^2|\bar{\mathbf{Z}}^*\right],$$

within the class of all estimators $\mathbf{t}(\bar{\mathbf{Z}}^*) = \mathbf{t} = (t_1,\ldots,t_m)$ of $x = (x_1,\ldots,x_m)$

that satisfy the following two conditions

$$\mathrm{E}(\bar{x}|\bar{\mathbf{Z}}^*) = \tfrac{1}{m}\sum_{i=1}^{m} t_i(\bar{\mathbf{Z}}^*) = \bar{t}(\bar{\mathbf{Z}}^*),$$

$$\mathrm{E}(\sum_{i=1}^{m}(x_i - \bar{x})^2|\bar{\mathbf{Z}}^*) = \sum_{i=1}^{m}(t_i(\bar{\mathbf{Z}}^*) - \bar{t}(\bar{\mathbf{Z}}^*))^2.$$

These conditions will help to obtain modified Bayes estimates of the values of the true area-specific covariate that have a histogram with the same mean and spread as the posterior mean and variance of the histogram of the true area-specific covariate. Now, using the Lagrange's method of undetermined multipliers, we obtain the CB estimators of x_i's as follow

$$x_{iCB}(\bar{\mathbf{Z}}^*) = \nu x_{iB}(\bar{Z}_i^*) + (1-\nu)\bar{x}_B(\bar{\mathbf{Z}}^*), \tag{3.1.2}$$

where

$$\nu \equiv \nu(\bar{\mathbf{Z}}^*) = \left(1 + \frac{H_1(\bar{\mathbf{Z}}^*)}{H_2(\bar{\mathbf{Z}}^*)}\right)^{\frac{1}{2}}, \tag{3.1.3}$$

and

$$H_1(\bar{\mathbf{Z}}^*) = (1-\tfrac{1}{m})\sum_{i=1}^{m}\frac{\tau^2\sigma_i^2}{\tau^2+\sigma_i^2} = (1-\tfrac{1}{m})\sum_{i=1}^{m}\tau^2 C_i$$

$$H_2(\bar{\mathbf{Z}}^*) = \sum_{i=1}^{m}(x_{iB}(\bar{\mathbf{Z}}^*) - \bar{x}_B(\bar{\mathbf{Z}}^*))^2. \tag{3.1.4}$$

Due to (3.1.3), ν has the stochastic nature. In Lemma 1, we present an almost sure asymptotic value of ν, say $\widehat{\nu}$. Using $\widehat{\nu}$, we show the optimality of CB estimators of x_i's over their corresponding ML estimators in Theorem 2.

Lemma 1. *Suppose the model parameters are known. Considering ν defined*

in (3.1.3), as $m \to \infty$, $\frac{\nu}{\widehat{\nu}}$ *almost surely converges to 1 where*

$$\widehat{\nu} = \left(1 + \frac{(1-\frac{1}{m})\sum_{i=1}^{m} C_i}{\sum_{i=1}^{m}(1-C_i)}\right)^{\frac{1}{2}}. \tag{3.1.5}$$

Proof. Using the Kolmogorov's strong law of large numbers under the independence (Sen and Singer, 1993), an approximation of $H_2(\bar{\mathbf{Z}}^*)$, as $m \to \infty$, is given by

$$\sum_{i=1}^{m} \tau^2(1-C_i).$$

Therefore, an almost sure approximation of ν, $\widehat{\nu}$, is obtained by

$$\left(1 + \frac{(1-\frac{1}{m})\sum_{i=1}^{m} C_i}{\sum_{i=1}^{m}(1-C_i)}\right)^{\frac{1}{2}}. \tag{3.1.6}$$

We keep using the asymptotic value of ν defined in (3.1.5) throughout the chapter due to the simplicity it introduces into the analysis. In Theorem 2, the aim is to show the superiority of the CB estimate of x_i over the ML estimate in terms of the Bayes risk when we plug in $\widehat{\nu}$ in (3.1.2) as an estimate of ν. To this end, the Bayes risk is defined as

$$r(\pi, \lambda(\mathbf{y}, \mathbf{X})) = \mathrm{E}_{\mathbf{x}}\left[\mathrm{E}_{(\mathbf{y},\mathbf{X})}\left(L(x_i, \lambda(\mathbf{y}, \mathbf{X}))\right)\right].$$

Theorem 2. *Consider the models (2.0.1) and (2.0.2) with known model parameters. Suppose the prior distribution on* $\mathbf{x} = (x_1, \ldots, x_m)$ *is* $\pi \sim N(\mu, \tau^2)$ *and the squared error loss function is used as the underlying loss function. Using the asymptotic value of* ν *defined in Lemma 1, the CB estimators,*

$\mathbf{x}_{CB} = (x_{1CB}, \ldots, x_{mCB})$, *of true area-specific covariates,* $\mathbf{x}$, *dominate the corresponding ML estimators,* $\bar{\mathbf{Z}}^* = (\bar{Z}_1^*, \bar{Z}_2^*, \ldots, \bar{Z}_m^*)$, *in terms of the Bayes risk, that is* $r(\pi, \mathbf{x}_{CB}) < r(\pi, \bar{\mathbf{Z}}^*)$.

Proof. Let $r(\pi, \delta)$ denote the Bayes risk of an arbitrary estimator of δ. We aim to show

$$r(\pi, \mathbf{x}_{CB}) < r(\pi, \bar{\mathbf{Z}}^*), \tag{3.1.7}$$

where $\bar{\mathbf{Z}}^* = (\bar{Z}_1^*, \bar{Z}_2^*, \ldots, \bar{Z}_m^*)$ is the vector of the ML estimates of the true area-specific covariates and $\mathbf{x}_{CB} = \mathbf{x}_{CB}(\bar{\mathbf{Z}}^*) = (x_{1CB}(\bar{\mathbf{Z}}^*), \ldots, x_{mCB}(\bar{\mathbf{Z}}^*))$. For the simplicity of the notation, we use $x_{iCB} = x_{iCB}(\bar{\mathbf{Z}}^*)$ and $x_{iB} = x_{iB}(\bar{Z}_i^*)$. Standard Bayesian calculations result in

$$r(\pi, \mathbf{x}_{CB}) - r(\pi, \bar{\mathbf{Z}}^*) = \sum_{i=1}^{m} \mathrm{E}_{\bar{Z}_i^*}(x_{iCB} - x_{iB})^2 - \sum_{i=1}^{m} \mathrm{E}_{\bar{Z}_i^*}(\bar{Z}_i^* - x_{iB})^2. \tag{3.1.8}$$

In order to show that $\mathbf{x}_{CB}$ dominates $\bar{\mathbf{Z}}^*$, it is enough to prove (3.1.8) is negative. Using the marginal distribution of ${\bar{Z}_i}^* \sim N(\mu, \sigma_{0i}^2 + \tau^2)$, $i = 1, \ldots, m$, we obtain

$$\sum_{i=1}^{m} \mathrm{E}_{\bar{Z}_i^*}(\bar{Z}_i^* - x_{iB})^2 = \sum_{i=1}^{m} C_i \sigma_i^2. \tag{3.1.9}$$

Further,

$$\sum_{i=1}^{m} \mathrm{E}_{\bar{Z}_i^*}(x_{iCB} - x_{iB})^2 = \sum_{i=1}^{m} \mathrm{E}_{\bar{Z}_i^*}((1-\nu)^2 (x_{iB} - \bar{x}_B)^2) \tag{3.1.10}$$

Another representation of $H_2(\bar{\mathbf{Z}}^*)$ in (3.1.4) is given by

$$\mathbf{x}_B' (\mathbf{I}_m - \tfrac{1}{m}\mathbf{J}_m)\mathbf{x}_B,$$

where $\mathbf{I}_m$ is the identity matrix, $\mathbf{J}_m$ is an $m \times m$ matrix of ones, and the quantity $x_{iB} \sim N(\mu, (1-C_i)^2(\sigma_{0i}^2+\tau^2))$, $i = 1, \ldots, m$. Due to the nonidentical distribution of x_{iB}'s, $i = 1, \ldots, m$, finding the exact distribution of the term $\mathbf{x}'_B(\mathbf{I}_m - \frac{1}{m}\mathbf{J}_m)\mathbf{x}_B$ is troublesome. Using Lemma 1, replacing $\widehat{\nu}$ in (3.1.10), and following straightforward mathematical calculations, we get the following equivalent expression for (3.1.7). Therefore, it is enough to prove

$$\sum_{i=1}^{m}\sigma_i^2 + 2(m-1)\tau^2\left((1-\tfrac{1}{m}\sum_{i=1}^{m}C_i)^{1/2}(1-\tfrac{1}{m^2}\sum_{i=1}^{m}C_i)^{1/2}-1\right) - \tau^2\tfrac{1}{m^2}\sum_{i=1}^{m}C_i > 0. \tag{3.1.11}$$

Using the Taylor expansion of $(1+x)^{1/2}$, we have

$$\begin{aligned}
&\sum_{i=1}^{m}\sigma_i^2 + 2(m-1)\tau^2\left((1-\tfrac{1}{m}\sum_{i=1}^{m}C_i)^{1/2}(1-\tfrac{1}{m^2}\sum_{i=1}^{m}C_i)^{1/2}-1\right) - \tau^2\tfrac{1}{m^2}\sum_{i=1}^{m}C_i \\
&\geq \sum_{i=1}^{m}\left(C_i^2\sigma_i^2 + \sigma_i^2(1-C_i)(C_i - \tfrac{1}{4}\bar{C})\right) + \tau^2\tfrac{1}{4m}(\sum_{i=1}^{m}C_i)^2 \\
&= \sum_{i=1}^{m}C_i\sigma_i^2 > 0,
\end{aligned}$$

which completes the proof.

Similar to Ghosh and Sinha (2007), Datta et al. (2010), and Torkashvand et al. (2015), we introduce a PB predictor of small area means based on the CB estimate of the true area-specific covariate by replacing x_{iJS} with x_{iCB} in (2.1.9), i.e.

$$\widehat{\gamma}_{iCB}^{PB} = \widehat{\gamma}_i^{PB}(x_{iCB}, \phi) = (1-f_iB_i)\bar{y}_i + f_iB_i(b_0 + b_1x_{iCB}). \tag{3.1.12}$$

In Theorem 3, we show that $\widehat{\gamma}_{iCB}^{PB}$ dominates $\widehat{\gamma}_{iML}^{PB}$ in terms of the Bayes risk.

Theorem 3. *The PB predictor of small area mean based on the CB estimate of the true area-specific covariate dominates the PB predictor of small area mean based on the ML estimate of the true area-specific covariate in terms of the Bayes risk*

$$\sum_{i=1}^{m} r(\pi, \widehat{\gamma}_{iCB}^{PB}) \leq \sum_{i=1}^{m} r(\pi, \widehat{\gamma}_{iML}^{PB}).$$

Proof. Following the notation given in Theorem 2, we aim to show that

$$\sum_{i=1}^{m} r(\pi, \widehat{\gamma}_{iCB}^{PB}) \leq \sum_{i=1}^{m} r(\pi, \widehat{\gamma}_{iML}^{PB}). \tag{3.1.13}$$

To this end, we first note that the Bayes risk of $\widehat{\gamma}_{iML}^{PB}$ can be written as follow $r(\pi, \widehat{\gamma}_{iML}^{PB}) = \mathrm{EE}(\widehat{\gamma}_{iML}^{PB} - \gamma_i)^2$

$$= \mathrm{EE}\left(f_i B_i b_1 (x_{iML} - x_i) - f_i B_i u_i + (1 - f_i B_i) \frac{\sum_{j=1}^{n_i} e_{ij}}{n_i} - \frac{\sum_{j=1}^{N_i} e_{ij}}{N_i} \right)^2$$

$$= \mathrm{EE}\left(f_i B_i b_1 (x_{iML} - x_i) \right)^2 + \mathrm{EE}(f_i B_i u_i)^2 + (1 - f_i B_i)^2 \mathrm{EE}\left(\frac{\sum_{j=1}^{n_i} e_{ij}}{n_i} \right)^2$$

$$+ \mathrm{EE}\left(\frac{\sum_{j=1}^{N_i} e_{ij}}{N_i} \right)^2 - 2\mathrm{EE}((f_i B_i)^2 b_1 u_i (x_{iML} - x_i))$$

$$+ 2\mathrm{EE}\left(f_i B_i (1 - f_i B_i) b_1 (x_{iML} - x_i) \frac{\sum_{j=1}^{n_i} e_{ij}}{n_i} \right)$$

$$- 2\mathrm{EE}\left(f_i B_i b_1 (x_{iML} - x_i) \frac{\sum_{j=1}^{N_i} e_{ij}}{N_i} \right)$$

$$
\begin{aligned}
&- 2\mathrm{EE}\left((1-f_iB_i)\frac{\sum_{j=1}^{n_i} e_{ij}}{n_i}\frac{\sum_{j=1}^{N_i} e_{ij}}{N_i}\right) \\
&= (f_iB_ib_1)^2\mathrm{EE}(x_{iML}-x_i)^2 + (f_iB_i)^2\sigma_u^2 + (1-f_iB_i)^2\frac{\sigma_e^2}{n_i} + \frac{\sigma_e^2}{N_i} \\
&- 2\mathrm{EE}((f_iB_i)^2b_1u_i(x_{iML}-x_i)) + 2\mathrm{EE}\left(f_iB_i(1-f_iB_i)b_1(x_{iML}-x_i)\frac{\sum_{j=1}^{n_i} e_{ij}}{n_i}\right) \\
&- 2\mathrm{EE}\left(f_iB_ib_1(x_{iML}-x_i)\frac{\sum_{j=1}^{N_i} e_{ij}}{N_i}\right) - 2(1-f_iB_i)\frac{\sigma_e^2}{N_i}. \qquad (3.1.14)
\end{aligned}
$$

Using the models (2.0.1) and (2.0.2), we have

$$
\begin{aligned}
x_{iML} &= \bar{X}_i + \frac{b_1\sigma_\eta^2\left(b_1(x_i-\bar{X}_i)+u_i+\sum_{i=1}^{n_i} e_{ij}/n_i\right)}{\sigma_e^2+n_i\sigma_u^2+b_1^2\sigma_\eta^2} \\
&= x_i + \bar{\eta}_i + \frac{b_1\sigma_\eta^2\left(-b_1\bar{\eta}_i+u_i+\sum_{i=1}^{n_i} e_{ij}/n_i\right)}{\sigma_e^2+n_i\sigma_u^2+b_1^2\sigma_\eta^2}.
\end{aligned}
$$

Therefore

$$
x_{iML}-x_i = \bar{\eta}_i + \frac{b_1\sigma_\eta^2\left(-b_1\bar{\eta}_i+u_i+\sum_{i=1}^{n_i} e_{ij}/n_i\right)}{\sigma_e^2+n_i\sigma_u^2+b_1^2\sigma_\eta^2}. \qquad (3.1.15)
$$

Equation (3.1.14) can be simplified using (3.1.15) and the independency between e_{ij}, u_i, and η_{ij} to get

$$
\begin{aligned}
r(\pi,\widehat{\gamma}_{iML}^{PB}) &= (f_iB_ib_1)^2\mathrm{EE}(x_{iML}-x_i)^2 + (f_iB_i)^2\sigma_u^2 + (1-f_iB_i)^2\frac{\sigma_e^2}{n_i} + \frac{\sigma_e^2}{N_i} \\
&\quad - 2(f_iB_ib_1)^2\frac{\sigma_u^2\sigma_\eta^2}{\sigma_e^2+n_i\sigma_u^2+b_1^2\sigma_\eta^2} + 2f_iB_i(1-f_iB_i)\frac{b_1^2\sigma_\eta^2\sigma_e^2/n_i}{\sigma_e^2+n_i\sigma_u^2+b_1^2\sigma_\eta^2} \\
&\quad - 2f_iB_ib_1^2\frac{\sigma_\eta^2\sigma_e^2/N_i}{\sigma_e^2+n_i\sigma_u^2+b_1^2\sigma_\eta^2} - 2(1-f_iB_i)\frac{\sigma_e^2}{N_i}. \qquad (3.1.16)
\end{aligned}
$$

Similarly, for γ_{iCB}^{PB} one can easily show that

$$
\begin{aligned}
r(\pi, \widehat{\gamma}_{iCB}^{PB}) &= \mathrm{EE}(\widehat{\gamma}_{iCB}^{PB} - \gamma_i)^2 \\
&= (f_i B_i b_1)^2 \mathrm{EE}(x_{iCB} - x_i)^2 + (f_i B_i)^2 \sigma_u^2 + (1 - f_i B_i)^2 \frac{\sigma_e^2}{n_i} + \frac{\sigma_e^2}{N_i} \\
&\quad - 2\mathrm{EE}((f_i B_i)^2 b_1 u_i (x_{iCB} - x_i)) \\
&\quad + 2\mathrm{EE}\left(f_i B_i (1 - f_i B_i) b_1 (x_{iCB} - x_i) \frac{\sum_{j=1}^{n_i} e_{ij}}{n_i} \right) \\
&\quad - 2\mathrm{EE}\left(f_i B_i b_1 (x_{iCB} - x_i) \frac{\sum_{j=1}^{N_i} e_{ij}}{N_i} \right) - 2(1 - f_i B_i) \frac{\sigma_e^2}{N_i}. \qquad (3.1.17)
\end{aligned}
$$

Similar to (3.1.15), we have

$$
\begin{aligned}
x_{iCB} - x_i &= \mu(\nu C_i + (1 - \nu)\bar{C}) \\
&\quad + (1 - C_i)\left(\nu + \frac{1 - \nu}{m} \right) \left[x_i + \bar{\eta}_i + \frac{b_1 \sigma_\eta^2 \left(-b_1 \bar{\eta}_i + u_i + \sum_{j=1}^{n_i} e_{ij}/n_i \right)}{\sigma_e^2 + n_i \sigma_u^2 + b_1^2 \sigma_\eta^2} \right] \\
&\quad + (1 - \nu) \sum_{l \neq i} \left[\frac{1 - C_l}{m} \left(x_l + \bar{\eta}_l + \frac{b_1 \sigma_\eta^2 \left(-b_1 \bar{\eta}_l + u_l + \sum_{j=1}^{n_l} e_{lj}/n_l \right)}{\sigma_e^2 + n_l \sigma_u^2 + b_1^2 \sigma_\eta^2} \right) \right] - x_i. \qquad (3.1.18)
\end{aligned}
$$

Formula (3.1.17) can be simplified by using (3.1.18) and assuming the indepen-

dency between e_{ij}, u_i and $\bar{\eta}_i$ in different areas. Hence, we have

$$r(\pi, \widehat{\gamma}_{iCB}^{PB}) = (f_i B_i b_1)^2 \mathrm{EE}(x_{iCB} - x_i)^2 + (f_i B_i)^2 \sigma_u^2 + (1 - f_i B_i)^2 \frac{\sigma_e^2}{n_i} + \frac{\sigma_e^2}{N_i}$$

$$+ (1 - C_i) \left(\nu + \frac{1-\nu}{m}\right) \left[-2(f_i B_i b_1)^2 \frac{\sigma_u^2 \sigma_\eta^2}{\sigma_e^2 + n_i \sigma_u^2 + b_1^2 \sigma_\eta^2} \right.$$

$$\left. + 2 f_i B_i (1 - f_i B_i) \frac{b_1^2 \sigma_\eta^2 \sigma_e^2 / n_i}{\sigma_e^2 + n_i \sigma_u^2 + b_1^2 \sigma_\eta^2} - 2 f_i B_i b_1^2 \frac{\sigma_\eta^2 \sigma_e^2 / N_i}{\sigma_e^2 + n_i \sigma_u^2 + b_1^2 \sigma_\eta^2} \right]$$

$$- 2(1 - f_i B_i) \frac{\sigma_e^2}{N_i}. \tag{3.1.19}$$

In Theorem 2, we showed that the CB estimator of the true area-specific covariate dominates the ML estimator of the true area-specific covariate in terms of the Bayes risk. The comparison of (3.1.16) and (3.1.19) shows that if

$$\sum_{i=1}^{m} (1 - C_i) \left(\nu + \frac{1-\nu}{m}\right) \left[-2(f_i B_i b_1)^2 \frac{\sigma_u^2 \sigma_\eta^2}{\sigma_e^2 + n_i \sigma_u^2 + b_1^2 \sigma_\eta^2} \right.$$

$$\left. + 2 f_i B_i (1 - f_i B_i) \frac{b_1^2 \sigma_\eta^2 \sigma_e^2 / n_i}{\sigma_e^2 + n_i \sigma_u^2 + b_1^2 \sigma_\eta^2} - 2 f_i B_i b_1^2 \frac{\sigma_\eta^2 \sigma_e^2 / N_i}{\sigma_e^2 + n_i \sigma_u^2 + b_1^2 \sigma_\eta^2} \right]$$

$$\leq \left[-2(f_i B_i b_1)^2 \frac{\sigma_u^2 \sigma_\eta^2}{\sigma_e^2 + n_i \sigma_u^2 + b_1^2 \sigma_\eta^2} + 2 f_i B_i (1 - f_i B_i) \frac{b_1^2 \sigma_\eta^2 \sigma_e^2 / n_i}{\sigma_e^2 + n_i \sigma_u^2 + b_1^2 \sigma_\eta^2} \right.$$

$$\left. - 2 f_i B_i b_1^2 \frac{\sigma_\eta^2 \sigma_e^2 / N_i}{\sigma_e^2 + n_i \sigma_u^2 + b_1^2 \sigma_\eta^2} \right], \tag{3.1.20}$$

the result follows. Note that

$$\left[-2(f_i B_i b_1)^2 \frac{\sigma_u^2 \sigma_\eta^2}{\sigma_e^2 + n_i \sigma_u^2 + b_1^2 \sigma_\eta^2} \right.$$

$$\left. + 2 f_i B_i (1 - f_i B_i) \frac{b_1^2 \sigma_\eta^2 \sigma_e^2 / n_i}{\sigma_e^2 + n_i \sigma_u^2 + b_1^2 \sigma_\eta^2} - 2 f_i B_i b_1^2 \frac{\sigma_\eta^2 \sigma_e^2 / N_i}{\sigma_e^2 + n_i \sigma_u^2 + b_1^2 \sigma_\eta^2} \right] = 0,$$

for $i = 1, \ldots, m$, which completes the proof.

Similar to (2.1.9), $\widehat{\gamma}_{iCB}^{PB}$ depends on the unknown model parameters. In Section 3.3, we provide estimates of the parameters using the method of moments, empirical Bayes, and hierarchical Bayes methods. To compare the performance of these predictors, we use different measures of performance such as the sensitivity (Se), specificity (Sp), positive predictive value (PPV), and negative predictive value (NPV). These measures provide some necessary probabilities that are useful to evaluate the precision of the PB predictors of the small area means when the goal is to study whether the small area means are above or under a specified threshold (Lyles and Xu, 1999).

3.2. Performance measures

In Section 3.1, we introduced the CB estimator of the true area-specific covariate. Some optimal properties of the CB estimator of the true area-specific covariate were also discussed in the Bayesian set-up. In practice, researchers sometimes want to classify areas according to whether their means are above or below some meaningful thresholds. For example, in the diastolic blood pressure application, having a blood pressure above the threshold $t = 80$ indicates the pre-hypertension phase (Zhang and Li, 2011). In Sections 3.4 and 3.5, we use the proposed estimator to study such scenarios. To this end, small area means are predicted using different approaches. People inside each area are considered to be in a pre-hypertension phase if their area mean is greater than

80. To evaluate the performance of the proposed estimators, we use the criteria defined in Lyles and Xu (1999). Consider t as a disease diagnostic threshold. In this work, we assume that a "positive" test or suffering from a "disease" happens if "$\widehat{\gamma}_i^{PB} > t$" where we treat $\widehat{\gamma}_i^{PB}$ as the predictor of the small area mean and a diagnostic test. If it is below the threshold, it will be an indication of "disease" and "positive" test, some adjustments of these definitions are then required. Following Lyles and Xu (1999), some statistical properties of the candidate predictors are specified as

$$Se = \mathbb{P}(\text{"Positive" given "disease"}) = \mathbb{P}(\widehat{\gamma}_i^{PB} > t | \gamma_i > t),$$

$$Sp = \mathbb{P}(\text{"Negative" given "no disease"}) = \mathbb{P}(\widehat{\gamma}_i^{PB} < t | \gamma_i < t),$$

$$PPV = \mathbb{P}(\text{"disease" given "Positive"}) = \mathbb{P}(\gamma_i > t | \widehat{\gamma}_i^{PB} > t),$$

$$NPV = \mathbb{P}(\text{"no disease" given "Negative"}) = \mathbb{P}(\gamma_i < t | \widehat{\gamma}_i^{PB} < t).$$

Analytical expressions for these quantities can be obtained following (2.0.1) and (2.0.2) and the bivariate normal distribution of $(\gamma_i, \widehat{\gamma}_i^{PB})$. In particular, we have

$$\begin{aligned} Se_{iCB} &= \frac{\mathbb{P}(\widehat{\gamma}_{iCB}^{PB} > t, \gamma_i > t)}{\mathbb{P}(\gamma_i > t)} = \frac{\mathbb{P}(S_i > (t - \alpha_i u_i), u_i > d_i)}{p} \\ &= \frac{\int_{d_i}^{\infty} \Phi(\frac{\mu_{s_i} - t + \alpha_i u_i}{\sigma_{s_i}}) f(u) du}{p}, \end{aligned} \tag{3.2.1}$$

where $p = \mathbb{P}(\gamma_i > t)$, $d_i = t - b_0 - b_1 x_i$, and

$$S_i = (1 - f_i B_i)(b_0 + b_1 x_i) + f_i B_i b_1 A_i x_i + f_i B_i \sum_{j \neq i} D_j \bar{Z}_j^*$$

$$+ \alpha_i \bar{e}_i + F_i \bar{\eta}_i + f_i B_i \left(b_0 + b_1 \mu (\nu C_i + (1 - \nu) \tfrac{\sum_{j=1}^m C_j}{m}) \right).$$

Also, $A_i = (1 - C_i)(\nu + \frac{1-\nu}{m})$, $\alpha_i = 1 - f_i B_i \left(1 - A_i \frac{b_1^2 \sigma_\eta^2}{\sigma_e^2 + n_i \sigma_u^2 + b_1^2 \sigma_\eta^2}\right)$, $D_j = \frac{1}{m}(1-\nu)(1 - C_j)$ and $F_i = f_i B_i b_1 A_i \left(1 - \frac{b_1^2 \sigma_\eta^2}{\sigma_e^2 + n_i \sigma_u^2 + b_1^2 \sigma_\eta^2}\right)$. Moreover, we have $S_i \sim N(\mu_{s_i}, \sigma_{s_i}^2)$, where

$$\mu_{s_i} = (1 - f_i B_i)(b_0 + b_1 x_i) + f_i B_i b_1 A_i x_i$$

$$+ f_i B_i (b_0 + b_1 \mu (\nu C_i + (1 - \nu) \tfrac{\sum_{j=1}^m C_j}{m})) + f_i B_i \sum_{j \neq i} D_j x_j,$$

and,

$$\sigma_{s_i}^2 = \alpha_i^2 \tfrac{\sigma_e^2}{n_i} + F_i^2 \tfrac{\sigma_\eta^2}{n_i} + (f_i B_i)^2 \sum_{j \neq i} D_j^2 \sigma_j^2.$$

Similarly, we have

$$Sp_{iCB} = \frac{\mathbb{P}(\widehat{\gamma}_{iCB}^{PB} < t, \gamma_i < t)}{\mathbb{P}(\gamma_i < t)} = \frac{\mathbb{P}(S_i < (t - \alpha_i u_i), u_i < d_i)}{1 - p}$$

$$= \frac{\int_{-\infty}^{d_i} \Phi(\frac{t - \alpha_i u_i - \mu_{s_i}}{\sigma_{s_i}}) f(u) du}{1 - p}. \tag{3.2.2}$$

Also,

$$PPV_{iCB} = \frac{pSe_{iCB}}{\Phi\left(-\frac{(t - \mathrm{E}(\widehat{\gamma}_{iCB}^{PB}))}{\sigma_{\widehat{\gamma}_{iCB}^{PB}}}\right)} \quad \text{and} \quad NPV_{iCB} = \frac{(1-p)Sp_{iCB}}{\Phi\left(\frac{(t - \mathrm{E}(\widehat{\gamma}_{iCB}^{PB}))}{\sigma_{\widehat{\gamma}_{iCB}^{PB}}}\right)}, \tag{3.2.3}$$

where

$$\mathrm{E}(\widehat{\gamma}_{iCB}^{PB}) = (1 - f_i B_i)(b_0 + b_1 x_i) + f_i B_i (b_0 + b_1 \mathrm{E}(x_{iCB})), \tag{3.2.4}$$

$$var(\widehat{\gamma}_{iCB}^{PB}) = (\sigma_u^2 + \tfrac{\sigma_e^2}{n_i})[(1 - f_i B_i)^2 + 2(1 - f_i B_i)(f_i B_i) A_i \tfrac{b_1^2 \sigma_\eta^2}{\sigma_e^2 + n_i \sigma_u^2 + b_1^2 \sigma_\eta^2}]$$

$$+ (f_i B_i b_1)^2 var(x_{iCB}), \tag{3.2.5}$$

$$\mathrm{E}(x_{iCB}) = \mu(\nu C_i + (1 - \nu)\tfrac{\sum_{j=1}^m C_j}{m}) + A_i x_i + \sum_{j \neq i} D_j x_j, \tag{3.2.6}$$

$$var(x_{iCB}) = A_i^2 \sigma_i^2 + \sum_{j \neq i} D_j^2 \sigma_j^2, \tag{3.2.7}$$

and $\sigma_{\widehat{\gamma}_{iCB}^{PB}} = \sqrt{var(\widehat{\gamma}_{iCB}^{PB})}$.

As there are unknown parameters in (3.2.1), (3.2.2), and (3.2.3), the method of moments and the hierarchical Bayes estimates of the model parameters are used to obtain estimates of the model parameters, and consequently using the numerical method of integration, estimates of the performance measures introduced in this section are given (see Section 3.4 for more details).

3.3. The Constrained Empirical (and Hierarchical) Pseudo-Bayes Predictor of Small Area Means

In Sections 3.1 and 3.2, $\widehat{\gamma}_{iCB}^{PB}$, Se_i, Sp_i, PPV_i, and NPV_i were obtained under the assumption that model parameters are known. To estimate parameters, two scenarios are considered. First, similar to Torkashvand et al. (2015), we

follow Efron and Morris (1975) and Ghosh and Sinha (2007) to obtain the empirical Bayes (EB) estimates of τ^2 and μ as well as the method of moments estimates of the model parameters (GS), respectively. Consequently, the CEB of the area-level covariate and PEB predictor of small area means are given by replacing unknown parameters in equations (3.1.2) and (3.1.12).

In the second scenario, we propose the CHB estimator of the true area-specific covariate using priors on hyper parameters $\psi = (b_0, b_1, \sigma_u^2, \sigma_e^2, \sigma_\eta^2)$, μ, and τ^2. We consider informative prior distributions for the purpose of the analysis (more details are given in Sections 3.4 and 3.5). We also derive Pseudo Hierarchical Bayes (PHB) predictors of small area means based on the CHB using the posterior means of the components of ψ. In Sections 3.4 and 3.5, we also evaluate the performance of PEB and PHB predictors based on the EB, GS, and HB estimates of the model parameters, respectively.

Let ν_{HB} estimates the asymptotic value of ν using the HB estimates of the model parameters. Similarly, ν_{EB} shows the estimate of the asymptotic value of ν using the GS and EB estimates of the model parameters. Ghosh and Maiti (1999) pointed out that $\nu_{HB} > \nu_{EB}$ in their set-up, showing that the individual estimates shrink toward the overall average to a lesser extend by using the hierarchical Bayes method. They also found that the hierarchical and empirical Bayes methods asymptotically show the same behaviour in estimating ν. In our nested error linear regression model with the functional measurement error, we are not able to compare the behaviour of ν_{HB} and ν_{EB} mathematically due

to their complicated forms. In Section 3.5, we evaluate their performances using simulation studies.

3.4. An Application

Similar to Section 2.5, we analyze the cross-sectional data from the New Zealand population as an application of the proposed approach. The aim is to predict the diastolic blood pressure using cholesterol level as the covariate and to determine groups of people (small areas) who are in danger of hypertension corresponding to areas with means above the specified threshold, 80, as the prehypertension phase. As the aim is to find the proportion of small areas in the upper tail of the distribution of γ_i's, it is reasonable to choose $\widehat{\gamma}_{iCEB}^{PEB}$ or $\widehat{\gamma}_{iCHB}^{PEB}$ over $\widehat{\gamma}_{iEB}^{PEB}$ and $\widehat{\gamma}_{iHB}^{PEB}$ due to the overshrinkage of either $\widehat{\gamma}_{iEB}^{PEB}$ or $\widehat{\gamma}_{iHB}^{PEB}$ towards the prior mean. Using this criterion, groups of people who are likely to be in hypertension phase may be recognized for possible medical treatments.

Under the same set-up as Section 2.5, the empirical Bayes estimates of μ and τ^2 are obtained as $\widehat{\mu}_{EB} = 5.06$ and $\widehat{\tau}_{EB}^2 = 0.15$. In order to apply the hierarchical Bayes method, we use the GS estimates of the model parameters (Section 2.5) and the empirical Bayes estimates of μ and τ^2 to define the prior distributions as Unif$(0, 2(9.66))$, Unif$(0, 2(5.1))$, Unif$(0, 2(0.98))$, Unif$(0, 2(0.38))$, $N(24.62, 4)$, $N(9.86, 4)$, and $N(5.063, 9)$ for σ_e, σ_u, σ_η, τ, b_0, b_1, and μ, respectively. The hierarchical Bayes method results in $\widehat{b}_{0HB} = 24.50$, $\widehat{b}_{1HB} = 9.93$,

$\widehat{\sigma}^2_{eHB} = 97.60$, $\widehat{\sigma}^2_{uHB} = 2.77$, $\widehat{\sigma}^2_{\eta HB} = 1.03$, $\widehat{\mu}_{HB} = 5.04$ and $\widehat{\tau}^2_{HB} = 0.46$. It is worth mentioning that small values of $\widehat{\sigma}^2_{\eta}$ (either $\widehat{\sigma}^2_{\eta GS}$ or $\widehat{\sigma}^2_{\eta HB}$) refer to the variance of X_{ij}'s (ranging between 2 and 10) while $\widehat{\sigma}^2_{e}$ (either $\widehat{\sigma}^2_{eGS}$ or $\widehat{\sigma}^2_{eHB}$) refers to the variance of Y_{ij}'s (ranging from 40 to 100). Further, we get $\nu_{HB} = 1.28$ and $\nu_{EB} = 1.47$. Figure 3.1 presents the histograms of the average cholesterol level, x_i, using different approaches. As Figure 3.1 shows, the histograms of the CEB and CHB estimates of x_i's have larger variance than the variance of the EB and HB estimates of the x_i's. As we show in the simulation study in Section 3.5, the CEB and CHB result in ensemble estimates of the true area-specific covariate with the histograms being close to the true histogram of x_i's in comparison with the EB and HB methods. The estimated values of diastolic blood pressure means, $\widehat{\gamma}_i$'s, using different approaches are given in Figure 3.2. We use $\widehat{\gamma}_i$'s to determine areas that are in danger of hypertension. Finally, Se, Sp, PPV, and NPV values of different predictors of small area means are given in Figure 3.3. In order to obtain these values, we used `integrate` function in `R` to approximate the integrals. If the obtained values of the measures of performance are greater than one or less than zero due to the approximation, they will be adjusted by projecting those values into $[0, 1]$.

As being in pre-hypertension phase is equivalent to having the diastolic blood pressure greater than 80 (Zhang and Li, 2011), Se and PPV are more important measures of performance for the current application. Figure 3.3 shows the pseudo estimates of Se, Sp, PPV, and NPV based on $\widehat{\gamma}^{PEB}_{iML}$, $\widehat{\gamma}^{PEB}_{iCEB}$,

$\widehat{\gamma}_{iEB}^{PEB}$, $\widehat{\gamma}_{iCHB}^{PEB}$, and $\widehat{\gamma}_{iHB}^{PEB}$. We observe that $\widehat{\gamma}_{iCEB}^{PEB}$ shows the best performance in terms of the Se and PPV. But, in terms of the Sp and NPV, they have the worst performance.

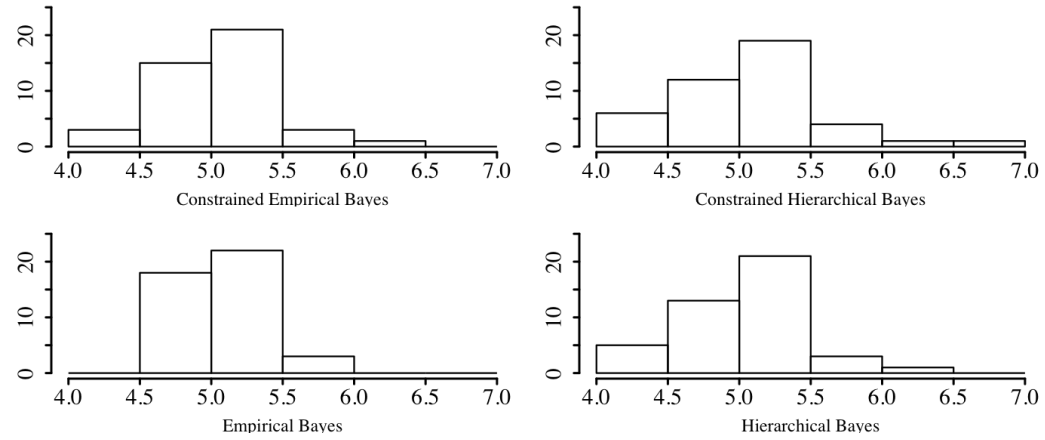

Figure 3.1: Average Cholesterol Level Estimates Using Different Approaches

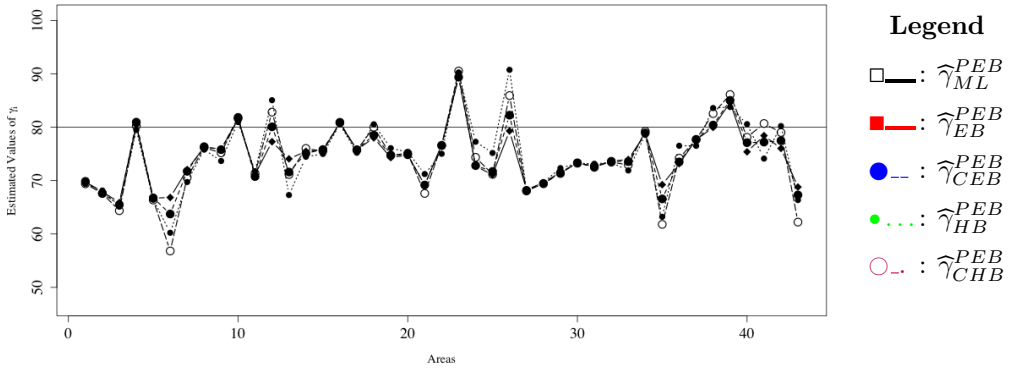

Figure 3.2: The estimated values of the small area mean.

In the case of areas with no sample units, the ML estimate of the true area-specific covariate cannot be defined, and we have $\widehat{\gamma}_{iML}^{PEB} = \widehat{b}_0$. However,

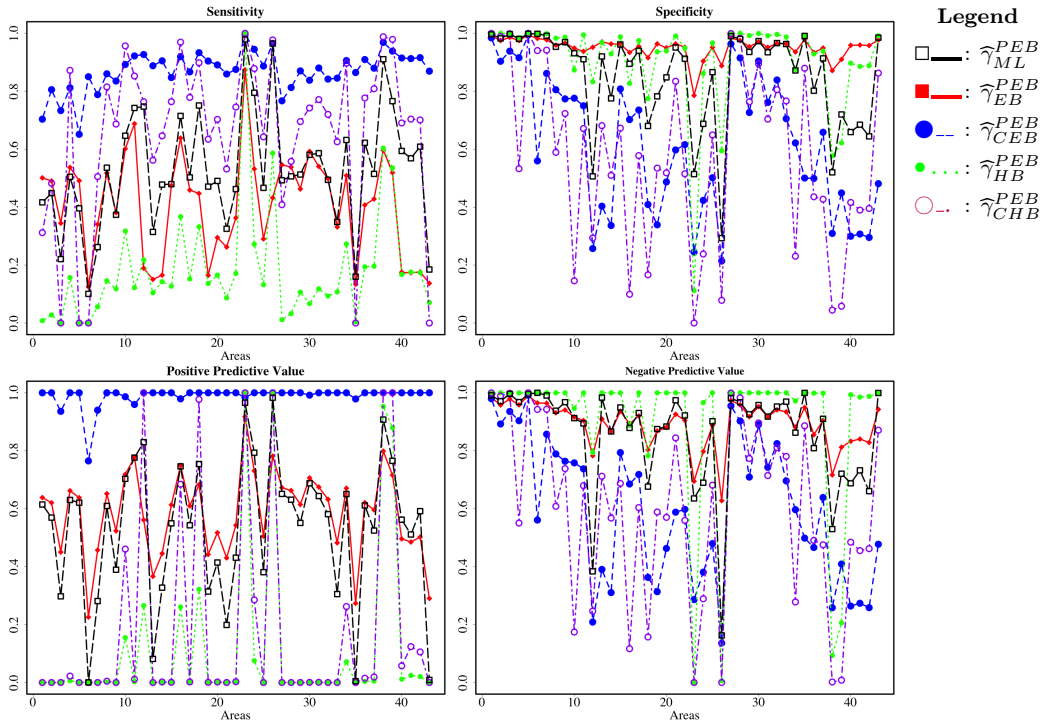

Figure 3.3: The estimated values of the measures of performance of different areas.

in the EB, CEB, HB, and CHB methods, to estimate the true area-specific covariate, we use the information from other areas as well. As (2.1.6) shows, for areas with no sample units we have $C_i = 1$ and so $x_{iB} = \mu$. Consequently, x_{iCB} can be defined using formula (3.1.2) for small areas with no sample units. The EB, CEB, HB and CHB estimates of the true area-specific covariate are calculated by replacing the EB and GS, and also HB estimates of model parameters, respectively. The PEB predictors of small area means can be found accordingly. In the application, there are 21 small areas with no sampled units. The predicted values of small area means for areas with no sample units are

$\widehat{\gamma}_{iML}^{PEB}$ = 24.62, $\widehat{\gamma}_{iEB}^{PEB}$ = 74.54, $\widehat{\gamma}_{iHB}^{PEB}$ = 74.54, $\widehat{\gamma}_{iCEB}^{PEB}$ = 106.97, and $\widehat{\gamma}_{iCHB}^{PEB}$ = 85.23. As detecting people who suffer from hypertension is important and also the simulation studies indicate that the CEB shows the best performance in terms of the Se and PPV, we recommend to use $\widehat{\gamma}_{iCEB}^{PEB}$ to obtain estimates of small area means.

Based on our results, areas with high diastolic blood pressure belong to overweight women. Based on our dataset, we conclude that the smoking status dose not have a significant contribution on the diastolic blood pressure. Our analysis also detects the age as an influential factor as old women with high BMI have high diastolic blood pressure.

3.5. Simulation Study

In this section, we implement a simulation study to evaluate the performance of the proposed method. The estimates of model parameters obtained from Section 3.4 are used for the purpose of simulation studies. To this end, we generate R = 5000 simulations with m = 43 areas in each simulation with observed values $\{(y_{ij}^{(r)}, X_{ij}^{(r)}); i = 1, \ldots, 43, j = 1, \ldots, n_i, r = 1, \ldots, 5000\}$ where b_0 = 24.62, b_1 = 9.86, σ_e^2 = 93.39, σ_u^2 = 26.07, σ_η^2 = 0.97, $\mathbf{x}$ = (4.47, 4.8, 4.47, 4.83, 4.31, 4.54,4.64, 5.01, 4.85, 5.18, 5.34, 5.35, 4.91, 5.09, 4.95, 5.33, 5.03, 5.45, 5.08, 5.10, 4.88, 5.04, 6.34, 5.54, 5.07, 5.84, 4.67, 4.83, 5.05, 4.93, 5.11, 4.93, 4.86, 5.25, 4.68, 5.26,5.08, 5.80, 5.49, 5.20, 5.18, 5.22, 4.72) where x's are the estimates of true area-specific cholesterol level obtained in Section 3.4, and the

sample size $n = (13, 8, 5, 10, 15, 1, 4, 7, 4, 9, 12, 1, 1, 1, 6, 9, 5, 3, 1, 2, 2, 3, 10, 4, 2, 2, 13, 10, 5, 12, 7, 7, 3, 5, 1, 3, 4, 4, 4, 1, 1, 1, 1)$. Further, we set $N_i = 100n_i$ for $i = 1, \ldots, m$.

In Section 3.4, we obtained different estimates of the model parameters using GS and HB methods. Table 3.1 gives the mean squared error (MSE) of the model parameters. Based on the MSE, it seems that we get more reliable estimates of σ_u^2 and σ_e^2 using the GS method. Our result (not shown here) indicates that using the GS method, τ^2 is estimated close to the true one while the HB method overestimates τ^2 significantly.

Table 3.1: The MSE of the model parameters using GS and HB methods

	b_0	b_1	σ_e^2	σ_u^2	σ_η^2
MSE GS	42147.00	1647.23	101.49	214.07	0.01
MSE HB	42040.84	1643.08	217.07	507.76	0.02

As the stochastic nature of ν introduces some difficulties in explaining the optimal properties of the CB estimator of the true area-specific covariate, we obtained an asymptotic expression for it in Lemma 1. In this section, we evaluate the precision of this asymptotic value. Figure 3.4 shows the asymptotic values of ν versus its exact values using the hierarchical and empirical Bayes estimates of the model parameters. We observe that the empirical Bayes estimates of the model parameters perform better than the corresponding HB estimates in terms of the asymptotic behaviour of ν. It is worth mentioning

that since (3.1.5) is a decreasing function of τ^2 and the HB method tends to overestimate τ^2 we expect to see $\nu_{HB} \leq \nu_{EB}$. This is also confirmed in our simulation study presented in Figure 3.4.

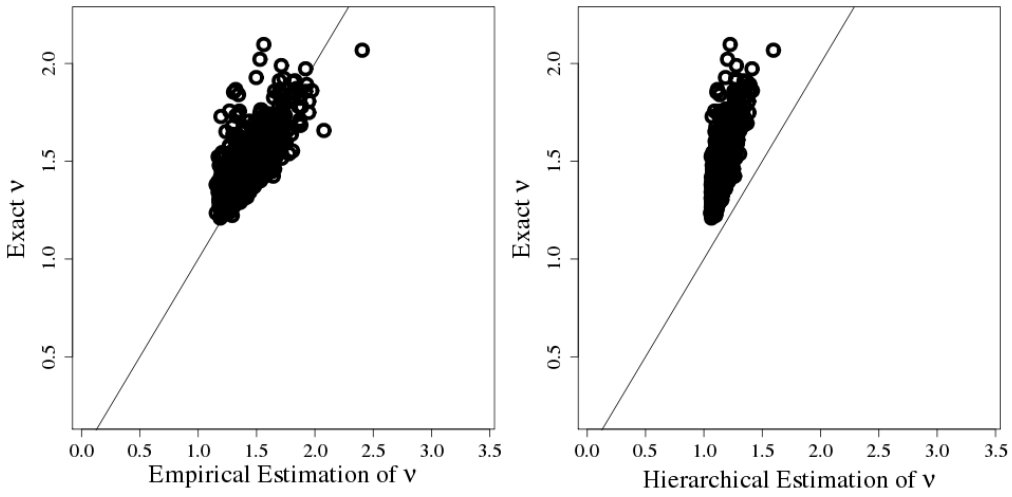

Figure 3.4: Exact versus asymptotic values of ν using hierarchical and empirical Bayes estimates of the model parameters

We are also interested to evaluate the performance of different methods in estimating the true area-specific covariate. In order to find the HB and CHB estimates of the x_i's, we apply the informative prior distributions on the model parameters. We use the MM estimates of the model parameters introduced in Ghosh and Sinha (2007) and the empirical Bayes estimates of μ and τ^2 to define the prior distribution on ψ, μ, and τ^2. To this end, we consider the prior distributions on the standard deviations, σ_e, σ_u, and σ_η

as the uniform distribution between zero and twice the GS estimates of the model parameters. For τ, we consider a uniform distribution between zero and twice the empirical Bayes estimate of τ obtained from Efron and Morris (1975). For b_0, b_1, and μ, we consider normal prior distributions with the means equal to GS estimates of b_0 and b_1 and the empirical Bayes estimate of μ, respectively, and the corresponding variances as 4, 4, and 9, respectively. It is worth mentioning that the simulation studies (not shown here) confirm that the hierarchical model is quite sensitive to non-informative prior distributions and also to the choice of hyper parameters. In order to address this problem, we consider the GS and EB estimates of the model parameters to define the priors. The densities of the estimated values of the true area-specific covariate using the Bayes, CB, EB, CEB, HB, and CHB methods are shown in Figure 3.5. As we expect, the CB, CEB, and the CHB methods lead to an ensemble of estimates with a histogram more similar to the true histogram of x_i's in comparison with the Bayes, EB, and HB methods, respectively. As Figure 3.5 shows, the CHB gives the closest ensemble of the estimates of the x_i's to the true area-specific covariate because the CHB makes use of the data twice - once in deriving the prior distribution using the GS estimates of the model parameters, and the second time in analyzing data.

We also evaluated the performance of different estimated values of the small area means using the performance measures obtained in Section 3.2. To this end, the Se, Sp, PPV, and NPV of the small area mean predictors are calculated for each simulation using the threshold of 80. We report the results

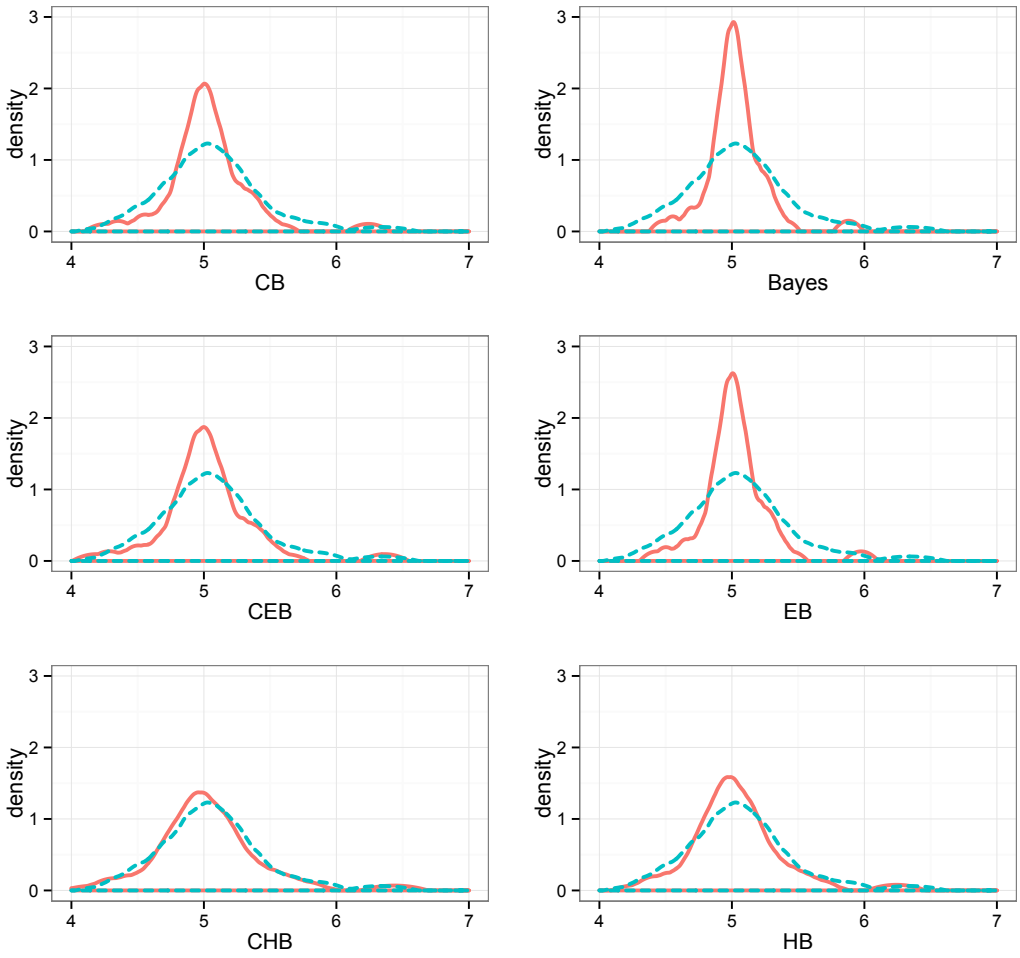

Figure 3.5: The density of the estimated values of the true area-specific covariate using different approaches (solid line) vs the density of the true area-specific covariate (dashed line)

averaged over $R = 5000$ simulations (Figure 3.6). The result illustrates that in terms of the Se and PPV measures, the CEB method performs the best. However, when it comes to the Sp, the HB method has the best performance while the CEB also shows a reasonable behaviour. The ML estimator shows

the worst performance in terms of the Se and PPV.

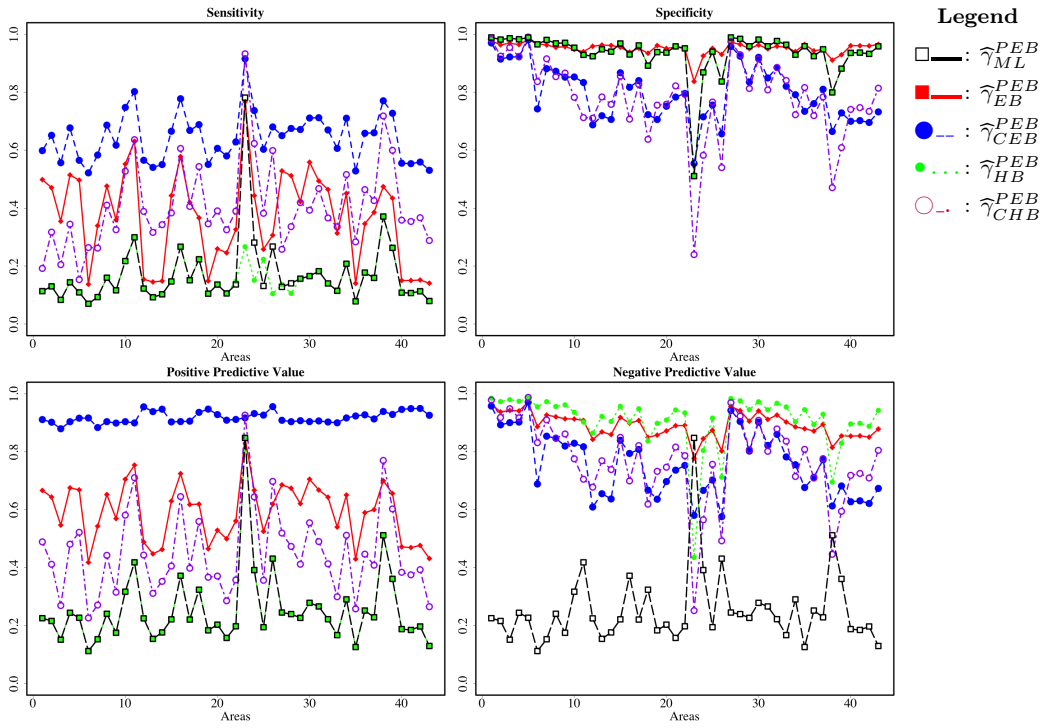

Figure 3.6: The simulated values of the measures of performance of different areas.

The histograms of the predicted values of the first nine small area means over $R = 5000$ simulations are shown in Figure 3.7. The small area means are mostly below the threshold 80 such that the probability of getting small area means over 80 is small using the model parameters in the simulation study. Note that we observe similar behaviour in other small areas as well.

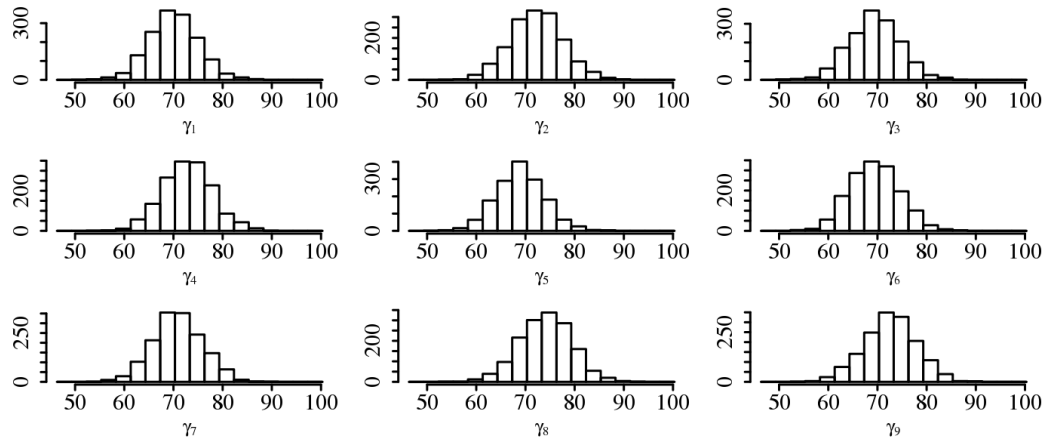

Figure 3.7: The histograms of the predicted small area means for the first nine areas

The MSPE of $\hat{\gamma}_{iCB}^{PB}$ is given by

$$\mathrm{E}(\hat{\gamma}_{iCB}^{PB} - \gamma_i)^2 = \mathrm{E}(\hat{\gamma}_{iCB}^{PB} - \hat{\gamma}_i^B)^2 + \mathrm{E}(\hat{\gamma}_i^B - \gamma_i)^2$$

$$= (f_i B_i b_1)^2 \left[\mathrm{E}(x_{iCB} - x_i)^2\right] + f_i^2 \left(\sigma_e^2 \left(\frac{(1-B_i)^2}{n_i} + \frac{1}{N_i - n_i}\right) + B_i^2 \sigma_u^2\right), \quad (3.5.1)$$

where (3.2.6) and (3.2.7) give $\mathrm{E}(x_{iCB})$ and $var(x_{iCB})$, respectively. Furthermore, Datta et al. (2010) obtained the $MSPE(\hat{\gamma}_{iML}^{PB})$ as

$$MSPE(\hat{\gamma}_{iML}^{PB}) = \frac{f_i^2 \sigma_e^2 (1 - A_i)}{n_i} + \frac{1}{N_i} f_i \sigma_e^2,$$

where $A_i = \sigma_e^2/(\sigma_e^2 + n_i \sigma_u^2 + b_1^2 \sigma_\eta^2)$. The $MSPE(\hat{\gamma}_{iB}^{PB})$ is also given by

$$\mathrm{E}(\hat{\gamma}_{iB}^{PB} - \gamma_i)^2 = \mathrm{E}(\hat{\gamma}_{iB}^{PB} - \hat{\gamma}_i^B)^2 + \mathrm{E}(\hat{\gamma}_i^B - \gamma_i)^2$$

$$= (f_i B_i b_1)^2 \left[\mathrm{E}(x_{iB} - x_i)^2 + f_i^2 \left(\sigma_e^2 \left(\frac{(1-B_i)^2}{n_i} + \frac{1}{N_i - n_i}\right) + B_i^2 \sigma_u^2\right), \quad (3.5.2)\right.$$

where $\mathrm{E}(x_{iB}) = C_i \mu + (1 - C_i) x_i$ and $var(x_{iB}) = (1 - C_i)^2 \sigma_i^2$. In order to evaluate

$MSPE(\hat{\gamma}_{iCB}^{PB})$, $MSPE(\hat{\gamma}_{iB}^{PB})$, and also $MSPE(\hat{\gamma}_{iML}^{PB})$, true values of the model parameters are used. Table 3.2 gives the MSPE of $\hat{\gamma}_{iB}^{PB}$, $\hat{\gamma}_{iCB}^{PB}$, $\hat{\gamma}_{iML}^{PB}$. As Table 3.2 shows, $\hat{\gamma}_{iB}^{PB}$ and $\hat{\gamma}_{iCB}^{PB}$ perform better than $\hat{\gamma}_{iML}^{PB}$ especially when the sample size is small. This happens because (2.1.6) and (3.1.2) use information from other areas to estimate the true area-specific covariate. In some areas, the MSPE of PEB estimators are smaller than the corresponding MSPE of the PB estimators. This is because there are cross-product terms involved in the MSPE of these estimators which are not negligible and indeed they are negative for some areas. We have planned to further investigate this issue in a separate research project when we deal with the MSPE estimation of the predictors of small area means.

Similar to Torkashvand et al. (2015), we also assess the performance of predictors of small area means in terms of the empirical MSPE (EMSPE). The EMSPE of $\hat{\gamma}_i^{PEB}$ for different methods ($\hat{\gamma}_{iEB}^{PEB}$, $\hat{\gamma}_{iCEB}^{PEB}$, $\hat{\gamma}_{iHB}^{PEB}$, $\hat{\gamma}_{iCHB}^{PEB}$, and $\hat{\gamma}_{iML}^{PEB}$) is defined as

$$EMSPE(\hat{\gamma}_i^{PEB}) = \frac{1}{R}\sum_{r=1}^{R}(\hat{\gamma}_i^{PEB(r)} - \gamma_i^{(r)})^2.$$

In Torkashvand et al. (2015), the superiority of $\hat{\gamma}_{iEB}^{PEB}$, the PEB predictor of the small area mean based on the JS estimate of the true area-specific covariate, in terms of the EMSPE over the $\hat{\gamma}_{iML}^{PEB}$ and some competitive predictors of the small area mean was shown (see Chapter 2). Table 3.2 gives the EMSPE of $\hat{\gamma}_{iEB}^{PEB}$, $\hat{\gamma}_{iCEB}^{PEB}$, $\hat{\gamma}_{iHB}^{PEB}$, $\hat{\gamma}_{iCHB}^{PEB}$, and $\hat{\gamma}_{iML}^{PEB}$. Our findings indicate that the PEB of

small area means based on ML estimates of x_i's has the largest EMSPE for the small areas with one, two, or three sample units while $\hat{\gamma}_{iEB}^{PEB}$ gives the minimum one in almost all small areas. The $\hat{\gamma}_{iCEB}^{PEB}$ and $\hat{\gamma}_{iCHB}^{PEB}$ show similar behaviour to $\hat{\gamma}_{iEB}^{PEB}$ and $\hat{\gamma}_{iHB}^{PEB}$, respectively, in terms of the EMSPE. As we expected the EMSPE of $\hat{\gamma}_{iCEB}^{PEB}$ and $\hat{\gamma}_{iCHB}^{PEB}$ are larger than $\hat{\gamma}_{iEB}^{PEB}$ and $\hat{\gamma}_{HB}^{PEB}$.

Table 3.2: Numerical values of the MSPE of $\hat{\gamma}_{iB}^{PB}$, $\hat{\gamma}_{iCB}^{PB}$ and $\hat{\gamma}_{iML}^{PB}$ and the EMSPE of $\hat{\gamma}_{iEB}^{PEB}$, $\hat{\gamma}_{iCEB}^{PEB}$, $\hat{\gamma}_{iHB}^{PEB}$, $\hat{\gamma}_{iCHB}^{PEB}$, and $\hat{\gamma}_{iML}^{PEB}$

		MSPE			*EMSPE*				
Area i	n_i	$\hat{\gamma}_{iB}^{PB}$	$\hat{\gamma}_{iCB}^{PB}$	$\hat{\gamma}_{iML}^{PB}$	$\hat{\gamma}_{iEB}^{PEB}$	$\hat{\gamma}_{iCEB}^{PEB}$	$\hat{\gamma}_{iHB}^{PEB}$	$\hat{\gamma}_{iCHB}^{PEB}$	$\hat{\gamma}_{iML}^{PEB}$
1	13	5.89	5.89	6.98	6.21	6.53	8.44	8.36	6.30
2	8	8.73	8.74	11.28	8.83	9.25	11.15	11.34	9.36
3	5	12.32	12.48	17.94	13.97	13.86	16.71	16.69	14.59
4	10	7.32	7.32	9.05	7.29	7.57	9.76	9.75	7.54
5	15	5.21	5.22	6.06	5.48	6.12	7.50	7.54	5.65
6	1	24.94	27.27	88.36	32.09	31.90	36.47	37.94	54.74
7	4	14.22	14.44	22.36	15.02	15.52	18.08	18.58	17.17
8	7	9.67	9.68	12.87	9.87	10.31	12.29	12.51	10.51
9	4	14.21	14.34	22.36	14.21	15.07	17.23	18.03	16.74
10	9	7.97	7.96	10.04	8.12	8.54	10.18	10.40	8.57
11	12	6.30	6.30	7.56	6.12	6.50	8.41	8.53	6.37
12	1	24.94	26.88	88.36	26.19	27.44	32.08	34.53	53.72
13	1	25.07	26.72	88.36	25.28	27.14	32.75	35.55	55.88
14	1	25.01	26.66	88.36	23.87	25.90	30.93	33.88	55.07
15	6	10.82	10.85	14.99	10.91	11.56	13.47	13.93	12.03
16	9	7.98	7.96	10.04	8.26	8.64	10.66	10.87	8.66
17	5	12.29	12.34	17.94	12.35	13.14	14.79	15.37	13.97
18	3	16.78	17.25	29.72	17.58	18.29	20.20	21.29	21.70
19	1	24.98	26.62	88.36	24.73	26.77	32.41	35.36	55.85
20	2	20.23	20.90	44.40	18.47	20.11	22.58	24.47	28.73
21	2	20.26	20.95	44.40	20.21	21.56	24.09	25.76	29.54
22	3	16.82	17.02	29.72	16.09	17.40	19.34	20.65	21.09
23	10	7.32	7.33	9.05	8.73	9.35	11.07	11.59	8.95
24	4	14.17	14.50	22.36	16.05	16.22	17.93	18.43	17.61
25	2	20.29	20.91	44.40	19.34	21.01	23.98	25.91	29.80
26	2	20.22	22.37	44.40	30.95	28.45	30.26	30.06	31.79
27	13	5.89	5.89	6.98	6.08	6.40	8.26	8.21	6.23
28	10	7.32	7.32	9.05	7.67	8.05	9.90	9.93	8.00
29	5	12.31	12.34	17.94	12.69	13.52	15.19	15.90	14.45
30	12	6.30	6.30	7.56	6.38	6.65	8.68	8.61	6.57
31	7	9.68	9.68	12.87	9.48	10.07	11.66	12.08	10.32
32	7	9.68	9.68	12.87	10.03	10.55	12.65	12.92	10.77
33	3	16.76	17.06	29.72	16.96	18.03	20.02	21.13	21.43
34	5	12.31	12.36	17.94	12.78	13.51	15.37	15.98	14.40
35	1	24.97	26.93	88.36	29.80	30.66	36.00	38.10	55.87
36	3	14.23	14.31	22.36	17.10	18.39	20.48	21.85	22.25
37	4	14.23	14.31	22.36	13.80	14.78	16.82	17.67	16.44
38	4	14.20	14.77	22.36	17.42	16.67	19.02	19.06	17.92
39	4	14.22	14.46	22.36	15.04	15.45	17.47	18.18	17.06
40	1	24.97	26.73	88.36	25.90	27.71	32.33	35.08	55.55
41	1	24.98	26.71	88.36	24.51	26.17	30.69	33.33	53.12
42	1	25.02	26.78	88.36	24.83	26.45	31.14	33.80	53.42
43	1	24.98	26.85	88.36	28.32	29.27	34.26	36.45	54.42

Chapter 4

Clustering in Small Area Estimation with Area-Level Linear Mixed Models

Recently, Datta et al. (2011) and Datta and Mandal (2015) studied whether or not the presence of the random effects in the small area estimation is necessary, especially, when the main concern is the prediction of small area means. They developed methodologies in order to evaluate the presence of the random effects in the area-level small area model based on both frequentist and Bayesian approaches, respectively.

Datta et al. (2011) introduced a statistical test to decide whether the inclusion of the random effect in the small area models is necessary. They studied that including the random effects in the model decreases the rate of convergence to the true value of the parameter in each area. The decrease is significant specially when the sample size in areas, n_i's, are large. They also pointed out that dropping the random effects can lead to more accurate

point and/or interval estimators, although the flexibility and adaptivity of the area-level (also called Fay-Herriot) model might be lost. The disadvantage of this methodology is that the random effects will be eliminated from all areas while it might be necessary to keep it for some areas. Datta and Mandal (2015) considered a Bayesian view towards the presence of the random effects in the area-level small area estimation. Using the spike and slab prior distribution for the area-level random effect, they were able to investigate the inclusion of the random effects in each area. In particular, they assumed that the random effects follow a non-degenerate and unique distribution with probability p in each small area while it is absent from the modelling with probability $1 - p$.

In this chapter, clustering small areas using the hierarchical clustering technique based on the Euclidean distance between their corresponding covariates is proposed to construct different groups of small areas where inside each group, areas are homogeneous and areas from different groups are non-homogenous. We start with an assumption that the random effect in each cluster follows a Normal distribution with a different variance component to give more flexibility to the behaviour of the random effects. We introduce a test statistic to test the null hypothesis of the equality of variance components of the Normal distributions. If the null hypothesis is rejected, we implement a modified version of the Tukey's method (Tukey, 1949) to combine some clusters. We assess the effect of different distributions of random effects on the precision of the small area means predictors using the MSPE and study situations where the proposed methodology results in more reliable predictors of small area means.

To this end, in Section 4.1, we review the general area-level model and study clustering in small area estimation. In Section 4.2, we introduce a test statistic in order to evaluate the assumption of homogeneity of variance components and prove some of its properties. Moreover, we show how a modification of Tukey's method can be used to combine some clusters. Using the new distributional form of the random effects, we find the EBLUP of small area means in Section 4.3. We also obtain an approximation to the MSPE of the EBLUP and derive an unbiased estimator of the MSPE up to a second order of approximation. In Section 4.4, we use the National Health and Nutrition Examination Survey (NHANES) for 2011-2012 to evaluate our proposed methodolgy. Implementing the simulation studies, we assess the performance of our proposed test statistic under different scenarios in Section 4.5. The precision of our proposed approach in predicting the small area means in terms of the MSPE is also assessed in this section. In addition, we study the relative bias (RB) of the estimator of MSPE.

4.1. Clustering in Area-Level Small Area Model

Consider the commonly used area-level regression model (1.1.1). Random effects u_i's in model (1.1.1) explain the lack of information provided by covariates about small area means, $\theta_i = \mathbf{X}_i\beta + u_i$, for $i = 1, \ldots, m$. In (1.1.1), it is usually assumed that the variance component, σ_u^2, is the same for all areas. The magnitude of the random effect part, u_i, depends on how good the

covariates explain small area means. In many applications, however, one might expect random effects for areas with similar covariates in terms of the Euclidian distance, $\|\mathbf{X}_i - \mathbf{X}_j\|_2 = \sqrt{\sum_{t=1}^{p}(X_{it} - X_{jt})^2}$, to show similar behaviour compared with random effects associated with other areas. This motivates to use clustering to form different groups that contain similar small areas. So, using the hierarchical clustering approach, we put small areas into different clusters, C_l, $l = 1, \ldots, k$, such that each cluster contains the most similar small areas in terms of the Euclidian distance between the values of their corresponding covariates. In this chapter, we also suggest to use a different variance component, $\sigma^2_{u_l}$, for $l = 1, \ldots, k$, in different clusters. In other words, we assume that $u_{j_l} \overset{i.i.d}{\sim} N(0, \sigma^2_{u_l})$ for $j_l \in C_l$, the l'th cluster, and $j_l = 1, \ldots, n_{c_l}$, where n_{c_l} is the number of small areas in the l'th cluster. Under this setting, model (1.1.1) can be represented as follows

$$y_{j_l} = \mathbf{X}_{j_l}\beta + u_{j_l} + e_{j_l}, \quad j_l = 1, \ldots, n_{c_l},\ l = 1, \ldots, k, \tag{4.1.1}$$

where y_{j_l} and $\mathbf{X}_{j_l}$ are the response variable and the covariate vector associated with the j'th element in the l'th cluster, respectively. Also, $e_{j_l} \sim N(0, D_{j_l})$ with known D_{j_l}'s. Throughout the chapter, we call this setting as the complete clustering approach where an optimal number of clusters, say k, is chosen such that there is no more significant changes in the distance between clusters for $k^* > k$. This makes the corresponding Fay-Herriot model more flexible in order to catch the true behaviour of the random effects.

Although complete clustering separates small areas into more homogeneous

clusters, still variance components might be equal. If this happens, then (4.1.1) reduces to (1.1.1), and we refer to this as the simple clustering. In Section 4.2, we test the assumption of equality of the variance components in different clusters by introducing a test statistic and showing some of its asymptotic properties. We might also have a situation where variance components are equal for some clusters, but not all. In this case, we combine clusters with equal variance components and the method will be referred to as the combined clustering approach. The MSPE of the simple, combined and complete methods will be calculated in Section 4.5. As shown in Section 4.5, the complete and combined clustering approaches lead to more precise prediction of $\theta_{j_l} = \mathbf{X}_{j_l}\beta + u_{j_l}$, for $j_l = 1, \ldots, n_{c_l}$ and $l = 1, \ldots, k$ in terms of the MSPE compared with the one based on the usual area-level small area model (1.1.1).

4.2. Testing the Equality of Variance Components

In this section, we introduce a test statistic to test the following hypothesis regarding the equality of variance components in the model (4.1.1)

$$H_0 : \sigma^2_{u_1} = \sigma^2_{u_2} = \cdots = \sigma^2_{u_k}, \quad vs. \quad H_a : \sim H_0.$$

To this end, because of the difference between the variance components and sampling variances under the model (4.1.1), the Kolmogrov's strong law of large numbers is used in order to define the test statistic. Due to the complexity that the weighted least square estimate of β introduces into the method of moments estimates of variance components, we use the ordinary least

square (OLS) to estimate β, where the consistency of $\hat{\beta}_{OLS}$ under the model (4.1.1) is shown in Lemma 2. Using $\hat{\beta}_{OLS}$, the modified method of moments estimators of variance components (MMM) are introduced in Theorem 4. The asymptotic distributions of these estimators are also found, and the test statistic is constructed accordingly.

Lemma 2. *In model (4.1.1), let* $\mathbf{X} = (\mathbf{X}_1', \ldots, \mathbf{X}_m')'$ *and assume that* $\mathbf{X}'\mathbf{X}$ *is full-rank, the columns of* $\mathbf{X}$ *are independent, and the covariate matrix of* $\mathbf{X}$ *is independent of* u_{j_l}*'s and* e_{j_l}*'s. Then, the OLS estimator of the regression coefficient,* $\hat{\beta}_{OLS}$*, is a consistent estimator of* β*.*

Proof. Note that

$$\hat{\beta}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y},$$

where $\mathbf{y} = (y_1, \ldots, y_m)$. First, we rewrite $\hat{\beta}_{OLS}$ as follows

$$\hat{\beta}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \delta)$$

$$= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\delta,$$

where $\delta = (u_1 + e_1, \ldots, u_m + e_m)'$. Now, $\hat{\beta}_{OLS}$ is a consistent estimator of β, if

$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\delta \xrightarrow{p} 0$ as $m \longrightarrow \infty$. To show this, note that

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\delta = (\tfrac{1}{m}\mathbf{X}'\mathbf{X})^{-1}(\tfrac{1}{m}\mathbf{X}'\delta)$$

$$= (\tfrac{1}{m}\mathbf{X}'\mathbf{X})^{-1}\begin{bmatrix} \frac{1}{m}\sum_{l=1}^{k}\sum_{j_l=1}^{n_{c_l}}\delta_{j_l} \\ \frac{1}{m}\sum_{l=1}^{k}\sum_{j_l=1}^{n_{c_l}}X_{j_l1}\delta_{j_l} \\ \vdots \\ \frac{1}{m}\sum_{l=1}^{k}\sum_{j_l=1}^{n_{c_l}}X_{j_lp}\delta_{j_l} \end{bmatrix}.$$

Let $S_{j_li} = X_{j_l}\delta_{j_l}$ and $i = 1, \ldots, p$, and assume that $X_{j_li} \sim X_i$, for $j_l = 1, \ldots, n_{c_l}$, $l = 1, \ldots, k$, where $\mathrm{E}(X_i) = \mu_i$ and $var(X_i) = \sigma_i^2$. Since $\mathbf{X}$ and δ are independent, we have

$$\mathrm{E}(S_{j_li}) = \mathrm{E}(X_{j_li})\mathrm{E}(\delta_{j_l}) = 0,$$

and,

$$var(S_{j_li}) = var(\mathrm{E}(S_{j_li}|X_{j_li})) + \mathrm{E}(var(S_{j_li}|X_{j_li}))$$

$$= 0 + \mathrm{E}(X_{j_li}^2(\sigma_{u_l}^2 + D_{j_l}))$$

$$= (\mu_i^2 + \sigma_i^2)(\sigma_{u_l}^2 + D_{j_l}).$$

On the other hand, $\exists M > 0$ such that $\sigma_{u_l}^2 + D_j \leq M$ for $j_l = 1, \ldots, n_{c_l}$ and $l = 1, \ldots, k$. One can easily show that

$$\tfrac{1}{m^2}\sum_{l=1}^{k}\sum_{j_l=1}^{n_{c_l}} var(S_{j_li}) = \tfrac{1}{m^2}(\mu_i^2 + \sigma_i^2)(\sum_{l=1}^{k} n_{c_l}\sigma_{u_l}^2 + \sum_{l=1}^{k}\sum_{j_l=1}^{n_{c_l}} D_{j_l}) < \infty,$$

Using the Kolmogrov's strong law of large numbers, we have

$$\tfrac{1}{m}\sum_{l=1}^{k}\sum_{j_l=1}^{n_{c_l}} S_{j_li} \xrightarrow{a.s.} 0.$$

This implies $\frac{1}{m}\sum_{l=1}^{k}\sum_{j_l=1}^{n_{c_l}} X_{j_l i}\delta_{j_l} \xrightarrow{p} 0$ as $m \longrightarrow \infty$. Using similar arguments, it is easy to show $(\frac{1}{m}\mathbf{X}'\mathbf{X})^{-1} \xrightarrow{p} constant$ as $m \longrightarrow \infty$, which completes the proof.

Theorem 4. *Let* $\hat{\sigma}^2_{u_l} = \frac{1}{n_{c_l}}\sum_{j_l=1}^{n_{c_l}}\left((y_{j_l} - \mathbf{X}_{j_l}\hat{\beta}_{OLS})^2 - D_{j_l}\right)$ *and assume* $n_{c_l} \longrightarrow \infty$ *as* $m \longrightarrow \infty$*, for* $l = 1, \ldots, k$*. Under the assumptions of Lemma 2 and the general model (4.1.1) the asymptotic distribution of* $\hat{\sigma}^2_{u_l}$ *as* $m \longrightarrow \infty$ *is given by*

$$\hat{\sigma}^2_{u_l} \sim N\left(\sigma^2_{u_l}, \frac{2}{n^2_{c_l}}\sum_{j_l=1}^{n_{c_l}}(\sigma^2_{u_l} + D_{j_l})^2\right), \quad l = 1, \ldots, k. \tag{4.2.1}$$

Proof. Note that

$$y_{j_l} \sim N(\mathbf{X}_{j_l}\beta, \sigma^2_{u_l} + D_{j_l}), \tag{4.2.2}$$

for $j_l = 1, \ldots, n_{c_l}$, $l = 1, \ldots, k$. Considering small areas that belong to the l'th cluster, we have

$$\mathrm{E}(y_{j_l} - \mathbf{X}_{j_l}\beta)^2 = \sigma^2_{u_l} + D_{j_l},$$

for $j_l = 1, \ldots, n_{c_l}$, $l = 1, \ldots, k$ which leads to the following estimator of $\sigma^2_{u_l}$

$$\hat{\sigma}^2_{u_l} = \frac{1}{n_{c_l}}\sum_{j_l=1}^{n_{c_l}}[(y_{j_l} - \mathbf{X}_{j_l}\beta)^2 - D_{j_l}].$$

Let $Z_{j_l} = (y_{j_l} - \mathbf{X}_{j_l}\beta)^2 - D_{j_l}$ for $j_l = 1, \ldots, n_{c_l}$, $l = 1, \ldots, k$. It is easy to show that

$$\mathrm{E}(Z_{j_l}) = \sigma^2_{u_l} \quad \text{and} \quad var(Z_{j_l}) = 2(\sigma^2_{u_l} + D_{j_l})^2.$$

As Z_{j_l}'s are not identically distributed, in order to find the asymptotic distribution of $\hat{\sigma}^2_{u_l}$, we check the Lindeberg's condition (Billingsley, 2008). Let

$s^2_{n_{c_l}} = \sum_{j_l=1}^{n_{c_l}} var(Z_{j_l})$. The interest is to show the following

$$\lim_{n_{c_l}\longrightarrow\infty} \frac{1}{s^2_{n_{c_l}}} \sum_{j_l=1}^{n_{c_l}} \mathrm{E}\left((Z_{j_l} - \sigma^2_{u_l})^2 \mathbf{1}_{|Z_{j_l}-\sigma^2_{u_l}|>\epsilon s_{n_{c_l}}}\right) = 0, \tag{4.2.3}$$

where $\epsilon > 0$ and $\mathbf{1}$ is the indicator function. To this end, we first expand the expectation term. Let $t = (y_{j_l} - \mathbf{X}_{j_l}\beta)$. Based on (4.2.2), we calculate $\mathrm{E}\left((Z_{j_l} - \sigma^2_{u_l})^2 \mathbf{1}_{|Z_{j_l}-\sigma^2_{u_l}|>\epsilon s_{n_{c_l}}}\right)$ as

$$\begin{aligned}
&2\int_{\sqrt{\epsilon s_{n_{c_l}}}+\sigma^2_{u_l}-\mathbf{X}_{j_l}\beta}^{\infty} \frac{\left(t^2 - D_{j_l} - \sigma^2_{u_l}\right)^2}{\sqrt{2\pi(\sigma^2_{u_l} + D_{j_l})}} \exp\left(-\frac{t^2}{2(\sigma^2_{u_l} + D_{j_l})}\right) dt \\
&= 2\int_{\sqrt{\epsilon s_{n_{c_l}}}+\sigma^2_{u_l}-\mathbf{X}_{j_l}\beta}^{\infty} t^4 \frac{1}{\sqrt{2\pi(\sigma^2_{u_l} + D_{j_l})}} \exp\left(-\frac{t^2}{2(\sigma^2_{u_l} + D_{j_l})}\right) dt \\
&- 4(\sigma^2_{u_l} + D_{j_l}) \int_{\sqrt{\epsilon s_{n_{c_l}}}+\sigma^2_{u_l}-\mathbf{X}_{j_l}\beta}^{\infty} t^2 \frac{1}{\sqrt{2\pi(\sigma^2_{u_l} + D_{j_l})}} \exp\left(-\frac{t^2}{2(\sigma^2_{u_l} + D_{j_l})}\right) dt \\
&+ (\sigma^2_{u_l} + D_{j_l})^2 \int_{\sqrt{\epsilon s_{n_{c_l}}}+\sigma^2_{u_l}-\mathbf{X}_{j_l}\beta}^{\infty} \frac{1}{\sqrt{2\pi(\sigma^2_{u_l} + D_{j_l})}} \exp\left(-\frac{t^2}{2(\sigma^2_{u_l} + D_{j_l})}\right) dt \\
&= 2\left[-t^3\sqrt{\frac{(\sigma^2_{u_l} + D_{j_l})}{2\pi}} \exp\left(-\frac{t^2}{2(\sigma^2_{u_l} + D_{j_l})}\right)\right]_{\sqrt{\epsilon s_{n_{c_l}}}+\sigma^2_{u_l}-\mathbf{X}_{j_l}\beta}^{\infty} \\
&+ 2(\sigma^2_{u_l} + D_{j_l})\left[-t\sqrt{\frac{(\sigma^2_{u_l} + D_{j_l})}{2\pi}} \exp\left(-\frac{t^2}{2(\sigma^2_{u_l} + D_{j_l})}\right)\right]_{\sqrt{\epsilon s_{n_{c_l}}}+\sigma^2_{u_l}-\mathbf{X}_{j_l}\beta}^{\infty} \\
&+ 3(\sigma^2_{u_l} + D_{j_l})^2(1 - \Phi(\sqrt{\epsilon s_{n_{c_l}}} + \sigma^2_{u_l} - \mathbf{X}_{j_l}\beta)).
\end{aligned} \tag{4.2.4}$$

As it was mentioned, D_{j_l}'s and $\sigma^2_{u_l}$ are bounded in the small area estimation

for $j_l = 1, \ldots, n_{c_l}$ and $l = 1, \ldots, k$. Let

$$M_0 = \max\{D_{j_l} \text{ and } \sigma^2_{u_l}; j_l = 1, \ldots, n_{c_l}, l = 1, \ldots, k\}.$$

So, (4.2.3) is less than

$$\lim_{n_{c_l} \longrightarrow \infty} \frac{1}{s^2_{n_{c_l}}} \left(\exp(-\tfrac{1}{4M_0}\epsilon s_{n_{c_l}}) \left[\sum_{j_l=1}^{n_{c_l}} \exp\left((\sigma^2_{u_l} - \mathbf{X}_{j_l}\beta)^2 \right.\right.\right.$$
$$\left. + 2\sqrt{\epsilon s_{n_{c_l}}}(\sigma^2_{u_l} - \mathbf{X}_{j_l}\beta) \right) \left(\sqrt{\tfrac{M_0}{\pi}} \left(\sqrt{\epsilon s_{n_{c_l}}} + \sigma^2_{u_l} - \mathbf{X}_{j_l}\beta \right)^3 \right.$$
$$\left.\left.\left. + 4\sqrt{\frac{M_0^3}{2\pi}} \left(\sqrt{\epsilon s_{n_{c_l}}} + \sigma^2_{u_l} - \mathbf{X}_{j_l}\beta \right) \right) \right] \right)$$
$$+ \lim_{n_{c_l} \longrightarrow \infty} \frac{1}{s^2_{n_{c_l}}} 12 M_0^2 \sum_{j_l=1}^{n_{c_l}} \left(1 - \Phi(\sqrt{\epsilon s_{n_{c_l}}} + \sigma^2_{u_l} - \mathbf{X}_{j_l}\beta) \right). \tag{4.2.5}$$

Noting that $s_{n_{c_l}} \longrightarrow \infty$ as $n_{c_l} \longrightarrow \infty$, (4.2.5) goes to zero. Thus, (4.2.3) holds. Since $\hat{\sigma}^2_{u_l} = \frac{1}{n_{c_l}} \sum_{j_l=1}^{n_{c_l}} Z_{j_l}$, the asymptotic distribution of $\hat{\sigma}^2_{u_l}$ easily obtained as follows

$$\hat{\sigma}^2_{u_l} \sim N\left(\sigma^2_{u_l}, \frac{2}{n^2_{c_l}} \sum_{j_l=1}^{n_{c_l}} (\sigma^2_{u_l} + D_{j_l})^2 \right),$$

as $n_{c_l} \longrightarrow \infty$.

Remark 3. *It is worth noting the estimator of σ^2_u in Theorem 4 is asymptotically unbiased. Due to the positive nature of the variance component, $\sigma^2_{u_l}$, for $l = 1, \ldots, k$, the interest lies in getting positive estimates of the variance component. As the underlying distribution in (4.2.1) is a Normal distribution, it is likely that $\hat{\sigma}^2_u$ becomes negative. However, one can easily show that for large n_{c_l} (where $m \longrightarrow \infty$) the probability of observing a negative estimate is*

negligible, i.e.,

$$\Phi\Big(-\frac{\sigma_{u_l}^2 n_{c_l}}{\sqrt{2\sum_{j_l=1}^{n_{c_l}}(\sigma_{u_l}^2 + D_{j_l})^2}}\Big) \longrightarrow 0, \tag{4.2.6}$$

provided that $\sigma_{u_l}^2$ *and* D_{j_l}*, for* $j_l = 1, \ldots, n_{c_l}$ *and* $l = 1, \ldots, k$*, are bounded.*

Let $\hat{\sigma}_u^2 = (\hat{\sigma}_{u_1}^2, \ldots, \hat{\sigma}_{u_k}^2)$. The null hypothesis, $H_0 : \sigma_{u_1}^2 = \sigma_{u_2}^2 = \cdots = \sigma_{u_k}^2 = \sigma_u^2$, implies

$$\hat{\sigma}_u^2 \sim N_k\left(\sigma_0, \Sigma_0\right), \tag{4.2.7}$$

where $\sigma_0 = (\sigma_u^2, \ldots, \sigma_u^2)$ and

$$\Sigma_0 = \operatorname{diag}\left(\tfrac{2}{n_{c_1}^2}\sum_{j_1=1}^{n_{c_1}}(\sigma_u^2 + D_{j_1})^2, \ldots, \tfrac{2}{n_{c_k}^2}\sum_{j_k=1}^{n_{c_k}}(\sigma_u^2 + D_{j_k})^2\right).$$

Thus,

$$(\hat{\sigma}_u^2 - \sigma_0)'\Sigma_0^{-1}(\hat{\sigma}_u^2 - \sigma_0) \sim \chi_{k-p-1}^2, \tag{4.2.8}$$

and the P-value of the test is approximately equal to

$$\text{P-value} = \mathbb{P}(\chi_{k-p-1}^2 > \chi_0^2), \tag{4.2.9}$$

where $\chi_0^2 = (\hat{\sigma}_u^2 - \hat{\sigma}_0)'\hat{\Sigma}_0^{-1}(\hat{\sigma}_u^2 - \hat{\sigma}_0)$, $\hat{\sigma}_0 = (\hat{\sigma}_{0u}^2, \ldots, \hat{\sigma}_{0u}^2)$, $\hat{\sigma}_{0u}^2 = \frac{1}{k}\sum_{l=1}^{k}\hat{\sigma}_{u_l}^2$, and

$$\hat{\Sigma}_0 = \operatorname{diag}\left(\tfrac{2}{n_{c_1}^2}\sum_{j_1=1}^{n_{c_1}}(\hat{\sigma}_{0u}^2 + D_{j_1})^2, \ldots, \tfrac{2}{n_{c_k}^2}\sum_{j_k=1}^{n_{c_k}}(\hat{\sigma}_{0u}^2 + D_{j_k})^2\right).$$

In Section 4.5, we evaluate the performance of (4.2.8) under different scenarios. The results indicate that the test has a large power. However, the simulated values of the type I error are larger than the significance level. Note that the distribution of (4.2.1) highly depends on the number of small

areas in each cluster. For larger clusters, it is expected to get more accurate estimates of variance components. Due to Theorem 4, estimated values might be significantly different from estimates obtained from smaller clusters. In other words, the difference between the information provided from clusters based on different small areas might result in the rejection of the null hypothesis even though it is correct, which explains inflated observed type I error in our simulation studies.

If the null hypothesis of the equality of variance components is rejected, there might exist some clusters that have the same variance components. In order to estimate fewer number of parameters, one might consider combining such clusters. This is similar to what happens after rejecting the null hypothesis in the ANOVA context. Tukey (1949) proposed a solution to this problem by using the T-test. In this chapter, we modify his approach to combine the clusters with the same variance components. To this end, we take the following steps

- We first sort the MMM estimates of the variance components from the minimum to the maximum.
- Considering (4.2.1) and conducting the T-test under the unequal variance set-up and using the same significance level as in (4.2.9), we make groups of clusters. To this end, starting from the cluster with smallest variance estimate, we compare it with the one with the second smallest variance estimates. If the null hypothesis of the equality of the related variance

components (using (4.2.1)) is not rejected, we make a new group consisting of corresponding clusters. Otherwise, we keep the cluster corresponding to the smallest variance component as a group with a single element. Then, the second smallest number is compared with the third smallest one. Similarly, if the null hypothesis of the equality is not rejected, we add the corresponding cluster to the group that the second cluster belongs to. This process stops if all groups are constructed by one or two clusters. If not, we go to the next step.

- For a group with the number of clusters larger than two, we find the maximum and the average values of the MMM estimates of the variance components, $\hat{\sigma}^2_{max}$ and $\hat{\sigma}^2_{mean}$, respectively. Depending on the number of clusters in each group, say k', we construct a test statistic W as follows

$$W = \frac{1}{3}\left(0.25 + \frac{1}{n_{c_{max}}}\right)^{-1}\left(\frac{\hat{\sigma}^2_{max}-\hat{\sigma}^2_{mean}}{var(\hat{\sigma}^2_{max}-\hat{\sigma}^2_{mean})} - 1.2\log_{10} k'\right), \quad k' > p+1,$$

$$W = \frac{1}{3}\left(0.25 + \frac{1}{n_{c_{max}}}\right)^{-1}\left(\frac{\hat{\sigma}^2_{max}-\hat{\sigma}^2_{mean}}{var(\hat{\sigma}^2_{max}-\hat{\sigma}^2_{mean})} - 0.5\log_{10} k'\right), \quad k' \leq p+1,$$

where $n_{c_{max}}$ is the number of small areas inside the cluster corresponding to $\hat{\sigma}^2_{max}$. Here

$$var(\hat{\sigma}^2_{max} - \hat{\sigma}^2_{mean}) = (1-\tfrac{1}{k'})^2 var(\hat{\sigma}^2_{max}) + (\tfrac{1}{k'})^2 \sum_{j \neq j_{max}} var(\hat{\sigma}^2_j)$$

where j_{max} is the index related to the $\hat{\sigma}^2_{max}$ and $var(\hat{\sigma}^2_j)$'s are obtained using the estimated values of the variance in (4.2.1).

The aim is to see whether or not we can split a group of clusters into smaller ones. If W is larger than the critical value of the standard Normal

distribution for the two-sided test with the level of significance in (4.2.9), we put the cluster corresponding to the maximum variance component into a new group. We repeat this step for the new maximum if $k' > 2$. If the new maximum is separated as well, we put it in the same group as the old one.

- When the number of clusters in a group remains the same in the previous step, depending on the number of clusters inside a group, we test the assumption of the equality of the variance components for the group with the size larger than two. If the size of the group is larger than $p+1$, we implement the test statistic proposed in this chapter. Otherwise, simultaneous T-tests with the same significance level as in (4.2.9) for a small p are conducted. If the null hypothesis is rejected, we split the group with an even number of clusters into subgroups of two clusters by starting from the smallest MMM estimate and moving forward to the largest one. For the odd number size, we let the last subgroup have three clusters and then test the assumption of the equality of the variance components one more time. If the null hypothesis is rejected, we make two new subgroups of two and one clusters starting from the cluster corresponding to the smallest MMM estimate.

After implementing the above algorithm, some of the clusters will be combined. Then, in order to estimate the MSPE, denoted by mspe, one needs to deal with fewer number of parameters. This results in less variability due to the

estimation of the model parameters into the estimation of the small area means and the corresponding MSPEs. In Section 4.5, it is shown that when there are clusters with the same variance component, the combined clustering has the same performance as the complete one in estimating small area means, but, significantly better than the simple clustering. However, the results of Section 4.5 indicate that if the difference between variance components is huge, the complete clustering performs better in estimating the small area means compared with combined and simple clustering approaches regardless of the number of parameters to be estimated.

In the following section the EBLUP of small area means are given. We also provide an approximation to the MSPE of the EBLUP and derive an estimator of MSPE of the EBLUP that is nearly unbiased up to the second order of approximation.

4.3. EBLUP and its MSPE Estimation

In this section, the BLUP of small area means and consequently, the EBLUP for the new distributional form of random effects are given. Following Henderson (1950), the BLUP of small area means is given by

$$\tilde{\theta}_{j_l} = \mathbf{X}_{j_l}\tilde{\beta} + \tilde{u}_{j_l}, \tag{4.3.1}$$

where $\tilde{u}_{j_l} = \mathbf{G}\mathbf{V}^{-1}(y_{j_l} - \mathbf{X}_{j_l}\tilde{\beta})$, $\tilde{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{y})$, $\mathbf{y} = (y_1, \ldots, y_m)$, $\mathbf{V} = \text{diag}(\mathbf{V}_1, \ldots, \mathbf{V}_k)$, $\mathbf{V}_l = \text{diag}(\sigma^2_{u_l}+D_{1_l}, \ldots, \sigma^2_{u_l}+D_{n_{c_l}})$, $\mathbf{G} = \text{diag}(\mathbf{G}_1, \ldots, \mathbf{G}_k)$,

$\mathbf{G}_l = \sigma^2_{u_l}\mathbf{I}_{n_{c_l}\times n_{c_l}}$, and $\mathbf{I}$ is the identity matrix for $j_l = 1,\ldots,n_{c_l}$ and $l = 1,\ldots,k$. The MSPE of $\tilde{\theta}_{j_l}$ can be written as follows

$$\begin{aligned}\text{MSPE}(\tilde{\theta}_{j_l}) &= \text{E}(\tilde{\theta}_{j_l} - \theta_{j_l})^2 \\ &= g_{1j_l}(\gamma) + g_{2j_l}(\gamma),\end{aligned}$$

where $\gamma = (\sigma^2_{u_1},\ldots,\sigma^2_{u_k})$, $g_{1j_l}(\gamma)$ is the j'th element of the l'th cluster on the diagonal of $\mathbf{G} - \mathbf{GV^{-1}G}$, $g_{2j_l}(\gamma) = \mathbf{d}'_{j_l}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{d}_{j_l}$, $\mathbf{d}'_{j_l} = \mathbf{X}_{j_l} - \mathbf{b}'_{j_l}\mathbf{X}_{j_l}$, and $\mathbf{b}'_{j_l}$ is the j'th row of the l'th cluster of $\mathbf{GV}^{-1}$.

Due to unknown γ, $\hat{\theta}_{j_l}$'s, the EBLUP of θ_{j_l}'s, are obtained for $j_l = 1,\ldots,n_{c_l}$ and $l = 1,\ldots,k$. To this end, the MMM estimates of variance components are used in formula (4.3.1). The MSPE of the EBLUP can be decomposed as

$$\begin{aligned}\text{MSPE}(\hat{\theta}_{j_l}) &= \text{E}(\hat{\theta}_{j_l} - \theta_{j_l})^2 \\ &= \text{E}(\tilde{\theta}_{j_l} - \theta_{j_l})^2 + \text{E}(\hat{\theta}_{j_l} - \tilde{\theta}_{j_l})^2 + 2\text{E}(\hat{\theta}_{j_l} - \tilde{\theta}_{j_l})(\tilde{\theta}_{j_l} - \theta_{j_l}).\end{aligned}$$

Under the normality assumption for random effects as well as the sampling error, the cross product term is zero (Rao and Molina, 2015). Therefore

$$\text{MSPE}(\hat{\theta}_{j_l}) = g_{1j_l}(\gamma) + g_{2j_l}(\gamma) + g_{3j_l}(\gamma),$$

where

$$g_{3j_l}(\gamma) = \text{trace}\left((\frac{\partial \mathbf{b}'_{j_l}}{\partial\gamma})\mathbf{V}(\frac{\partial \mathbf{b}'_{j_l}}{\partial\gamma})'\mathbf{var}(\gamma)\right), \tag{4.3.2}$$

$$\frac{\partial \mathbf{b}'_{j_l}}{\partial \gamma} = \begin{bmatrix} \frac{\partial \mathbf{b}'_{j_l 1}}{\partial \sigma^2_{u_1}} & \cdots & \frac{\partial \mathbf{b}'_{j_l m}}{\partial \sigma^2_{u_1}} \\ \frac{\partial \mathbf{b}'_{j_l 1}}{\partial \sigma^2_{u_2}} & \cdots & \frac{\partial \mathbf{b}'_{j_l m}}{\partial \sigma^2_{u_2}} \\ & \vdots & \\ \frac{\partial \mathbf{b}'_{j_l 1}}{\partial \sigma^2_{u_k}} & \cdots & \frac{\partial \mathbf{b}'_{j_l m}}{\partial \sigma^2_{u_k}} \end{bmatrix}_{k \times m},$$

and $var(\gamma) = \Sigma_0$.

Prasad and Rao (1990) gave the second order MSPE estimation of small area means as the measure of the variability of the EBLUP as follows

$$\text{mspe}(\hat{\theta}_{j_l}) \approx g_{1j_l}(\hat{\gamma}) + g_{2j_l}(\hat{\gamma}) + 2g_{3j_l}(\hat{\gamma}), \tag{4.3.3}$$

where $\hat{\gamma}$ is the MMM estimates of γ and $g_{1j_l}(\hat{\gamma})$, $g_{2j_l}(\hat{\gamma})$, $g_{3j_l}(\hat{\gamma})$, $\hat{\mathbf{d}}_{j_l}$, $\widehat{\partial \mathbf{b}'_{j_l}/\partial \gamma}$, and $\widehat{var(\gamma)} = \hat{\Sigma}_0$ are obtained by substituting the MMM estimates of variance components in their original definition. Let $\hat{\mathbf{V}}_l = \text{diag}(\hat{\sigma}^2_{u_l} + D_{1_l}, \ldots, \hat{\sigma}^2_{u_l} + D_{n_{c_l}})$ and $\hat{\mathbf{G}}_l = \hat{\sigma}^2_{u_l} \mathbf{I}_{n_{c_l} \times n_{c_l}}$, for $l = 1, \ldots, k$. Accordingly, let $\hat{\mathbf{V}} = \text{diag}(\hat{\mathbf{V}}_1, \ldots, \hat{\mathbf{V}}_k)$ and $\hat{\mathbf{G}} = \text{diag}(\hat{\mathbf{G}}_1, \ldots, \hat{\mathbf{G}}_k)$. Note that Prasad and Rao (1990) proposed (4.3.3) for the consistent estimators of γ. As the MMM estimates of the variance components are consistent, we use them for the purpose of data analysis and simulation studies.

It is worth mentioning that the magnitude of g_{3j_l} depends on the number of variance components in the model. So, the more parameters we have in the model, the larger g_{3j_l} will be. This indicates that by introducing more parameters into the model, we will get a larger MSPE because of the variability

induced by the estimation of variance components. To evaluate the performance of the estimator of MSPE, we use the relative bias (RB) as the measure of the precision.

4.4. Real Data Analysis

In this section, we use the National Health and Nutrition Examination Survey (NHANES) for 2011-2012 as a unit-level dataset to predict the waist circumference based on the Body Mass Index (BMI). Vague (1947) showed that there is a high association between the waist circumference and cardiovascular disease, type 2 diabetes, and also hypertension. He concluded that "apple shaped obesity" observed in men is a high-risk obesity while the "gynoid obesity", often found in women, has a lower risk.

The data is categorized to small domains based on age, gender, education, ethnic, and poverty. We consider the mean of the waist circumference of people belonging to a small area as the response variable, y_i, while the mean of their BMI's as the covariate, x_i. The units inside each small area are used to obtain the sampling variance, D_i. To this end, we propose the following method which is an extension of the method in Wang and Fuller (2003).

Suppose we have enough sample units, n_i, such that the regression line $y_{ij} = \beta_0 + \beta_1 x_{ij} + u_i + e_{ij},\ j = 1, \ldots, n_i,\ n_i > 1$ is estimable. Let

$$y_i = \beta_0 + \beta_1 x_i + u_i + e_i,$$

where $e_i = \frac{1}{n_i}\sum_{j=1}^{n_i} e_{ij}$, $x_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$, and $y_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$, and $i = 1, \ldots, 1362$. An unbiased estimate of the sampling variance inside the small area, $\hat{\sigma}^2_{e_i}$ is given by

$$\frac{1}{n_i - 1}\sum_{j=1}^{n_i}\left((y_{ij} - y_i)^2 - \hat{\beta}_1^2(x_{ij} - x_i)^2\right). \tag{4.4.1}$$

In order to fit a regression line in each area, at least two sample units are required. As there are small areas with one sample unit, we use the overall regression line obtained from the complete data to estimate $\hat{\sigma}^2_e$. Consequently, the sampling variance for the mean of the response variable in each area is given by

$$D_i = \frac{\hat{\sigma}^2_{e_i}}{n_i}. \tag{4.4.2}$$

D_i's ranges from 4.58×10^{-16} to 61.63. There are 519 small areas with $n_i = 1$ where the overall regression line is used to estimate $D_i = 61.63$.

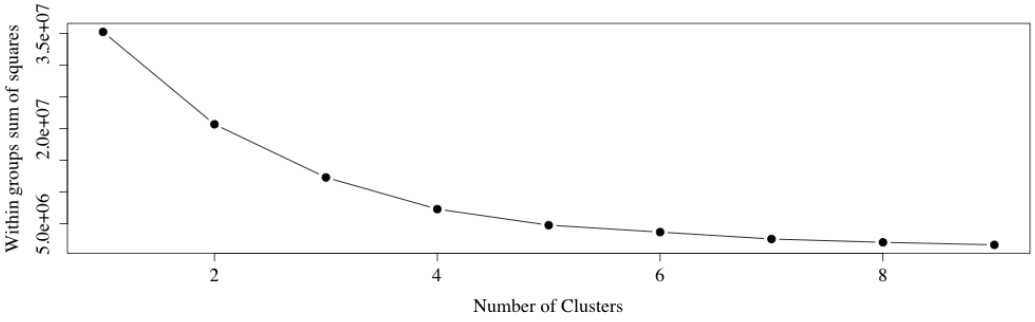

Figure 4.1: The effect of number of clusters on within groups sums of squares in the k-means clustering

There are a few approaches to cluster small areas. For example, the k-means clustering is an approach that minimizes the Euclidian distance of the covariate from the mean of the covariates belonging to the cluster (Hartigan and Wong, 1979). Another approach is to use the hierarchical clustering (Ward Jr, 1963) based on the Lance-Williams algorithm through the squared Euclidian distance between covariates. The Silhouette method (Rousseeuw, 1987) also gives the suitable number of clusters based on the average of the silhouettes (see (1.3.1)). It is worth mentioning that we do not look for the optimal number of clusters (i.e., the smallest number of clusters) as the aim is to have small areas inside a cluster as similar as possible. So, we choose the number of clusters such that the distance between clusters does not change significantly after that. Even if the number of clusters are far from the optimum value, using the Tukey's method, we can always combine clusters which are not significantly different. So, as Figure 4.1 shows, we cluster the small areas in 7 groups using the hierarchical clustering and model the data as

$$y_{j_l} = \beta_0 + \beta_1 x_{j_l} + u_{j_l},$$

where $j_l = 1, \ldots, n_{c_l}$, $l = 1, \ldots, 7$ and $n_c = (521, 333, 64, 252, 165, 9, 14, 4)$. The MMM estimates of the variance components are $(0.36, 2.87, 4.05, 1.76, 11.40, 420.77, 36.17)$. The assumption of the equality of the variance components is rejected with $\chi^2_4 = 329.4864$ with the corresponding p-value of zero. By implementing the Tukey's method, we have 5 combined clusters with the variance components $(0.36, 2.89, 11.40, 420.77, 36.17)$.

Using the complete clustering approach, we observe more than 49% improvements over the simple approach in the mspe in 70% of small areas. In 10.00% of small areas, the mspe of the simple clustering approach is up to 62.24 times more than the mspe of the complete clustering approach. On average we get the mspe of the simple clustering approach up to 12.43 times more than the complete clustering approach. After implementing the Tukey's method, there is over 75% improvement in 70% of small areas. The complete and combined clustering approaches have the same performance in over 80% of small areas in terms of the mspe. In order to compare the performance of the proposed methods with the simple clustering approach, we first calculate the mspe of the complete clustering, mspe_c, combined clustering, mspe_{cb}, and the simple approach, mspe_s. Then, $\frac{\text{mspe}_s}{\text{mspe}_c}$, $\frac{\text{mspe}_s}{\text{mspe}_{cb}}$, and $\frac{\text{mspe}_{cb}}{\text{mspe}_c}$ are calculated. An overal comparison of three methods is given in Table 4.1.

As the MMM estimates of the variance components are significantly different, the usage of either the complete or combined clustering schemes are justifiable. However, the MMM estimates of the variance components for three clusters, 2.87, 4.05, 1.76, are very close. Because of this similarity, the combined clustering scheme merges the corresponding clusters.

Figures 4.2 and 4.3 present the mspe and predicted values of small area means, respectively.

Table 4.1: The comparison between the complete, combined, and simple clustering in terms of the quantiles of their corresponding mspe

	complete over simple	combined over simple	complete over combined
Minimum	0.00	0.00	0.28
1-decile	0.86	1.00	0.90
2-decile	1.00	1.06	0.94
3-decile	1.49	1.75	0.98
4-decile	2.59	2.95	1.00
5-decile	3.06	3.14	1.00
6-decile	5.48	7.29	1.00
7-decile	10.95	11.08	1.00
8-decile	15.09	11.14	1.00
9-decile	62.24	62.22	1.17
Maximum	62.65	62.64	1.60
Mean	12.43	12.39	0.98

4.5. Simulation Studies

In this section, we consider different scenarios of the sampling variances, D_{j_l}'s, and the variance components, $\sigma^2_{u_l}$'s, for $j_l = 1, \ldots, n_{c_l}$ and $l = 1, \ldots, k$ to evaluate the performance of the test statistic. We also design simulation studies to see the performance of the proposed method in the reduction of the MSPE. To this end, the empirical MSPE (EMSPE) of small area means using different clustering schemes is calculated. Further, we evaluate the performance of the complete and combined clustering approaches on the estimation of MSPE,

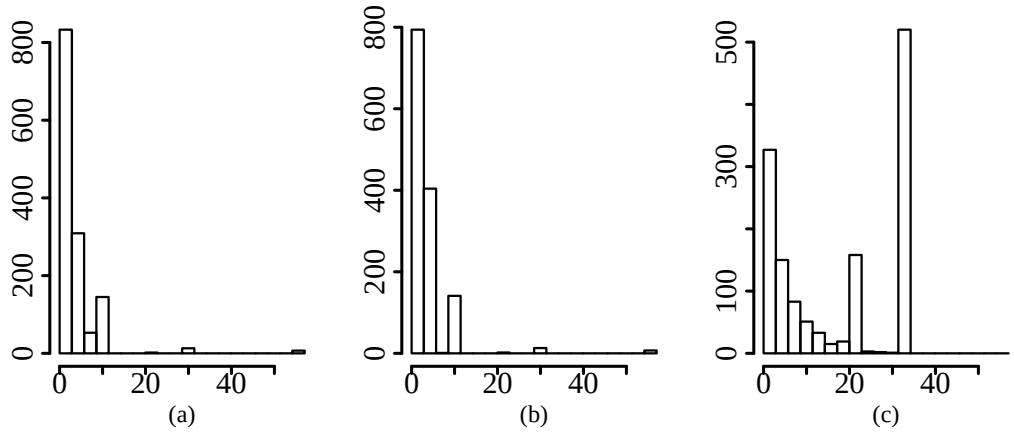

Figure 4.2: Histogram of the mspe's: (a) the complete clustering method, (b) the combined clustering method, (c) the simple method

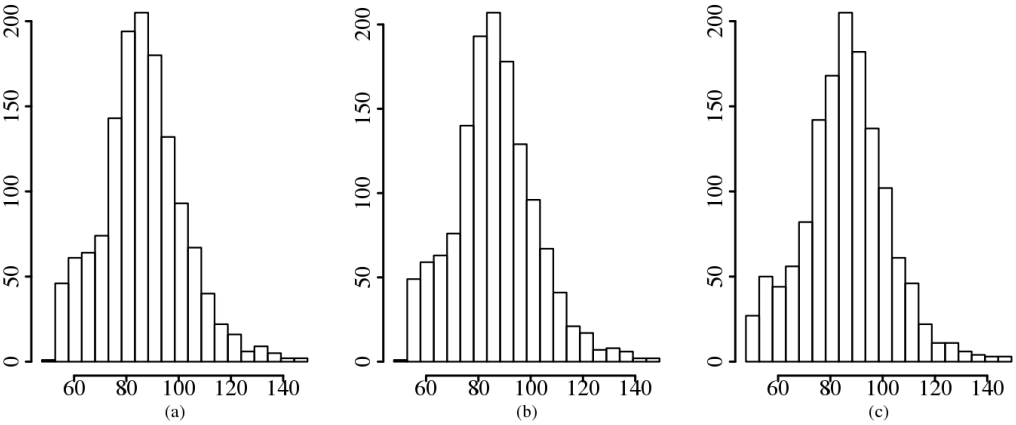

Figure 4.3: Histogram of predicted small area means: (a) the complete clustering method, (b) the combined clustering method, (c) the simple method

mspe, using RB. In Section 4.4, an unbiased estimator of the sampling variance inside the small area was proposed. We assess the performance of this estimator

using simulation studies.

4.5.1. Evaluating the test statistic and the effect of the complete and combined clustering on EMSPE and mspe

For the purpose of simulation studies, we use an area-level dataset regarding the prescription costs from Union Régionale des Caisses d'Assurance Maladie (URCAM) of the Midi-Pyrénées Region in the south-west of France, during the period January-December, 1999. The dataset consists of $m = 268$ cantons (a type of administrative division of a country) that are considered as small areas and the goal is to predict the average prescription cost in each area. In general cantons are relatively small in terms of population size when compared to counties, departments, or provinces. Because of the confidentiality problem and the privacy concerns, the data set is only available in the area-level format. This dataset has been used by Cressie et al. (2005, 2006) to assess the performance of their proposed estimators in the context of the spatial model. Kang et al. (2009) also considered this dataset in a spatial analysis to predict the average prescription amount in each canton.

In this work, we consider small area estimation to conduct the simulation studies using this dataset. The area-level small area model provides an appropriate link between different small areas in order to aggregate the information from all small areas to predict the small area mean. The idea is different from Kang et al. (2009) as in the spatial model, cantons are considered to be

dependent while in our set-up we assume that they are independent. Following Kang et al. (2009), we use the percentage of patients over 70 years as the covariate in each canton. We consider the average prescription amount in each canton as the response variable.

Cantons can be clustered based on the percentage of patients over 70 years. We expect patients in similar aging groups to have reasonably similar prescription costs as they need similar number of follow-up visits, etc. Another important factor is the type of medication that is taken by the patient. Apparently, people in similar aging groups need similar supplementary treatments (Speros, 2009).

Similar to the approach in Section 4.4, we initially obtain $k = 10$ number of clusters. However, using the Tukey's method, at the end of the analysis, we get an updated number of clusters.

Consider $n_c = (33, 37, 32, 14, 34, 65, 4, 13, 18, 18)$. Let x_{j_l} denote j'th covariate in the l'th cluster for $j_l = 1, \ldots, n_{c_l}$ and $l = 1, \ldots, k$. Throughout this section, different scenarios for D_{j_l}'s and $\sigma^2_{u_l}$'s are considered. Generally, D_{j_l}'s are generated, from a uniform distribution with different ranges, for $j_l = 1, \ldots, n_{c_l}$ and $l = 1, \ldots, k$. The response random variable is simulated following two steps. First, let $\theta_{j_l} \sim N(\mathbf{X}_{j_l}\beta, \sigma^2_{u_l})$ where $\mathbf{X}_{j_l} = (1, x_{j_l})$ and $\beta = (2.8093, 0.0059)$ that is obtained by regressing the prescription cost on the percentage of patients over 70 years. Then, $y_{j_l} \sim N(\theta_{j_l}, D_{j_l})$, for $j_l = 1, \ldots, n_{c_l}$ and $l = 1, \ldots, k$. We generate $R = 5000$ simulations of the response random variables.

In Theorem 4, the MMM estimate of the variance components is introduced. Throughout this chapter, we work with the MMM estimates of the variance components rather than their Restricted Maximum Likelihood (REML) estimates to predict small area means and also estimate the MSPE. We performed simulation studies in order to compare our proposed estimates of the variance components with the REML estimates of the variance components (Cressie, 1992) in terms of mean squared error (MSE). Tables 4.2 and 4.3 show that the two methods have almost the same performance using the MSE criterion. As Table 4.3 shows the MSE for the same value of the variance components are different since we have different cluster sizes even for two clusters with the same variance components. According to Tables 4.2 and 4.3, we expect to get a smaller MSE of variance components for larger cluster sizes.

Table 4.2: MSE of the REML and MMM estimates for different variance components

σ_u^2	1	8	25	5	64	10	49	36	40	13
n_c	33	37	32	14	34	65	4	13	18	18
MSE REML	0.07	2.38	27.00	2.67	159.23	2.03	767.33	123.32	118.05	12.91
MSE MMM	0.09	2.37	26.57	2.73	155.73	2.03	702.52	119.69	112.03	12.80

In order to calculate the power of the proposed test, the value of the test statistic is calculated for each simulated response variable. After finding the corresponding p-value, we reject the null hypothesis of the equality of variance components if p-value is less than the predetermined level of significance

Table 4.3: MSE of the REML and MMM estimates for similar variance components in some clusters

σ_u^2	1	8	25	25	64	25	49	40	40	13
n_c	33	37	32	14	34	65	4	13	18	18
MSE REML	0.07	2.39	27.01	58.25	159.25	12.15	772.09	152.07	118.75	13.02
MSE MMM	0.10	2.40	26.60	55.83	155.78	12.08	704.98	147.54	112.08	12.83

($\alpha = 0.05$). The average of the number of times that the null hypothesis is rejected when the alternative hypothesis is correct is used to evaluate the power of the test. Table 4.4 gives the power of the test under different set-ups.

As it is shown in Table 4.4, when the difference between D_{jl}'s and $\sigma_{u_l}^2$'s gets larger, the test becomes more powerful. Moreover, when the similarities between the $\sigma_{u_l}^2$'s, for $l = 1, \ldots, k$, increase, the power of the test decreases. Generally, removing clusters from the analysis may increase or decrease the power of the test (Table 4.5). For instance, looking at the second row of Table 4.5, removing the smallest cluster with four small areas increases the power. As this cluster has the same variance component as the eighth one, we expect to have a higher power because of more distinct values of the variance components for the remaining clusters. Obviously, removing a large cluster decreases the power of the test more significantly. The simulation study indicates that in this scenario the test rejects the null hypothesis more than the predetermined significance level $\alpha = 0.05$ (Table 4.6). As it is explained in Section 4.1, this is

due to the fact that the precision of the estimate of the $\sigma^2_{u_l}$'s, for $l = 1, \ldots, k$ depends highly on the number of small areas in each cluster. In our set-up, the 7'th cluster contains four small areas while the 6'th cluster contains 65 small areas. So, even though the variance components are the same in both, the null hypothesis of the equality might be rejected because of the difference between n_{c_6} and n_{c_7}.

Table 4.4: The power of the test for different values of $\sigma^2_{u_l}$'s and D_j's

	D_{j_l}			
σ^2_u	U(0.25, 0.5)	U(0.25, 1.5)	U(0.25, 2.5)	U(0.25, 3.5)
$(1, 8, 25, 5, 64, 10, 49, 36, 40, 13)$	1	1	1	1
$(1, 8, 25, 25, 64, 25, 49, 40, 40, 13)$	1	1	1	1
$(1, 1, 2, 4, 2.5, 4, 3, 3, 5, 6)$	1	0.97	0.90	0.86
$(1.5, 1.5, 2.4, 6, 4, 4, 5.9, 4, 6, 4)$	0.96	0.89	0.79	0.74
$(0.1, 0.5, 0.2, 0.5, 0.25, 0.4, 0.49, 0.3, 0.4, 0.13)$	0.50	0.26	0.19	0.22
$(0.1, 0.1, 0.2, 0.15, 0.25, 0.3, 0.12, 0.3, 0.09, 0.13)$	0.36	0.21	0.17	0.20

We also perform simulation studies in order to evaluate the performance of our proposed method in terms of the EMSPE. To this end, we calculate the EBLUP of small area means, $\hat{\theta}^{(r)}_{j_l}$'s, by finding the MMM estimates of the variance components and substituting them in (4.3.1). The EMSPE is given as follows

$$\text{EMSPE}(\theta_{j_l}) = \frac{1}{R}\sum_{r=1}^{R}(\hat{\theta}^{(r)}_{j_l} - \theta^{(r)}_{j_l})^2, \text{ for } j_l = 1, \ldots, n_{c_l} \text{ and } l = 1, \ldots, k.$$

Table 4.5: The effect of removing clusters on the power of the test for the case of $\sigma_u^2 = (1, 1, 2, 4, 2.5, 4, 3, 3, 5, 6)$

	D_{jl}			
	U(0.25, 0.5)	U(0.25, 1.5)	U(0.25, 2.5)	U(0.25, 3.5)
Removing the largest cluster	0.99	0.95	0.87	0.82
Removing the smallest cluster	1	0.98	0.92	0.87
Removing two largest clusters	0.94	0.87	0.75	0.70
Removing three largest clusters	0.95	0.89	0.78	0.72
Removing four largest clusters	0.69	0.60	0.53	0.48

Table 4.6: The percentage of times the null hypothesis is rejected by mistake

	D_{jl}			
σ_u^2	U(0.25, 0.5)	U(0.25, 1.5)	U(0.25, 2.5)	U(0.25, 3.5)
$(0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1)$	0.14	0.15	0.13	0.13
$(1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$	0.15	0.16	0.15	0.18
$(1.2, 1.2, 1.2, 1.2, 1.2, 1.2, 1.2, 1.2, 1.2, 1.2)$	0.15	0.16	0.15	0.18
$(2, 2, 2, 2, 2, 2, 2, 2, 2, 2)$	0.15	0.15	0.14	0.15
$(3, 3, 3, 3, 3, 3, 3, 3, 3, 3)$	0.15	0.16	0.15	0.18
$(20, 20, 20, 20, 20, 20, 20, 20, 20, 20)$	0.15	0.16	0.15	0.16

where $\theta_{jl}^{(r)}$ is the small area mean in the r'th iteration.

We compare EMSPE of the complete clustering, EMSPE_c, with the simple method, EMSPE_s, by finding the following ratio for each small area

$$\frac{\text{EMSPE}_s}{\text{EMSPE}_c}.$$

Values larger than one indicate that the complete clustering reduces the true MSPE for small areas. We find the ratio of $\frac{\text{EMSPE}_s}{\text{EMSPE}_{cb}}$ and $\frac{\text{EMSPE}_{cb}}{\text{EMSPE}_c}$ to compare EMSPE of the combined version of the clustering, EMSPE_{cb}, after implementing the Tukey's method, with the simple method and the complete clustering.

We observe that when σ_l^2's are highly different, the improvement in either EMSPE_c or EMSPE_{cb} over the simple linear mixed model become larger. Dealing with $\sigma_u^2 = (1, 8, 25, 5, 64, 10, 49, 36,\ 40, 13)$, the improvement for the complete clustering gets up to 67% with the minimum of −8% compared to the simple approach. We get larger improvements for areas belonging to clusters with variance components far from their overal average (25.1) while the negative improvement shows a scattered pattern. For combined clusters, we gain up to 49% improvement with the minimum of −10% compared to the simple approach. Also, the complete clustering performs better than the combined one for up to 18% and the minimum of −2%. Similar to complete clustering scheme, in the case of combined clustering, the larger improvement happens for areas belonging to clusters with variance components far from the overal average. Table 4.7 summarizes the results for $\sigma_u^2 = (1, 8, 25, 5, 64, 10, 49, 36,\ 40, 13)$.

Table 4.7: Comparison of EMSPE of predictors of small area means using different approaches based on their deciles: (a) $\frac{\text{EMSPE}_s}{\text{EMSPE}_c}$, (b) $\frac{\text{EMSPE}_s}{\text{EMSPE}_{cb}}$, (c) $\frac{\text{EMSPE}_{cb}}{\text{EMSPE}_c}$ when $\sigma_u^2 = (1, 8, 25,\ 5, 64, 10, 49, 36, 40, 13)$ and $D_{j_l} \sim \text{Uniform}(0.25, 0.5)$. For columns (a) and (b), values larger than 1 indicate the complete and combined clustering approaches estimate the small area mean with a higher precision in comparison with the simple approach, respectively. Values larger than 1 in column (c) indicate the complete clustering approach performs better than the combined clustering approach in estimating the small area mean.

	(a)	(b)	(c)
Minimum	0.92	0.90	0.97
1-decile	0.99	0.98	0.99
2-decile	0.99	0.99	1.00
3-decile	1.00	0.99	1.00
4-decile	1.00	1.00	1.00
5-decile	1.01	1.01	1.00
6-decile	1.02	1.01	1.00
7-decile	1.03	1.03	1.01
8-decile	1.04	1.04	1.02
9-decile	1.19	1.14	1.04
Maximum	1.67	1.49	1.18
Mean	1.05	1.03	0.92

In order to have a numerical evaluation of the performance of (4.3.3), we use the RB

$$\text{RB}_{j_l} = \frac{\text{E}(\text{mspe}_{j_l})}{\text{EMSPE}_{j_l}} - 1 \text{ for } j_l = 1, \ldots, n_{c_l} \text{ and } l = 1, \ldots, k. \tag{4.5.1}$$

where $\text{E}(\text{mspe}_{j_l})$, for $j_l = 1, \ldots, n_{c_l}$ and $l = 1, \ldots, k$, is the average of obtained

values from (4.3.3) over $R = 5000$ iterations. Table 4.8 gives the summary statistics of the RB of the estimation of the MSPE for different approaches when $\sigma_u^2 = (1, 8, 25, 5, 64, 10, 49, 36, 40, 13)$ and $D_{jl} \sim \text{Uniform}(0.25, 0.5)$. Our findings indicate that we get small RB for the three approaches. In particular, the value of $|RB|$ is less than 0.55, 0.57, 0.57 for the complete, combined, and simple clustering methods.

Table 4.8: The summary statistics for the RB of the estimator of MSPE of small area means using different approaches for $\sigma_u^2 = (1, 8, 25, 5, 64, 10, 49, 36, 40, 13)$: (a) the complete clustering approach, (b) the combined clustering approach, and (c) the simple approach

	(a)	(b)	(c)
Minimum	-0.34	-0.34	-0.34
1-decile	-0.16	-0.17	-0.16
2-decile	-0.12	-0.12	-0.11
3-decile	-0.07	-0.08	-0.08
4-decile	-0.04	-0.05	-0.03
5-decile	0.00	-0.01	0.00
6-decile	0.02	0.02	0.04
7-decile	0.08	0.08	0.07
8-decile	0.12	0.12	0.12
9-decile	0.20	0.20	0.20
Maximum	0.55	0.56	0.57
Mean	0.01	0.01	0.01

We now consider the simulation studies for $D_{jl} \sim \text{Uniform}(0.25, 0.5)$ and

$\sigma_u^2 = (1, 8, 25, 25, 64, 25, 49, 40, 40, 13)$ to evaluate to what extend the difference between the variance components affects estimation of MSPE. We gain improvement in terms of EMSPE by using clustering based on the covariate (Table 4.9). The maximum improvement of 66% for the complete clustering and the minimum of −4% are obtained compared to the simple approach. For the combined version, the maximum improvement of 62% and the minimum of −4% are obtained compared to the simple approach. Also, the complete clustering performs better than the combined one for up to 10% and the minimum of −4%. Similar to Table 4.7, we are not able to determine a specific trend for the negative improvement. The highest amount of improvement happens for small areas with the variance component far from the overal average ($\sigma_{u_1}^2 = 1$). Table 4.10 gives the summary statistics for the RB of the estimation of the MSPE using different approaches when $\sigma_u^2 = (1, 8, 25, 25, 64, 25, 49, 40, 40, 13)$. The complete and combined clustering have similar performance with the $|RB| \leq 0.56$ while the simple approach results in $|RB| \leq 0.57$.

4.5.2. Assessing the performance of the proposed estimator of D_i

In Formula (4.4.2), we proposed an estimator of D_i's for the unit level data. In order to evaluate the performance of this estimator, we implement simulation studies. The estimated values of the parameters from Section 4.4 and also its covariate matrix are used to generate the response variable (waist circumference). We have 1362 small areas with 519 of them having one

sample unit. We consider the MMM estimates of the variance components, $(0.36, 2.87, 4.05, 1.76, 11.40, 420.77, 36.17)$, as well as the estimates of D_i's to generate the data. In each iteration, the sampling variance of the mean of the response variable in each small area is calculated. As we explained in Section 4.4, for small areas where the regression line cannot be defined, we use the overal regression line obtained from all sample units. Mathematically, (4.4.1) is an unbiased estimator of σ_e^2. Expanding (4.4.1) results in

$$\frac{1}{n_i-1}\left[2b_1\sum_{j=1}^{n_i}(x_{ij}-x_i)(e_{ij}-e_i)+\sum_{j=1}^{n_i}(e_{ij}-e_i)^2\right]. \tag{4.5.2}$$

We perform simulation studies to check the magnitude of the first term in (4.5.2), $\frac{1}{n_i-1}[2b_1\sum_{j=1}^{n_i}(x_{ij}-x_i)(e_{ij}-e_i)]$. Figure 4.4 shows the histogram of this term for all small areas which is almost zero. This indicates that the error of neglecting the first term in (4.5.2) is negligible. Figure 4.5 displays that Formula (4.4.2) underestimates the sampling variance, but, in general, it gives reliable estimates.

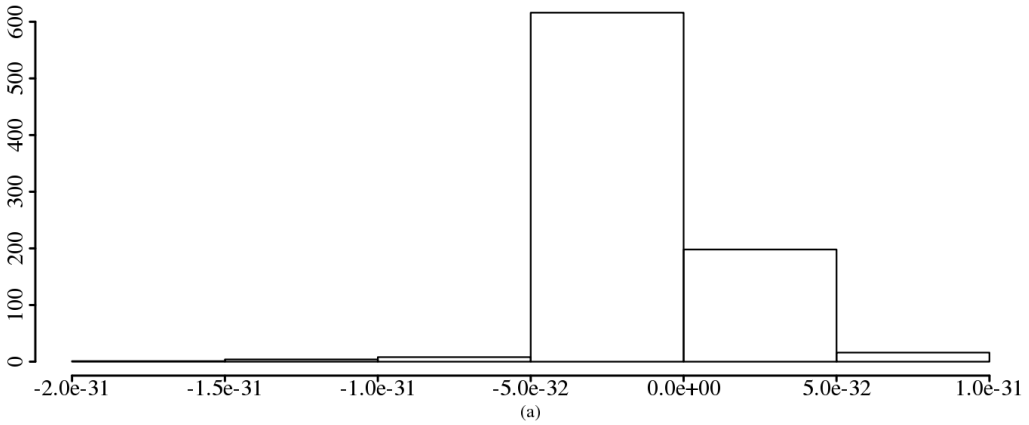

Figure 4.4: Histogram of the neglected terms in estimation of $\sigma^2_{e_i}$'s given in (4.5.2).

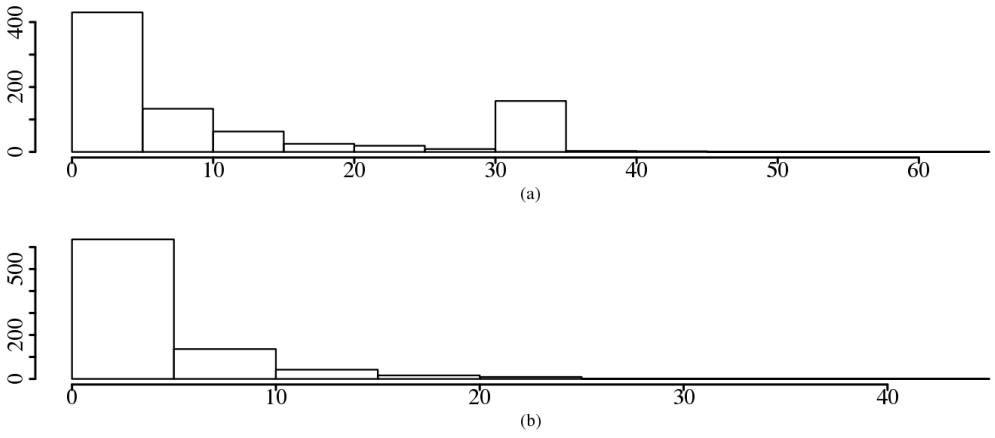

Figure 4.5: (a) histogram of the true D_i's vs. (b) histogram of the estimates of D_i's

Table 4.9: Comparison of EMSPE of predictors of small area means using different approaches based on their deciles: (a) $\frac{\text{EMSPE}_s}{\text{EMSPE}_c}$, (b) $\frac{\text{EMSPE}_s}{\text{EMSPE}_{cb}}$, (c) $\frac{\text{EMSPE}_{cb}}{\text{EMSPE}_c}$ when $\sigma_u^2 = (1, 8, 25,\ 25, 64, 25, 49, 40, 40, 13)$ and $D_{jl} \sim \text{Uniform}(0.25, 0.5)$. For columns (a) and (b), values larger than 1 indicate the complete and combined clustering approaches estimate the small area mean with a higher precision in comparison with the simple approach, respectively. Values larger than 1 in column (c) indicate the complete clustering approach performs better than the combined clustering approach in estimating the small area mean.

	(a)	(b)	(c)
Minimum	0.96	0.96	0.96
1-decile	0.99	0.99	0.99
2-decile	0.99	0.99	1.00
3-decile	1.00	1.00	1.00
4-decile	1.00	1.00	1.00
5-decile	1.00	1.00	1.00
6-decile	1.01	1.01	1.00
7-decile	1.01	1.01	1.00
8-decile	1.03	1.02	1.01
9-decile	1.20	1.17	1.01
Maximum	1.66	1.62	1.10
Mean	1.04	1.04	0.96

Table 4.10: The summary statistics for the RB of the estimator of MSPE of small area means using different approaches for $\sigma_u^2 = (1, 8, 25, 25, 64, 25, 49, 40, 40, 13)$: (a) the complete clustering approach, (b) the combined clustering approach, and (c) the simple approach

	(a)	(b)	(c)
Minimum	-0.34	-0.34	-0.34
1-decile	-0.16	-0.16	-0.16
2-decile	-0.12	-0.11	-0.12
3-decile	-0.07	-0.07	-0.08
4-decile	-0.04	-0.04	-0.04
5-decile	0.00	-0.01	0.01
6-decile	0.02	0.03	0.04
7-decile	0.08	0.08	0.07
8-decile	0.13	0.12	0.12
9-decile	0.20	0.20	0.20
Maximum	0.56	0.56	0.57
Mean	0.01	0.01	0.01

Chapter 5

Summary and Future Work

In this thesis, we developed some new and efficient statistical techniques in the context of small area estimation. In the first two research problems, we consider the functional measurement error in the observed covariate and account for the error made in observing the true area-specific covariate when the goal is to predict small area means. We predicted small area means using the James-Stein (JS) as well as constrained Bayes (CB) estimation methods in order to get more precise point predictors and also a closer histogram of the predictors of small area means to its true underlying histogram, respectively. In the third research problem, we clustered small areas into homogenous groups and considered different variance components for each group to give more insight into the behaviour of the random effects.

In Chapter 2, a new pseudo empirical Bayes (PEB) predictor of small area means based on the JS estimate of the true area-specific covariate for a unit level regression model with the functional measurement error was introduced.

Our findings showed that the PEB of small area means using the JS estimate of the true covariate performs better than the PEB of γ_i using the maximum likelihood (ML) estimate and also method of moments in terms of the weighted and unweighted mean squared prediction error estimation. Specifically, if the range of the true area-specific covariate is small, the PEB based on the JS performs very well which can be easily justified using the structure of the JS estimator. We performed several simulation studies and observed that the PEB predictor based on the JS estimate always dominates the corresponding PEB predictors (GS and DRT) in terms of MSPE($\widehat{\gamma}_i^{\text{PEB}}$) in small areas with one sample unit even if the range of the true area-specific covariate is large. In addition, the JS method gives estimates of the true area specific covariate for areas with no sample units. Consequently, PEB predictors of the small area means based on JS method are obtained for theses areas. However, at least one sampled unit is required to find the ML and method of moment estimates of the true area-specific covariates. For small areas with no sample units we obtain $\widehat{\gamma}_{i\text{GS}}^{\text{PEB}} = \widehat{\gamma}_{i\text{DRT}}^{\text{PEB}} = \hat{b}_0$.

We used the jackknife method to get weighted and unweighted jackknife estimators of MSPE($\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}$). Our simulation results showed that these estimates perform well in terms of the relative bias of mspe($\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}$). Also, in our simulation studies we observed that the jackknife estimation of the MSPE has a large variance. Thus, we suggested to use this method with caution. This is specially important if one uses other measures of performance for the jackknife estimators

such as the coefficient of variation. It is worth mentioning that, although the weighted and unweighted jackknife estimators of the MSPE have almost the same performance in terms of RB, we suggested using the weighted jackknife method as it results in a smaller variance. Further, using the result obtained from the simulation studies, we observed that different PEB methods show similar results in the case of large sample sizes within areas. But, the total number of the areas does not have significant effect on the MSPE or EMSPE of the PEB predictors. Based on our simulation studies, we observed that the estimates of the model parameters using the method of moments (Ghosh and Sinha, 2007) do not perform well. We assessed the proposed approach using a cross sectional study in New Zealand to predict the diastolic blood pressure using the cholesterol level.

In Chapter 3, following Ghosh and Sinha (2007), Datta et al. (2010), and Torkashvand et al. (2015), a linear mixed model with the functional measurement error in the true area-specific covariate was considered. Using the general paradigm of Ghosh (1992), the CB estimate of the true area-specific covariate was introduced in order to get a more precise estimate of the true underlying histogram of the population parameter. We showed that the CB estimate of the true area-specific covariate dominated the ML estimate in terms of the Bayes risk. Also, the pseudo-Bayes (PB) predictor of the small area means based on the CB estimate of the true area-specific covariate dominated the PB predictor of the small area means using the ML estimate in terms of

Bayes risk. In order to evaluate the performance of different predictors of the small area means, the sensitivity (Se), specificity (Sp), positive predictive value (PPV), and negative predictive value (NPV) of the PB predictors of the small area means were obtained.

As an application, a cross-sectional data from the New Zealand population was analyzed. A simulation study was also conducted to evaluate the performance of the proposed approach. The simulation study showed that the histograms of the estimated values of the true area-specific covariates using the CB, the constrained empirical Bayes (CEB), and the constrained hierarchical Bayes (CHB) methods closely followed the true underlying histogram of the area-specific covariates. In addition, the pseudo empirical Bayes (PEB) predictor based on the CEB performed the best in terms of the Se and the PPV. We also observed desirable behaviour of our proposed estimators in terms of the Sp and the NPV measures. We noted that the PEB predictor based on the ML estimator of the true area-specific covariate performed the worst in terms of the Se and the PPV measures.

The MSPE and EMSPE of different predictors of small area means were also calculated. The MSPE of the small area mean predictor based on the CB estimate of the true area-specific covariate dominated the MSPE of the small area mean predictor based on the ML estimator of the true area-specific covariate. Estimation of the MSPE using the jackknife method, similar to Datta et al. (2010) and Torkashvand et al. (2015), remains as a future research

project.

In Chapters 2 and 3, we considered the case where we assumed there is only one covariate in the study that is subject to the functional measurement error. There are many situations where there exists more than one covariate available in the study subject to the functional measurement error. Developing methodology for these situations can be considered as an extension of the current work.

In Chapter 4, the main interest lied in predicting small area means while the precision of the predictor is quantified using MSPE. We proposed to use clustering in small area estimation set-up based on covariates in order to get more accurate predictions of small area means. To this end, we clustered small areas based on covariates and conducted a hypothesis test to check the assumption of the equality of the variance components in different clusters. Our results indicated that the test has a high power with inflated type I error. Following Tukey (1949), we combined some clusters with similar variance components. The small area means were predicted by implementing the idea of clustering either with complete or combined clustering schemes. In order to compare the performance of the new predictors with simple predictors of small area means, a unique variance component for all small areas, EMSPE of three methods was calculated. The results showed improvement in terms of the EMSPE specially when the difference between variance components was significant.

A real data set was analyzed corresponding to the unit level scenario. Using the complete clustering approach, mspe was on average 12.43 times smaller than mspe obtained using the simple method. The estimated values of variance components for this dataset were significantly different and large. Due to similarities between variance components for couple of clusters, they were merged by the Tukey's method. We proposed to consider the combined clustering for this data set.

Although the approach we proposed in this chapter works well, it suffers from some limitations. As Theorem 4 shows, there is a possibility of finding a negative value for the variance component. We have planned to address this issue in our future study.

In Chapter 4, we considered the simplest case of normal distribution for the random effects in each cluster. Clustering enables us to consider different types of distributions for the random effects in order to get an accurate picture of their behaviour in each cluster. Using the concept of the clustering for generalized linear mixed models is another research area and will also be proceeded in our future work.

In an ongoing research project, we consider clustering in small area estimation with unit-level linear mixed models with functional measurement error in the area-specific covariate. In this work, we consider models (2.0.1) and (2.0.2) to study the effect of clustering on the precision of the prediction of small area means. Estimates of x_i's, JS, CEB, and CHB, are obtained. We cluster small areas using hierarchical clustering based on estimates of the

true area-specific covariates. Due the complicated mathematical formulas, we propose a bootstrap test to evaluate the assumption of equality of the variance components in different clusters. Despite the test statistic we proposed in Chapter 4, our bootstrap test works for an arbitrary number of small areas. The role of clustering on EMSPE and also the estimation of MSPE, mspe, is evaluated. We apply our methodology to a real dataset from the USDA's Continuing Survey of Food Intakes by Individuals 1985 and 1986 (CSFII 85-86) core sample survey in order to study the effect of BMI on the calorie intake in the small area estimation context. The aim is to determine areas with large values of BMI and small value of the calorie intake. This can identify groups that suffer from hormone imbalance problems or improper eating habit. It is worth mentioning that this requires an innovative grouping in order to assume BMI variation in each category is negligible.

In this thesis research, we focused on the linear mixed model with the functional measurement error. In practice, the generalized linear mixed model are used more widely. Examples of these models are logistic and Poisson mixed models. For instance, researchers in biology deal with the situation that a disease occurs or not. This can be categorized as a Bernoulli distribution where we work with the logistic regression model. The other scenario happens when the researcher counts the number of rare cases of interest (e.g., cancer) in each small area which is an application of the Poisson regression model. The implementation of the measurement error model, either functional or structural, will improve our prediction of the small area means. Benchmarking based

on either internal or external conditions in this setup is another interesting area that we will study. Datta et al. (2011) considered benchmarked Bayes estimators and developed some preliminary results. This technique lets the researcher combine the information from the previous surveys in the study as the external benchmarking. In our setup, we studied the internal benchmarking for the linear mixed model with the functional measurement error in Chapter 3. Evaluating the performance of the (internal or external) benchmarked estimator for the generalized linear mixed model with measurement error is also important and will introduce another research area. In Chapter 3, some criteria were introduced for the evaluation of different predictors of the small area means. Calculating those criteria in the context of the generalized linear mixed model with functional measurement errors is another research problem. This problem will be both computationally and theoretically challenging.

So far, the squared error loss is considered in the estimation of small area means. There are some disadvantages associated with the squared error loss. First, this loss function is not bounded. In real qualification examples, there is a boundary for the loss that should be taken into account. Second, it penalizes departures from the true value in either direction equally. Sometimes, the penalty we pay for overestimation is more serious than the underestimation. And finally, we might look for a parameter to control the behaviour of the loss function near the true parameter with more precision. The reflected normal loss function is a capable candidate to address these problems (Spiring, 1993). An important research direction for future study is to assess the problem of

small area estimation with or without the measurement error with this loss function similar to Ghosh and Sinha (2007) and Prasad and Rao (1990).

In this thesis, our main interest was to predict small area means. There are set-ups where we are interested in the estimation of positive small area parameters. As an illustrative example Ghosh et al. (2015) considered estimation of the average spending on education. The squared error loss and weighted squared error loss might result in negative estimation of the small area parameters. We currently work to introduce a new loss function for the positive parameters in the context of small area estimation. The goal is to show the new loss function addresses some of the disadvantages of loss functions used in the past. Similar to Prasad and Rao (1990) and Ghosh et al. (2015), we need to approximate the risk function mathematically. This will open the door for a new research direction in this context and under asymmetric loss functions.

Bibliography

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. (Cited on pages 4, 5, 9 and 40.)

Billingsley, P. (2008). *Probability and measure*. John Wiley & Sons. (Cited on page 94.)

Buonaccorsi, J. P. (2010). *Measurement error: models, methods, and applications*. CRC Press. (Cited on page 8.)

Carroll, R. J., K. Roeder, and L. Wasserman (1999). Flexible parametric measurement error models. *Biometrics 55*(1), 44–54. (Cited on pages 23 and 58.)

Carroll, R. J., D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu (2010). *Measurement error in nonlinear models: a modern perspective*. CRC press. (Cited on pages 23, 26 and 58.)

Chen, S. and P. Lahiri (2002). On mean squared prediction error estimation in small area estimation problems. *In: Proceeding of the Survey Research Methods Section. American Statistical Association, Washington*, 473–477. (Cited on pages 38 and 40.)

Cressie, N., O. Perrin, and C. Thomas-Agnan (2005). Likelihood-based estimation for gaussian MRFs. *Statistical Methodology 2*(1), 1–16. (Cited on page 110.)

Cressie, N. A. (1992). REML estimation in empirical bayes smoothing of census undercount. *Faculty of Informatics –Papers (Archive)*, 75–94. (Cited on page 112.)

Cressie, N. A., O. Perrin, and C. Thomas-Agnan (2006). Doctors prescribing patterns in the midi-pyrénées rregion of france: Point-process aggregation. In *Case Studies in Spatial Point Process Modeling*, pp. 183–195. Springer. (Cited on page 110.)

Crowder, M. (1986). On consistency and inconsistency of estimating equations. *Econometric Theory*, 305–330. (Cited on page 36.)

DasGupta, A. (2008). *Asymptotic theory of statistics and probability*. Springer. (Cited on page 29.)

Datta, G. S., M. Ghosh, R. Steorts, and J. Maples (2011). Bayesian bench-

marking with applications to small area estimation. *Test 20*(3), 574–588. (Cited on pages 56 and 132.)

Datta, G. S., P. Hall, and A. Mandal (2011). Model selection by testing for the presence of small-area effects, and application to area-level data. *Journal of the American Statistical Association 106*(493), 362–374. (Cited on page 87.)

Datta, G. S. and A. Mandal (2015). Small area estimation with uncertain random effects. *Journal of the American Statistical Association 110*(512), 1735–1744. (Cited on pages 87 and 88.)

Datta, G. S., J. Rao, and M. Torabi (2010). Pseudo-empirical bayes estimation of small area means under a nested error linear regression model with functional measurement errors. *Journal of Statistical Planning and Inference 140*(11), 2952–2962. (Cited on pages 11, 28, 37, 38, 39, 40, 63, 83, 127 and 128.)

Efron, B. and C. Morris (1972). Limiting the risk of bayes and empirical bayes estimators part ii: The empirical bayes case. *Journal of the American Statistical Association 67*(337), 130–139. (Cited on pages 22 and 24.)

Efron, B. and C. Morris (1975). Data analysis using stein's estimator and its generalizations. *Journal of the American Statistical Association 70*(350), 311–319. (Cited on pages 24, 25, 72 and 80.)

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. (Cited on pages 4 and 5.)

Fuller, W. A. (2009). *Measurement error models*, Volume 305. John Wiley & Sons. (Cited on pages 9 and 22.)

Ghosh, M. (1992). Constrained bayes estimation with applications. *Journal of the American Statistical Association 87*(418), 533–540. (Cited on pages 56, 59 and 127.)

Ghosh, M., T. Kubokawa, and Y. Kawakubo (2015). Benchmarked empirical bayes methods in multiplicative area-level models with risk evaluation. *Biometrika 102*(3), 647–659. (Cited on page 133.)

Ghosh, M. and T. Maiti (1999). Adjusted bayes estimators with applications to small area estimation. *Sankhyā: The Indian Journal of Statistics, Series B*, 71–90. (Cited on page 72.)

Ghosh, M. and K. Sinha (2007). Empirical bayes estimation in finite population sampling under functional measurement error models. *Journal of statistical planning and inference 137*(9), 2759–2773. (Cited on pages 8, 28, 32, 33, 35, 37, 38, 40, 63, 72, 79, 127 and 133.)

Ghosh, M. and R. C. Steorts (2013). Two-stage benchmarking as applied to small area estimation. *Test 22*(4), 670–687. (Cited on page 57.)

Gonzalez, M. E. (1973). Use and evaluation of synthetic estimates. In *Proceedings of the Social Statistics Section*, pp. 33–36. (Cited on page 4.)

Ha, N. S. (2013). Hierarchical bayesian estimation of small area means using complex survey data. (Cited on page 57.)

Ha, N. S. and P. Lahiri (2014). Comments on: Single and two-stage cross-sectional and time series benchmarking procedures for small area estimation. *Test 23*(4), 670. (Cited on page 57.)

Hartigan, J. A. and M. A. Wong (1979). Algorithm as 136: A k-means clustering algorithm. *Applied statistics*, 100–108. (Cited on page 106.)

Henderson, C. (1950). Estimation of genetics parameters (abstract). *Annals of Mathematical Statistics 21*, 309–310. (Cited on pages 6 and 101.)

James, G., D. Witten, T. Hastie, and R. Tibshirani (2013). *An introduction to statistical learning*, Volume 112. Springer. (Cited on page 14.)

Jiang, J. (2010). *Large sample techniques for statistics*. Springer Science & Business Media. (Cited on page 4.)

Jiang, J. and P. Lahiri (2006). Mixed model prediction and small area estimation. *Test 15*(1), 1–96. (Cited on pages 4 and 57.)

Jiang, J., P. Lahiri, and S.-M. Wan (2002). A unified jackknife theory for

empirical best prediction with m-estimation. *The Annals of Statistics 30*(6), 1782–1810. (Cited on pages 38 and 40.)

Kang, E. L., D. Liu, and N. Cressie (2009). Statistical analysis of small-area data based on independence, spatial, non-hierarchical, and hierarchical models. *Computational Statistics & Data Analysis 53*(8), 3016–3032. (Cited on pages 110 and 111.)

Kaufman, L. and P. J. Rousseeuw (2009). *Finding groups in data: an introduction to cluster analysis*, Volume 344. John Wiley & Sons. (Cited on page 13.)

Kubokawa, T., W. E. Strawderman, et al. (2013). Dominance properties of constrained bayes and empirical bayes estimators. *Bernoulli 19*(5B), 2200–2221. (Cited on page 57.)

Lahiri, P. (1990). " adjusted" bayes and empirical bayes estimation in finite population sampling. *Sankhyā: The Indian Journal of Statistics, Series B*, 50–66. (Cited on page 56.)

Lehmann, E. L. and G. Casella (1998). *Theory of point estimation*, Volume 31. Springer. (Cited on page 22.)

Louis, T. A. (1984). Estimating a population of parameter values using bayes and empirical bayes methods. *Journal of the American Statistical Association 79*(386), 393–398. (Cited on pages 56 and 59.)

Lyles, R. H., A. K. Manatunga, R. H. Moore, F. DuBois Bowman, and C. B. Cook (2007). Improving point predictions of random effects for subjects at high risk. *Statistics in medicine 26*(6), 1285–1300. (Cited on pages 55 and 59.)

Lyles, R. H. and J. Xu (1999). Classifying individuals based on predictors of random effects. *Statistics in medicine 18*(1), 35–52. (Cited on pages 68 and 69.)

Maiti, T., H. Ren, S. C. Dass, C. Lim, and K. S. Maier (2011). Clustering-based small area estimation: An application to meap data. *Calcutta Statistical Association Bulletin 66*, 73–93. (Cited on page 12.)

Pfeffermann, D., A. Sikov, and R. Tiller (2014). Single-and two-stage cross-sectional and time series benchmarking procedures for small area estimation. *Test 23*(4), 631–666. (Cited on page 57.)

Prasad, N. and J. Rao (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American statistical association 85*(409), 163–171. (Cited on pages 7, 103 and 133.)

Purcell, N. J. and L. Kish (1979). A biometrics invited paper. estimation for small domains. *Biometrics*, 365–384. (Cited on page 4.)

Rao, J. N. and I. Molina (2015). *Small area estimation, Second Edition*. Wiley, New York. (Cited on pages 4, 5, 7 and 102.)

Robbins, H. (1956). An empirical bayes approach to statistics. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, 157–163. (Cited on pages 37 and 38.)

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics 20*, 53–65. (Cited on pages 15 and 106.)

Schaible, W. L. (1978). Choosing weights for composite estimators for small area statistics. (Cited on page 4.)

Sen, P. K. and J. M. Singer (1993). *Large sample methods in statistics: An introduction with applications*. New York: Chapman & Hall. (Cited on page 61.)

Small, C. G. and Z. Yang (1999). Multiple roots of estimating functions. *Canadian Journal of Statistics 27*(3), 585–598. (Cited on page 36.)

Speros, C. (2009). More than words: Promoting health literacy in older adults. *OJIN: The Online Journal of Issues in Nursing 14*(3), Manuscript 5. (Cited on page 111.)

Spiring, F. A. (1993). The reflected normal loss function. *Canadian Journal of Statistics 21*(3), 321–330. (Cited on page 132.)

Spjøtvoll, E. and I. Thomsen (1987). Application of some empirical bayes methods to small area statistics. *Bulletin of the International statistical Institute 4*, 435–450. (Cited on page 55.)

Torkashvand, E., M. Jafari Jozani, and M. Torabi (2015). Pseudo-empirical bayes estimation of small area means based on the james-stein estimation in linear regression models with functional measurement errors. *Canadian Journal of Statistics 43*, 265–287. (Cited on pages 55, 63, 71, 84, 127 and 128.)

Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, 99–114. (Cited on pages 88, 98 and 129.)

Ugarte, M., A. Militino, and T. Goicoa (2009). Benchmarked estimates in small areas using linear mixed models with restrictions. *Test 18*(2), 342–364. (Cited on page 57.)

Vague, J. (1947). Sexual differentiation, a factor affecting the forms of obesity. *Presse Med 30*, 339–340. (Cited on page 104.)

Wang, J. and W. A. Fuller (2003). The mean squared error of small area predictors constructed with estimated area variances. *Journal of the American Statistical Association 98*(463), 716–723. (Cited on page 104.)

Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association 58*(301), 236–244. (Cited on pages 14 and 106.)

Whittemore, A. S. (1989). Errors-in-variables regression using stein estimates. *The American Statistician 43*(4), 226–228. (Cited on page 22.)

Ybarra, L. M. and S. L. Lohr (2008). Small area estimation when auxiliary information is measured with error. *Biometrika 95*(4), 919–931. (Cited on page 39.)

Zhang, W. and N. Li (2011). Prevalence, risk factors, and management of prehypertension. *International journal of hypertension 2011*. (Cited on pages 68 and 74.)