CIMAT

Centro de Investigación en Matemáticas, A.C.

SMALL AREA ESTIMATION BASED ON A TWO-FOLD NESTED ERROR LOGNORMAL MODEL

TESIS

Que para obtener el Grado de:

Doctor en Ciencias con Orientación en Probabilidad y Estadística

P R E S E N T A:

Georges Bucyibaruta

Director:

Dr. Rogelio Ramos Quiroga

Guanajuato, Gto, México, diciembre de 2018

Integrantes del Jurado.

Lector Especial: Dr. Elas Rodríguez Muñoz (U. de Guanajuato)

Presidente: Dr. Enrique Raúl Villa Diharce (CIMAT)

Secretaria: Dra. Lilia Leticia Ramírez Ramírez (CIMAT)

Vocal: Dr. Paulino Pérez Rodríguez (Colegio de Postgraduados)

Vocal: Dra. Anna Sikov (Universidad Federal Rural de Pernambuco)

Vocal y Director de la Tesis: Dr. Rogelio Ramos Quiroga (CIMAT).

Asesor:

Dr. Rogelio Ramos Quiroga

Sustentante:

Georges Bucyibaruta

Abstract

The demand for reliable small area estimates derived from survey data has increased greatly in recent years due to, among other factors, their growing use in formulating policies and programs, allocation of government funds, regional planning, small area business decisions and other applications.

Following the definition given by Rao (2003), the term small area or small domain refers to a subpopulation for which the domain-specific sample is not large enough to produce direct estimates with reliable precision. This subpopulation can be a small geographical area (county, state, district, etc.), a demographic group within a geographical region (specific sex-age group, etc.) or any subdivision of the population. Most surveys provide little information on individual small areas since they are generally designed to produce accurate estimates at a higher level of aggregation. Small area estimation methods are well suited for settings that involve many domains, with small (or no) samples from individual domains. In this setting, traditional design-based direct survey estimates based only on samples from individual small areas are not reliable. In order to improve on the traditional estimates based on individual area sample, one may "borrow strength" from neighboring or related small areas, or other correlated dependent variables and relevant covariate information available from other sources, such as administrative records, to produce accurate small area estimates (Molina and Rao, 2015).

Most of the methods proposed in literature, to estimate small-area quantities, assume that the variable of interest follows a linear model and the linking covariates are available at population element (or observational unit) level or area level (Pereira and Coelho, 2012; Marhuenda et al., 2013; Pfeffermann, D., 2013; Petrucci and Pratesi, 2014; Berg and Chandra, 2014; Rao and Molina , 2015). This assumption is not common, as it is plausible that some variables of interest in various surveys can be skewed distributed (Molina and Rao, 2010; Karlberg, 2014). Besides, it is not always easy to link the covariates obtained from other sources (censuses and/or other surveys) to those associated with the characteristic of interest (Datta and Ghosh, 1991).

In this thesis, we consider small area estimation (SAE) techniques focusing primarily on estimating and predicting skewed (lognormal) distributed data. A brief review on the theory of Linear Mixed Models is given. Estimation of a small area population mean under a two-fold nested error lognormal model is proposed. Closed form expressions for an empirical Bayes (EB) predictor and its bias-corrected estimator, as well as approximated analytical expression of the associated mean squared error (MSE) together with its bootstrap bias-corrected estimator, are obtained. To improve the performance of the simple bootstrap, we propose a double parametric bootstrap method for bias-correction. Under simulation experiments, the bias of empirical Bayes predictors is studied. We demonstrate that the suggested predictor, under the assumption of lognormal model, behaved well according to prediction.

Dedication

This Thesis is dedicated
To
God Almighty,
my mother,
my sisters and brothers,
and the memory of my father.

Acknowledgements

Throughout the period of my study, I have had cause to be grateful for the advice, support and understanding of many people. In particular I would like to express my boundless gratitude to my supervisor Dr. Rogelio Ramos Quiroga for helping me get into the problem of small area estimation and kindly direct this work. He has guided me in my research through the use of his profound knowledge, insightful thinking, thoughtful suggestions as well as his unflagging enthusiasm. I really appreciate the enormous amount of time and patience he has spent on me.
This work has also benefited greatly from the advice, comments and observations of Dr. Lilia Letitia Ramírez Ramírez, Dr. Anna Sikov, Dr. Paulino Pérez Rodrguez, Dr. Enrique Raúl Villa Diharce and Dr. Elas Rodríguez Muñoz.

As well as the people mentioned above, my special thanks go to members of department of probability statistics of CIMAT for the interesting lectures they offered to us and their consistent support and encouragement throughout my study. A mention must also go to my wonderful classmates, current and former students and the CIMAT community in general who provided pleasant atmosphere and useful distraction during my stay in Guanajuato.

I gratefully acknowledge the admission to CIMAT and financial support from El Consejo Nacional de Ciencia y Tecnología (CONACYT), México with which it has been possible to realize this achievement.

Finally but certainly not least, I want to remember my entire family in Rwanda whose unlimited support, prayers, and encouragement carried me through times of despondency and threatening despair.

My sincere thanks to all.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This preface introduces the general idea of small area estimation, focusing on different works of various authors and shows the importance of small area estimation in a wide variety of situations.

1.1 Background and current status of the topic

Public and private sectors use survey information, provided by statistical agencies, to support government policies and business decisions, position a product on the market, etc. Typically these surveys are designed to produce information for the target sampling population and for large population subgroups. The standard sampling estimates for large population subgroups of a finite population are called design-based estimates or direct estimates because the estimation is based only on the sampling data and the selection probabilities for the sample in the subgroup of interest. Unlike what happens in most areas of statistics, the statistical inference under a sampling design does not depend on the validity of a statistical model (Sarndal et al., 1992). This situation is advantageous in its estimation, since it is not necessary to suppose an a prior model of the data. However, inference under a sampling design becomes problematic when the sample size in the subgroups of interest is very small or even zero.

Following the definition given by Rao (2003), the term small area or small domain refers to a subpopulation for which the domain-specific sample is not large enough to produce direct estimates with reliable precision. This subpopulation can be a small geographical area (county, state, district, etc.), a demographic group within a geographical region (specific sex-age group, etc.) or any subdivision of the population.

Most surveys provide little information on individual small areas since they are generally designed to produce accurate estimates at a higher level of aggregation. Small area estimation methods are well suited for settings that involve many domains, with small (or no) samples from individual domains. In this setting, traditional design-based direct survey estimates based only on samples from individual small areas are not reliable. In order to improve on the traditional estimates based on individual area sample, one may "borrow strength" from neighboring or related small areas, or other correlated dependent variables and relevant covariate information available from other sources, such as administrative records, to produce accurate small area estimates (Molina and Rao, 2015).

Due to the lack of precision of the direct estimators of small area parameters, new estimation procedures have been developed. A number of model-based methods have been proposed and are increasingly needed to adjust the variables of interest and associated covariates to produce so-called small area estimators or indirect estimators. Rao (2003), Pfeffermann (2013) and Rao and Molina (2015) give comprehensive and detailed description on the theory of small area estimation.

The statistical techniques of small area estimation (SAE) have been a topic of great interest and focus for various authors in the context of sample surveys. There is an over-growing demand for reliable estimates of small area populations of different types. Over the past three decades, this demand has increased in different areas of application, including income and expenditure, poverty, education, health, and agriculture (Rao and Molina, 2015; Pfeffermann, 2013; Esteban et al., 2012; Molina and Rao, 2010; Rao, 2003; Battese et al., 1988; Fay and Herriot, 1979). The main reason of this growth in need is the recent trend involving social objective policies and economic programs of different countries at a more local level. Some of these policies and programs include determination of state funding allocations, formation and evaluation of policies and programs, regional planning, administrative planning, and disease mapping (Rao and Molina, 2015).

One of the objectives of SAE is to provide estimators that, without increasing the size of the sample, allow to obtain good estimators; i.e, the estimation error is smaller than that of the direct estimators. An appropriate consideration of the response variable given the available covariates is extremely important. When covariates related to the variable of interest are available at the unit level, a model, widely used in the literature of small area estimation and in numerous applications, is the basic Battese et al. (1988) model. This model is a one-fold nested error linear model. It was used to obtain empirical best linear unbiased prediction (EBLUP) estimates of areas under corn and soyabeans for each of the 12 counties in North-Central Iowa under normal distribution assumption.

Small area methods based on the ideas of non-normal distributions have been considered by some authors. Slud and Maiti (2006) proposed an empirical Bayes (EB) or best predictor for a small area mean assuming that the area-level direct estimators have a lognormal distribution. Ghosh and Maiti (2004) discussed a small area unit-level model based on natural exponential quadratic variance function families, where they assumed that the covariates are the same across units in a single small area. Chandra and Chambers (2011) considered a lognormal distribution as a basis for constructing a model-based direct estimator for a small area mean. Their model is a weighted sum of sampled units in which the weights are defined to give the minimum mean squared error linear predictor of the population mean when the parameters of the lognormal distribution were known. Berg and Chandra (2014) proposed a model-based indirect estimator that follows the framework of Chandra and Chambers (2011) under the assumption that the covariates are available at the unit-level.

Small area estimation procedures under two-fold nested error linear models, appropriate for two-stage sampling in each of the small areas, were considered in different literature. Datta and Ghosh (1991) and Pfeffermann and Barnard (1991) used the two-fold model for the special case of cluster-specific covariates. In estimating the model mean squared error, Stukel and Rao (1999) extended the results of Prasad and Rao (1990) and Stukel (1991) to general two-fold nested error regression models, considering the unit-level covariates to be available.

The work presented in this thesis can be divided into two broad categories. In the first, under a log-normal distribution, we propose a two-fold nested error lognormal model. This model extends the number of models previously considered in this area (Stukel and Rao 1999, Datta and Ghosh 1991, Pfeffermann and Barnard 1991, Berg and Chandra 2014). The proposed model is different from that of Berg and Chandra (2014) because we work with a two-fold nested error instead of one-fold nested error data. It is also different from Stukel and Rao (1999), since their model is a two-fold nested error under a linear case. Furthermore, considering that we need to combine information from different sources where it is not possible to find an connector item (identifier) between information regarding each observational unity from those sources, we adapt the case of cluster-specific covariates as in Datta and Ghosh

(1991) and Pfeffermann and Barnard (1991). In the second category, we derive the closed-form expression for mean squared error (MSE). We derive its bias-corrected estimator based on a bootstrap approach and we provide its approximated expression under a double parametric bootstrap. The suggested methods are computationally simple because the predictor and mean squared error estimator have closed form expressions.

1.2 Objectives of the thesis

1.2.1 General objectives of the thesis

Most of the methods proposed in literature, to estimate small-area quantities, assume that the variable of interest follows a linear model and the linking covariates are available at population element (or observational unit) level or area level (Pereira and Coelho, 2012; Marhuenda et al., 2013; Pfeffermann, D., 2013; Petrucci and Pratesi, 2014; Berg and Chandra, 2014; Rao and Molina , 2015). This assumption is not common, as it is plausible that some variables of interest in various surveys can be skewed distributed (Molina and Rao, 2010; Karlberg, 2014). Besides, it is not always easy to link the covariates obtained from other sources (censuses and/or other surveys) to those associated with the characteristic of interest (Datta and Ghosh, 1991).

Under the assumption that a population charateristic of interest is skewed (log-normal) distributed, and focusing on the model-based theory of estimation in small areas, the research carried out in this thesis aims to propose a methodology for obtaining a predictor of population mean at small-area level. Because of the potentially large effect of decisions that are made by using survey results, it is important that estimates be reported together with their precisions, and the mean-squared error (MSE) is the measure of precision which has become standard in the small area field (Rao, 2003). For bias-corrected empirical Bayes predictor small area point estimators, mean-squared error formulae and estimators are provided. The behavior of these mean-squared error estimators is illustrated by a simulation study.

1.2.2 Specific objectives

The objectives on which we will focus our work are the following:

i. To estimate population mean of the characteristic of interest at small-area level when the sample data was obtained from a two-stage sampling design.

ii. In the case of skewed data, to extend a two-fold nested error model by including random effects explaining heterogeneity at the two levels of aggregation.

iii. To propose a small area estimation method for skewed data under the assumption that a lognormal model is a reasonable distribution of the original response variable given cluster-level covariates.

iv. To derive the closed form expressions for an empirical Bayes (EB) predictor and the associated mean squared error estimator.

1.3 Content of the thesis

This thesis investigates the applicability of mixed lognormal models with two nested random factors to estimate small area parameters. Public statistics particularly socioeconomic statistics are interested in the use of this type of models. However, there are also relevant applications in the field of environmental statistics, in modeling of agricultural data, as mentioned above.

The present document consists of six chapters and appendices distributed as follows:

In Chapter 1, we introduce the notion of small area estimation, providing a comprehensive description of the background and current status of the topic. We also highlight the general as well as the specific objectives of the thesis.

In Chapter 2 we give an overview of approaches to small area estimation, explaining the two types of small area estimation: direct and indirect estimation.

In Chapter 3, we briefly review general concepts pertaining to the theory of linear mixed models necessary to address later chapters, and stress the general theory of prediction, in the methodology of fitting a linear mixed models and obtaining mean square error of predictors.

In Chapter 4, a new model for small area estimation, two-fold nested error lognormal, is developed and the restricted maximum likelihood approach is proposed to estimate the regression parameters. We adapt the Fisher-scoring algorithm to estimate the vector parameter of components of variance. We apply the general theorem of prediction to obtain the optimal predictor of the variable of interest, known as best linear unbiased predictor (BLUP) or EBLUP when the components of variance are known or unknown respectively.

We perform a simulation experiment that allows us to analyze the behavior of these estimators.

In Chapter 5, we apply the corresponding theory to the estimation of mean square error of the EBLUP. Extending this theory to the model under study, we obtain the closed form expressions of mean squared error (MSE), and provide its bias-corrected estimator based on a bootstrap approach and we propose its approximated expression under a double parametric bootstrap. In simulation studies we compare those two approaches.

The thesis concludes with a discussion regarding the methods presented and possible future research projects inspired by Chapter 6 of this work.

The appendices detail the intermediate calculations which lead to the expressions described in Chapters 4 and 5 respectively.

Chapter 2

Approaches to small area estimation

Up to now, in the context of SAE, different methods of estimation have been developed, designed and model based (see Rao and Molina, 2015). Traditionally there are two types of small area estimation namely direct and indirect estimation. The direct small area estimation is based on survey design and includes the Horvitz - Thompson (HT) estimator, generalized regression (GREG) estimator and modified direct estimator. On the other hand, indirect approaches are mainly based on different statistical models and techniques. Implicit model based approaches include synthetic and composite estimations; whereas explicit models are categorized as area level and unit level models.

2.1 Direct estimators

Following the definition given by Rao (2003), a small area estimator is direct when it uses the sample values of study variables from the specified area only. In general, the direct estimators are unbiased estimators, however, due to the small sample size, such estimators might have unacceptably large standard errors. Direct estimates are classical design-based estimators that are obtained by applying survey weights to the sample units in each small area, Saei and Chambers (2003). For direct estimation all small areas must be sampled in order to produce these kinds of estimates. The following two estimators are common in direct estimation.

2.1.1 Horvitz-Thompson (HT) estimator

The Horvitz-Thompson is the simplest direct estimator. Suppose that a finite survey population U, consists of N distinct elements identified through the labels $1, 2, \ldots, N$ is divided into M disjoint sub-populations (or small areas), U_i, of size N_i each, and $\sum_{i=1}^{M} N_i = N$. Consider a sample s ($s \subseteq U$) drawn from U with a given probability design $p(.)$, and s_i ($s_i \subset s$) is the set of individuals that have been selected in the sample from small area i. Suppose that the inclusion Probability $\pi_k = P(k \in s_i)$ is strictly positive and known. For the elements $k \in s_i$, let (y_{ik}, x_{ik}) be a set of sample observations, where y_{ik} is the value of the characteristic (or variable) of interest for the k^{th} unit in the small area i and x_{ik} is the vector of covariates associated with y_{ik}.

Now, if Y_i and X_i represent the target variable and the available covariates for small area i respectively, the Horvitz-Thompson estimator, Sarndal et al. (1992), of the population total for i^{th} small area can be defined as

$$\hat{Y}_{i,HT} = \sum_{k \in s_i} w_{ik} y_{ik},$$

where $w_{ik} = \frac{1}{\pi_{ik}}$, $k \in s_i \subset s$, are design weights depending on the given probability sampling design $p(.)$. In the context of small area estimation problems, with an inadequate sample, HT estimator can be design-biased and more unreliable (Petrucci and Pratesi, 2014).

2.1.2 Generalized regression (GREG) estimator

The generalized regression (GREG) estimator is obtained by combining the individual sample information from the survey data with covariates (X_i). The GREG estimator of population total is defined as follows

$$\hat{Y}_{i,GREG} = X_i^T \hat{\beta} + \sum_{k \in s_i} w_{ik}(y_{ik} - x_{ik}^T \hat{\beta}),$$

with

$$\hat{\beta} = \Big(\sum_{k \in s_i} w_{ik} x_{ik} x_{ik}^T\Big)^{-1} \Big(\sum_{k \in s_i} w_{ik} x_{ik} y_{ik}^T\Big).$$

Where $\hat{\beta}$ is the sample weighted least square estimates of generalized regression, (Sarndal et al. 1992, Rao 2003). The GREG estimator is approximately design-unbiased for small area estimation but it is not consistent because of high residuals

(Rao, 2003, chap.2). Rao (1999), gives a detailed discussion about the GREG estimator in the context of small area estimation.

2.1.3 Modified direct estimator

If the covariates X_i in the i^{th} small area are available, the modified direct estimator, known also as Regression estimator, can be used to improve the reliability of direct estimators by borrowing strength for estimating the regression coefficient over small area. The modified direct estimator of population total is given by

$$\hat{Y}_{i,R} = \hat{Y}_i + (t_{X_i} - \hat{X}_i)^T \hat{\beta},$$

where $\hat{Y}_i$, $\hat{X}_i$ are the HT estimators of the target variable Y_i and covariates X_i respectively, t_{X_i} is the population total of covariates X_i for the small area i, and

$$\hat{\beta} = \Big(\sum_{k\in s} w_k x_k x_k^T\Big)^{-1}\Big(\sum_{k\in s} w_k x_k y_k^T\Big),$$

is the overall sample weighted least square estimates of the regression coefficients. The modified direct estimator is approximately design-unbiased as the overall sample size increases, even if the regional sample size is small. Although the modified direct estimator borrows strength for estimating the overall regression coefficients, it does not increase the effective sample size, unlike indirect small area estimators, Rao,(2003, chap. 2). This estimator is also referred to in Battese et al. (1988) as the modified GREG estimator or the "survey regression" estimator.

2.2 Indirect estimation

Indirect, or model-based small area estimators, rely on statistical models to provide estimates for all small areas. Once the model is chosen, its parameters are estimated using the data obtained in the survey. An important issue in indirect small area estimation is that covariates are needed (Rao, 2003, chap. 4).

2.2.1 Synthetic estimation

The synthetic estimator is an example of an estimator, which can be considered either model-based or design-based model-assisted. In both cases the specified linear relationship between y (study variable) and the covariates, described with the parameter β (vector of regression coefficients) plays an important role.

The name synthetic estimator derives from the fact that these estimators borrow strength by synthesising data from many different areas. Gonzalez (1973) defines an estimator as synthetic when a reliable direct estimator for a large area is used to derive an indirect estimator for a small area belonging to the large area under the assumption that all small areas have the similar characteristics as the large area. In addition, Rao (2003) provides extensive overviews on various synthetic estimation approaches in small areas estimation.

- Under the assumption that covariates are not available, the synthetic estimator of population total is given by

  $$\hat{Y}_{iS} = \hat{Y},$$

  where $\hat{Y}$ is the direct estimator of the overall population total, $\hat{Y} = \sum_s w_k y_k$.

- If domain-specific covariates are available in the form of known totals X_i, the regression-synthetic estimator of the population total is defined as

  $$\hat{Y}_{iGRS} = X_i^T \hat{\beta},$$

  where

  $$\hat{\beta} = \Big(\sum_{k \in s} c_k^{-1} x_k x_k^T \Big)^{-1} \Big(\sum_{k \in s} c_k^{-1} x_k y_k^T \Big),$$

- A special case of the regression-synthetic is the ratio-synthetic estimator in the case of a single covariate. By letting $c_k = x_k$, the ratio-synthetic estimator of population total is given by

  $$\hat{Y}_{iRS} = X_i \frac{\hat{Y}_i}{\hat{X}_i},$$

  where $\hat{Y}_i$, $\hat{X}_i$ are the HT estimators of the target variable Y_i and covariates X_i respectively and t_{X_i} is the population total of covariates X_i for the small area i.

2.2.2 Composite estimation

As we mentioned, as the sample size in a small area increases, a direct estimator becomes more desirable than a synthetic estimator. This means, when area level sample sizes are relatively small the synthetic estimator outperforms the traditional direct estimator. Synthetic estimators have a big influence of information from the other areas, thus they may have small variance but a large bias in the case where the hypothesis of homogeneity is not satisfied.

According to Rao (2003), to avoid the potential bias of a synthetic estimator, say $\hat{Y}_{iS}$ and the instability of the direct estimator, say $\hat{Y}_{iD}$, we consider a convex combination of both, known as the composite estimator.

$$\hat{Y}_{iC} = \omega_i \hat{Y}_{iS} + (1 - \omega_i)\hat{Y}_{iD},$$

for a suitable chosen weight ω_i ($0 \leq \omega_i \leq 1$), where c, s, and d stand for composite, synthetic and direct, respectively. The choice of optimal weight ω_i^{opt} can be obtained by minimizing the mean square error (MSE) of the composite estimator,$\hat{Y}_{ic}$, with respect to ω_i,(Ghosh and Rao, 1994; Rao, 2003). This yields

$$\omega_i^{opt} = \frac{MSE(\hat{Y}_{iS})}{MSE(\hat{Y}_{iD}) + MSE(\hat{Y}_{iS})}.$$

A number of estimators proposed in literature have the form of composite estimators, for instance the James-Stein estimator proposed by James and Stein (1961) which considers common weight ω. Efron and Morris (1975) have generalized the James-Stein estimator. Composite estimators are biased and they may have improved precision depending on the selection of the weight.

2.3 Small area models

2.3.1 Introduction

Traditional methods of indirect estimators, mentioned above, are based on implicit models (synthetic and composite). We now turn to explicit linking models which provide significant improvements in techniques for indirect estimation. Based on mixed model methodology, these techniques incorporate random effects into the model. The random effects account for the between-area variation that cannot be explained by including covariates. Most small area models can be defined as an area-level model or a unit-level model (Rao, 2003, Chap. 5).

2.3.2 Basic area level model

The area level model relates the small area information on the response variable to area-specific covariates. One of the most widely used area level models for small area estimation was proposed by Fay and Herriot (1979). According to the Fay-Herriot model, a basic area level model assumes that the small area parameter of interest η_i is related to area-specific covariates x_i through a linear model

$$\eta_i = x_i^t \beta + \upsilon_i, \; i = 1, 2, \ldots, m, \tag{2.1}$$

where m is the number of small areas, $\beta = (\beta_1, \beta_2, \ldots, \beta_p)^t$ is $p \times 1$ vector of regression coefficients, and the υ_i's are area-specific random effects assumed to be independent and identically distributed (iid) with $\mathrm{E}(\upsilon_i) = 0$ and $\mathrm{Var}(\upsilon_i) = \sigma_\upsilon^2$, model expectation and model variance respectively. Normality of the random effects υ_i is also often used, but it is possible to make robust inferences by relaxing the normality assumption (Rao, 2003).

The area level model assumes that there exists a direct survey estimator $\hat{\eta}_i$ for the small area parameter η_i such that

$$\hat{\eta}_i = \eta_i + \epsilon_i, \; i = 1, 2, \ldots, m, \tag{2.2}$$

where the ϵ_i is the sample error associated with the direct estimator $\hat{\eta}_i$, with the assumptions that the ϵ_i's are independent normal random variables with mean $\mathrm{E}(\epsilon_i) = 0$ and $\mathrm{Var}(\epsilon_i) = \tau_i$. Combining these two equations, yields the area level linear mixed model

$$\hat{\eta}_i = x_i^t \beta + \upsilon_i + \epsilon_i. \tag{2.3}$$

2.3.3 Basic unit level model

Unit level models relate the unit values of the study variable to unit-specific covariates. These variables are related to the unit level values of response through a linear mixed model known as nested error linear regression model. This type of model can be represented by the following equation

$$y_{ij} = x_{ij}^t \beta + \nu_i + \epsilon_{ij}, \tag{2.4}$$

where y_{ij} is the response of unit j, $j = 1, 2, \ldots, n_i$, in area i, $i = 1, 2, \ldots, m$, x_{ij} is the vector of covariates, β is the vector of regression parameters, ν_i is the random effect of area i and ϵ_{ij} is the individual unit error term. The area effects ν_i are assumed independent with mean zero and variance σ^2_ν. The errors ϵ_{ij} are independent with mean zero and variance σ^2_ϵ. In addition, the ν_i and ϵ_{ij}'s are assumed to be independent.

The nested error unit level regression model (2.4) was first used to model county crop areas in North-Central Iowa, USA (Battese et al., 1988).

2.4 Theory of prediction under the general linear model

The model-based approach to finite population theory treats the population vector $\mathbf{y} = (y_1, y_2, \ldots, y_N)^T$ as a realization of a random variable Y. Let a model ξ characterizes the probability distribution of Y. The focus is to estimate the value of a population quantity of interest, which can be seen as function $h(\mathbf{y})$ of $\mathbf{y}$, typically a linear combination $\mathbf{c}^T\mathbf{y}$. If $\mathbf{c} = \mathbf{1}$, where $\mathbf{1}$ is an unit vector, then $h(\mathbf{y})$ is the population total, and if $\mathbf{c} = \mathbf{1}/N$, then $h(\mathbf{y})$ is the population mean.

Now suppose that the interest is to estimate the total population of the variable of interest for area i

$$Y_i = \sum_{j \in U_i} y_{ij}.$$

Following Valliant et al. (2000) and Royall (1976), let s denote a sample of size n from a finite population U and let r denote the nonsampled remainder of U so that $U = s \cup r$. Correspondingly, let $\mathbf{y}_s$ be the vector of observations in the sample and $\mathbf{y}_r$ the rest of $\mathbf{y}$. We have the following decomposition:

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_s \\ \mathbf{y}_r \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} \mathbf{c}_s \\ \mathbf{c}_r \end{pmatrix}.$$

The population quantity to be estimated is now

$$h(\mathbf{y}) = \mathbf{c}_s^T \mathbf{y}_s + \mathbf{c}_r^T \mathbf{y}_r$$

a realization of the random variable

$$h(Y) = \mathbf{c}_s^T Y_s + \mathbf{c}_r^T Y_r.$$

Note that the first term $\mathbf{c}_s^T\mathbf{y}_s$ is observed from the sample, whereas the second term must be estimated (or predicted, since it is a function of random variables, not a fixed parameter). Thus, estimating $h(\mathbf{y})$, or predicting $h(Y)$, is essentially predicting the value $\mathbf{c}_r^T\mathbf{y}_r$ of the unobserved random variable $\mathbf{c}_r^T Y_r$. The information needed in the prediction will come from the sample vector Y_s, and the predictor (or estimator) of $h(Y)$ can be written

$$\hat{h}(Y) = c_s^T Y_s + a^T Y_s,$$

where a is some $n \times 1$ vector defining the predictor $a^T Y_s$ of $c_r^T Y_r$.
The estimator $\hat{h}$(y)is model unbiased, if

$$E_\xi[\hat{h}(y) - h(y)] = 0,$$

and the model-based error variance is

$$var_\xi[\hat{h}(y) - h(y)] = E_\xi[\hat{h}(y) - h(y)]^2.$$

The general prediction theorem gives the best linear unbiased predictor (BLUP) of $c^T Y$ as well as its error variance under the general linear model, in the case of finite population. The best linear unbiased predictor means here a model-unbiased predictor, which is linear in Y_s and has the minimum model-based error variance among all linear unbiased predictors. The theorem serves as a general basis of the BLUP approach to small area estimation with unit level models.
Either we may consider only the population U_i of an area i, which leads to the direct estimation, or we may consider the overall population U (and a sample from it), but define the coefficient vector c so that it picks only those elements of y, which come from the area i. This leads to the indirect estimation with ability to borrow strength. At area level it is more straightforward to develop the BLUP approach within the standard theory of linear mixed models (Henderson, 1975).

Lets consider ξ to be a general linear model in a such way that

$$\mathrm{E}_\xi(Y) = X\beta$$

and

$$\mathrm{Var}_\xi(Y) = V.$$

In accordance with the partition $U = s \cup r$ we can arrange X and V so that

$$X = \begin{pmatrix} X_s \\ X_r \end{pmatrix}$$

and

$$V = \begin{pmatrix} V_s & V_{sr} \\ V_{rs} & V_r \end{pmatrix},$$

where X contains covariates, β is a vector of unknown parameters and V is an arbitrary positive definite covariance matrix. The model ξ covers a variety of special cases, including the linear mixed models. The theorem does not require normality.

The general prediction theorem.

Under the model ξ for a finite population U the best linear model-unbiased predictor of $h(Y) = c^T Y$ is

$$\text{BLUP}(c^T Y) = c_s^T \mathbf{y}_s + c_r^T [X_r \hat{\beta} + V_{rs} V_s^{-1} (\mathbf{y}_s - X_s \hat{\beta})], \tag{2.5}$$

and the error variance is

$$\begin{aligned} \text{Var}_\xi[\text{BLUP}(c^T Y) - c^T Y] =& c_r^T (V_r - V_{rs} V_s^{-1} V_{sr}) c_r + \\ & c_r^T (X_r - V_{rs} V_s^{-1} X_s)(X_s^T V_s^{-1} X_s)^{-1} (X_r - V_{rs} V_s^{-1} X_s)^T c_r, \end{aligned}$$

where

$$\hat{\beta} = (X_s^T V_s^{-1} X_s)^{-1} X_s^T V_s^{-1} \mathbf{y}_s$$

is the general least squares (GLS) estimator of β.

Note that the GLS estimator is also the best linear unbiased estimator (BLUE) of β, i.e., it has the minimum variance among linear unbiased estimators (McCulloch and Searle, 2001).

Proof:
As we are interested in predicting the part which is not observed, the information needed will come from observed variables (sample vector $\mathbf{y}_s$). Considering the decomposition of the population quantity of interest

$$h(Y) = \mathbf{c}_s^T Y_s + \mathbf{c}_r^T Y_r,$$

the random term is $\mathbf{c}_r^T Y_r$. Now suppose the predictor of this term is a linear function of the data, i.e. the predictor $\mathbf{c}_r^T Y_r$ is of the form $\mathbf{a}^T Y_s$ where $\mathbf{a}$ is a vector to be specified. In addition the estimator $\hat{h}(Y)$ is assumed to be unbiased. Now, define

$$\begin{aligned} \mathrm{E}_\xi(a^T Y_s - c_r^T Y_r)^2 &= \mathrm{Var}(a^T Y_s - c_r^T Y_r) + (\mathrm{E}_\xi(a^T Y_s - c_r^T Y_r))^2 \\ &= a^T V_s a + c_r^T V_r c_r - 2a^T V_{sr} c_r + [(a^T X_s - c_r^T X_r)\beta]^2, \end{aligned} \tag{2.6}$$

the last expression is obtained from the expression of V and the fact that $\mathrm{E}(Y) = X\beta$.

Now the BLUP$(c^T Y)$ will be found by minimizing the $\mathrm{E}_\xi(a^T Y_s - c_r^T Y_r)^2$ with respect to a under the constraint of model unbiasedness,

$$\mathrm{E}_\xi(a^T Y_s - c_r^T Y_r) = (a^T X_s - c_r^T X_r)\beta = 0,$$

for all β, which is equivalent to

$$a^T X_s - c_r^T X_r = 0.$$

Lets consider the following Lagrangian function

$$\mathcal{L}(a, \lambda) = a^T V_s a - 2a^T V_{sr} c_r + 2(a^T X_s - c_r^T X_r)\lambda,$$

where λ is the vector of Lagrange multipliers.

The first partial derivative respect a a is

$$\frac{\partial \mathcal{L}(a, \lambda)}{\partial a} = 2V_s a - 2V_{sr} + 2X_s\lambda,$$

equating to zero, yields

$$X_s\lambda = V_s a - V_{sr}, \tag{2.7}$$

and then

$$a = V_s^{-1}(V_{sr} c_r - X_s\lambda). \tag{2.8}$$

Now multiply (2.7) by $X_s^T V_s^{-1}$ yields $X_s^T V_s^{-1} X_s\lambda = X_s^T V_s^{-1} V_s a - V_{sr}$, using the unbiasedness constraint $a^T X_s = c_r^T X_r$ and then solving for λ, we get

$$\lambda = (X_s^T V_s^{-1} X_s)^{-1}(X_s^T V_s^{-1} V_{sr} - X_r^T)c_r,$$

then after substituting this expression in (2.8) and making some simplifications, we get

$$a = V_s^{-1}[V_{sr} - X_s(X_s^T V_s^{-1} X_s)^{-1}(X_s^T V_s^{-1} V_{sr} - X_r^T)]c_r. \tag{2.9}$$

Recall that $\hat{h}(Y) = c_s^T Y_s + a^T Y_s$. Then substituting a with (2.9), yields

$$\begin{aligned}
\text{BLUP}(c^T Y) &= c_s^T Y_s + (V_s^{-1}[V_{sr} - X_s(X_s^T V_s^{-1} X_s)^{-1}(X_s^T V_s^{-1} V_{sr} - X_r^T)]c_r)^T Y_s \\
&= c_s^T Y_s + c_r^T [V_{sr} - X_s(X_s^T V_s^{-1} X_s)_{-1}(X_s^T V_s^{-1} V_{sr} - X_r^T)]^T V_s^{-1} Y_s \\
&= c_s^T Y_s + c_r^T [V_{sr}^T V_s^{-1} Y_s - (X_s^T V_s^{-1} V_{sr} - X_r^T)^T (X_s^T V_s^{-1} X_s)^{-1} X_s^T V_s^{-1} Y_s] \\
&= c_s^T Y_s + c_r^T [V_{sr}^T V_s^{-1} Y_s - (X_s^T V_s^{-1} V_{sr} - X_r^T)^T \hat{\beta}] \\
&= c_s^T Y_s + c_r^T [X_r \hat{\beta} + V_{sr}^T V_s^{-1}(Y_s - X_s \hat{\beta})],
\end{aligned}$$

where $\hat{\beta} = (X_s^T V_s^{-1} X_s)^{-1} X_s^T V_s^{-1} Y_s$.

The expression of error variance is found in the same way by inserting (2.9) into (2.6) •

The inference concerning the BLUP of $c^T y$ usually appeals to the normal distribution. Valliant et al., (2000) give fairly reasonable conditions, under which the BLUP is asymptotically normal.

Chapter 3

Review of mixed model theory

3.1 Introduction

It is not intended to be an all-encompassing exposition on the subject, the rationale is to briefly explore the methods used for parameter estimation throughout the thesis. The model-based small area estimation largely employs linear mixed models involving random area effects. The covariates are introduced in the fixed part of the model as covariates.
Linear mixed models have a wide range of applications. In particular, the ability to predict linear combination of fixed and random effects is one the more attractive properties of such models. In a series of papers, Henderson (1975) developed the best linear unbiased prediction (BLUP) method for mixed models. However, the BLUP methods described in Henderson (1975) assumed that the variances associated with random effects in the mixed model (the variance components) are known. In practice such variance components are unknown and have to be estimated from the data. There are several methods for estimating variance components reviewed in Harville (1977). The predictor obtained from the BLUP when unknown variance components are replaced by associated estimators is called the empirical best linear unbiased predictor (EBLUP) and is described in Robinson (1991).

3.2 Linear mixed model

The ordinary fixed effects linear model is usually written as

$$y = X\beta + e, \tag{3.1}$$

where y is an $n \times 1$ random vector of response variable, β is a $p \times 1$ vector of regression coefficients, X is a known $n \times p$ model matrix containing the values of the explanatory (or predictors) variables and e is the vector of random errors. The normal distribution is assumed to e, with $E(e) = 0$ and $cov(e) = \sigma_e^2 I_n$, where I_n is the $n \times n$ identity matrix, to make

$$y \sim N_n(X\beta, \sigma_e^2 I_n).$$

Hence the vector y contains random variables that are independent with equal variability.
The linear mixed model is obtained by incorporating a $q \times 1$ vector v of random effects, i.e. effects that are considered random variables instead of fixed constants, with an appropriate model matrix Z into the fixed effects model 3.1

$$y = X\beta + Zv + e. \tag{3.2}$$

The model matrix Z is often an incidence matrix (design matrix) of zeros and ones only, but it may also contain explanatory variables (that usually are present also in X). In the latter case the model (3.2) is often called random coefficient regression model. If Z is an incidence matrix and the random effects in v are uncorrelated, we have a special case called variance component model.
For the random vectors v and e we make the following assumptions:

$$\begin{aligned} E(v) = 0, \quad & cov(v) = G \\ E(e) = 0, \quad & cov(e) = R \\ & cov(v, e) = 0, \end{aligned}$$

where, in principle, G and R can be arbitrary positive definite covariance matrices. In variance component models the matrices G and R are diagonal.
Under these assumptions the expected value of y is

$$E(y) = X\beta,$$

and if v is given

$$E(y|v) = X\beta + Zv.$$

The covariance matrix of y is

$$cov(y) = V = ZGZ^T + R,$$

and if v is given,

$$cov(y|v) = R.$$

Note that the fixed part $X\beta$ of model (3.2) defines the mean structure of y, whereas the random part $Zv + e$ defines the covariance structure. The covariance matrices G and R are functions of a set of variance parameters (σ). Therefore we can write $G = G(\sigma)$, $R = R(\sigma)$ and $V = V(\sigma)$.
The normal distribution is usually assigned to both random terms v and e so that

$$\begin{pmatrix} v \\ e \end{pmatrix} \sim \mathbf{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} G & 0 \\ 0 & R \end{pmatrix}\right).$$

The model with the above assumptions on v and e implies that y are multivariate normal random vectors of dimension n with a particular form of covariance matrix. That is

$$y \sim \mathbf{N}(X\beta, V). \tag{3.3}$$

This is called the marginal formulation of linear mixed model (Verbeke and Molenberghs, 2000). Note that although the marginal model (3.3) follows from the linear mixed model (3.2), the models are not equivalent, because the marginal model does not explicitly define the random effect structure in (3.2). The marginal model, however, gives the basis of the maximum likelihood estimation of the model parameters.

3.3 Estimation of model parameters

3.3.1 ANOVA method

For variance component models, where all the parameters in σ are variances, the classical method for estimating σ is so-called ANOVA estimation, based on equating the mean squares of the analysis of variance to their expected values and solving the estimates from the resulting equations. The ANOVA estimation, originally meant for balanced data, was adapted to unbalanced data by Henderson in 1950s. The Henderson 3 method, also known as method of fitting constants, has still fairly recently been suggested to be used with small area models (Prasad and Rao, 1990). ANOVA methods are non-iterative and therefore easy to implement, give unbiased variances estimates (which sometimes appear negative, though) and they require no normality of random effects in the model. Their major drawback is that they only

apply to a limited choice of models.
The regression coefficients β can be estimated by generalized least square

$$\hat{\beta}_{GLS} = (X^T V^{-1} X)^{-1} X^T V^{-1} y. \tag{3.4}$$

The estimator $\hat{\beta}_{GLS}$ is generally the best linear unbiased estimator (BLUE) of β. If the covariance matrix $V(\sigma) = ZG(\sigma)Z^T + R(\sigma)$ is unknown, it will be replaced with its estimate $\hat{V} = V(\hat{\sigma})$, where $\hat{\sigma}$ is obtained by the ANOVA method, for example. Anova methods are only applicable to limited choice of models (Searle et al. 1992). In order to overcome this, we need to use maximum likelihood (ML) or restricted maximum likelihood (REML) estimation methods, which are applicable to more general models and boast of attractive properties like consistency, efficiency and asymptotic normality of the estimators. These solutions have two troublesome properties. First, unlike simple cases of models, the ML vector of fixed effects $\hat{\beta}$ is a function of the variance matrix $\hat{V}$, which in turn contains the variance components that we wish to estimate. Second, because these solutions involve the inverse of $\hat{V}$, they are nonlinear functions of the variance components. As a consequence, there is no simple one-step solution. Detailed discussion of it can be found e.g. in McCulloch and Searle (2001) and Searle, Casella and McCulloch (1992).

3.3.2 Maximum likelihood (ML) estimatiom

The likelihood function to be maximized for the estimates of β and σ comes from the marginal model (3.3), where the covariance matrix V is a function of variance parameters σ, $V(\sigma)$. Its density function is given by

$$f(y;\beta,V) = (2\pi)^{-\frac{n}{2}} |V|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(y - X\beta)^T V^{-1}(y - X\beta)\right\}, \tag{3.5}$$

and leads to the log likelihood function

$$\log L(\beta,\sigma) = -\frac{1}{2}(Y - X\beta)^t V^{-1}(Y - X\beta) - \frac{1}{2}\log(|V|) - \frac{n}{2}\log(2\pi). \tag{3.6}$$

The partial derivative of (3.6) with respect to β is

$$\frac{\partial \log L(\beta,\sigma)}{\partial \beta} = X^T V(\sigma)^{-1}(y - X\beta),$$

and setting this to zero leads to the maximum likelihood estimator $\hat{\beta}_{ML}$ for a given V, which also is equivalent to the GLS estimator (3.4). For unknown V the ML estimator of β is

$$\hat{\beta}_{ML} = (X^T \hat{V}_{ML}^{-1} X)^{-1} X^T \hat{V}_{ML}^{-1},$$

where $\hat{V}_{ML} = V(\hat{\sigma}_{ML})$ is the ML estimator of V. The ML estimator of β is unbiased under the normality of y (Kackar and Harville 1981, 1984). The maximum likelihood estimates of the variance parameters are biased downwards (Verbeke and Molenberghs, 1997, 2000). The bias arises because estimation of the fixed effects is not taken into consideration when estimating the variance parameters. An unbiased method for estimating the variance parameters is obtained if the observed data are partitioned by a linear transformation. This idea was introduced by Patterson and Thompson (1971)and further developed by Harville (1977) and is called restricted maximum likelihood (REML).

3.3.3 Restricted maximum likelihood (REML) estimation

The REML method is based on such linear transformation of the data y that the resulting distribution does not depend on the fixed effects parameter vector β. Hence β is eliminated from the log likelihood, but at the same time the loss of degrees of freedom involved in estimating β is taken into account in the estimation of $V(\sigma)$.
The REML method follows the likelihood principle and has the same merits, like consistency, efficiency and asymptotic normality, as ML. Since the REML estimators produce unbiased or nearly unbiased variance estimates, have the same desirable properties as the ML estimators and do not require computations that are essentially more complex than those needed in ML estimation, the REML method is now a widely preferred approach to estimate variance parameters in mixed models (Searle et al. 1992, Pinheiro and Bates 2000, Verbeke and Molenberghs 2000, McCulloch, C.E. and Searle, S.R. 2001, Diggle et al. 2002).
Introduce a linear transformation $z = K^T y$ of the normal response vector y, where K is a $n \times (n-p)$ matrix of full rank, for which $K^T X = 0$. The distribution of z is then $N_{n-p}(0, K^T V K)$, which does not depend on β. The elements of z are sometimes referred as error contrasts (e.g. Harville 1977, Verbeke and Molenberghs 2000). The REML estimators of the variance parameters σ are obtained by maximizing the likelihood function associated with the error contrasts z instead of the original data y. The fixed parameter vector β is then estimated by applying the GLS formula (3.4), the covariance matrix V being replaced with its estimate $\hat{V}_{REML} = V(\hat{\sigma}_{REML})$.
An appropiate K is found by selecting $n-p$ columns from the projection matrix

$$Q = I - X(X^T X)^{-1} X^T,$$

which transforms y to the usual OLS residuals. However, the resulting likelihood function and inference do not depend on which columns are used, and nor even the choice of K, as Harville (1977) has shown (see also Diggle et al. 2002). Instead, any full-rank $n \times (n-p)$ matrix K giving the property $E(z) = 0$ for all β will do.
A convenient expression for REML likelihood function is found defining a non-zero $n \times (n-p)$ matrix A such that $AA^T = Q$ and $A^T A = I$. The matrix A is also an appropriate choice for K and we define $z = A^T y$, with $E(z) = 0$ and $var(z) = A^T V A$. As $A^T X = 0$, we can also write $z = A^T y = A^T(y - X\hat{\beta}_{GLS})$, where $\hat{\beta}_{GLS}$ is defined in (3.4). The density function of z is

$$\begin{aligned} f(z;\sigma) =& (2\pi)^{-\frac{(n-p)}{2}} |A^T V A|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} z^T (A^T V A)^{-1} z \right\} \\ =& (2\pi)^{-\frac{(n-p)}{2}} |A^T V A|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (y - X\hat{\beta}_{GLS})^T A (A^T V A)^{-1} A^T (y - X\hat{\beta}_{GLS}) \right\}. \end{aligned}$$

Applying the Result 33 in Rao (1973, p.77), yields

$$A^T (A^T V A)^{-1} A = V^{-1} - V^{-1} X (X^T V^{-1} X)^{-1} X^T V^{-1},$$

which leads straightforwardly to

$$(y - X\hat{\beta}_{GLS})^T A^T (A^T V A)^{-1} A (y - X\hat{\beta}_{GLS}) = (y - X\hat{\beta}_{GLS})^T V^{-1} (y - X\hat{\beta}_{GLS}). \tag{3.7}$$

To find an expression for the determinant $|A^T V A|$ we define $H = [A \quad D]$, where $D = V^{-1} X (X^T V^{-1} X)^{-1}$ (note that $\hat{\beta}_{GLS} = D^T y$). Then

$$|H^T V H| = \begin{vmatrix} A^T V A & A^T V D \\ D^T V A & D^T V D \end{vmatrix} = \begin{vmatrix} A^T V A & 0 \\ 0 & D^T V D \end{vmatrix} = |A^T V A| |D^T V D|,$$

so that

$$|A^T V A| = |H^T V D|_{-1}. \tag{3.8}$$

By a straightforward calculation

$$|D^T V D| = |X^T V^{-1} X|^{-1}. \tag{3.9}$$

For $|H^T V H|$ we note that

$$|H^T V H| = |H^T| |V| |H| = |V| |H^T H|, \tag{3.10}$$

because H is a square matrix. By using $AA^T = Q$ and $A^TA = I$ and the standard result for block determinants we get

$$|H^TH| = \begin{vmatrix} A^TA & A^TD \\ D^TA & D^TD \end{vmatrix} = |I||D^TD - D^TAA^TD| = |D^TD - D^TQD| = |X^TX|^{-1}. \tag{3.11}$$

Collecting (3.7)-(3.11) together leads to the following expression of the density of the error contrasts $z = A^Ty$,

$$f(z) = (2\pi)^{-\frac{(n-p)}{2}}|X^TX|^{\frac{1}{2}}|X^TV^{-1}X|^{-\frac{1}{2}}|V|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(y - X\hat{\beta}_{GLS})^TV^{-1}(y - X\hat{\beta}_{GLS})\right\}.$$

Now it is seen that the matrix A defining the error contrasts z does not explicitly appear in the density. It is implicitly related to the matrix X^TX, but this does not depend on V and therefore does not affect the maximization of the likelihood. The consequence is that the REML estimation does not depend on the choice of $n - p$ error contrasts.
The density of z leads to the restricted or residual log likelihood, which is written here using the notation of (3.6)

$$\log L_{RELM}(\sigma) = -\frac{1}{2}(y - X\hat{\beta}_{GLS})^TV^{-1}(y - X\hat{\beta}_{GLS}) - \frac{1}{2}\log|V| - \frac{1}{2}\log|X^TV^{-1}X| + const. \tag{3.12}$$

Maximizing (3.12) produces the REML estimate $\hat{\sigma}_{REML}$. Searle et al. (1992, Ch. 6.6) provide the REML estimation equations for variance component models. Because these equations for the REML solutions are highly nonlinear in the case of models with complex structure, closed analytical solutions are not available. In general the maximization requires numerical optimization methods like Newton-Raphson, Fisher scoring or EM algorithms. These algorithms are discussed e.g. by Lindstrom and Bates (1988) and Longford (1993, 2005). If we compare the ML log likelihood (3.6) and the REML log likelihood (3.12), we note that their only difference is the penalty term $\frac{1}{2}\log|X^TV(\sigma)^{-1}X|$.
Substituting the estimator $\hat{V}_{REML} = V(\hat{\sigma}_{REML})$ into the GLS formula (3.4) gives the REML estimator of β.

3.4 Prection of random effects

3.4.1 Best linear unbiased predictor (BLUP)

Technically speaking, the random effects v in model (4.16) are not model parameters like β and σ. However, as Pinheiro and Bates (2000) point out, in a way they behave like parameters and since they are unobservable, there often is interest in obtaining estimates of their values. For example, estimates of random area effects are needed in the estimation of small area means. In the frequentist theory the concept of estimation is usually reserved only for the fixed parameters, and since the vector v contains random variables, not unknown constants, we say that we do not estimate but predict their values (for opposite points of view, see Robinson 1991).
It can be shown that the best predictor BP of v, in the sense that it minimizes the mean squared prediction error, is the conditional mean

$$\tilde{v} = BP(v) = E(v|y).$$

The normality assumptions for model (4.16) imply that v and y have a joint multivariate normal distribution

$$\begin{bmatrix} v \\ y \end{bmatrix} \sim N_{q+n}\left(\begin{bmatrix} 0 \\ X\beta \end{bmatrix}, \begin{bmatrix} G & GZ^T \\ ZG & V \end{bmatrix} \right),$$

and under the normal theory, the mean of v given y is

$$\begin{aligned} E(v|y) =& E(v) + cov(v,y)[var(y)]^{-1}(y - E(y)) \\ =& GZ^T V^{-1}(y - X\beta). \end{aligned} \tag{3.13}$$

This is the best predictor of v, and being a linear function of y it also is the best linear predictor (BLP) of v.
In practice the unknown β in (3.13) is replaced with its estimator $\hat{\beta}$, which is the BLUE of β, yielding the Best Linear Unbiased Predictor (BLUP) of v

$$\tilde{v} = GZ^T V^{-1}(y - X\hat{\beta}). \tag{3.14}$$

3.4.2 Empirical best linear unbiased predictor (EBLUP)

Usually the covariance matrices V, G and R are unknown. Then, in predicting v bu the BLUP formula (3.15) they will be replaced with their RELM or ML estimators to yield

$$\hat{v} = \hat{G}Z^T \hat{V}^{-1}(y - X\hat{\hat{\beta}}). \tag{3.15}$$

The predictor $\hat{v}$ is called Empirical Best Linear Unbiased Predictor (EBLUP) of v, the term "empirical" referring to the fact that the values of G and V have been obtained from the observed data (cf. Empirical Bayes). The estimator $\hat{\beta}$ is now the GLS estimator, where V is replaced with its estimate, and it is sometimes called empirical BLUE of β.

In small area estimation the approach, where e.g. small area totals are estimated by utilizing the empirical BLU predictors of random area effects, is often referred as EBLUP method or approach (Ghosh and Rao 1994, Pfeffermann 2002, Rao 2003).

Chapter 4

Two-fold nested error lognormal model

4.1 Introduction

Multistage sampling designs are used in many practical cases, when a design involves two different aggregation levels, domain (small area) and sub-domains (sub-small areas or clusters), it is reasonable to assume a twofold nested error model including random effects explaining the heterogeneity at the two levels of aggregation. In this chapter, we consider the two-fold nested error lognormal regression model for estimating small area means. This model includes small area and sub-small area (cluster) effects to account for the unexplained between-area and between-cluster heterogeneity, respectively.

Some variables of interest are skewed distributed and there is a need to provide small area estimates for these variables. The problem of highly skewed data is, according to Barnett and Lewis (1994), particularly common in business and social surveys. Usual standard estimation methods, under a linear model, for the characteristic of interest (mean in this case) of a skewed variable can be inappropriate.

The model considered by Berg and Chandra (2014) is referred to the one-fold nested error model since only one aggregated level, the small area, is modeled. In this study, we propose a two-fold lognormal model which is different from the Berg and Chandra (2014) because we work with a two-fold unit-level data instead of one-fold unit-level data. Furthermore, we consider the case of cluster-specific covariates as in Datta and Ghosh (1991) and Pfeffermann and Barnard (1991).

The description of a two-fold model is provided in Section 4.2 and the prediction of small area means based on the proposed two-fold lognormal model are shown in Section 4.3.

4.2 Two-fold nested error linear model

Different authors have worked on the one-fold nested error model where only one aggregated level, the small area, is modeled. However, in many real applications, it may be of interest to incorporate additional aggregated levels in the model to account for extra variability or to reflect the sampling design. In our model, we are interested in modeling data where the sample was selected under a two-stage sampling design. At the first stage, primary sampling units (PSU) or clusters are selected. Within each PSU, secondary sampling units (SSU) are selected where in some situations are considered as observational sampling units or individuals units. Fuller and Battese (1973) proposed a two-fold model, that can be used to model data from such complex design in order to capture variability from both the PSU and SSU levels, and its transformation where the transformed quantities are the differences between the original observations and multiples of averages of subsets of obsevations. Later, Datta and Ghosh (1991, under a Bayesian framework, and Pfeffermann and Barnard (1991) used the two-fold model for the special case of cluster-specific covariates. Stukel and Rao (1999) extended the results of Datta and Ghosh (1991) and Pfeffermann and Barnard (1991) to general two-fold nested error regression models, considering the unit-level covariates to be available.
For the classical model-based approach, the characteristics of interest, y, and the covariates, X, are available at the unit level and the linear mixed models (LMM) are used to represent the assumed stochastic relationship between the quantities. The two-fold nested error linear model is formally defined as

$$Y_{ijk} = x_{ijk}^T\beta + v_i + u_{ij} + e_{ijk}; \quad i = 1, \ldots, M; j = 1, \ldots, M_i; k = 1, \ldots, N_{ij}, \qquad (4.1)$$

where the value y_{ijk} is the observed characteristic of interest associated to unit k from cluster j within small area i, the covariates $x_{ijk}^T = (x_{ijk1}, \ldots, x_{ijkp})$ are a $1 \times p$ vector of known variables, β is a $p \times 1$ vector of unknown regression parameters, and the area effects v_i, the cluster effects u_{ij} and the residual errors e_{ijk} are assumed to be mutually independent. Furthermore,

$$\begin{aligned} v_i &\sim N(0, \sigma_v^2), \\ u_{ij} &\sim N(0, \sigma_u^2), \\ e_{ijk} &\sim N(0, \sigma_e^2). \end{aligned}$$

The quantity of interest is the small-area population mean

$$\bar{Y}_i = \frac{1}{N_i} \sum_{j=1}^{M_i} \sum_{k=1}^{N_{ij}} y_{ijk}.$$

In the settings of the theory of prediction presented in section 2.4 and the reviewed theory on linear mixed models in Chapter 3, the following theorem gives the form of the best predictor (BP) and shows why it has a minimum mean squared error.

Theorem 1. Under the two-fold nested error linear model (4.1), the best predictor of the small area mean $\bar{Y}_i$, $i = 1, \dots, M$ is given by $\bar{Y}_i^{BP} = E(\bar{Y}_i|y_s)$ where y_s is the observed characteristic of interest, with

$$\hat{\bar{Y}}_i^{BP}(\theta) = \frac{1}{N_i} \left[\sum_{j,k \in s_i} y_{ijk} + \sum_{j \in s_i} \sum_{k \in \bar{s}_{ij}} \tilde{y}_{ijk}^{*BP} + \sum_{j \in \bar{s}_i} \sum_{k=1}^{N_{ij}} \tilde{y}_{ijk}^{**BP} \right], \tag{4.2}$$

where $\bar{s}_{ij}$ is the set of nonsampled units in the j^{th} sampled cluster and $\bar{s}_i$ is the set of nonsampled sampled clusters in small area i. Also, the predictors $\tilde{y}_{ijk}^{*BP}$ and $\tilde{y}_{ijk}^{**BP}$, from (4.2), are defined as follows

$$\begin{aligned} \tilde{y}_{ijk}^{*BP} &= x_{ijk}^T \beta + \tilde{v}_i^{BP} + \tilde{u}_{ij}^{BP} \\ \tilde{y}_{ijk}^{**BP} &= x_{ijk}^T \beta + \tilde{v}_i^{BP}, \end{aligned} \tag{4.3}$$

where $\tilde{v}_i^{BP} = E(v_i|y_s)$ and $\tilde{u}_{ij}^{BP} = E(u_{ij}|y_s)$.

Proof. Let's consider another estimator $\hat{\bar{Y}}_i$ of $\bar{Y}_i$ function of y_s then

$$\begin{aligned} MSE(\hat{\bar{Y}}_i) =& E(\hat{\bar{Y}}_i - \bar{Y}_i)^2 \\ =& E(\hat{\bar{Y}}_i - \bar{Y}_i^{BP} + \bar{Y}_i^{BP} - \bar{Y}_i)^2 \\ =& E(\hat{\bar{Y}}_i - \bar{Y}_i^{BP})^2 + E(\bar{Y}_i^{BP} - \bar{Y}_i)^2 + 2E(\hat{\bar{Y}}_i - \bar{Y}_i^{BP})(\bar{Y}_i^{BP} - \bar{Y}_i) \\ =& MSE(\bar{Y}_i^{BP}) + E(\bar{Y}_i^{BP} - \bar{Y}_i)^2 + 2E(\hat{\bar{Y}}_i - \bar{Y}_i^{BP})(\bar{Y}_i^{BP} - \bar{Y}_i). \end{aligned}$$

Since $E(\bar{Y}_i^{BP} - \bar{Y}_i)^2$ is positive, it suffices to show that $E(\hat{\bar{Y}}_i - \bar{Y}_i^{BP})(\bar{Y}_i^{BP} - \bar{Y}_i) = 0$. Note that $\hat{\bar{Y}}_i - \bar{Y}_i^{BP}$ is a function of y_s say $f(y_s)$, it follows that

$$\begin{aligned} E(\hat{\bar{Y}}_i - \bar{Y}_i^{BP})(\bar{Y}_i^{BP} - \bar{Y}_i) =& E(f(y_s)(\bar{Y}_i^{BP} - \bar{Y}_i)) \\ =& E(f(y_s)\bar{Y}_i^{BP}) - E(f(y_s)\bar{Y}_i) \\ =& E(f(y_s)\bar{Y}_i) - E(f(y_s)\bar{Y}_i) \\ =& 0. \end{aligned}$$

Which means that $MSE(\hat{\bar{Y}}_i) \geq MSE(\bar{Y}_i^{BP})$.
The above equality is the direct application of the following conditional expectation property

$$E(f(X)E(Y|X)) = E(f(X)Y).$$

The best predictor $\bar{Y}_i^{BP}$ can be written as:

$$\begin{aligned} \bar{Y}_i^{BP} =& E(\bar{Y}_i|y_s) \\ =& \frac{1}{N_i}\left[\sum_{j,k\in s_i} y_{ijk} + \sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}} E[y_{ijk}|y_s] + \sum_{j\in \bar{s}_i}\sum_{k=1}^{N_{ij}} E[y_{ijk}|y_s]\right]. \end{aligned} \quad (4.4)$$

4.3 Empirical Bayes predictors for a two-fold nested error lognormal model

4.3.1 Two-fold nested error lognormal model

Suppose that the ith small area contains M_i first-stage units or primary sampling units (or clusters) and that the jth cluster in the ith area contains N_{ij} second-stage units or observational (or simple) sampling units (elements). Let (Y_{ijk}, X_{ij}) be the y-value and x-value for the the kth element in the jth cluster from the ith area $(k = 1, 2, \ldots, N_{ij}; j = 1, 2, \ldots, M_i; i = 1, 2, \ldots, M)$. Under this population structure, we consider a two-stage sampling in each small area, where a sample s_i, of m_i clusters is selected from the ith sampled small area and, if the jth cluster is sampled, then a subsample, s_{ij}, of n_{ij} elements is selected from it. Without loss of generality, the sample values are denoted by (y_{ijk}, x_{ij}), $(k = 1, 2, \ldots, n_{ij}; j = 1, 2, \ldots, m_i; i = 1, 2, \ldots, m)$.

Under the aforementioned population structure, we obtain $\hat{\bar{Y}}_i^{MMSE}$ using the following nested error two-fold regression model on the logarithm of the variable of interest. The proposed model in its general form for all population units, is given by

$$log(y_{ijk}) \equiv l_{ijk} = x_{ij}^T\beta + v_i + u_{ij} + e_{ijk}, \quad \begin{array}{l} k = 1, 2, \ldots, N_{ij} \\ j = 1, 2, \ldots, M_i \\ i = 1, 2, \ldots, M, \end{array} \tag{4.5}$$

where v_i is the effect of the area i , u_{ij} is effect for cluster j within the domain i and e_{ijk} is the individual model error. Domain and cluster effects, and individual model errors are all assumed to be mutually independent. Furthermore,

$$\begin{aligned} v_i &\sim N(0, \sigma_v^2), \\ u_{ij} &\sim N(0, \sigma_u^2), \\ e_{ijk} &\sim N(0, \sigma_e^2). \end{aligned}$$

The objective is to predict the value of small area population mean:

$$\bar{Y}_i = \frac{1}{N_i} \sum_{j=1}^{M_i} \sum_{k=1}^{N_{ij}} y_{ijk}. \tag{4.6}$$

4.3.2 Minimun model MSE predictor

Since the variance of a small area estimator based on the direct small area sample is excessively large, there is a need for constructing model based estimators with low mean squared prediction error (MSPE). This section introduces the minimum mean squared predicted error (MMSE), known also as Best/Bayes predictor (BP) of a function of a random vector in a finite population.
We assume that the sample values have the mentioned structure and follow the assumed model (4.5). Thus the sample model may be written as

$$log(y_{ijk}) \equiv l_{ijk} = x_{ij}^T\beta + v_i + u_{ij} + e_{ijk}, \quad \begin{array}{l} k = 1, 2, \ldots, n_{ij} \\ j = 1, 2, \ldots, m_i \\ i = 1, 2, \ldots, m, \end{array} \tag{4.7}$$

where, for notational simplicity, the sample clusters, s_i, are denoted as $j = 1, 2, \ldots, m_i$ and sample elements s_{ij} as $k = 1, 2, \ldots, n_{ij}$.
Following **Theorem 1**, the minimum MSE predictor of the $\bar{Y}_i$ is $E[\bar{Y}_i|(y, x)]$, where $(y, x) = \{y_{ijk}, i \in s, j \in s_i, k \in s_{ij}\} \cup \{x_{ij}, i = 1, \ldots, m; j = 1, \ldots, M_i\}$. Where s is the set of indices of those small areas that are in the sample.

Theorem 2. Under the assumed model, (4.5), the expression for the minimum MSE predictor is

$$\begin{aligned}\hat{\bar{Y}}_i^{MMSE} =& E[\bar{Y}_i|(y,x)]\\ =& \frac{1}{N_i}\left[\sum_{j,k\in s_i} y_{ijk} + \sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}} E[y_{ijk}|(y,x)] + \sum_{j\in \bar{s}_i}\sum_{k=1}^{N_{ij}} E[y_{ijk}|(y,x)]\right]\\ =& \frac{1}{N_i}\left[\sum_{j,k\in s_i} y_{ijk} + \sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}} E[\exp\{l_{ijk}\}|(y,x)] + \sum_{j\in \bar{s}_i}\sum_{k=1}^{N_{ij}} E[\exp\{l_{ijk}\}|(y,x)]\right].\end{aligned} \tag{4.8}$$

This expression reflects two cases to be discussed, first: The sub-small area within the small area sampled and contains some observations from the sample, which corresponds to the second term to the right (4.8); second: The sub-small area within the small area sampled, and does not contain any observations from the sample. This corresponds to the third term to the right of (4.8) .

The model (4.7) in matrix form for each $j \in s_i$ is as follows:

$$l_{ij} = X_{ij}\beta + v_i 1_{n_{ij}} + u_{ij}1_{n_{ij}} + e_{ij}, \tag{4.9}$$

where

$$l_{ij} = \begin{bmatrix} l_{ij1}\\ l_{ij2}\\ \vdots\\ l_{ijn_{ij}}\end{bmatrix}_{n_{ij}\times 1} ; \quad X_{ij} = \begin{bmatrix} x_{ij}^T\\ x_{ij}^T\\ \vdots\\ x_{ij}^T\end{bmatrix}_{n_{ij}\times p} = 1_{n_{ij}} \otimes x_{ij}^T ; \quad e_{ij} = \begin{bmatrix} e_{ij1}\\ e_{ij2}\\ \vdots\\ e_{ijn_{ij}}\end{bmatrix}_{n_{ij}\times 1} .$$

From (4.9), the variance of the vector of transformed variable, l_{ij}, is given by

$$\begin{aligned} var(l_{ij}) =& \sigma_v^2 1_{n_{ij}}1_{n_{ij}}^T + \sigma_u^2 1_{n_{ij}}1_{n_{ij}}^T + \sigma_e^2 I_{n_{ij}}\\ =& \sigma_v^2 J_{n_{ij}} + \sigma_u^2 J_{n_{ij}} + \sigma_e I_{n_{ij}}.\end{aligned}$$

Then by a given i we have the vector, l_i, that combines the expressions represented in (4.9)

$$l_i = \begin{bmatrix} l_{i1}^T\\ l_{i2}^T\\ \vdots\\ l_{im_i}^T\end{bmatrix} = \begin{bmatrix} X_{i1}\\ X_{i2}\\ \vdots\\ X_{im_i}\end{bmatrix}\beta + \begin{bmatrix} 1_{n_{i1}}\\ 1_{n_{i2}}\\ \vdots\\ 1_{n_{im_i}}\end{bmatrix} v_i + \begin{bmatrix} 1_{n_{i1}} & & & \\ & 1_{n_{i2}} & & \\ & & \ddots & \\ & & & 1_{n_{im_i}}\end{bmatrix}\begin{bmatrix} u_{i1}\\ u_{i2}\\ \vdots\\ u_{im_i}^T\end{bmatrix} + \begin{bmatrix} e_{i1}\\ e_{i2}\\ \vdots\\ e_{im_i}\end{bmatrix},$$

and its variance is given by

$var(l_i) =$

$$V_i = \begin{bmatrix} (\sigma_v^2+\sigma_u^2)J_{n_{i1}} + \sigma_e^2 I_{n_{i1}} & \sigma_v^2 J_{n_{i1}n_{i2}} & \dots & \sigma_v^2 J_{n_{i1}n_{im_i}} \\ \sigma_v^2 J_{n_{i2}n_{i1}} & (\sigma_v^2+\sigma_u^2)J_{n_{i2}} + \sigma_e^2 I_{n_{i2}} & \dots & \sigma_v^2 J_{n_{i2}n_{im_i}} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_v^2 J_{n_{im_i}n_{i1}} & \sigma_v^2 J_{n_{im_i}n_{i2}} & \dots & (\sigma_v^2+\sigma_v^2)J_{n_{im_i}} + \sigma_e^2 I_{n_{im_i}} \end{bmatrix}.$$

So, from the previous, we have a joint distribution of the expressions represented by (4.7) for a given small area i.

$$l_i = \begin{bmatrix} l_{i1} \\ l_{i2} \\ \vdots \\ l_{im_i} \end{bmatrix} \sim N\Big(\begin{bmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{im_i} \end{bmatrix} \beta, V_i \Big) \equiv N(X_i\beta, V_i).$$

Taking the mean of each cluster, it follows that (see appendix A)

$$\bar{l}_i \sim N(x_i\beta, \bar{V}_i),$$

where

$$\bar{V}_i = W_i V_i W_i^T = \begin{bmatrix} (\sigma_v^2+\sigma_u^2) + \frac{1}{n_{i1}}\sigma_e^2 & \sigma_v^2 & \dots & \sigma_v^2 \\ \sigma_v^2 & (\sigma_v^2+\sigma_u^2) + \frac{1}{n_{i2}}\sigma_e^2 & \dots & \sigma_v^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_v^2 & \sigma_v^2 & \dots & (\sigma_v^2+\sigma_u^2) + \frac{1}{n_{im_i}}\sigma_e^2 \end{bmatrix}.$$

- Let's consider the first case where the sub-small area, j, i.e, for $j \in s_i$ and $k \in \bar{s}_{ij}$. Under the assumed log-transformed model, the joint distribution is

$$\begin{bmatrix} \bar{l}_i \\ u_{ij} \\ v_i \end{bmatrix} \sim N\Big(\begin{bmatrix} x_i\beta \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \bar{V}_i & \alpha^{(j)} & \gamma \\ \alpha^{(j)T} & \sigma_u^2 & 0 \\ \gamma^T & 0 & \sigma_v^2 \end{bmatrix} \Big), \tag{4.10}$$

where

$$\begin{aligned} cov(u_{ij}, \bar{l}_i) &= cov(u_{ij}, W_i l_i) \\ &= C_1^{(j)T} W_i^T = W_i C_1^{(j)} = \sigma_u^2 \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{(m_i \times 1)} = \alpha^{(j)}, \\ cov(v_i, \bar{l}_i) &= cov(v_i, W_i l_i) = W_i cov(l_i, v_i) \\ &= \sigma_v^2 W_i \begin{bmatrix} 1_{n_{i1}} \\ 1_{n_{i2}} \\ \vdots \\ 1_{n_{ij}} \\ \vdots \\ 1_{n_{im_i}} \end{bmatrix}_{(n_i \times 1)} = \sigma_v^2 \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{(m_i \times 1)} = \gamma. \end{aligned}$$

After showing that (see appendix A)

$$E(u_{ij}|l_i) = E(u_{ij}|\bar{l}_i)$$

and

$$var(u_{ij}|l_i) = var(u_{ij}|\bar{l}_i),$$

letting $(y, x) \equiv \bar{l}_i$ and calculating the conditional distribution from a multivariate normal distribution, it follows that

$$E(\begin{bmatrix} u_{ij} \\ v_i \end{bmatrix} \mid \bar{l}_i) = \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}_i^{-1}(\bar{l}_i - x_i\beta) \equiv \mu_{1j},$$

and

$$var(\begin{bmatrix} u_{ij} \\ v_i \end{bmatrix} \mid \bar{l}_i) = \begin{bmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_v^2 \end{bmatrix} - \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}_i^{-1} \begin{bmatrix} \alpha^{(j)} & \gamma \end{bmatrix} \equiv \Sigma_{1j}.$$

Under the model (4.5),

$$y_{ijk} = \exp\{x_{ij}^T\beta + v_i + u_{ij} + e_{ijk}\},$$

by (4.10) and the moment generating function of the lognormal distribution we have

$$E(\exp\{v_i + u_{ij}\} \mid \bar{l}_i) = \exp\{1^T \mu_{1j} + \frac{1}{2} 1^T \Sigma_{1j} 1\}.$$

Now the expression of the second term in (4.8) is given by

$$\tilde{y}^*_{ijk} \equiv E(y_{ijk} \mid \bar{l}_i) = \exp\{x_{ij}^T \beta + 1^T \mu_{1j} + \frac{1}{2} 1^T \Sigma_{1j} 1 + \frac{1}{2} \sigma_e^2\}. \quad (4.11)$$

- Next we consider the second case with a sub-small area, r, in the sampled small area, but does not have any observation in the sample, i.e, for $r \in \bar{s}_i$. Under the assumed log-transformed model, the joint distribution is

$$\begin{bmatrix} \bar{l}_i \\ u_{ir} \\ v_i \end{bmatrix} \sim N\Big(\begin{bmatrix} x_i \beta \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \bar{V}_i & 0 & \gamma \\ 0^T & \sigma_u^2 & 0 \\ \gamma^T & 0 & \sigma_v^2 \end{bmatrix} \Big).$$

Proceeding the same way as in the previous case, we have

$$E(\begin{bmatrix} u_{ir} \\ v_i \end{bmatrix} \mid \bar{l}_i) = \begin{bmatrix} 0^T \\ \gamma^T \end{bmatrix} \bar{V}_i^{-1} (\bar{l}_i - x_i \beta) \equiv \mu_2,$$

and

$$var(\begin{bmatrix} u_{ir} \\ v_i \end{bmatrix} \mid \bar{l}_i) = \begin{bmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_v^2 \end{bmatrix} - \begin{bmatrix} 0^T \\ \gamma^T \end{bmatrix} \bar{V}_i^{-1} \begin{bmatrix} 0 & \gamma \end{bmatrix} \equiv \Sigma_2.$$

Using the above results, we have

$$E(\exp\{v_i + u_{ir}\} \mid \bar{l}_i) = \exp\{1^T \mu_2 + \frac{1}{2} 1^T \Sigma_2 1\}.$$

Now the expression of the third term in (4.8) is given by

$$\tilde{y}^{**}_{ijk} \equiv E(y_{ijk} \mid \bar{l}_i) = \exp\{x_{ij}^T \beta + 1^T \mu_2 + \frac{1}{2} 1^T \Sigma_2 1 + \frac{1}{2} \sigma_e^2\}. \quad (4.12)$$

Substituting the expressions (4.11) and (4.12) in (4.8), the minimum MSE predictor, under the assumption that $\theta = (\beta, \sigma_v^2, \sigma_u^2, \sigma_e^2)^T$ is known, is given by

$$\begin{aligned} \hat{\bar{Y}}_i^{MMSE}(\theta) =& E[\bar{Y}_i | (y, x)] \\ =& \frac{1}{N_i} \Bigg[\sum_{j,k \in s_i} y_{ijk} + \sum_{j \in s_i} \sum_{k \in \bar{s}_{ij}} \tilde{y}^*_{ijk}(\theta) + \sum_{j \in \bar{s}_i} \sum_{k=1}^{N_{ij}} \tilde{y}^{**}_{ijk}(\theta) \Bigg]. \quad (4.13) \end{aligned}$$

4.3.3 Empirical Bayes predictor

In practice, θ is not unknown, so it is not possible to calculate (4.13). We replace the true value of θ with its consistent estimator to obtain the Emprical Bayes (EB) predictor. Let $\hat{\theta}^T = (\hat{\beta}, \hat{\sigma})$ be a restricted maximum likelihood (REML) estimator. By substituting the true θ in (A.11) with an estimator, we obtain

$$\hat{\bar{Y}}_i^{EB} = \frac{1}{N_i}\left\{\sum_{j,k\in s_i} y_{ijk} + \sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}} \tilde{y}^*_{ijk}(\hat{\theta}) + \sum_{j\in \bar{s}_i}\sum_{k=1}^{N_{ij}} \tilde{y}^{**}_{ijk}(\hat{\theta})\right\}, \tag{4.14}$$

where
$\hat{\bar{Y}}_i^{EB} = \hat{\bar{Y}}_i^{MMSE}(\hat{\theta})$, $\hat{y}^{*EB}_{ijk} = \tilde{y}^*_{ijk}(\hat{\theta})$, and $\hat{y}^{**EB}_{ijk} = \tilde{y}^{**}_{ijk}(\hat{\theta})$.

4.4 Restricted maximum likelihood

The sample model (4.7) may be seen as a special case of a general linear mixed model with block diagonal covariance structure, involving fixed and random effects, and a small-area mean can be expressed as a linear combination of fixed effects and realized values of random effects i.e., a model composed by m independent submodels:

$$l_{ij} = col_{1\le k\le n_{ij}}(l_{ijk}),\ l_i = col_{1\le j\le m_i}(l_{ij}), l = col_{1\le j\le m}(l_i),$$

where

$$l_{ij} = 1_{n_{ij}}(x_{ij}^T\beta) + 1_{n_{ij}}v_i + 1_{n_{ij}}u_{ij} + e_{ij},$$

and

$$l_i = diag_{1\le j\le m_i}(1_{n_{ij}} \otimes x_{ij}^T\beta) + 1_{n_i}v_i + diag_{1\le j\le m_i}(1_{n_{ij}})u_i + e_i,$$

with $1_{n_i} = col_{1\le j\le m_i}(1_{n_{ij}})$.
Then, the matrix form of the model is

$$l = X\beta + Z_1 v + Z_2 u + e, \tag{4.15}$$

where
$X = X_{n\times p}$, $Z_1 = diag_{1\le i\le m}(1_{n_i})$, $Z_2 = diag_{1\le i\le m}(diag_{1\le j\le m_i}(1_{n_{ij}}))_{n\times d}$,
$n = \sum_{i=1}^m n_i$, $n_i = \sum_{j=1}^{m_i} n_{ij}$, $d = \sum_{i=1}^m m_i$.
The model (4.15) can be rewritten in the following form

$$l = X\beta + Zw + e, \tag{4.16}$$

where, $Z = (Z_1, Z_2)$ and $w = (v^T, u^T)^T$.
The variance, V, of l is given by $V(\sigma) = diag_{1\le i\le m}(V_i(\sigma))$,
where V_i is defined in the previous section and $\sigma = (\sigma_v^2, \sigma_u^2, \sigma_e^2)^T$ is the vector of unkown parameters involved in the covariance structure of the model.
Following Henderson (1975), the best linear unbiased estimator (BLUE) of β in (4.16) is given by

$$\tilde{\beta}(\sigma) = (X^T V^{-1}(\sigma)X)^{-1} X^T V^{-1}(\sigma)l.$$

Replacing an estimator $\hat{\sigma}$ for σ in previous equation we obtain the so called empirical BLUE (EBLUE) $\hat{\beta} = \tilde{\beta}(\hat{\sigma})$.

4.4.1 Fisher-scoring algorithm under restricted maximum likelihood

The restricted maximum likelihood (REML) method maximizes the joint probability density of $n-p$ linear independent contrasts $\omega = Al$, where A^T is an $n \times (n-p)$ full column rank matrix satisfying $AA^T = I_{n-p}$ and $BX = 0$. Thus, the probability density function of ω does not depend on β and is given by

$$L(\omega|\sigma) = (2\pi)^{\frac{-(n-p)}{2}} |X^T X|^{1/2} |V(\sigma)|^{-1/2} |X^T V^{-1}(\sigma)X|^{-1/2} \exp\left[-\frac{1}{2} l^T P(\sigma) l\right],$$

where

$$P(\sigma) = V^{-1}(\sigma) - V^{-1}(\sigma)X(X^T V^{-1}(\sigma)X)^{-1} X^T V^{-1}(\sigma).$$

Note that $P(\sigma)$ satisfies $P(\sigma)V(\sigma)P(\sigma) = P(\sigma)$ and $P(\sigma)X = 0_n$.

The REML estimator of σ is the maximizer of $l_{REML}(\sigma) = logL(\omega|\sigma)$. As mentioned in the reviewing chapter, the fact that the REML equations are nonlinear, they do not have closed analytical solutions. We adapt the iterative technique for solving the REML equations. A common variant of Newton-Raphson (NR) algorithm is Fisher-scoring method, which is appears to be slightly more robust to initial values than strict NR (Jennrich and Sampson, 1976), which replaces the inverse of the Hessian matrix by its expected value, which after allowing for a change in sign, turns out to be defined by the inverse of Fisher's information matrix.

Let $S(\sigma) = \partial l_{REML}(\sigma)/\partial\sigma = (S_1(\sigma), \ldots, S_3(\sigma))$ and $F(\sigma) = -E[\partial l_{REML}(\sigma)/\partial\sigma\partial\sigma^T] = (F_{qr}(\sigma))$ be the scores vector and the Fisher information matrix respectively. Using the fact that

$$\frac{\partial P(\sigma)}{\partial \sigma_s} = -P(\sigma)\frac{\partial V(\sigma)}{\partial \sigma_s}P(\sigma), \quad q = 1, 2, 3,$$

the first order partial derivative of $l_{REML}(\sigma)$ with respect to σ_s is given by

$$S_q(\sigma) = -\frac{1}{2}trace\left[P(\sigma)\frac{\partial V(\sigma)}{\partial \sigma_q}\right] + \frac{1}{2}l^T P(\sigma)\frac{\partial V(\sigma)}{\sigma_q}P(\sigma)l, \;\; q = 1, 2, 3.$$

Then, taking the negative expectation of second order partial derivative of $l_{REML}(\theta)$ with respect to σ_q and σ_r, the element (q, r) of the Fisher information matrix is obtained by

$$F_{qr}(\sigma) = \frac{1}{2}trace\left[P(\sigma)\frac{\partial V(\sigma)}{\partial \sigma_q}P(\sigma)\frac{\partial V(\sigma)}{\partial \sigma_r}\right], \;\; q, r = 1, \dots, 4.$$

Then, assuming σ^s to be the value of the estimator at iteration s, the updating expression of the Fisher-scoring algorithm is

$$\sigma^{s+1} = \sigma^s + [F(\sigma^s)]^{-1}S(\sigma^s).$$

Chapter 5

Model MSE estimation for the EB predictor

Since our goal is to use the predictor of the small-area population mean in practice, we need to compute the mean square error of MMSE predictor, as well as its estimator. The MSE of an EB predictor, or EBLUP under normal assumption, can be written as a sum of two terms (Kackar and Harville, 1984; Prasad and Rao, 1990):

$$MSE(\hat{\bar{Y}}_i^{EB}) = M_{i1}(\theta) + M_{i2}(\theta),$$

where
$M_{1i} = E[(\bar{Y}_i - \hat{\bar{Y}}_i^{MMSE})^2]$ and $M_{2i} = E[(\hat{\bar{Y}}_i^{MMSE} - \hat{\bar{Y}}_i^{EB})^2]$.
The first term, $M_{i1}(\theta)$, is the variance of the error in the minimum MSE predictor (4.13), the predictor obtained under the true (unknown) θ. The second term accounts for variability of the predictor due to estimation of the parameters in $\theta = (\beta^T, \sigma^T)^T$. In the next two sections, we give a closed form expression for $M_{i1}(\theta)$ and a linear approximation for $M_{i2}(\theta)$ respectively.

5.1 MSE of the MMSE predictor and the EB predictor correction

In this section we derive the expression of the MSE of the minimum MSE predictor (A.11)(see appendix B for more details).

$$
\begin{aligned}
& M_{1i} = E[(\bar{Y}_i - \hat{\bar{Y}}_i^{MMSE})^2] \\
=& E[(\bar{Y}_i - E[\bar{Y}_i|(y,x)])^2] \\
=& E\{E[(\bar{Y}_i - E[\bar{Y}_i|(y,x)])^2|(y,x)]\} \\
=& E[var(\bar{Y}_i|(y,x))] \\
=& E\left[\frac{1}{N_i^2}\left(var\left(\sum_{j\in s_i}\sum_{k\in s_{ij}} y_{ijk} + \sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}} y_{ijk} + \sum_{j\in \bar{s}_i}\sum_{k=1}^{N_{ij}} y_{ijk}|(y,x)\right)\right)\right] \\
=& E\left[\frac{1}{N_i^2}\left(var\left(\sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}} y_{ijk}|(y,x)\right) + var\left(\sum_{j\in \bar{s}_i}\sum_{k=1}^{N_{ij}} y_{ijk}|(y,x)\right) + \right.\right. \\
& \left.\left. 2cov\left(\sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}} y_{ijk}, \sum_{r\in \bar{s}_i}\sum_{p=1}^{N_{ir}} y_{irp}|(y,x)\right)\right)\right] \\
=& \frac{1}{N_i^2}\left[E(V_1) + E(V_2) + 2E(C_1)\right].
\end{aligned} \tag{5.1}
$$

Now computing those three terms in the right-hand side of (5.1) separately, the expression that corresponds to the first term is given by

$$
\begin{aligned}
E(V_1) = &\sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}}\sum_{r\in s_i}\sum_{p\in \bar{s}_{ir}} v_{11j} I(p=k)I(r=j) + \sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}}\sum_{r\in s_i}\sum_{p\in \bar{s}_{ir}} v_{12j} I(r=j) + \\
&\sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}}\sum_{r\in s_i}\sum_{p\in \bar{s}_{ir}} v_{13jr} I(r\neq j) \\
= &\sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}} v_{11j}\left[\sum_{r\in s_i}\sum_{p\in \bar{s}_{ir}} I(p=k)I(r=j)\right] + \sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}} v_{12j}\left[\sum_{r\in s_i}\sum_{p\in \bar{s}_{ir}} I(r=j)\right] + \\
&\sum_{j\in s_i}\sum_{r\in s_i} v_{13jr} I(r\neq j)\left[\sum_{k\in \bar{s}_{ij}}\sum_{p\in \bar{s}_{ir}} 1\right] \\
= &\sum_{j\in s_i}(N_{ij}-n_{ij})v_{11j} + \sum_{j\in s_i}(N_{ij}-n_{ij})(N_{ir}-n_{ir})v_{12j} + \sum_{j\in s_i}\sum_{r\in s_i}(N_{ij}-n_{ij})(N_{ir}-n_{ir})v_{13jr}I(r\neq j),
\end{aligned}
\tag{5.2}
$$

where,

$$
\begin{aligned}
v_{11j} =& \exp\{2(x_{ij}^T\beta + 1^T V_{\mu_{1j}} 1) + 1^T\Sigma_{1j}1 + \sigma_e^2\}(\exp\{1^T\Sigma_{1j}1 + \sigma_e^2\} - \exp\{1^T\Sigma_{1j}1\}) \\
v_{12j} =& \exp\{2(x_{ij}^T\beta + 1^T V_{\mu_{1j}} 1) + 1^T\Sigma_{1j}1 + \sigma_e^2\}(\exp\{1^T\Sigma_{1j}1\} - 1). \\
v_{13jr} =& \exp\{x_{ij}^T\beta + x_{ir}^T\beta + \frac{1}{2}1^T V_{\mu_{1jr}} 1 + \frac{1}{2}1^T\Sigma_{1jr}1 + \sigma_e^2\} - \exp\{x_{ij}^T\beta + x_{ir}^T\beta + \frac{1}{2}1^T(V_{\mu_{1j}} + V_{\mu_{1r}})1 + \\
&\frac{1}{2}1^T(\Sigma_{1j} + \Sigma_{1r})1 + \sigma_e^2\}.
\end{aligned}
$$

Then the expression that corresponds to the second term in (B.1) is given by

$$
\begin{aligned}
E(V_2) = &\sum_{j\in \bar{s}_i}\sum_{k=1}^{N_{ij}}\sum_{r\in \bar{s}_i}\sum_{p=1}^{N_{ir}} v_{21j} I(p=k)I(r=j) + \sum_{j\in \bar{s}_i}\sum_{k=1}^{N_{ij}}\sum_{r\in \bar{s}_i}\sum_{p=1}^{N_{ir}} v_{22j} I(r=j) + \\
&\sum_{j\in \bar{s}_i}\sum_{k=1}^{N_{ij}}\sum_{r\in \bar{s}_i}\sum_{p=1}^{N_{ir}} v_{23jr} I(r\neq j) \\
= &\sum_{j\in \bar{s}_i} N_{ij}v_{21j} + \sum_{j\in \bar{s}_i} N_{ij}N_{ir}v_{22j} + \sum_{j\in \bar{s}_i}\sum_{r\in \bar{s}_i} N_{ij}N_{ir}v_{23jr}I(r\neq j),
\end{aligned}
\tag{5.3}
$$

where

$$
\begin{aligned}
v_{21j} &= \exp\{2(x_{ij}^T\beta + 1^T V_{\mu_2} 1) + 1^T\Sigma_2 1 + \sigma_e^2\}(\exp\{1^T\Sigma_2 1 + \sigma_e^2\} - \exp\{1^T\Sigma_2 1\}) \\
v_{22j} &= \exp\{2(x_{ij}^T\beta + 1^T V_{\mu_2} 1) + 1^T\Sigma_2 1 + \sigma_e^2\}(\exp\{1^T\Sigma_2 1\} - 1) \\
v_{23jr} &= \exp\{x_{ij}^T\beta + x_{ir}^T\beta + 1^T V_{\mu_3} 1 + \frac{1}{2}1^T\Sigma_3 1 + \sigma_e^2\} - \exp\{x_{ij}^T\beta + x_{ir}^T\beta + 21^T V_{\mu_2} 1 + 1^T\Sigma_2 1 + \sigma_e^2\}.
\end{aligned}
$$

And finally the third term in (5.1) is equal to

$$
\begin{aligned}
E(C_1) &= \sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}}\sum_{r\in \bar{s}_i}\sum_{p=1}^{N_{ir}}(c_{1ijr} - c_{2ijr}) \\
&= \sum_{j\in s_i}\sum_{r\in \bar{s}_i}(N_{ij} - n_{ij})N_{ir}(c_{1ijr} - c_{2ijr}),
\end{aligned} \tag{5.4}
$$

where

$$
\begin{aligned}
c_{1jr} &= \exp\{x_{ij}^T\beta + x_{ir}^T\beta + 1^T V_{\mu_{4j}} 1 + \frac{1}{2}1^T\Sigma_{4j} 1 + \sigma_e^2\} \\
c_{2jr} &= \exp\{x_{ij}^T\beta + x_{ir}^T\beta + 1^T(V_{\mu_{1j}} + V\mu_2) + \frac{1}{2}1^T(\Sigma_{1j} + \Sigma_2)1 + \sigma_e^2\}.
\end{aligned}
$$

Therefore, from (5.1), (5.2), (5.3) and (5.4) we have

$$
\begin{aligned}
M_{1i} = \frac{1}{N_i^2}\Bigg[&\sum_{j\in s_i}(N_{ij} - n_{ij})v_{11j} + \sum_{j\in s_i}(N_{ij} - n_{ij})(N_{ir} - n_{ir})v_{12j} + \\
&\sum_{j\in s_i}\sum_{r\in s_i}(N_{ij} - n_{ij})(N_{ir} - n_{ir})v_{13jr}I(r \neq j) + \\
&\sum_{j\in \bar{s}_i}N_{ij}v_{21j} + \sum_{j\in \bar{s}_i}N_{ij}N_{ir}v_{22j} + \sum_{j\in \bar{s}_i}\sum_{r\in \bar{s}_i}N_{ij}N_{ir}v_{23jr}I(r \neq j) + 2\sum_{j\in s_i}\sum_{r\in \bar{s}_i}(N_{ij} - n_{ij})N_{ir}(c_{1jr} - c_{2jr})\Bigg].
\end{aligned} \tag{5.5}
$$

.

5.2 MSE of the empirical Bayes predictor

In this section we derive the expression of the MSE of the empirical Bayes predictor (A.11)(see appendix C for more details).

$$
\begin{aligned}
&M_{2i}(\theta) = E[(\bar{Y}_i^{MMSE}(\theta) - \bar{Y}_i^{MMSE}(\hat{\theta}))]\\
=&\frac{1}{N_i^2}E\left[\left(\sum_{j\in s_i}\sum_{k\in\bar{s}_{ij}} y_{ijk}^{MMSE}(\theta) + \sum_{r\in\bar{s}_i}\sum_{k=1}^{N_{ir}} y_{irk}^{MMSE}(\theta)\right) - \left(\sum_{j\in s_i}\sum_{k\in\bar{s}_{ij}} y_{ijk}^{MMSE}(\hat{\theta}) + \sum_{r\in\bar{s}_i}\sum_{k=1}^{N_{ir}} y_{irk}^{MMSE}(\hat{\theta})\right)\right]^2\\
=&\frac{1}{N_i^2}E\left[\sum_{j\in s_i}\sum_{k\in\bar{s}_{ij}}(y_{ijk}^{MMSE}(\theta) - y_{ijk}^{MMSE}(\hat{\theta})) + \sum_{r\in\bar{s}_i}\sum_{k=1}^{N_{ir}}(y_{irk}^{MMSE}(\theta) - y_{irk}^{MMSE}(\hat{\theta}))\right]^2\\
=&\frac{1}{N_i^2}E\Bigg[\sum_{j\in s_i}\sum_{k\in\bar{s}_{ij}}\sum_{q\in s_i}\sum_{p\in\bar{s}_{iq}}(y_{ijk}^{MMSE}(\theta) - y_{ijk}^{MMSE}(\hat{\theta}))(y_{iqp}^{MMSE}(\theta) - y_{iqp}^{MMSE}(\hat{\theta}))+\\
&2\sum_{j\in s_i}\sum_{k\in\bar{s}_{ij}}\sum_{r\in\bar{s}_i}\sum_{p=1}^{N_{ir}}(y_{ijk}^{MMSE}(\theta) - y_{ijk}^{MMSE}(\hat{\theta}))(y_{irp}^{MMSE}(\theta) - y_{irp}^{MMSE}(\hat{\theta}))+\\
&\sum_{g\in\bar{s}_i}\sum_{k=1}^{N_{ig}}\sum_{r\in\bar{s}_i}\sum_{p=1}^{N_{ir}}(y_{igk}^{MMSE}(\theta) - y_{igk}^{MMSE}(\hat{\theta}))(y_{irp}^{MMSE}(\theta) - y_{irp}^{MMSE}(\hat{\theta}))\Bigg].\\
=&\frac{1}{N_i^2}\sum_{j\in s_i}\sum_{k\in\bar{s}_{ij}}\sum_{q\in s_i}\sum_{p\in\bar{s}_{iq}} E\left[(y_{ijk}^{MMSE}(\theta) - y_{ijk}^{MMSE}(\hat{\theta}))(y_{iqp}^{MMSE}(\theta) - y_{iqp}^{MMSE}(\hat{\theta}))\right]+\\
&2\frac{1}{N_i^2}\sum_{j\in s_i}\sum_{k\in\bar{s}_{ij}}\sum_{r\in\bar{s}_i}\sum_{p=1}^{N_{ir}} E\left[(y_{ijk}^{MMSE}(\theta) - y_{ijk}^{MMSE}(\hat{\theta}))(y_{irp}^{MMSE}(\theta) - y_{irp}^{MMSE}(\hat{\theta}))\right]+\\
&\frac{1}{N_i^2}\sum_{g\in\bar{s}_i}\sum_{k=1}^{N_{ig}}\sum_{r\in\bar{s}_i}\sum_{p=1}^{N_{ir}} E\left[(y_{igk}^{MMSE}(\theta) - y_{igk}^{MMSE}(\hat{\theta}))(y_{irp}^{MMSE}(\theta) - y_{irp}^{MMSE}(\hat{\theta}))\right]\\
=&\frac{1}{N_i^2}\Bigg[H_1 + 2H_2 + H_3\Bigg].
\end{aligned} \tag{5.6}
$$

After applying the Taylor series approximation of the second order of the expressions that involve $\hat{\theta}$, we have

$$H_1 \simeq \sum_{j\in s_i}(N_{ij} - n_{ij})^2 h_{1ij} + \sum_{j\in s_i}\sum_{q\in s_i}(N_{iq} - n_{ij})(N_{iq} - n_{iq})h_{1ijq}I(q \neq j), \tag{5.7}$$

where

$$h_{1ij} = \exp\{2(\Omega_1(\theta) + 1^T V_{\mu_{1j}} 1)\} - 2\exp\{2\Omega_1(\theta) + \frac{1}{2}\varphi_1(\theta)\} + \exp\{2(\Omega_1(\theta) + \Phi_1(\theta))\}$$

$$h_{1ijq} = \exp\{\Omega_1^{(j)}(\theta) + \Omega_1^{(q)}(\theta)\}[\exp\{\frac{1}{2}1^T(V_{\mu_{1j}} + V_{\mu_{1q}})1\} - \exp\{\frac{1}{2}(1^T V_{\mu_{1j}} 1 + \Phi_1^{(q)}(\theta))\} - \exp\{\frac{1}{2}(1^T V_{\mu_{1q}} 1 + \Phi_1^{(j)}(\theta))\} + \exp\{\frac{1}{2}(\Phi_1^{(j)}(\theta) + \Phi_1^{(q)}(\theta))\}],$$

$$H_2 \simeq \sum_{j \in s_i} \sum_{r \in \bar{s}_i} (N_{ij-n_{ij}}) N_{ir} h_{2ijr}, \tag{5.8}$$

where

$$h_{2ijr} = \exp\{\Omega_1(\theta) + \Omega_2(\theta)\}[\exp\{\frac{1}{2}1^T(V_{\mu_{1j}} + V_{\mu_2})1\} - \exp\{\frac{1}{2}(1^T V_{\mu_{1j}} 1 + \Phi_2(\theta))\} - \exp\{\frac{1}{2}(1^T V_{\mu_2} 1 + \Phi_1(\theta))\} + \exp\{\frac{1}{2}(\Phi_1(\theta) + \Phi_2(\theta))\}],$$

and

$$H_3 \simeq \sum_{g \in \bar{s}_i} N_{ig}^2 h_{3ig} + \sum_{g \in \bar{s}_i} \sum_{r \in \bar{s}_i} N_{ig} N_{ir} h_{3igr} I(r \neq g), \tag{5.9}$$

where

$$h_{3ig} = \exp\{2(\Omega_2(\theta) + 1^T V_{\mu_2} 1)\} - 2\exp\{2\Omega_2(\theta) + \frac{1}{2}\varphi_2(\theta)\} + \exp\{2(\Omega_2(\theta) + \Phi_2(\theta))\}$$

$$h_{3igr} = \exp\{\Omega_2^{(g)}(\theta) + \Omega_2^{(r)}(\theta)\}[\exp\{\frac{1}{2}1^T(V_{\mu_{2g}} + V_{\mu_{2r}})1\} - \exp\{\frac{1}{2}(1^T V_{\mu_{2g}} 1 + \Phi_2^{(r)}(\theta))\} - \exp\{\frac{1}{2}(1^T V_{\mu_{2r}} 1 + \Phi_2^{(g)}(\theta))\} + \exp\{\frac{1}{2}(\Phi_2^{(g)}(\theta) + \Phi_2^{(r)}(\theta))\}].$$

Then substituting the expressions (5.7), (5.8) and (5.9) into (5.6) we get

$$M_{2i} \approx \frac{1}{N_i^2}\Bigg[\sum_{j \in s_i} (N_{ij} - n_{ij})^2 h_{1ij} + \sum_{j \in s_i} \sum_{q \in s_i} (N_{ij} - n_{ij})(N_{iq} - n_{iq}) h_{1ijq} I(q \neq j) + 2\sum_{j \in s_i} \sum_{r \in \bar{s}_i} (N_{ij-n_{ij}}) N_{ir} h_{2ijr} + \sum_{g \in \bar{s}_i} N_{ig}^2 h_{3ig} + \sum_{g \in \bar{s}_i} \sum_{r \in \bar{s}_i} N_{ig} N_{ir} h_{3igr} I(r \neq g)\Bigg]. \tag{5.10}$$

5.3 Derivation of the EB predictor correction

By the fact that the predictor (4.14) is a nonlinear transformation of estimators of parameters,

$$E[\hat{y}_{ijk}^{EB}] \neq E[y_{ijk}^{MMSE}(\theta)]. \tag{5.11}$$

The aim of this section is to find approximately unbiased predictor of the non-sampled value of y_{ijk}.
Now considering the two cases separately and using the expressions calculated in the section 5.2 it follows that (see appendix C for more details):

case 1 : $j \in s_i$

$$E[\tilde{y}_{ijk}^{*}(\theta)] = \exp\{\Omega_1(\theta) + \frac{1}{2}1^T V_{\mu_{1j}} 1\} \tag{5.12}$$

and

$$E[\tilde{y}_{ijk}^{*}(\hat{\theta})] = \exp\{\Omega_1(\theta) + \frac{1}{2}(1^T V_{\mu_{1j}} 1 + \lambda_{1j})\} \tag{5.13}$$

Therefore,

$$\frac{E[\tilde{y}_{ijk}^{*}(\hat{\theta})]}{E[\tilde{y}_{ijk}^{*}(\theta)]} \approx \exp\{\frac{1}{2}\lambda_{1j}\}. \tag{5.14}$$

Now from (5.14), we define the multiplicative approximately bias-corrected predictor (BCP)

$$\hat{y}_{ijk}^{*EB.BCP} = \hat{y}_{ijk}^{*EB} \exp\{-\frac{1}{2}\hat{\lambda}_{1j}\}, \tag{5.15}$$

where $\hat{\lambda}_{1j} = \lambda_{1j}(\hat{\theta})$, with $\lambda_{1j}(\theta) = \delta_{1j}^T var(\hat{\beta})\delta_{1j} + trace[E(\rho_{1j}\rho_{1j}^T)var(\hat{\sigma})]$.

case 2 : Case $j \in \bar{s}_i$

$$E[\tilde{y}_{ijk}^{**}(\theta)] = \exp\{\Omega_2(\theta) + \frac{1}{2}1^T V_{\mu_2} 1\} \tag{5.16}$$

and

$$E[\tilde{y}_{ijk}^{**}(\hat{\theta})] = \exp\{\Omega_2(\theta) + \frac{1}{2}(1^T V_{\mu_2} 1 + \lambda_2)\} \tag{5.17}$$

Therefore,

$$\frac{E[\tilde{y}^{**}_{ijk}(\hat{\theta})]}{E[\tilde{y}^{**}_{ijk}(\theta)]} \approx \exp\{\frac{1}{2}\lambda_2\}. \tag{5.18}$$

From (5.18), we define the multiplicative approximately bias-corrected predictor (BCP)

$$\hat{y}^{**EB.BCP}_{ijk} = \hat{y}^{**EB}_{ijk}\exp\{-\frac{1}{2}\hat{\lambda}_2\}, \tag{5.19}$$

where $\hat{\lambda}_2 = \lambda_2(\hat{\theta})$, with $\lambda_2(\theta) = \delta_2^T var(\hat{\beta})\delta_2 + trace[E(\rho_2\rho_2^T)var(\hat{\sigma})]$.

Now from(5.15) and (5.19) the approximately corrected-bias predictor for $\bar{Y}_i^{MMSE}$ is given by:

$$\bar{Y}_i^{EB.BCP} = \frac{1}{N_i}\left\{\sum_{j,k\in s_i} y_{ijk} + \sum_{j\in s_i}\sum_{k\in\bar{s}_{ij}} \hat{y}^{*EB.BCP}_{ijk} + \sum_{j\in\bar{s}_i}\sum_{k=1}^{N_{ij}} \hat{y}^{**EB.BCP}_{ijk}\right\}. \tag{5.20}$$

.

5.4 Parametric bootstrap for MSE estimation

The parametric bootstrap that we propose here to estimate the MSE of EB bias-corrected predictors $\bar{Y}_i^{EB.BCP}$, is an extension of the parametric bootstrap method for finite population proposed by González-Manteiga et al., 2008, Molina and Rao, 2010. This parametric procedure is described as below:

1. Fit model (4.5) to sample data and obtain model parameters estimates $\hat{\beta}$, $\hat{\sigma}_v^2$, $\hat{\sigma}_u^2$, and $\hat{\sigma}_e^2$.

2. Generate bootstrap random area effects as $v_i^* \sim N(0, \hat{\sigma}_v^2)$, $i = 1, \ldots, M$.

3. Generate, independently of random area effects v_i^*, bootstrap random cluster effects
$u_{ij}^* \sim N(0, \hat{\sigma}_u^2)$, $i = 1, \ldots, M$, $j = 1, \ldots, M_i$.

4. Generate, independently of random area effects v_i^* and random cluster effects u_{ij}^*, bootstrap random errors $e_{ijk}^* \sim N(0, \hat{\sigma}_e^2)$, $i = 1, \ldots, M$, $j = 1, \ldots, M_i$,$j = 1, \ldots, N_{ij}$.

5. Construct a bootstrap population using the estimated model

$$log(y^*_{ijk}) = l^*_{ijk} = x^T_{ij}\hat{\beta} + v^*_i + u^*_{ij} + e^*_{ijk}, \tag{5.21}$$

   and calculate the small area population mean

$$\bar{Y}^*_i = \frac{1}{N_i}\sum_{j=1}^{M_i}\sum_{k=1}^{N_{ij}} y^*_{ijk}. \tag{5.22}$$

6. select the elements l^*_{ijk} that correspond to the indices contianed in the sample s, denote l^*_s. Fit the model to l^*_s obtining new model parameters estimates $\hat{\beta}^*$, $\hat{\sigma}^{2*}_v$, $\hat{\sigma}^{2*}_u$, and $\hat{\sigma}^{2*}_e$.

7. Using the bootstrap sample data l^*_s and the known matrix X, apply the EB method with its correction as it was described in subsections 4.3 and 5.3 and calculate bootstrap EB prodictors, $\bar{Y}^{EB*}_i$, $i = 1, \ldots, M$.

Note that the bootstrap population model, given the original sample data, preserve properties of the original population model. This can be observed as follows

$$E_*(v^*_i|l) = E_*(u^*_{ij}|l) = E_*(e^*_{ijk}|l) = 0, \;\; var_*(v^*_i|l) = \hat{\sigma}^2_v, \;\; var_*(u^*_{ij}|l) = \hat{\sigma}^2_u \;\; , var_*(e^*_{ijk}|l) = \hat{\sigma}^2_e, \tag{5.23}$$

where E_* and var_* represent conditional expectation and variance with respect to the distribution defined by the bootstrap model (5.35) given the sample data l_s. Thereby, the distribution of the bootstrap population l^* (given sample data l_s) mimics that of the original population l. Then an estimator of $MSE(\bar{Y}^{EB.BCP}_i)$ is the bootstrap MSE of the bootstrap EB.BCP, defined as

$$MSE_*(\bar{Y}^{EB.BCP*}_i) = E_*[(\bar{Y}^{EB.BCP*}_i - \bar{Y}^*_i)^2]. \tag{5.24}$$

In practice, this expression can be approximated through a Monte Carlo simulation, by repeating steps $2 - 7$ a large number of times, B, and then taking the mean over the the B replicates as follows:
Let $\bar{Y}^{*(b)}_i$ and $\bar{Y}^{EB.BCP*(b)}_i$ be the area population mean and its corresponding EB bias-corrected predictor for the bootstrap replicate b, for $b = 1, \ldots, B$. Then, the estimator of the MSE is calculated as

$$MSE(\bar{Y}^{EB.BCP}_i) = \frac{1}{B}\sum_{b=1}^{B}(\bar{Y}^{EB.BCP^{*(b)}}_i - \bar{Y}^{*(b)}_i)^2. \tag{5.25}$$

5.5 Bias-corrected MSE estimator based on single bootstrap

A naive estimator of MSE is

$$\widehat{MSE}_i = M_{1i}(\hat{\theta}) + M_{2i}(\hat{\theta}), \tag{5.26}$$

where $M_{1i}(\hat{\theta})$ and $M_{2i}(\hat{\theta})$ are the expressions (5.5) and (5.10), respectively, evaluated at the estimator of θ. In general, it is known that $M_{1i}(\hat{\theta})$ is an asymptotically unbiased estimator (Prasad and Rao, 1990). Furthermore, $M_{1i}(\theta)$ is a nonlinear function of θ, the naive estimator $M_{1i}(\hat{\theta})$ in (5.26) is biased, so that we need to correct the bias. Given the complexity of the expression of $M_{1i}(\theta)$, in subsection 5.1, it is not possible to correct the bias using analytical approach. The alternative solution is the bootstrap method. We derive the single bootstrap bias-corrected estimator of $M_{1i}(\theta)$ in two steps (Butar and Lahir, 2003; Rachida, O., 2011; Kubokawa and Nagashima, 2012). At the first step, under the assumption of known parameters the derivation of MSE is presented. At the second step, a parametric bootstrap approach, that described in subsection 5.4, is proposed for bias correction and approximation of the uncertainty due to the estimation of θ.

Definition: The single bootstrap bias corrected estimator is defined as

$$M_{1i}^{BC}(\hat{\theta}) = M_{1i}(\hat{\theta}) + b_{1i}(\hat{\theta}), \tag{5.27}$$

where $b_{1i}(\hat{\theta}) = M_{1i}(\hat{\theta}) - E_{\hat{\theta}}(M_{1i}(\hat{\theta}^*))$.
Below, we present a second stage parametric bootstrap algorithm for bias-correction of the MSE estimator:

1. Fit model (4.5) to sample data and obtain model parameters estimates $\hat{\theta} = (\hat{\beta}, \hat{\sigma}_v^2, \hat{\sigma}_u^2, \hat{\sigma}_e^2)^T$.
2. Generate bootstrap random area effects as $v_i^* \sim N(0, \hat{\sigma}_v^2)$, $i = 1, \ldots, m$.
3. Generate, independently of random area effects v_i^*, bootstrap random cluster effects
   $u_{ij}^* \sim N(0, \hat{\sigma}_u^2)$, $i = 1, \ldots, m$, $j = 1, \ldots, m_i$.
4. Generate, independently of random area effects v_i^* and random cluster effects u_{ij}^*, bootstrap random errors $e_{ijk}^* \sim N(0, \hat{\sigma}_e^2)$, $i = 1, \ldots, m$, $j = 1, \ldots, m_i$,$j = 1, \ldots, n_{ij}$.

5. Construct a bootstrap samples using the estimated model

$$log(y^*_{ijk}) = l^*_{ijk} = x^T_{ij}\hat{\beta} + v^*_i + u^*_{ij} + e^*_{ijk}, \tag{5.28}$$

and for each bootstrap replicate b, for $b = 1, \ldots, B$ we calculate the bootstrap version $M_{1i}(\hat{\theta})^{(b)}$. Then a Monte Carlo estimate of M_{1i} is given by

$$M^{BC}_{1i}(\hat{\theta}) = 2M_{1i}(\hat{\theta}) - \frac{1}{B}\sum_{1=b}^{B} M_{1i}(\hat{\theta})^{(b)}. \tag{5.29}$$

Furthermore, the unbiased estimator of the MSE based on the parametric bootstrap is given by

$$\widehat{mse}_i = M^{BC}_{1i}(\hat{\theta}) + M_{2i}(\hat{\theta}). \tag{5.30}$$

From the described algorithm, we set out the justification behind this approach as it was introduced by Butar and Lahiri (2003) and presented in Kubokawa and Nagashima (2012):
Let $f(\theta)$ be a smooth function. In the spite of the fact that $f(\hat{\theta})$ is an asymptotically unbiased estimator of $f(\theta)$, in general, there exists a second-order bias. Then, we need to approximate the expectation $E[f(\hat{\theta})]$. It is supposed that the approximation is given by

$$E[f(\hat{\theta})] = f(\theta) + b(\theta), \tag{5.31}$$

where $b(\theta)$ is a smooth function. Then,

$$\begin{aligned} E[f(\hat{\theta}) - b(\hat{\theta})] =& E[f(\hat{\theta})] - E[b(\hat{\theta})] \\ =& \{f(\theta) + b(\theta)\} - b(\theta) \\ =& f(\theta). \end{aligned} \tag{5.32}$$

Using model (5.35), it follows that

$$E_{\hat{\theta}}[f(\hat{\theta}^*)|l] = f(\hat{\theta}) + b(\hat{\theta}), \tag{5.33}$$

where $E_{\hat{\theta}}[.|l]$ is the conditional expectation with respect to the model (5.35) given l, and the calculation of $\hat{\theta}^*$ is the same as that of $\hat{\theta}$ except that $\hat{\theta}^*$ is calculated based on l^* instead of l. Hence from (5.32), we have

$$\begin{aligned} E[2f(\hat{\theta}) - E_{\hat{\theta}}[f(\hat{\theta}^*)|l]] =& E[f(\hat{\theta}) - E_{\hat{\theta}}[f(\hat{\theta}^*) - f(\hat{\theta})|l]] \\ =& E[f(\hat{\theta}) - b(\hat{\theta})] \\ =& f(\theta). \end{aligned} \tag{5.34}$$

Therefore, $2f(\hat{\theta}) - E_{\hat{\theta}}[f(\hat{\theta}^*)|l]$ is the second-order unbiased estimator of $f(\theta)$.

5.6 Double parametric bootstrap for bias-correction

Following Hall and Maiti (2006) and adopting a double parametric bootstrap to bias-correction, we provide a population double bootstrap bias adjustment to the MSE estimator of EB bias-corrected predictors $\bar{Y}_i^{EB.BCP}$, but in the setting of the parametric bootstrap method for finite population proposed by González-Manteiga et al., 2008, Molina and Rao, 2009. The double parametric procedure is described as below:

1. Fit model (4.5) to sample data and obtain model parameters estimates $\hat{\beta}$, $\hat{\sigma}_v^2$, $\hat{\sigma}_u^2$, and $\hat{\sigma}_e^2$.
2. Generate bootstrap random area effects as $v_i^* \sim N(0, \hat{\sigma}_v^2)$, $i = 1, \ldots, M$.
3. Generate, independently of random area effects v_i^*, bootstrap random cluster effects
   $u_{ij}^* \sim N(0, \hat{\sigma}_u^2)$, $i = 1, \ldots, M$, $j = 1, \ldots, M_i$.
4. Generate, independently of random area effects v_i^* and random cluster effects u_{ij}^*, bootstrap random errors $e_{ijk}^* \sim N(0, \hat{\sigma}_e^2)$, $i = 1, \ldots, M$, $j = 1, \ldots, M_i$,$j = 1, \ldots, N_{ij}$.
5. Construct a bootstrap population using the estimated model

$$log(y_{ijk}^*) = l_{ijk}^* = x_{ij}^T\hat{\beta} + v_i^* + u_{ij}^* + e_{ijk}^*, \tag{5.35}$$

   and calculate the small area population mean

$$\bar{Y}_i^* = \frac{1}{N_i}\sum_{j=1}^{M_i}\sum_{k=1}^{N_{ij}} y_{ijk}^*. \tag{5.36}$$

6. Select the elements l_{ijk}^* that correspond to the indices contianed in the sample s, denote l_s^*. Fit the model to l_s^* obtining new model parameters estimates $\hat{\beta}^*$, $\hat{\sigma}_v^{2*}$, $\hat{\sigma}_u^{2*}$, and $\hat{\sigma}_e^{2*}$.
7. Using the bootstrap sample data l_s^* and the known matrix X, apply the EB method as described in Section 2 and calculate bootstrap EB prodictors, $\bar{Y}_i^{EB*}$, $i = 1, \ldots, M$. Then an estimator of $MSE(\bar{Y}_i^{EB.BCP})$ is the bootstrap MSE of the bootstrap EB.BCP, defined as

$$MSE_*(\bar{Y}_i^{EB.BCP*}) = E_{\hat{\theta}}[(\bar{Y}_i^{EB.BCP*} - \bar{Y}_i^*)^2]. \tag{5.37}$$

Let's note $\bar{Y}_i^{*(b_1)}$ and $\bar{Y}_i^{EB.BCP*(b_1)}$ as the area population mean and its corresponding EB bias-corrected predictor for the bootstrap replicate b_1, for $b_1 = 1,\ldots,B_1$. Then, the estimator of the MSE is calculated as

$$MSE_*(\bar{Y}_i^{EB.BCP*}) = B_{i1} = \frac{1}{B_1}\sum_{b_1=1}^{B_1}(\bar{Y}_i^{EB.BCP^{*(b_1)}} - \bar{Y}_i^{*(b_1)})^2. \quad (5.38)$$

8. For each bootstrap replicate $b_1 = 1,\ldots,B_1$, obtain parameters estimates $\hat{\beta}^{*(b_1)}$, $\hat{\sigma}_v^{2*(b_1)}$, $\hat{\sigma}_u^{2*(b_1)}$, and $\hat{\sigma}_e^{2*(b_1)}$, and generate for $b_2 = 1,\ldots,B_2$:

$$v_i^{**} \sim N(0,\hat{\sigma}_v^{2*(b_1)}), \quad i = 1,\ldots,M$$
$$u_{ij}^{**} \sim N(0,\hat{\sigma}_u^{2*(b_1)}), \quad i = 1,\ldots,M, j = 1,\ldots,M_i$$
$$e_{ijk}^{**} \sim N(0,\hat{\sigma}_e^{2*(b_1)}), \quad i = 1,\ldots,M, \quad j = 1,\ldots,M_i, \quad j = 1,\ldots,N_{ij}$$

9. Constructing a new bootstrap populations using

$$log(y_{ijk}^{**}) = l_{ijk}^{**} = x_{ij}^T\hat{\beta}^{*(b_1)} + v_i^{**} + u_{ij}^{**} + e_{ijk}^{**}, \quad (5.39)$$

and calculate the small area population mean

$$\bar{Y}_i^{**} = \frac{1}{N_i}\sum_{j=1}^{M_i}\sum_{k=1}^{N_{ij}} y_{ijk}^{**}. \quad (5.40)$$

10. Select the elements l_{ijk}^{**} that correspond to the indices contianed in the sample s, denote l_s^{**}. Fit the model to l_s^{**} obtining new model parameters estimates $\hat{\beta}^{(b_2)}$, $\hat{\sigma}_v^{2(b_2)}$, $\hat{\sigma}_u^{2(b_2)}$, and $\hat{\sigma}_e^{2(b_2)}$.

11. Using the bootstrap sample data l_s^{**} and the known matrix X, apply the EB method as described in chapter 4 and calculate bootstrap EB prodictors, $\bar{Y}_i^{EB**}$, $i = 1,\ldots,M$. Then an estimator of $MSE(\bar{Y}_i^{EB.BCP})$ is the bootstrap MSE of the bootstrap EB.BCP, defined as

$$MSE_{**}(\bar{Y}_i^{EB.BCP**}) = E_{\hat{\theta}^*}[(\bar{Y}_i^{EB.BCP**} - \bar{Y}_i^{**})^2] \quad (5.41)$$

Noting $\bar{Y}_i^{**(b_2(b_1))}$ and $\bar{Y}_i^{EB.BCP**(b_2(b_1))}$ as the area population mean and its corresponding EB bias-corrected predictor for the bootstrap replicate b_2, for $b_2 = 1,\ldots,B_2$. Then, the estimator of the MSE is calculated as

$$MSE_{**}(\bar{Y}_i^{EB.BCP**}) = B_{i2} = \frac{1}{B_1}\sum_{b_1=1}^{B_1}\frac{1}{B_2}\sum_{b_2=1}^{B_2}(\bar{Y}_i^{EB.BCP^{**(b_2(b_1))}} - \bar{Y}_i^{**(b_2(b_1))})^2. \quad (5.42)$$

From (5.38) and (5.42) we get the population bias-corrected MSE estimator

$$M\hat{S}E(\bar{Y}_i^{EB.BCP}) = 2MSE_*(\bar{Y}_i^{EB.BCP*}) - MSE_{**}(\bar{Y}_i^{EB.BCP**}). \quad (5.43)$$

Note that this expression was obtained under the guidelines presented in subsection 5.5, and is detailed in appendix D.

5.7 Bias-corrected MSE estimator based on double bootstrap

The population bias-corrected MSE estimator, (5.43), presented in subsection 5.6 can not be calculated in practical settings since it depends on population quantities, in this subsection we derive a bias-corrected MSE estimator of (5.26) based on a double bootstrap. As Davison and Hinkley (1997) pointed out, the bootstrap does not provide exact solution, the same as in most statistical methods, regarding the bias correction. However, it is helpful to have available a general technique for making a bias correction to a bootstrap calculation. That technique is the bootstrap itself. The bias-corrected estimator of $M_{1i}(\theta)$ based on double bootstrap is given by (Rachida O., 2011; Chang and Hall, 2015)

$$\hat{m}_{1i}^{bcc} = 3M_{1i}(\hat{\theta}) - 3E_{\hat{\theta}}(M_{1i}(\hat{\theta}^*)|l) + E_{\hat{\theta}^*}(M_{1i}(\hat{\theta}^{**})|l^*), \quad (5.44)$$

where $E_{\hat{\theta}}[.|l]$ is the conditional expectation with respect to the model (5.35) given l, and the calculation of $\hat{\theta}^*$ is the same as that of $\hat{\theta}$ except that $\hat{\theta}^*$ is calculated based on l^* instead of l, and $E_{\hat{\theta}^*}[.|l^*]$ is the conditional expectation with respect to the model (5.39) given l^*, and the calculation of $\hat{\theta}^{**}$ is the same as that of $\hat{\theta}^*$ except that $\hat{\theta}^{**}$ is calculated based on l^{**} instead of l^*.

Applying the bootstrap algorithm of the subsection 5.6, the Monte Carlo approximation to the quantity $\hat{m}_{1i}^{bcc}$ is given by

$$\tilde{m}_{1i}^{bcc} = 3M_{1i}(\hat{\theta}) - \frac{3}{B_1}\sum_{b_1=1}^{B_1} M_{1i}(\hat{\theta}^{*(b_1)}) + \frac{1}{B_1 B_2}\sum_{b_1}^{B_1}\sum_{b_2}^{B_2} M_{1i}(\hat{\theta}^{**(b_2(b_1))}), \quad (5.45)$$

where $M_{1i}(\hat{\theta})^{*(b_1)}$ is the version of $M_{1i}(\hat{\theta})$ calculated from (5.35) for each bootstrap replicate b_1, for $b_1 = 1, \ldots, B_1$, and $M_{1i}(\hat{\theta})^{**(b_2(b_1))}$ is the version of $M_{1i}(\hat{\theta})$ obtained from (5.39) for each b_2, for $b_2 = 1, \ldots, B_2$ for each b_1.

Therefore, from (5.26) and (5.45) the bias-corrected MSE estimator based on double bootstrap is given by

$$\widehat{mse}_i^{bcc} = \tilde{m}_{1i}^{bcc} + M_{2i}(\hat{\theta}). \quad (5.46)$$

5.8 Simulation study

As Molina and Rao (2010), we use a model-based superpopulation approach. A central assumption is that the sampling is ignorable. A consequence of this is that the sampled values of the survey variable follow the superpopulation model. Generally, the preferred sampling design would depend on the known auxiliary information. Here, however, we only focus on the model and not on the sample inclusion probabilities; as a consequence, the complexities in the calculation process stem from the assumed superpopulation model and not from the sampling design. In our simulation studies, we have used simple random sampling in both two steps (without replacement).

For the purpose of evaluating the performance of the proposed EB predictors, a simulation experiment is conducted in order to investigate the bias of $MSE(\bar{Y}_i^{EB})$, obtained under a studied model, comparing the derived naive estimator, its proposed bootstrap estimator, and the double bootstrap estimator of the MSE of EB estimators. Note that this experiment will be repited $K = 100$ times.

Under the model (4.5), we generate the response variable for the population units $\log(Y_{ijk})$, similarly to Molina and Rao (2010) but including an indicator of clusters within small area. The transformed variable of the original variable of interest for the population units Y_{ijk} were generated from proposed model considering as auxiliary variables two dummies $X_1 \in \{0,1\}$ and $X_2 \in \{0,1\}$, where the indicator variables mimic the real case where only categorical variables are available, plus an intercept. Regarding this settings, we assume that the mean value of the characteristic of interest increases when moving from the case $(X_1 = 0, X_2 = 0)$ to $(X_1 = 1, X_2 = 0)$, but decreases when moving from $(X_1 = 0, X_2 = 0)$ to $(X_1 = 0, X_2 = 1)$. We consider a clustered finite population from which samples are drawn in two stages using simple random sampling at each stage.

In summary, the specifications of the model for the k^{th} simulation, for $k = 1, \ldots, K$, is:

1. We consider a balanced two-fold model, with a population size $N = 120000$ partitioned into $M = 30$ small areas, with small area population size of $N_i = 4000$, $i = 1, \ldots, M$, and each small area is composed of $M_i = 40$ clusters, $i = 1, \ldots, M$. Cluster population sizes are $N_{ij} = 100$, $j = 1, \ldots, M_i$, $i = 1, \ldots, M$.

2. Two dummy variables are used as covariates plus intercept. The population values of these indicators for the units are generated from Bernoulli distributions $Ber(p_{hij})$,$h = 1, 2$, with probabilities of success $p_{1ij} = 0.3 + 0.5i/M + 0.1j/M_i$ and $p_{2ij} = 0.2$. The covariates are held fixed across the simulated populations.

3. The fixed effects are $\beta = (6, 0.03, -4)^T$.

4. The small area effects, cluster effects and individual errors are independent; with $v_i \sim N(0, \sigma_v^2)$, $u_{ij} \sim N(0, \sigma_u^2)$ and $e_{ijk} \sim N(0, \sigma_e^2)$. To imitate different situations that can be existe in real cases, simulation experiments are repeated for various combinations of variance components: small area (Domain) variability, σ_v^2, and cluster (subdomain) variability, σ_u^2. Sixteen tests of the experiment are carried out, for the sixteen possible combinations of the values $\sigma_e^2 = 0.025$, $\sigma_u^2 = \{0.05, 0.1, 0.15, 0.2\}$, and $\sigma_v^2 = \{0.05, 0.1, 0.15, 0.2\}$, according to the following table

r	1	2	3	4	5	6	7	8
$\sigma_u^{2,(r)}$	0.05	0.05	0.05	0.05	0.1	0.1	0.1	0.1
$\sigma_v^{2,(r)}$	0.05	0.1	0.15	0.2	0.05	0.1	0.15	0.2
r	9	10	11	12	13	14	15	16
$\sigma_u^{2,(r)}$	0.15	0.15	0.15	0.15	0.2	0.2	0.2	0.2
$\sigma_v^{2,(r)}$	0.05	0.1	0.15	0.2	0.05	0.1	0.15	0.2

Table 5.1: Combinations of σ_u^2 and σ_v^2 for a simulation experiment

5. Within each small area i, a sample of $m_i = 5$ clusters is selected using simple random sampling (SRS), and a simple random sample of size $n_{ij} = 10$ is drawn from each sampled cluster. The small area sample sizes are equal $n_i = 50$.

We generate a bootstrap population as it is described in subsection 5.4. We draw a sample from each Bootstrap population and we fit the model and we compute the MSE estimator (5.46) and double parametric bootstrap MSE (5.43).

5.8.1 Simulation experiments

This experiment of simulation is motivated by the fact that practical usage of EB predictors requires, of course, estimates of variance components. It consists of currying out several runs of the simulation study, described in Subsection 5.7, keeping constant the sample sizes, the population sizes and the number of levels and sublevels of the random factors, and varying the values of σ_v^2 and σ_u^2.
Below are presented the graphical results in which square root of the versions of the mean squared error are represented. In these figures, each contains four graphics where we fix the value of σ_v^2 and we vary the values of σ_u^2, boxplot of each of the

efficiency measures are given. For each of the diagrams, three boxes are presented where each box represents the variability for the 30 different values of the domain or small area.Furthermore, we compare three versions of MSE estimator: naive estimator (mse.naive), bias-corrected estimators (based on simple (mse.1bbc)) and simple bootstrap (mse.1bc). The appendix E presents the tables with the numerical values corresponding to the realization of the simulation experiment in the case $\sigma_v^2 = 0.05$.

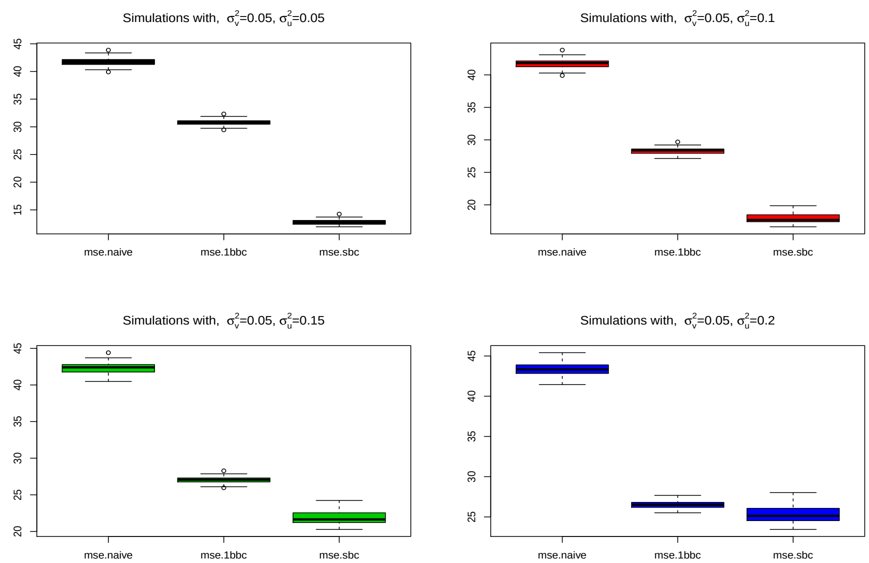

Figure 5.1: Sqrt MSE for all 30 domains, obtained by keeping $\sigma_v^2 = 0.05$ fixed and changing $\sigma_u^2 = \{0.05, 0.1, 0.15, 0.2\}$

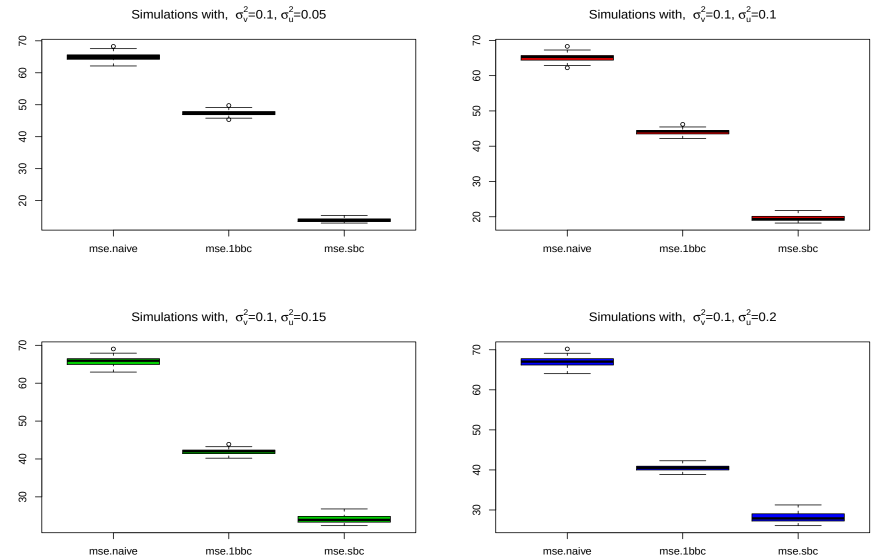

Figure 5.2: Sqrt MSE for all 30 domains, obtained by keeping constant $\sigma_v^2 = 0.1$ and changing $\sigma_u^2 = \{0.05, 0.1, 0.15, 0.2\}$

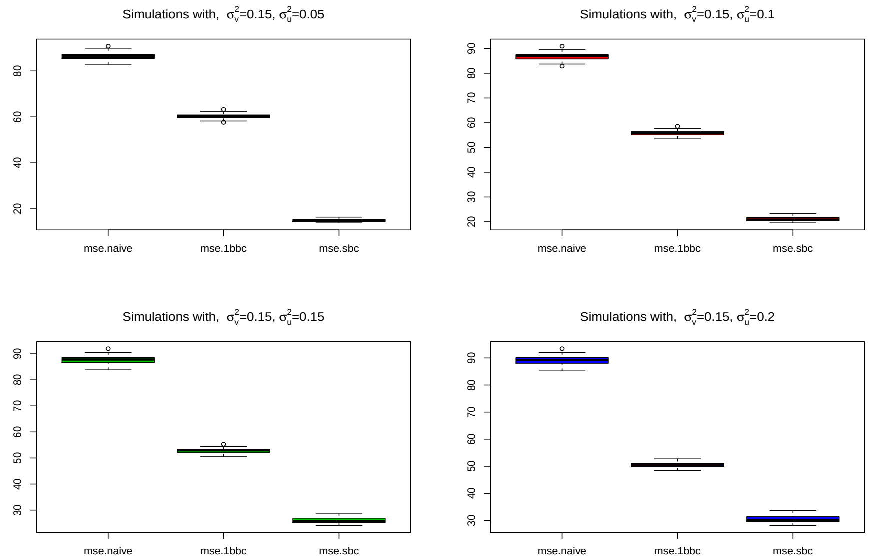

Figure 5.3: Sqrt MSE for all 30 domains, obtained by keeping $\sigma_v^2 = 0.15$ fixed and changing $\sigma_u^2 = \{0.05, 0.1, 0.15, 0.2\}$

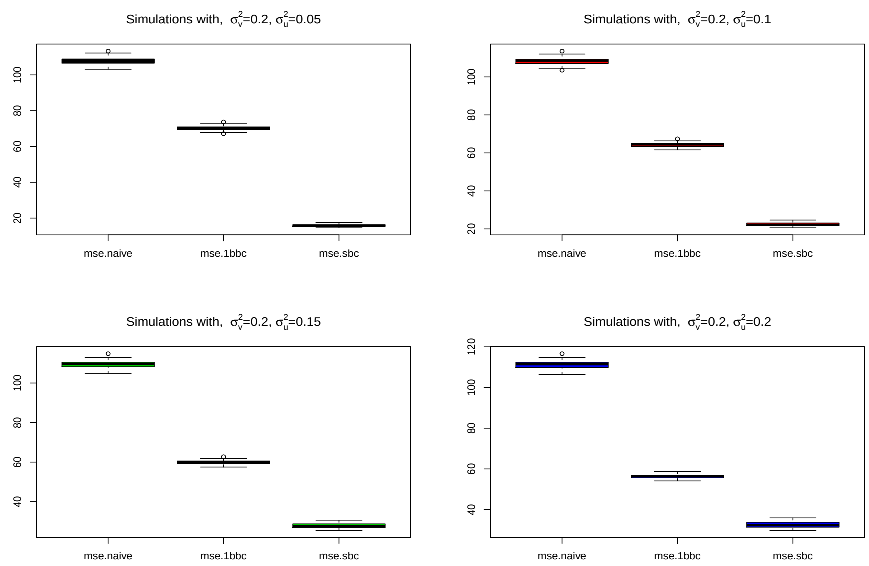

Figure 5.4: Sqrt MSE for all 30 domains, obtained by keeping $\sigma_v^2 = 0.2$ fixed and changing $\sigma_u^2 = \{0.05, 0.1, 0.15, 0.2\}$

In Figure 5.1(a), we show the behavior of those MSEs when the domain and cluster variances increase simultaneously, while the three remaining plots show the behavior of MSEs when we fix the cluster variance, σ_u^2, and varying the domain variability, σ_v^2.
The following plots represent the average of the square roots of the four versions of MSE across the domains with respect to the variance components. In Figure 5.1(a), we show the behavior of those MSEs when the domain and cluster variances increase simultaneously, while the three remaining plots show the behavior of MSEs when we fix the cluster variance, σ_u^2, and varying the domain variability, σ_v^2.

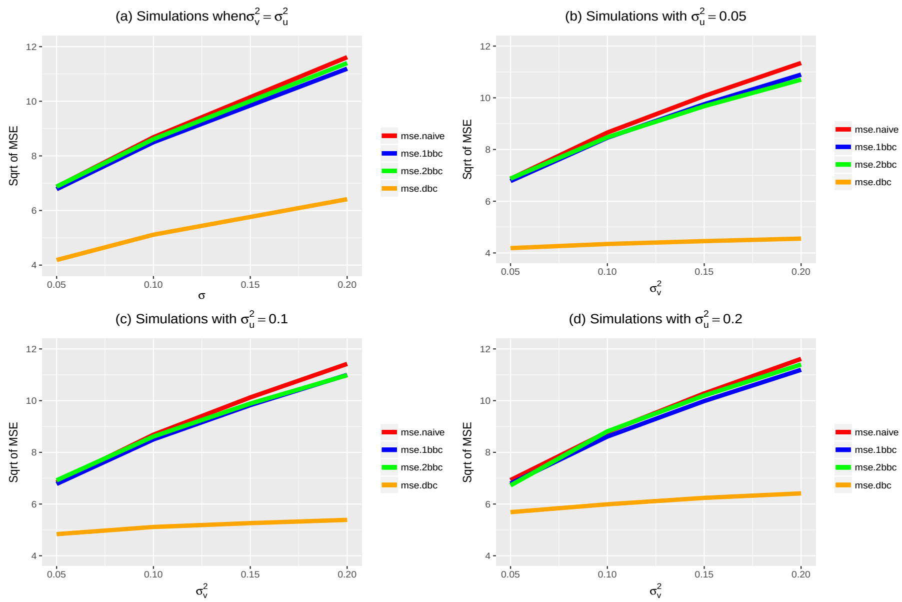

Figure 5.5: Average sqrt MSE across the domains with respect to the change in variance components.

5.8.2 Simulation results

The assessment of the estimated mean squared error was based on variance components analysis and was centered on three versions of MSE: naive estimator (mse.naive), bias-corrected estimators (based on simple (mse.1bbc) and double bootstrap (mse.2bbc)) and double bootstrap (mse.dbc) expressions. Examining the results obtained during the simulation experiments and presented by means of plots (Figure 5.5), it shows that the MSE along the domains increases in magnitude when the values of σ_u^2 and σ_v^2 increase or decrease simultaneously (Figure 5.5(a)). In addition, the bias is moderately reduced in magnitude as the values of σ_u^2 and σ_v^2 increased and decreased respectively (Figure 5.5(b), (c), (d)); that is, the larger the variance between clusters (sub-domains) and the smaller the variance between domains are, the corrected MSE of the EB predictor becomes closer to the one obtained under the ideal double bootstrap MSE. In summary, as the cluster variability σ_u^2 increases compared with the domain variability, σ_v^2, the corrected MSE estimators and bootstrap MSE versions are getting closer.

After those simulation experiments we considered the case $\sigma_v^2 = 0.05$, $\sigma_u^2 = 0.2$, $\sigma_e^2 = 0.025$, where the MSE estimates are close to those calculated under bootstrap setting. Under this consideration, we compare the predictors obtained under our proposed model with the synthetic estimator. The simulation experiments were repeated 100 times and the following average estimates were obtained:

$$\hat{\beta} = (5.99, 0.04, -3.99), \quad \hat{\sigma} = (0.0499, 0.199, 0.025).$$

In terms of prediction, the plot below, (Figure: 5.6), compares the average by domain of population values to their predicted values under the proposed model and the synthetic estimators. We consider the case that the population elements $\log(Y_{ijk})$ following a linear model without domain and cluster effects.

$$\begin{aligned} \log(Y_{ijk}) = l_{ijk} = x_{ij}^T\beta + e_{ijk},\ k = 1, \dots, N_{ij},\ j = 1, \dots, N_i,\ i = 1, \dots, M \\ e_{ijk} \sim N(0, \sigma^2). \end{aligned} \tag{5.47}$$

Taking the average of (5.47) over the elements in the domain i,

$$\bar{l}_i = \bar{X}_i^T\beta + \bar{e}_i,$$

where

$$\bar{X}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} x_{ij}^T, \quad \bar{e}_i = \frac{1}{N_i}\sum_{j=1}^{N_i}\sum_{k=1}^{N_{ij}} e_{ijk}.$$

Following Molina and Rao (2010), and assuming that all model parameters are known. The synthetic estimator, obtained by predicting all the Y_{ijk} and then taking the domain mean, is given by

$$\hat{\bar{Y}}_i^{syn} = \exp(\bar{X}_i^T\beta). \tag{5.48}$$

As Molina and Rao (2010) pointed out, the synthetic estimator is a good estimator when the true is model (5.47). When the true model is a two-fold nested error lognormal model, the synthetic estimation does not take into account the domain and cluster effects. Thus, the synthetic estimator of MSE can lead to serious underestimation of MSE. The MSE of the synthetic estimator was calculated using a bootstrap method.

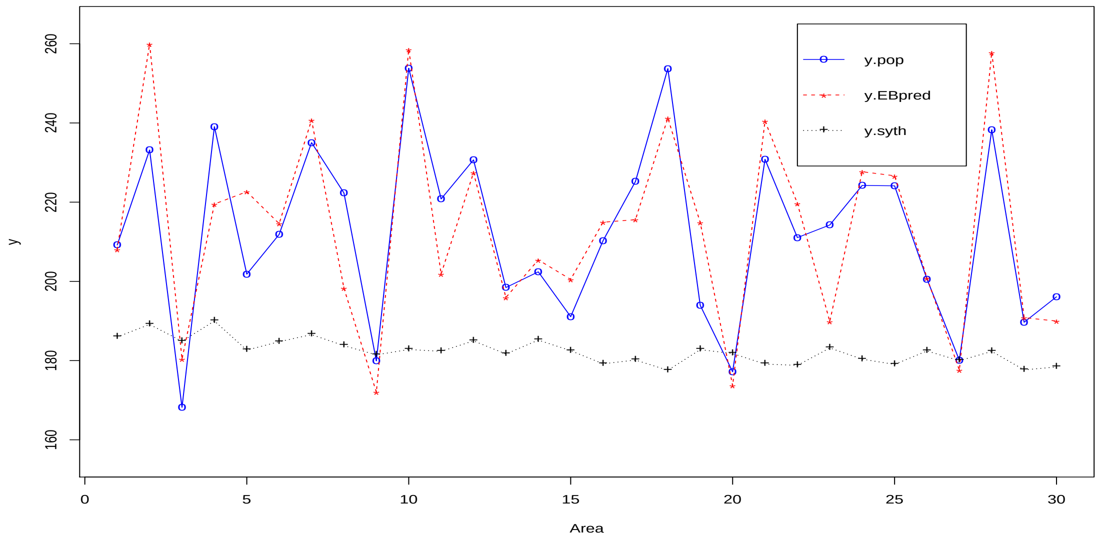

Figure 5.6: Averages of population, EB predicted, and Synthetic estimators values obtained after 100 simulations, with $\sigma_v^2 = 0.05$ and $\sigma_u^2 = 0.2$.

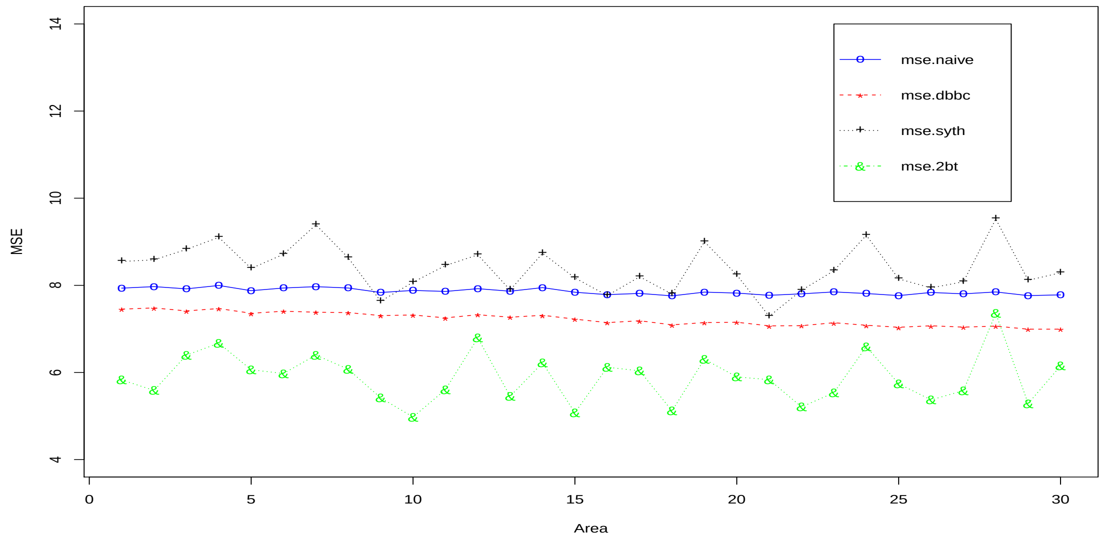

Figure 5.7: Square roots of MSE estimators values obtained after 100 simulations, with $\sigma_v^2 = 0.05$ and $\sigma_u^2 = 0.2$.

The above plots represent the EB predictors obtained under the proposed model and the Synthetic predictors (Figure 5.6) and the MSE estimator of the synthetic estimator with the others presented in the thesis (Figure 5.7) respectively. From the above simulations, we see that the estimates, as well as predictors are close to the true values and thus, the proposed estimators support the theoretical results. The appendix F presents the tables with the numerical values corresponding to the realization of the simulation results.

Chapter 6

Concluding remarks and future research

This concluding chapter resumes the work presented in this thesis and suggests some future directions.
The minimum mean squared error predictor under the proposed model for small area estimation was developed. To obtain the empirical Bayes predictors of pupulation means for small areas, the Scoring-Fisher algorithm based on restricted maximum likelihood to estimate the variance components was used. Following Prasad and Rao (1990), the estimation theory of MSE for the EB predictor, was adapted to the model under study and the closed form expressions of MSE were obtained. Furthermore, we proposed the bias-corrected estimator of MSE under a parametric bootstrap, as well as a double bootstrap method. We studied the prediction capacity of our model under simulation experiments. The simulation studies established clearly the positive performance of using the proposed model in terms of prediction.
In this study, the assessment of the estimated mean squared error was based on variance components analysis and was centered on three versions of MSE: naive estimator, bias-corrected estimators (based on simple and double bootstrap) and double bootstrap expressions. Examining the results obtained during the simulation experiments and presented by means of plots (Figure 5.5), it shows that the MSE along the domains increases in magnitude when the values of σ_u^2 and σ_v^2 increase or decrease simultaneously (Figure 5.5(a)). In addition, the bias is moderately reduced in magnitude as the values of σ_u^2 and σ_v^2 increased and decreased respectively (Figure 5.5(b),(c),(d)); that is, the larger the variance between clusters (sub-domains) and the smaller the variance between domains are, the corrected MSE of the EB predictor becomes closer to the one obtained under the ideal double bootstrap MSE. In sum-

mary, as the cluster variability σ_u^2 increases compared with the domain variability, σ_v^2, the corrected MSE estimators and bootstrap MSE versions are getting closer.
For the sake of illustration of our methodology, the simulation experiments were performed only for the balanced case, that is, when the number of samples was the same for each cluster. For further analysis, the experiment can be extended to the unbalanced case. This work confined the attention to the framework of mixed models with homogeneous random area-specific effects. However, in real life, this assumption may not always be justified. The assessment of the performance of the proposed models including spatial dependent random area effects, as well as a development of prediction intervals theoretically appropriate for lognormal data would be interesting avenues for future research. Furthermore, in terms of model misspecification, it should be interesting to study some other skewed distributions following the developed methodolog.

Appendices

Appendix A

Derivation of MMSE Predictor

The model (4.7) in matrix form for each $j \in s_i$ is as follows:

$$l_{ij} = X_{ij}\beta + v_i 1_{n_{ij}} + u_{ij} 1_{n_{ij}} + e_{ij}, \tag{A.1}$$

where

$$l_{ij} = \begin{bmatrix} l_{ij1} \\ l_{ij2} \\ \vdots \\ l_{ijn_{ij}} \end{bmatrix}_{n_{ij}\times 1} ; \quad X_{ij} = \begin{bmatrix} x_{ij}^T \\ x_{ij}^T \\ \vdots \\ x_{ij}^T \end{bmatrix}_{n_{ij}\times p} ; \quad e_{ij} = \begin{bmatrix} e_{ij1} \\ e_{ij2} \\ \vdots \\ e_{ijn_{ij}} \end{bmatrix}_{n_{ij}\times 1} ; \quad 1_{n_{ij}} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{n_{ij}\times 1} .$$

From (A.1), the variance of the vector of transformed variable, l_{ij}, is given by

$$\begin{aligned} var(l_{ij}) =& \sigma_v^2 1_{n_{ij}} 1_{n_{ij}}^T + \sigma_u^2 1_{n_{ij}} 1_{n_{ij}}^T + \sigma_e^2 I_{n_{ij}} \\ =& \sigma_v^2 J_{n_{ij}} + \sigma_u^2 J_{n_{ij}} + \sigma_e I_{n_{ij}} \end{aligned}$$

Considering all clusters in the sample for each i, it follows that

$$\begin{aligned} l_{i1} =& X_{i1}\beta + v_i 1_{n_{i1}} + u_{i1} 1_{n_{i1}} + 0u_{i2} + \ldots + 0u_{im_i} + e_{i1} \\ l_{i2} =& X_{i2}\beta + v_i 1_{n_{i2}} + 0u_{i1} + u_{i2} 1_{n_{i2}} + 0u_{i3} + \ldots + 0u_{im_i} + e_{i2} \\ &\vdots \\ l_{im_i} =& X_{im_i}\beta + v_i 1_{n_{im_i}} + 0u_{i1} + 0u_{i2} + \ldots + u_{im_i} 1_{n_{im_i}} + e_{im_i}. \end{aligned}$$

Then by a given i we have the vector, l_i, that combines the expressions represented in (A.1)

$$l_i = \begin{bmatrix} l_{i1}^T \\ l_{i2}^T \\ \vdots \\ l_{im_i}^T \end{bmatrix} = \begin{bmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{im_i} \end{bmatrix} \beta + \begin{bmatrix} 1_{n_{i1}} \\ 1_{n_{i2}} \\ \vdots \\ 1_{n_{im_i}} \end{bmatrix} v_i + \begin{bmatrix} 1_{n_{i1}} & & & \\ & 1_{n_{i2}} & & \\ & & \ddots & \\ & & & 1_{n_{im_i}} \end{bmatrix} \begin{bmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{im_i}^T \end{bmatrix} + \begin{bmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{im_i} \end{bmatrix},$$

and its variance is given by

$$var(l_i) = V_i = \sigma_v^2 \begin{bmatrix} 1_{n_{i1}} \\ 1_{n_{i2}} \\ \vdots \\ 1_{n_{im_i}} \end{bmatrix} \begin{bmatrix} 1_{n_{i1}} \\ 1_{n_{i2}} \\ \vdots \\ 1_{n_{im_i}} \end{bmatrix}^T + \sigma_u^2 \begin{bmatrix} 1_{n_{i1}} & & & \\ & 1_{n_{i2}} & & \\ & & \ddots & \\ & & & 1_{n_{im_i}} \end{bmatrix} \begin{bmatrix} 1_{n_{i1}} & & & \\ & 1_{n_{i2}} & & \\ & & \ddots & \\ & & & 1_{n_{im_i}} \end{bmatrix}^T + \sigma_e^2 I_i.$$

By the fact that,

$$\begin{bmatrix} A \\ B \end{bmatrix}^T = \begin{bmatrix} A^T & B^T \end{bmatrix},$$

it follows

$$\begin{bmatrix} 1_{n_{i1}} \\ 1_{n_{i2}} \\ \vdots \\ 1_{n_{im_i}} \end{bmatrix} \begin{bmatrix} 1_{n_{i1}}^T & 1_{n_{i2}}^T & \dots & 1_{n_{im_i}}^T \end{bmatrix} = \begin{bmatrix} J_{n_{i1}} & J_{n_{i1}n_{i2}} & \dots & J_{n_{i1}n_{im_i}} \\ J_{n_{i2}n_{i1}} & J_{n_{i2}} & \dots & J_{n_{im_i}n_{i1}} \\ \vdots & & & \\ J_{n_{im_i}n_{i1}} & J_{n_{im_i}n_{i2}} & \dots & J_{n_{im_i}} \end{bmatrix}$$

$$\begin{bmatrix} 1_{n_{i1}} & & & \\ & 1_{n_{i2}} & & \\ & & \ddots & \\ & & & 1_{n_{im_i}} \end{bmatrix} \begin{bmatrix} 1_{n_{i1}} & & & \\ & 1_{n_{i2}} & & \\ & & \ddots & \\ & & & 1_{n_{im_i}} \end{bmatrix}^T = \begin{bmatrix} J_{n_{i1}} & & & \\ & J_{n_{i2}} & & \\ & & \ddots & \\ & & & J_{n_{im_i}} \end{bmatrix}$$

$$I_i = \begin{bmatrix} I_{n_{i1}} & & & \\ & I_{n_{i2}} & & \\ & & \ddots & \\ & & & I_{n_{im_i}} \end{bmatrix}.$$

Then

$$
\begin{aligned}
V_i =&\sigma_v^2 \begin{bmatrix} J_{n_{i1}} & J_{n_{i1}n_{i2}} & \dots & J_{n_{i1}n_{im_i}} \\ J_{n_{i2}n_{i1}} & J_{n_{i2}} & \dots & J_{n_{im_i}n_{i1}} \\ \vdots & & & \\ J_{n_{im_i}n_{i1}} & J_{n_{im_i}n_{i2}} & \dots & J_{n_{im_i}} \end{bmatrix} + \sigma_u^2 \begin{bmatrix} J_{n_{i1}} & & & \\ & J_{n_{i2}} & & \\ & & \ddots & \\ & & & J_{n_{im_i}} \end{bmatrix} + \sigma_e^2 \begin{bmatrix} I_{n_{i1}} & & & \\ & I_{n_{i2}} & & \\ & & \ddots & \\ & & & I_{n_{im_i}} \end{bmatrix} \\
=& \begin{bmatrix} (\sigma_v^2+\sigma_u^2)J_{n_{i1}}+\sigma_e^2 I_{n_{i1}} & \sigma_v^2 J_{n_{i1}n_{i2}} & \dots & \sigma_v^2 J_{n_{i1}n_{im_i}} \\ \sigma_v^2 J_{n_{i2}n_{i1}} & (\sigma_v^2+\sigma_u^2)J_{n_{i2}}+\sigma_e^2 I_{n_{i2}} & \dots & \sigma_v^2 J_{n_{i2}n_{im_i}} \\ \vdots & & & \\ \sigma_v^2 J_{n_{im_i}n_{i1}} & \sigma_v^2 J_{n_{im_i}n_{i2}} & \dots & (\sigma_v^2+\sigma_v^2)J_{n_{im_i}}+\sigma_e^2 I_{n_{im_i}} \end{bmatrix}.
\end{aligned}
$$

So from the previous, we have a joint distribution of the expressions represented by (4.7) for a given small area i.

$$
l_i = \begin{bmatrix} l_{i1} \\ l_{i2} \\ \vdots \\ l_{im_i} \end{bmatrix} \sim N\Big(\begin{bmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{im_i} \end{bmatrix} \beta, V_i \Big) \equiv N(X_i\beta, V_i).
$$

Now considering the mean at cluster level

$$
\bar{l}_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} l_{ijk} = \frac{1}{n_{ij}} \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} l_{ij1} \\ l_{ij2} \\ \vdots \\ l_{ijn_{ij}} \end{bmatrix} = \frac{1}{n_{ij}} 1_{n_{ij}}^T l_{ij}. \tag{A.2}
$$

It follows that

$$
\begin{aligned}
\bar{l}_i = \begin{bmatrix} \bar{l}_{i1} \\ \bar{l}_{i2} \\ \vdots \\ \bar{l}_{im_i} \end{bmatrix} = \begin{bmatrix} \frac{1}{n_{i1}} 1_{n_{i1}}^T l_{i1} \\ \frac{1}{n_{i2}} 1_{n_{i2}}^T l_{i2} \\ \vdots \\ \frac{1}{n_{im_i}} 1_{n_{im_i}}^T l_{im_i} \end{bmatrix} = \begin{bmatrix} \frac{1}{n_{i1}} 1_{n_{i1}}^T & & & \\ & \frac{1}{n_{i2}} 1_{n_{i2}}^T & & \\ & & \ddots & \\ & & & \frac{1}{n_{im_i}} 1_{n_{im_i}}^T \end{bmatrix} \begin{bmatrix} l_{i1} \\ l_{i2} \\ \vdots \\ l_{im_i} \end{bmatrix} \\
= W_i l_i.
\end{aligned}
$$

Then

$$
\bar{l}_i = W_i l_i \sim N(W_i X_i \beta, W_i V_i W_i^T).
$$

Recall that

$$X_{ij} = \begin{bmatrix} x_{ij}^T \\ x_{ij}^T \\ \vdots \\ x_{ij}^T \end{bmatrix}$$

which implies that

$$\frac{1}{n_{ij}} 1_{n_{ij}}^T X_{ij} = \frac{1}{n_{ij}} \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} x_{ij}^T \\ x_{ij}^T \\ \vdots \\ x_{ij}^T \end{bmatrix} = x_{ij}^T.$$

Note that

$$W_i X_i \beta = \begin{bmatrix} x_{i1}^T \\ x_{i2}^T \\ \vdots \\ x_{in}^T \end{bmatrix} \beta = x_i \beta,$$

$$\frac{1}{n_{ij}} 1_{n_{ij}}^T J_{n_{ij}} = \frac{1}{n_{ij}} 1_{n_{ij}}^T 1_{n_{ij}} 1_{n_{ij}}^T = 1_{n_{ij}}^T.$$

Then

$$\bar{V}_i = W_i V_i W_i^T = \begin{bmatrix} \frac{1}{n_{i1}} 1_{n_{i1}}^T & & & \\ & \frac{1}{n_{i2}} 1_{n_{i2}}^T & & \\ & & \ddots & \\ & & & \frac{1}{n_{im_i}} 1_{n_{im_i}}^T \end{bmatrix} V_i \begin{bmatrix} \frac{1}{n_{i1}} 1_{n_{i1}} & & & \\ & \frac{1}{n_{i2}} 1_{n_{i2}} & & \\ & & \ddots & \\ & & & \frac{1}{n_{im_i}} 1_{n_{im_i}} \end{bmatrix}$$

$$= \begin{bmatrix} (\sigma_v^2 + \sigma_u^2) 1_{n_{i1}}^T + \sigma_e^2 1_{n_{i1}}^T & \sigma_v^2 1_{n_{i2}}^T & \dots & \sigma_v^2 1_{n_{im_i}}^T \\ \sigma_v^2 1_{n_{i1}}^T & (\sigma_v^2 + \sigma_u^2) 1_{n_{i2}}^T + \sigma_e^2 1_{n_{i2}}^T & \dots & \sigma_v^2 1_{n_{im_i}}^T \\ \vdots & & & \\ \sigma_v^2 1_{n_{i1}}^T & \sigma_v^2 1_{n_{i2}}^T & \dots & (\sigma_v^2 + \sigma_v^2) 1_{n_{im_i}}^T + \sigma_e^2 1_{n_{im_i}}^T \end{bmatrix} \times$$

$$\begin{bmatrix} \frac{1}{n_{i1}} 1_{n_{i1}} & & & \\ & \frac{1}{n_{i2}} 1_{n_{i2}} & & \\ & & \ddots & \\ & & & \frac{1}{n_{im_i}} 1_{n_{im_i}} \end{bmatrix}$$

$$= \begin{bmatrix} (\sigma_v^2 + \sigma_u^2) + \frac{1}{n_{i1}} \sigma_e^2 & \sigma_v^2 & \dots & \sigma_v^2 \\ \sigma_v^2 & (\sigma_v^2 + \sigma_u^2) + \frac{1}{n_{i2}} \sigma_e^2 & \dots & \sigma_v^2 \\ \vdots & & & \\ \sigma_v^2 & \sigma_v^2 & \dots & (\sigma_v^2 + \sigma_u^2) + \frac{1}{n_{im_i}} \sigma_e^2 \end{bmatrix}.$$

Then

$$\bar{l}_i \sim N(x_i \beta, \bar{V}_i),$$

which is the joint distribution of the elements represented by (A.2).
Now let figure out the joint distribution between l_i and the random effects. At the first step we consider the cluster random effects

$$\begin{bmatrix} l_i \\ u_{ij} \end{bmatrix} = \begin{bmatrix} l_{i1} \\ l_{i2} \\ \vdots \\ l_{ij} \\ \vdots \\ l_{im_i} \\ u_{ij} \end{bmatrix}.$$

Considering the case where the sub-small area, j, has some observations within the sample,

$$
\begin{aligned}
cov(l_{i1}, u_{ij}) =& cov(\begin{bmatrix} l_{i11} \\ l_{i12} \\ \vdots \\ l_{i1n_1} \end{bmatrix}, u_{ij}) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{(n_{ij} \times 1)}, \\
cov(l_{ij}, u_{ij}) =& cov(\begin{bmatrix} l_{ij1} \\ l_{ij2} \\ \vdots \\ l_{ijn_j} \end{bmatrix}, u_{ij}) = \sigma_u^2 1_{n_{ij}}.
\end{aligned}
$$

It follows that

$$
C_1^{(j)} = cov(l_i, u_{ij}) = \begin{bmatrix} cov(l_{i1}, u_{ij}) \\ cov(l_{i2}, u_{ij}) \\ \vdots \\ cov(l_{ij}, u_{ij}) \\ \vdots \\ cov(l_{in}, u_{ij}) \end{bmatrix} = \begin{bmatrix} 0_{n_{i1}} \\ 0_{n_{i2}} \\ \vdots \\ \sigma_u^2 1_{n_{ij}} \\ \vdots \\ 0_{n_{im_i}} \end{bmatrix}_{(n_i \times 1)},
$$

now we have

$$
var(\begin{bmatrix} l_i \\ u_{ij} \end{bmatrix}) = \begin{bmatrix} V_i & C_1^{(j)} \\ C_1^{(j)T} & \sigma_u^2 \end{bmatrix}.
$$

Then

$$
\begin{bmatrix} l_i \\ u_{ij} \end{bmatrix} \sim N\Big(\begin{bmatrix} X_i\beta \\ 0 \end{bmatrix} \begin{bmatrix} V_i & C_1^{(j)} \\ C_1^{(j)T} & \sigma_u^2 \end{bmatrix} \Big).
$$

The next step is to calculate the conditional expectations

$$
\begin{aligned}
E(u_{ij}|l_i) =& cov(u_{ij}, l_i) V_i^{-1} (l_i - X_i\beta) \\
=& C_1^{(j)T} V_i^{-1} (l_i - X_i\beta)
\end{aligned} \tag{A.3}
$$

and

$$\begin{aligned}
E(u_{ij} \mid \bar{l}_i) =& cov(u_{ij}, \bar{l}_i)(var(\bar{l}_i))^{-1}(\bar{l}_i - x_i\beta) \\
=& cov(u_{ij}, W_i l_i)(W_i V_i W_i^T)^{-1}(W_i l_i - W_i X_i \beta) \\
=& cov(u_{ij}, l_i) W_i^T (W_i V_i W_i^T)^{-1}(W_i l_i - W_i X_i \beta) \\
=& C_1^{(j)T} W_i^T W_i^{-T} V_i^{-1} W_i^{-1} W_i (l_i - X_i \beta) \\
=& C_1^{(j)T} V_i^{-1} (l_i - X_i \beta).
\end{aligned} \tag{A.4}$$

Therefore, from (A.3) and (A.4) we have

$$E(u_{ij}|l_i) = E(u_{ij}|\bar{l}_i).$$

Now we prove the same results but in case of conditional variance.

$$\begin{aligned}
var(u_{ij}|l_i) =& var(u_{ij}) - cov(u_{ij}, l_i) V_i^{-1} cov(l_i, u_{ij}) \\
=& \sigma_u^2 - C_1^{(j)T} V_i^{-1} C_1^{(j)}
\end{aligned} \tag{A.5}$$

and

$$\begin{aligned}
var(u_{ij}|\bar{l}_i) =& var(u_{ij}) - cov(u_{ij}, \bar{l}_i) \bar{V}^{-1} cov(\bar{l}_i, u_{ij}) \\
=& \sigma_u^2 - cov(u_{ij}, l_i) W_i^T (W_i V_i W_i^T)^{-1} W_i cov(l_i, u_{ij}) \\
=& \sigma_u^2 - C_1^{(j)T} W_i^T W_i^{-T} V_i^{-1} W_i^{-1} W_i C_1^{(j)} \\
=& \sigma_u^2 - C_1^{(j)T} V_i^{-1} C_1^{(j)}.
\end{aligned} \tag{A.6}$$

Therefore, from (A.5) and (A.6) we have

$$var(u_{ij}|l_i) = var(u_{ij}|\bar{l}_i).$$

From the above expressions we have

$$
\begin{aligned}
\bar{l}_i &\sim N(x_i\beta, \bar{V}_i) \\
cov(u_{ij}, \bar{l}_i) &= cov(u_{ij}, W_i l_i) \\
&= C_1^{(j)T} W_i^T = W_i C_1^{(j)} = \sigma_u^2 \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{(m_i \times 1)} = \alpha^{(j)}, \\
cov(v_i, \bar{l}_i) &= cov(v_i, W_i l_i) = \sigma_v^2 W_i cov(l_i, v_i) \\
&= \sigma_v^2 W_i \begin{bmatrix} 1_{n_{i1}} \\ 1_{n_{i2}} \\ \vdots \\ 1_{n_{ij}} \\ \vdots \\ 1_{n_{im_i}} \end{bmatrix}_{(n_i \times 1)} = \sigma_v^2 \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{(m_i \times 1)} = \gamma.
\end{aligned}
$$

Now the joint distribution is

$$
\begin{bmatrix} \bar{l}_i \\ u_{ij} \\ v_i \end{bmatrix} \sim N\Big(\begin{bmatrix} x_i\beta \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \bar{V}_i & \alpha^{(j)} & \gamma \\ \alpha^{(j)T} & \sigma_u^2 & 0 \\ \gamma^T & 0 & \sigma_v^2 \end{bmatrix} \Big). \tag{A.7}
$$

Recall that, given a multivariate normal distribution : $X \sim N(\mu, \Sigma)$, where

$$
X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\Big(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \Big)
$$

we have

$$
\begin{aligned}
E(X_1 | X_2 = x_2) &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) \\
var(X_1 | X_2 = x_2) &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.
\end{aligned}
$$

The moment generating function of X is

$$
M_X(t) = E[e^{t^T X}] = e^{t^T\mu + \frac{1}{2}t^T\Sigma t}.
$$

From (A.7), we have

$$E(\begin{bmatrix} u_{ij} \\ v_i \end{bmatrix} \mid \bar{l}_i) = \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}_i^{-1}(\bar{l}_i - x_i\beta) \equiv \mu_{1j}$$

and

$$var(\begin{bmatrix} u_{ij} \\ v_i \end{bmatrix} \mid \bar{l}_i) = \begin{bmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_v^2 \end{bmatrix} - \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}_i^{-1} \begin{bmatrix} \alpha^{(j)} & \gamma \end{bmatrix} \equiv \Sigma_{1j}.$$

Using the above results, we have

$$E(\exp\{v_i + u_{ij}\} \mid \bar{l}_i) = \exp\{1^T\mu_{1j} + \frac{1}{2}1^T\Sigma_{1j}1\}.$$

Now the expression of the second term in (4.8) is given by

$$\tilde{y}^*_{ijk} \equiv E(y_{ijk} \mid \bar{l}_i) = \exp\{x_{ij}^T\beta + 1^T\mu_{1j} + \frac{1}{2}1^T\Sigma_{1j}1 + \frac{1}{2}\sigma_e^2\}. \qquad \text{(A.8)}$$

Next we consider the second case with a sub-small area, r, in the sampled small area, but does not have any observation in the sample

$$cov(l_{i1}, u_{ir}) = cov(\begin{bmatrix} l_{i11} \\ l_{i12} \\ \vdots \\ l_{i1n_1} \end{bmatrix}, u_{ir}) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{(n_{ij}\times 1)}$$

$$cov(l_{ij}, u_{ir}) = cov(\begin{bmatrix} l_{ij1} \\ l_{ij2} \\ \vdots \\ l_{ijn_j} \end{bmatrix}, u_{ir}) = 0_{n_{ij}}.$$

It follows that

$$C_2 = cov(l_i, u_{ir}) = \begin{bmatrix} cov(l_{i1}, u_{ir}) \\ cov(l_{i2}, u_{ir}) \\ \vdots \\ cov(l_{ij}, u_{ir}) \\ \vdots \\ cov(l_{in}, u_{ir}) \end{bmatrix} = \begin{bmatrix} 0_{n_{i1}} \\ 0_{n_{i2}} \\ \vdots \\ 0_{n_{ij}} \\ \vdots \\ 0_{n_{im_i}} \end{bmatrix}_{(n_i\times 1)} \equiv 0.$$

Now we have

$$var(\begin{bmatrix} l_i \\ u_{ir} \end{bmatrix}) = \begin{bmatrix} V_i & 0 \\ 0^T & \sigma_u^2 \end{bmatrix}.$$

Then

$$\begin{bmatrix} l_i \\ u_{ir} \end{bmatrix} \sim N\Big(\begin{bmatrix} X_i\beta \\ 0 \end{bmatrix} \begin{bmatrix} V_i & 0 \\ 0^T & \sigma_u^2 \end{bmatrix} \Big).$$

So it follows that

$$E(u_{ir} \mid l_i) = E(u_{ir} \mid \bar{l}_i) = 0.$$

Now the joint distribution is

$$\begin{bmatrix} \bar{l}_i \\ u_{ir} \\ v_i \end{bmatrix} \sim N\Big(\begin{bmatrix} x_i\beta \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \bar{V}_i & 0 & \gamma \\ 0^T & \sigma_u^2 & 0 \\ \gamma^T & 0 & \sigma_v^2 \end{bmatrix} \Big). \tag{A.9}$$

Proceeding the same way as in the first case and using (A.9) we have

$$E(\begin{bmatrix} u_{ir} \\ v_i \end{bmatrix} \mid \bar{l}_i) = \begin{bmatrix} 0^T \\ \gamma^T \end{bmatrix} \bar{V}_i^{-1}(\bar{l}_i - x_i\beta) \equiv \mu_2$$

and

$$var(\begin{bmatrix} u_{ir} \\ v_i \end{bmatrix} \mid \bar{l}_i) = \begin{bmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_v^2 \end{bmatrix} - \begin{bmatrix} 0^T \\ \gamma^T \end{bmatrix} \bar{V}^{-1} \begin{bmatrix} 0 & \gamma \end{bmatrix} \equiv \Sigma_2.$$

Using the above results, we have

$$E(\exp\{v_i + u_{ir}\} \mid \bar{l}_i) = \exp\{1^T\mu_2 + \frac{1}{2}1^T\Sigma_2 1\}.$$

Now the expression of the third term in (4.8) is given by

$$\tilde{y}_{ijk}^{**} \equiv E(y_{ijk} \mid \bar{l}_i) = \exp\{x_{ij}^T\beta + 1^T\mu_2 + \frac{1}{2}1^T\Sigma_2 1 + \frac{1}{2}\sigma_e^2\}. \tag{A.10}$$

Substituting the expressions (A.8) and (A.10) in (4.8), the minimum MSE predictor, under the assumption that $\theta = (\beta, \sigma_v^2, \sigma_u^2, \sigma_e^2)^T$ is known, is given by

$$\begin{aligned} \hat{\bar{Y}}_i^{MMSE}(\theta) =& E[\bar{Y}_i|(y,x)] \\ =& \frac{1}{N_i}\Bigg[\sum_{j,k\in s_i} y_{ijk} + \sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}} \tilde{y}_{ijk}^*(\theta) + \sum_{j\in \bar{s}_i}\sum_{k=1}^{N_{ij}} \tilde{y}_{ijk}^{**}(\theta) \Bigg]. \end{aligned} \tag{A.11}$$

Appendix B

MSE of the MMSE predictor

By definition,

$$
\begin{aligned}
M_{1i} &= E[(\bar{Y}_i - \hat{\bar{Y}}_i^{MMSE})^2] \\
=&E[(\bar{Y}_i - E[\bar{Y}_i|(y,x)])^2] \\
=&E\{E[(\bar{Y}_i - E[\bar{Y}_i|(y,x)])^2|(y,x)]\} \\
=&E[var(\bar{Y}_i|(y,x))] \\
=&E\left[\frac{1}{N_i^2}\left(var\left(\sum_{j\in s_i}\sum_{k\in s_{ij}} y_{ijk} + \sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}} y_{ijk} + \sum_{j\in \bar{s}_i}\sum_{k=1}^{N_{ij}} y_{ijk}|(y,x)\right)\right)\right] \\
=&E\Bigg[\frac{1}{N_i^2}\Bigg(var\left(\sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}} y_{ijk}|(y,x)\right) + var\left(\sum_{j\in \bar{s}_i}\sum_{k=1}^{N_{ij}} y_{ijk}|(y,x)\right) + \\
&2cov\left(\sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}} y_{ijk}, \sum_{j\in \bar{s}_i}\sum_{k=1}^{N_{ij}} y_{ijk}|(y,x)\right)\Bigg)\Bigg] \\
=&\frac{1}{N_i^2}\Bigg[E(V_1) + E(V_2) + 2E(C_1)\Bigg]. \qquad \text{(B.1)}
\end{aligned}
$$

In below we show in detail the corresponding expressions of those terms of the right-hand side of (B.1).

Starting by the first term, it follows that

$$\begin{aligned} V_1 =& var\left(\sum_{j\in s_i}\sum_{k\in\bar{s}_{ij}} y_{ijk}|(y,x)\right) \\ =& cov\left(\sum_{j\in s_i}\sum_{k\in\bar{s}_{ij}} y_{ijk}, \sum_{q\in s_i}\sum_{p\in\bar{s}_{iq}} y_{iqp}|(y,x)\right) \\ =& \sum_{j\in s_i}\sum_{k\in\bar{s}_{ij}}\sum_{q\in s_i}\sum_{p\in\bar{s}_{ir}} cov(y_{ijk}, y_{irp}|(y,x)). \end{aligned} \tag{B.2}$$

The conditional covariance is defined by

$$cov(y_{ijk}, y_{irp}|(y,x)) = E[y_{ijk}y_{iqp}|(y,x)] - E[y_{ijk}|(y,x)]E[y_{iqp}|(y,x)].$$

We start by $E[y_{ijk}y_{iqp}|(y,x)]$:

- For $j = q$ and $k = p$:

$$\begin{aligned} E[y_{ijk}y_{iqp}|(y,x)] =& E[\exp\{2(x_{ij}^T\beta + v_i + u_{ij} + e_{ijk})\}|(y,x)], \\ =& \exp\{2(x_{ij}^T\beta + 1^T\mu_{1j} + 1^T\Sigma_{1j}1 + \sigma_e^2)\}. \end{aligned}$$

- For $j = q$ and $k \neq p$:

$$\begin{aligned} E[y_{ijk}y_{irp}|(y,x)] =& E[\exp\{2x_{ij}^T\beta + 2v_i + 2u_{ij} + e_{ijk} + e_{ijp}\}], \\ =& \exp\{2(x_{ij}^T\beta + 1^T\mu_{1j} + 1^T\Sigma_{1j}1) + \sigma_e^2\}. \end{aligned}$$

- For $j \neq q$:

$$E[y_{ijk}y_{iqp}|(y,x)] = E[\exp\{x_{ij}^T\beta + x_{iq}^T\beta + 2v_i + u_{ij} + u_{iq} + e_{ijk} + e_{iqp}\}]. \tag{B.3}$$

Note that

$$\begin{bmatrix} \bar{l}_i \\ 2v_i \\ u_{ij} \\ u_{iq} \end{bmatrix} \sim N\left(\begin{bmatrix} x_i\beta \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \bar{V}_i & 2\gamma & \alpha^{(j)} & \alpha^{(q)} \\ 2\gamma^T & 4\sigma_v^2 & 0 & 0 \\ \alpha^{(j)T} & 0 & \sigma_u^2 & 0 \\ \alpha^{(q)T} & 0 & 0 & \sigma_u^2 \end{bmatrix} \right). \tag{B.4}$$

From (B.4), it follows that

$$E\left(\begin{bmatrix} 2v_i \\ u_{ij} \\ u_{ir} \end{bmatrix} \mid \bar{l}_i \right) = \begin{bmatrix} 2\gamma^T \\ \alpha^{(j)T} \\ \alpha^{(r)T} \end{bmatrix} \bar{V}^{-1}(\bar{l}_i - x_i\beta) \equiv \mu_{1jq} \tag{B.5}$$

and

$$var\left(\begin{bmatrix} 2v_i \\ u_{ij} \\ u_{iq} \end{bmatrix} \mid \bar{l}_i \right) = \begin{bmatrix} 4\sigma_v^2 & 0 & 0 \\ 0 & \sigma_u^2 & 0 \\ 0 & 0 & \sigma_u^2 \end{bmatrix} - \begin{bmatrix} 2\gamma^T \\ \alpha^{(j)T} \\ \alpha^{(q)T} \end{bmatrix} \bar{V}_i^{-1} \begin{bmatrix} 2\gamma & \alpha^{(j)} & \alpha^{(q)} \end{bmatrix} \equiv \Sigma_{1jq}. \tag{B.6}$$

Using (B.5) and (B.6), the expression (B.3) becomes

$$E[y_{ijk}y_{iqp}|(y,x)] = \exp\{x_{ij}^T\beta + x_{iq}^T\beta + 1^T\mu_{1jq} + \frac{1}{2}1^T\Sigma_{1jq}1 + \sigma_e^2\}.$$

From those three cases,

$$\begin{aligned} &E[y_{ijk}y_{iqp}|(y,x)] = \\ &[\exp\{2(x_{ij}^T\beta + 1^T\mu_{1j} + 1^T\Sigma_{1j}1 + \sigma_e^2)\}I(k=p)+ \\ &\exp\{2(x_{ij}^T\beta + 1^T\mu_{1j} + 1^T\Sigma_{1j}1) + \sigma_e^2\}I(k \neq p)]I(j=q)+ \\ &[\exp\{x_{ij}^T\beta + x_{iq}^T\beta + 1^T\mu_{1jq} + \frac{1}{2}1^T\Sigma_{1jq}1 + \sigma_e^2\}]I(j \neq q). \end{aligned} \tag{B.7}$$

Now the next expression is $E[y_{ijk}|(y,x)]E[y_{iqp}|(y,x)]$:

- For $j = q$:

$$\begin{aligned} E[y_{ijk}|(y,x)]E[y_{irp}|(x,y)] &= (E[y_{ijk}|(y,x)])^2 \\ &= \exp\{2(x_{ij}^T\beta + 1^T\mu_{1j}) + 1^T\Sigma_{1j}1 + \sigma_e^2\}. \end{aligned}$$

- For $j \neq q$:

$$\begin{aligned} &E[y_{ijk}|(y,x)]E[y_{iqp}|(x,y)] = \\ &= E[\exp\{x_{ij}^T\beta + v_i + u_{ij} + e_{ijk}\}|(y,x)]E[\exp\{x_{iq}^T\beta + v_i + u_{iq} + e_{iqp}\}|(y,x)] \\ &= \exp\{x_{ij}^T\beta + 1^T\mu_{1j} + 1^T\Sigma_{1j}1 + \sigma_e^2\}\exp\{(x_{ir}^T\beta + 1^T\mu_{1q} + 1^T\Sigma_{1q}1 + \sigma_e^2\} \\ &= \exp\{x_{ij}^T\beta + x_{iq}^T\beta + 1^T(\mu_{1j} + \mu_{1q}) + \frac{1}{2}1^T(\Sigma_{1j} + \Sigma_{1q})1 + \sigma_e^2\}. \end{aligned}$$

It follows that

$$\begin{aligned} &E[y_{ijk}|(y,x)]E[y_{iqp}|(x,y)] = \\ &= [\exp\{2(x_{ij}^T\beta + 1^T\mu_{1j}) + 1^T\Sigma_{1j}1 + \sigma_e^2\}]I(j=q)+ \\ &[\exp\{x_{ij}^T\beta + x_{iq}^T\beta + 1^T(\mu_{1j} + \mu_{1q}) + \frac{1}{2}1^T(\Sigma_{1j} + \Sigma_{1q})1 + \sigma_e^2\}]I(j \neq q). \end{aligned} \tag{B.8}$$

We have from (B.7) and (B.8) that

$$
\begin{aligned}
&cov(y_{ijk}, y_{iqp} \mid \bar{l}_i) = \\
&[\exp\{2(x_{ij}^T\beta + 1^T\mu_{1j} + 1^T\Sigma_{1j}1 + \sigma_e^2)\}I(k=p)+ \\
&\exp\{2(x_{ij}^T\beta + 1^T\mu_{1j} + 1^T\Sigma_{1j}1) + \sigma_e^2\}I(k \neq p)]I(j=q)+ \\
&[\exp\{x_{ij}^T\beta + x_{iq}^T\beta + 1^T\mu_{1jq} + \frac{1}{2}1^T\Sigma_{1jq}1 + \sigma_e^2\}]I(j \neq q)- \\
&[\exp\{2(x_{ij}^T\beta + 1^T\mu_{1j}) + 1^T\Sigma_{1j}1 + \sigma_e^2\}]I(j=q)+ \\
&[\exp\{x_{ij}^T\beta + x_{iq}^T\beta + 1^T(\mu_{1j} + \mu_{1q}) + \frac{1}{2}1^T(\Sigma_{1j} + \Sigma_{1q})1 + \sigma_e^2\}]I(j \neq q) \\
=&[(\exp\{2(x_{ij}^T\beta + 1^T\mu_{1j} + 1^T\Sigma_{1j}1 + \sigma_e^2)\} - \exp\{2(x_{ij}^T\beta + 1^T\mu_{1j}) + 1^T\Sigma_{1j}1 + \sigma_e^2\})I(k=p)+ \\
&(\exp\{2(x_{ij}^T\beta + 1^T\mu_{1j} + 1^T\Sigma_{1j}1) + \sigma_e^2\} - \exp\{2(x_{ij}^T\beta + 1^T\mu_{1j}) + 1^T\Sigma_{1j}1 + \sigma_e^2\})I(k \neq p)]I(j=q)+ \\
&[\exp\{x_{ij}^T\beta + x_{iq}^T\beta + 1^T\mu_{1jq} + \frac{1}{2}1^T\Sigma_{1jq}1 + \sigma_e^2\} - \exp\{x_{ij}^T\beta + x_{iq}^T\beta + 1^T(\mu_{1j} + \mu_{1q})+ \\
&\frac{1}{2}1^T(\Sigma_{1j} + \Sigma_{1q})1 + \sigma_e^2\}]I(j \neq q).
\end{aligned}
$$

From the previous section,

$$
\begin{aligned}
E(1^T\mu_{1j}) =& 0 \\
var(1^T\mu_{1j}) =& 1^T V_{\mu_{1j}} 1,
\end{aligned}
$$

where

$$
V_{\mu_{1j}} = var(\mu_{1j}) = \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}_i^{-1} \begin{bmatrix} \alpha^{(j)} & \gamma \end{bmatrix}.
$$

Then we have

$$
\begin{aligned}
&E[cov(y_{ijk}, y_{iqp} \mid \bar{l}_i)] = \\
&[(\exp\{2(x_{ij}^T\beta + 1^T V_{\mu_{1j}} 1 + 1^T \Sigma_{1j} 1 + \sigma_e^2)\} - \exp\{2(x_{ij}^T\beta + 1^T V_{\mu_{1j}} 1) + 1^T \Sigma_{1j} 1 + \sigma_e^2\}) I(k = p) + \\
&(\exp\{2(x_{ij}^T\beta + 1^T V_{\mu_{1j}} 1 + 1^T \Sigma_{1j} 1) + \sigma_e^2\} - \exp\{2(x_{ij}^T\beta + 1^T V_{\mu_{1j}} 1) + 1^T \Sigma_{1j} 1 + \sigma_e^2\}) I(k \neq p)] I(j = q) \\
&+ [\exp\{x_{ij}^T\beta + x_{iq}^T\beta + \frac{1}{2} 1^T V_{\mu_{1jq}} 1 + \frac{1}{2} 1^T \Sigma_{1jq} 1 + \sigma_e^2\} - \exp\{x_{ij}^T\beta + x_{iq}^T\beta + \frac{1}{2} 1^T (V_{\mu_{1j}} + V_{\mu_{1q}}) 1 + \\
&\frac{1}{2} 1^T (\Sigma_{1j} + \Sigma_{1q}) 1 + \sigma_e^2\}] I(j \neq q) \\
=&[v_{11j} I(k = p) + v_{12j} I(k \neq p)] I(j = q) + v_{13jq} I(j \neq q) \\
=&\exp\{2(x_{ij}^T\beta + 1^T V_{\mu_{1j}} 1) + 1^T \Sigma_{1j} 1 + \sigma_e^2\} (\exp\{1^T \Sigma_{1j} 1 + \sigma_e^2\} - 1) I(k = p) + \\
&(\exp\{1^T \Sigma_{1j} 1\} - 1) I(k \neq p)] I(j = q) \\
&+ [\exp\{x_{ij}^T\beta + x_{iq}^T\beta + \frac{1}{2} 1^T V_{\mu_{1jq}} 1 + \frac{1}{2} 1^T \Sigma_{1jq} 1 + \sigma_e^2\} - \exp\{x_{ij}^T\beta + x_{iq}^T\beta + \frac{1}{2} 1^T (V_{\mu_{1j}} + V_{\mu_{1q}}) 1 + \\
&\frac{1}{2} 1^T (\Sigma_{1j} + \Sigma_{1q}) 1 + \sigma_e^2\}] I(j \neq q).
\end{aligned}
$$

Within the expression above we add and subtract $\exp\{1^T \Sigma_{1j} 1\} I(k = p)$

$$
\begin{aligned}
&E[cov(y_{ijk}, y_{iqp} \mid \bar{l}_i)] = \\
&\exp\{2(x_{ij}^T\beta + 1^T V_{\mu_{1j}} 1) + 1^T \Sigma_{1j} 1 + \sigma_e^2\} [(\exp\{1^T \Sigma_{1j} 1 + \sigma_e^2\} I(k = p) + \exp\{1^T \Sigma_{1j} 1\} I(k = p) - \\
&\exp\{1^T \Sigma_{1j} 1\} I(k = p) + \exp\{1^T \Sigma_{1j} 1\} I(k \neq p) - 1)] I(j = q) \\
&+ [\exp\{x_{ij}^T\beta + x_{iq}^T\beta + \frac{1}{2} 1^T V_{\mu_{1jq}} 1 + \frac{1}{2} 1^T \Sigma_{1jq} 1 + \sigma_e^2\} - \exp\{x_{ij}^T\beta + x_{iq}^T\beta + \frac{1}{2} 1^T (V_{\mu_{1j}} + V_{\mu_{1q}}) 1 + \\
&\frac{1}{2} 1^T (\Sigma_{1j} + \Sigma_{1q}) 1 + \sigma_e^2\}] I(j \neq q) \\
=&\exp\{2(x_{ij}^T\beta + 1^T V_{\mu_{1j}} 1) + 1^T \Sigma_{1j} 1 + \sigma_e^2\} [(\exp\{1^T \Sigma_{1j} 1 + \sigma_e^2\} - \exp\{1^T \Sigma_{1j} 1\}) I(k = p) + \\
&(\exp\{1^T \Sigma_{1j} 1\} - 1)] I(j = q) \\
&+ [\exp\{x_{ij}^T\beta + x_{iq}^T\beta + 1^T V_{\mu_{1jq}} 1 + \frac{1}{2} 1^T \Sigma_{1jq} 1 + \sigma_e^2\} - \exp\{x_{ij}^T\beta + x_{iq}^T\beta + 1^T (V_{\mu_{1j}} + V_{\mu_{1q}}) 1 + \\
&\frac{1}{2} 1^T (\Sigma_{1j} + \Sigma_{1q}) 1 + \sigma_e^2\}] I(j \neq q)
\end{aligned}
$$

$$
\begin{aligned}
=&[\exp\{2(x_{ij}^T\beta + 1^T V_{\mu_{1j}} 1) + 1^T\Sigma_{1j}1 + \sigma_e^2\}(\exp\{1^T\Sigma_{1j}1 + \sigma_e^2\} - \exp\{1^T\Sigma_{1j}1\})I(k=p)+ \\
&\exp\{2(x_{ij}^T\beta + 1^T V_{\mu_{1j}} 1) + 1^T\Sigma_{1j}1 + \sigma_e^2\}(\exp\{1^T\Sigma_{1j}1\} - 1)]I(j=q) \\
&+ [\exp\{x_{ij}^T\beta + x_{iq}^T\beta + \frac{1}{2}1^T V_{\mu_{1jq}} 1 + \frac{1}{2}1^T\Sigma_{1jq}1 + \sigma_e^2\} - \exp\{x_{ij}^T\beta + x_{iq}^T\beta + \frac{1}{2}1^T(V_{\mu_{1j}} + V_{\mu_{1q}})1 + \\
&\frac{1}{2}1^T(\Sigma_{1j} + \Sigma_{1q})1 + \sigma_e^2\}]I(j \neq q) \\
=&v_{11j}I(k=p)I(j=q) + v_{12j}I(j=q) + v_{13jq}I(j \neq q),
\end{aligned}
$$

where

$$
\begin{aligned}
v_{11j} =& \exp\{2(x_{ij}^T\beta + 1^T V_{\mu_{1j}} 1) + 1^T\Sigma_{1j}1 + \sigma_e^2\}(\exp\{1^T\Sigma_{1j}1 + \sigma_e^2\} - \exp\{1^T\Sigma_{1j}1\}) \\
v_{12j} =& \exp\{2(x_{ij}^T\beta + 1^T V_{\mu_{1j}} 1) + 1^T\Sigma_{1j}1 + \sigma_e^2\}(\exp\{1^T\Sigma_{1j}1\} - 1). \\
v_{13jr} =& \exp\{x_{ij}^T\beta + x_{iq}^T\beta + \frac{1}{2}1^T V_{\mu_{1jq}} 1 + \frac{1}{2}1^T\Sigma_{1jq}1 + \sigma_e^2\} - \exp\{x_{ij}^T\beta + x_{iq}^T\beta + \frac{1}{2}1^T(V_{\mu_{1j}} + V_{\mu_{1q}})1 + \\
&\frac{1}{2}1^T(\Sigma_{1j} + \Sigma_{1q})1 + \sigma_e^2\}.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
E(V_1) =& \sum_{j\in s_i}\sum_{k\in\bar{s}_{ij}}\sum_{q\in s_i}\sum_{p\in\bar{s}_{iq}} v_{11j}I(p=k)I(q=j) + \sum_{j\in s_i}\sum_{k\in\bar{s}_{ij}}\sum_{q\in s_i}\sum_{p\in\bar{s}_{iq}} v_{12j}I(q=j)+ \\
&\sum_{j\in s_i}\sum_{k\in\bar{s}_{ij}}\sum_{q\in s_i}\sum_{p\in\bar{s}_{iq}} v_{13jq}I(q \neq j) \\
=& \sum_{j\in s_i}\sum_{k\in\bar{s}_{ij}} v_{11j}\left[\sum_{q\in s_i}\sum_{p\in\bar{s}_{iq}} I(p=k)I(q=j)\right] + \sum_{j\in s_i}\sum_{k\in\bar{s}_{ij}} v_{12j}\left[\sum_{q\in s_i}\sum_{p\in\bar{s}_{ir}} I(q=j)\right] + \\
&\sum_{j\in s_i}\sum_{q\in s_i} v_{13jq}I(q \neq j)\left[\sum_{k\in\bar{s}_{ij}}\sum_{p\in\bar{s}_{ir}} 1\right] \\
=& \sum_{j\in s_i}(N_{ij} - n_{ij})v_{11j} + \sum_{j\in s_i}(N_{ij} - n_{ij})^2 v_{12j} + \sum_{j\in s_i}\sum_{q\in s_i}(N_{ij} - n_{ij})(N_{iq} - n_{iq})v_{13jq}I(q \neq j).
\end{aligned}
\tag{B.9}
$$

Here we calculate the expression of the second term of (B.1)

$$\begin{aligned} V_2 =& var\left(\sum_{g\in\bar{s}_i}\sum_{k=1}^{N_{ig}} y_{igk}|(y,x)\right) \\ =& cov\left(\sum_{g\in\bar{s}_i}\sum_{k=1}^{N_{ig}} y_{igk}, \sum_{r\in\bar{s}_i}\sum_{k=1}^{N_{ir}} y_{irk}|(y,x)\right) \\ =& \sum_{g\in\bar{s}_i}\sum_{k=1}^{N_{ig}}\sum_{r\in\bar{s}_i}\sum_{p=1}^{N_{ir}} cov(y_{ijk}, y_{irp}|(y,x)). \end{aligned} \tag{B.10}$$

The covarinace in (B.10) is given by

$$cov(y_{igk}, y_{irp}|(y,x)) = E[y_{igk}y_{irp}|(y,x)] - E[y_{igk}|(y,x)]E[y_{irp}|(y,x)].$$

We start by $E[y_{igk}y_{irp}|(y,x)]$:

- For $g = r$ and $k = p$:

$$\begin{aligned} E[y_{igk}y_{irp}|(y,x)] =& E[\exp\{2(x_{ig}^T\beta + v_i + u_{ig} + e_{igk})\}|(y,x)] \\ =& \exp\{2(x_{ig}^T\beta + 1^T\mu_2 + 1^T\Sigma_2 1 + \sigma_e^2)\}. \end{aligned}$$

- For $g = r$ and $k \neq p$:

$$\begin{aligned} E[y_{igk}y_{irp}|(y,x)] =& E[\exp\{2x_{ig}^T\beta + 2v_i + 2u_{ig} + e_{igk} + e_{igp}\}|(y,x)] \\ =& \exp\{2(x_{ig}^T\beta + 1^T\mu_2 + 1^T\Sigma_2 1) + \sigma_e^2)\}. \end{aligned}$$

- For $g \neq r$:

$$E[y_{igk}y_{irp}|(y,x)] = E[\exp\{x_{ig}^T\beta + x_{ir}^T\beta + 2v_i + u_{ig} + u_{ir} + e_{igk} + e_{irp}\}|(y,x)]. \tag{B.11}$$

Note that

$$\begin{bmatrix} \bar{l}_i \\ 2v_i \\ u_{ig} \\ u_{ir} \end{bmatrix} \sim N\Bigg(\begin{bmatrix} x_i\beta \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \bar{V}_i & 2\gamma & 0 & 0 \\ 2\gamma^T & 4\sigma_v^2 & 0 & 0 \\ 0^T & 0 & \sigma_u^2 & 0 \\ 0^T & 0 & 0 & \sigma_u^2 \end{bmatrix} \Bigg). \tag{B.12}$$

From (B.12), it follows that

$$E\left(\begin{bmatrix} 2v_i \\ u_{ig} \\ u_{ir} \end{bmatrix} \mid \bar{l}_i \right) = \begin{bmatrix} 2\gamma^T \\ 0^T \\ 0^T \end{bmatrix} \bar{V}_i^{-1}(\bar{l}_i - x_i\beta) \equiv \mu_3, \tag{B.13}$$

and

$$var\left(\begin{bmatrix} 2v_i \\ u_{ig} \\ u_{ir} \end{bmatrix} \mid \bar{l}_i \right) = \begin{bmatrix} 4\sigma_v^2 & 0 & 0 \\ 0 & \sigma_u^2 & 0 \\ 0 & 0 & \sigma_u^2 \end{bmatrix} - \begin{bmatrix} 2\gamma^T \\ 0^T \\ 0^T \end{bmatrix} \bar{V}_i^{-1} \begin{bmatrix} 2\gamma & 0 & 0 \end{bmatrix} \equiv \Sigma_3. \tag{B.14}$$

Using (B.13) and (B.14), the expression (B.11) becomes

$$E[y_{igk}y_{irp}|(y,x)] = \exp\{x_{ig}^T\beta + x_{ir}^T\beta + 1^T\mu_3 + \frac{1}{2}1^T\Sigma_3 1 + \sigma_e^2\}. \tag{B.15}$$

From those three cases,

$$\begin{aligned} &E[y_{igk}y_{irp}|(y,x)] = \\ &[\exp\{2(x_{ig}^T\beta + 1^T\mu_2 + 1^T\Sigma_2 1 + \sigma_e^2)\}I(k=p) + \\ &\exp\{2(x_{ig}^T\beta + 1^T\mu_2 + 1^T\Sigma_2 1) + \sigma_e^2\}I(k \neq p)]I(g=r) + \\ &[\exp\{x_{ig}^T\beta + x_{ir}^T\beta + 1^T\mu_3 + \frac{1}{2}1^T\Sigma_3 1 + \sigma_e^2\}]I(j \neq r). \end{aligned} \tag{B.16}$$

Now the next expression is $E[y_{igk}|(y,x)]E[y_{irp}|(y,x)]$

- For $g = r$:

$$\begin{aligned} E[y_{igk}|(y,x)]E[y_{irp}|(x,y)] &= (E[y_{igk}|(y,x)])^2 \\ &= \exp\{2(x_{ig}^T\beta + 1^T\mu_2) + 1^T\Sigma_2 1 + \sigma_e^2\}. \end{aligned}$$

- For $g \neq r$:

$$\begin{aligned} &E[y_{igk}|(y,x)]E[y_{irp}|(x,y)] = \\ &= E[\exp\{x_{ig}^T\beta + v_i + u_{ig} + e_{igk}\}|(y,x)]E[\exp\{x_{ir}^T\beta + v_i + u_{ir} + e_{irp}\}|(y,x)] \\ &= \exp\{x_{ig}^T\beta + 1^T\mu_2 + 1^T\Sigma_2 1 + \sigma_e^2\}\exp\{(x_{ir}^T\beta + 1^T\mu_2 + 1^T\Sigma_2 1 + \sigma_e^2\} \\ &= \exp\{x_{ig}^T\beta + x_{ir}^T\beta + 21^T\mu_2 + 1^T\Sigma_2 1 + \sigma_e^2\}. \end{aligned}$$

It follows that

$$
\begin{aligned}
&E[y_{igk}|(y,x)]E[y_{irp}|(x,y)] = \\
&[\exp\{2(x_{ig}^T\beta + 1^T\mu_2) + 1^T\Sigma_2 1 + \sigma_e^2\}]I(g=r)+ \\
&\exp\{x_{ig}^T\beta + x_{ir}^T\beta + 21^T\mu_2 + 1^T\Sigma_2 1 + \sigma_e^2\}]I(g \neq r).
\end{aligned} \tag{B.17}
$$

we have from (B.16) and (B.17) that

$$
\begin{aligned}
&cov(y_{igk}, y_{irp} \mid \bar{l}_i) = \\
&[\exp\{2(x_{ig}^T\beta + 1^T\mu_2 + 1^T\Sigma_2 1 + \sigma_e^2)\}I(k=p)+ \\
&\exp\{2(x_{ig}^T\beta + 1^T\mu_2 + 1^T\Sigma_2 1) + \sigma_e^2\}I(k \neq p)]I(g=r)+ \\
&[\exp\{x_{ig}^T\beta + x_{ir}^T\beta + 1^T\mu_3 + \frac{1}{2}1^T\Sigma_3 1 + \sigma_e^2\}]I(g \neq r)- \\
&[\exp\{2(x_{ig}^T\beta + 1^T\mu_2) + 1^T\Sigma_2 1 + \sigma_e^2\}]I(g=r)+ \\
&\exp\{x_{ig}^T\beta + x_{ir}^T\beta + 21^T\mu_2 + 1^T\Sigma_2 1 + \sigma_e^2\}]I(g \neq r) \\
=&[(\exp\{2(x_{ig}^T\beta + 1^T\mu_2 + 1^T\Sigma_2 1 + \sigma_e^2)\} - \exp\{2(x_{ig}^T\beta + 1^T\mu_2) + 1^T\Sigma_2 1 + \sigma_e^2\})I(k=p) \\
&+(\exp\{2(x_{ig}^T\beta + 1^T\mu_2 + 1^T\Sigma_2 1) + \sigma_e^2\} - \exp\{2(x_{ij}^T\beta + 1^T\mu_2) + 1^T\Sigma_2 1 + \sigma_e^2\})I(k \neq p)]I(g=r) \\
&+[\exp\{x_{ig}^T\beta + x_{ir}^T\beta + 1^T\mu_3 + \frac{1}{2}1^T\Sigma_3 1 + \sigma_e^2\} - \exp\{x_{ig}^T\beta + x_{ir}^T\beta + 21^T\mu_2 + 1^T\Sigma_2 1 + \sigma_e^2\}]I(g \neq r).
\end{aligned} \tag{B.18}
$$

Factorizing and adding and subtracting $\exp\{1^T\Sigma_2 1\}$ the expression under the $I(g = r)$ we have

$$
\begin{aligned}
&cov(y_{igk}, y_{irp} \mid \bar{l}_i) = \\
&\exp\{2(x_{ig}^T\beta + 1^T\mu_2) + 1^T\Sigma_2 1 + \sigma_e^2\}[(\exp\{1^T\Sigma_2 1 + \sigma_e^2\} - \exp\{1^T\Sigma_2 1\})I(k=p)+ \\
&(\exp\{1^T\Sigma_2 1\} - 1)]I(g=r) \\
&+[\exp\{x_{ig}^T\beta + x_{ir}^T\beta + 1^T\mu_3 + \frac{1}{2}1^T\Sigma_3 1 + \sigma_e^2\} - \exp\{x_{ig}^T\beta + x_{ir}^T\beta + 21^T\mu_2 + 1^T\Sigma_2 1 + \sigma_e^2\}]I(j \neq r).
\end{aligned}
$$

From the previous section note that

$$
\begin{aligned}
E(1^T\mu_2) =& 0 \\
var(1^T\mu_2) =& 1^T V_{\mu_2} 1,
\end{aligned}
$$

where

$$
V_{\mu_2} = var(\mu_2) = \begin{bmatrix} 0^T \\ \gamma^T \end{bmatrix} \bar{V}_i^{-1} \begin{bmatrix} 0 & \gamma \end{bmatrix}.
$$

Then we have

$$
\begin{aligned}
&E[cov(y_{igk}, y_{irp} \mid \bar{l}_i)] = \\
&\exp\{2(x_{ig}^T\beta + 1^T V_{\mu_2} 1) + 1^T\Sigma_2 1 + \sigma_e^2\}[(\exp\{1^T\Sigma_2 1 + \sigma_e^2\} - \\
&\exp\{1^T\Sigma_2 1\})I(k = p) + (\exp\{1^T\Sigma_2 1\} - 1)]I(g = r) \\
&+[\exp\{x_{ig}^T\beta + x_{ir}^T\beta + 1^T V_{\mu_3} 1 + \frac{1}{2}1^T\Sigma_3 1 + \sigma_e^2\} - \exp\{x_{ig}^T\beta + x_{ir}^T\beta + 21^T V_{\mu_2} 1 + 1^T\Sigma_2 1 + \sigma_e^2\}]I(g \neq r) \\
&= v_{21g} I(k = p)I(g = r) + v_{22g} I(g = r) + v_{23gr} I(g \neq r),
\end{aligned}
$$

where

$$
\begin{aligned}
v_{21g} &= \exp\{2(x_{ig}^T\beta + 1^T V_{\mu_2} 1) + 1^T\Sigma_2 1 + \sigma_e^2\}(\exp\{1^T\Sigma_2 1 + \sigma_e^2\} - \exp\{1^T\Sigma_2 1\}), \\
v_{22g} &= \exp\{2(x_{ig}^T\beta + 1^T V_{\mu_2} 1) + 1^T\Sigma_2 1 + \sigma_e^2\}(\exp\{1^T\Sigma_2 1\} - 1), \\
v_{23gr} &= \exp\{x_{ig}^T\beta + x_{ir}^T\beta + 1^T V_{\mu_3} 1 + \frac{1}{2}1^T\Sigma_3 1 + \sigma_e^2\} - \exp\{x_{ig}^T\beta + x_{ir}^T\beta + 21^T V_{\mu_2} 1 + 1^T\Sigma_2 1 + \sigma_e^2\}.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
E(V_2) =& \sum_{g\in\bar{s}_i}\sum_{k=1}^{N_{ig}}\sum_{r\in\bar{s}_i}\sum_{p=1}^{N_{ir}} v_{21g} I(p = k)I(r = g) + \sum_{g\in\bar{s}_i}\sum_{k=1}^{N_{ig}}\sum_{r\in\bar{s}_i}\sum_{p=1}^{N_{ir}} v_{22g} I(r = g) + \\
& \sum_{g\in\bar{s}_i}\sum_{k=1}^{N_{ig}}\sum_{r\in\bar{s}_i}\sum_{p=1}^{N_{ir}} v_{23jr} I(r \neq g) \\
=& \sum_{g\in\bar{s}_i} N_{ig} v_{21g} + \sum_{g\in\bar{s}_i} N_{ig}^2 v_{22g} + \sum_{g\in\bar{s}_i}\sum_{r\in\bar{s}_i} N_{ig} N_{ir} v_{23gr} I(r \neq g). \qquad \text{(B.19)}
\end{aligned}
$$

The expression of the third term of (B.1) is given by

$$
\begin{aligned}
C_1 =& cov\left(\sum_{j\in s_i}\sum_{k\in\bar{s}_{ij}} y_{ijk}, \sum_{r\in\bar{s}_i}\sum_{p=1}^{N_{ir}} y_{irp} |(y, x)\right) \\
=& \sum_{j\in s_i}\sum_{k\in\bar{s}_{ij}}\sum_{r\in\bar{s}_i}\sum_{p=1}^{N_{ir}} cov(y_{ijk}, y_{irp}|(y, x)). \qquad \text{(B.20)}
\end{aligned}
$$

The covariance is expressed as follows

$$
cov(y_{ijk}, y_{irp}|(y, x)) = E[y_{ijk} y_{irp}|(y, x)] - E[y_{ijk}|(y, x)]E[y_{irp}|(y, x)].
$$

Then, note that

$$\begin{bmatrix} \bar{l}_i \\ 2v_i \\ u_{ij} \\ u_{ir} \end{bmatrix} \sim N\left(\begin{bmatrix} x_i\beta \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \bar{V} & 2\gamma & \alpha^{(j)} & 0 \\ 2\gamma^T & 4\sigma_v^2 & 0 & 0 \\ \alpha^{(j)T} & 0 & \sigma_u^2 & 0 \\ 0^T & 0 & 0 & \sigma_u^2 \end{bmatrix} \right). \tag{B.21}$$

From (B.21), it follows that

$$E\left(\begin{bmatrix} 2v_i \\ u_{ij} \\ u_{ir} \end{bmatrix} \mid \bar{l}_i \right) = \begin{bmatrix} 2\gamma^T \\ \alpha^{(j)T} \\ 0^T \end{bmatrix} \bar{V}^{-1}(\bar{l}_i - x_i\beta) \equiv \mu_{4j}, \tag{B.22}$$

and

$$var\left(\begin{bmatrix} 2v_i \\ u_{ij} \\ u_{ir} \end{bmatrix} \mid \bar{l}_i \right) = \begin{bmatrix} 4\sigma_v^2 & 0 & 0 \\ 0 & \sigma_u^2 & 0 \\ 0 & 0 & \sigma_u^2 \end{bmatrix} - \begin{bmatrix} 2\gamma^T \\ \alpha^{(j)T} \\ 0^T \end{bmatrix} \bar{V}^{-1} \begin{bmatrix} 2\gamma & \alpha^{(j)} & 0 \end{bmatrix} \equiv \Sigma_{4j}. \tag{B.23}$$

Then by referring $j \in s_i$ and $r \in \bar{s}_i$ it follows

- $E[y_{ijk}y_{irp}|(y,x)]$, from (B.22) and (B.23) we have

$$\begin{aligned} &E[y_{ijk}y_{irp}|(y,x)] = \\ &E[\exp\{x_{ij}^T\beta + x_{ir}^T\beta + 2v_i + u_{ij} + u_{ir} + e_{ijk} + e_{irp}\}|(y,x)] \\ =&\exp\{x_{ij}^T\beta + x_{ir}^T\beta + 1^T\mu_{4j} + \frac{1}{2}1^T\Sigma_{4j}1 + \sigma_e^2\}. \end{aligned} \tag{B.24}$$

- $E[y_{ijk}|(y,x)]E[y_{irp}|(y,x)]$, from (A.8) and (A.10)

$$\begin{aligned} &E[y_{ijk}|(y,x)]E[y_{irp}|(y,x)] = \\ &\exp\{x_{ij}^T\beta + 1^T\mu_{1j} + \frac{1}{2}1^T\Sigma_{1j}1 + \frac{1}{2}\sigma_e^2\}\exp\{x_{ir}^T\beta + 1^T\mu_2 + \frac{1}{2}1^T\Sigma_2 1 + \frac{1}{2}\sigma_e^2\} \\ =&\exp\{x_{ij}^T\beta + x_{ir}^T\beta + 1^T(\mu_{1j} + \mu_2) + \frac{1}{2}1^T(\Sigma_{1j} + \Sigma_2)1 + \sigma_e^2\}. \end{aligned} \tag{B.25}$$

Then from (B.24) and (B.25) we have

$$\begin{aligned} cov(y_{ijk}, y_{irp}|\bar{l}_i) =& \exp\{x_{ij}^T\beta + x_{ir}^T\beta + 1^T\mu_{4j} + \frac{1}{2}1^T\Sigma_{4j}1 + \sigma_e^2\} - \\ &\exp\{x_{ij}^T\beta + x_{ir}^T\beta + 1^T(\mu_{1j} + \mu_2) + \\ &\frac{1}{2}1^T(\Sigma_{1j} + \Sigma_2)1 + \sigma_e^2\}. \end{aligned} \tag{B.26}$$

By (B.26) it follows

$$\begin{aligned}
&E[cov(y_{ijk}, y_{irp}|\bar{l}_i)] \\
=&\exp\{x_{ij}^T\beta + x_{ir}^T\beta + 1^T V_{\mu_{4j}} 1 + \frac{1}{2}1^T\Sigma_{4j}1 + \sigma_e^2\} - \\
&\exp\{x_{ij}^T\beta + x_{ir}^T\beta + 1^T(V_{\mu_{1j}} + V\mu_2)1 + \frac{1}{2}1^T(\Sigma_{1j} + \Sigma_2)1 + \sigma_e^2\} \\
=&c_{1jr} - c_{2jr},
\end{aligned} \tag{B.27}$$

where

$$\begin{aligned}
V_{\mu_{4j}} &= \begin{bmatrix} 2\gamma^T \\ \alpha^{(j)T} \\ 0^T \end{bmatrix} \bar{V}^{-1} \begin{bmatrix} 2\gamma & \alpha^{(j)} & 0 \end{bmatrix}, \\
c_{1jr} &= \exp\{x_{ij}^T\beta + x_{ir}^T\beta + 1^T V_{\mu_{4j}} 1 + \frac{1}{2}1^T\Sigma_{4j}1 + \sigma_e^2\}, \\
c_{2jr} &= \exp\{x_{ij}^T\beta + x_{ir}^T\beta + 1^T(V_{\mu_{1j}} + V\mu_2) + \frac{1}{2}1^T(\Sigma_{1j} + \Sigma_2)1 + \sigma_e^2\}.
\end{aligned}$$

From (B.27) it follows that

$$\begin{aligned}
E(C_1) &= \sum_{j\in s_i}\sum_{k\in\bar{s}_{ij}}\sum_{r\in\bar{s}_i}\sum_{p=1}^{N_{ir}}(c_{1ijr} - c_{2ijr}) \\
&= \sum_{j\in s_i}\sum_{r\in\bar{s}_i}(N_{ij} - n_{ij})N_{ir}(c_{1ijr} - c_{2ijr}).
\end{aligned} \tag{B.28}$$

Therefore, from (B.1), (B.9), (B.19) and (B.28) we have

$$\begin{aligned}
M_{1i} =& \frac{1}{N_i^2}\Bigg[\sum_{j\in s_i}(N_{ij} - n_{ij})v_{11j} + \sum_{j\in s_i}(N_{ij} - n_{ij})(N_{ir} - n_{ir})v_{12j} + \\
&\sum_{j\in s_i}\sum_{r\in s_i}(N_{ij} - n_{ij})(N_{ir} - n_{ir})v_{13jr}I(r \neq j) + \\
&\sum_{j\in\bar{s}_i}N_{ij}v_{21j} + \sum_{j\in\bar{s}_i}N_{ij}N_{ir}v_{22j} + \sum_{j\in\bar{s}_i}\sum_{r\in\bar{s}_i}N_{ij}N_{ir}v_{23jr}I(r \neq j) + \\
&2\sum_{j\in s_i}\sum_{r\in\bar{s}_i}(N_{ij} - n_{ij})N_{ir}(c_{1jr} - c_{2jr})\Bigg].
\end{aligned} \tag{B.29}$$

Appendix C

MSE of the empirical Bayes predictor

By definition,

$$M_{2i}(\theta) = E[(\bar{Y}_i^{MMSE}(\theta) - \bar{Y}_i^{MMSE}(\hat{\theta}))^2]$$

$$= \frac{1}{N_i^2} E\left[\left(\sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}} y_{ijk}^{MMSE}(\theta) + \sum_{r\in \bar{s}_i}\sum_{k=1}^{N_{ir}} y_{irk}^{MMSE}(\theta)\right) - \left(\sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}} y_{ijk}^{MMSE}(\hat{\theta}) + \sum_{r\in \bar{s}_i}\sum_{k=1}^{N_{ir}} y_{irk}^{MMSE}(\hat{\theta})\right)\right]^2$$

$$= \frac{1}{N_i^2} E\left[\sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}} (y_{ijk}^{MMSE}(\theta) - y_{ijk}^{MMSE}(\hat{\theta})) + \sum_{r\in \bar{s}_i}\sum_{k=1}^{N_{ir}} (y_{irk}^{MMSE}(\theta) - y_{irk}^{MMSE}(\hat{\theta}))\right]^2$$

$$= \frac{1}{N_i^2} E\Bigg[\sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}}\sum_{q\in s_i}\sum_{p\in \bar{s}_{iq}} (y_{ijk}^{MMSE}(\theta) - y_{ijk}^{MMSE}(\hat{\theta}))(y_{iqp}^{MMSE}(\theta) - y_{iqp}^{MMSE}(\hat{\theta})) +$$

$$2\sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}}\sum_{r\in \bar{s}_i}\sum_{p=1}^{N_{ir}} (y_{ijk}^{MMSE}(\theta) - y_{ijk}^{MMSE}(\hat{\theta}))(y_{irp}^{MMSE}(\theta) - y_{irp}^{MMSE}(\hat{\theta})) +$$

$$\sum_{g\in \bar{s}_i}\sum_{k=1}^{N_{ig}}\sum_{r\in \bar{s}_i}\sum_{p=1}^{N_{ir}} (y_{igk}^{MMSE}(\theta) - y_{igk}^{MMSE}(\hat{\theta}))(y_{irp}^{MMSE}(\theta) - y_{irp}^{MMSE}(\hat{\theta}))\Bigg]$$

$$= \frac{1}{N_i^2}\sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}}\sum_{q\in s_i}\sum_{p\in \bar{s}_{iq}} E\left[(y_{ijk}^{MMSE}(\theta) - y_{ijk}^{MMSE}(\hat{\theta}))(y_{iqp}^{MMSE}(\theta) - y_{iqp}^{MMSE}(\hat{\theta}))\right] +$$

$$2\frac{1}{N_i^2}\sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}}\sum_{r\in \bar{s}_i}\sum_{p=1}^{N_{ir}} E\left[(y_{ijk}^{MMSE}(\theta) - y_{ijk}^{MMSE}(\hat{\theta}))(y_{irp}^{MMSE}(\theta) - y_{irp}^{MMSE}(\hat{\theta}))\right] +$$

$$\frac{1}{N_i^2}\sum_{g\in \bar{s}_i}\sum_{k=1}^{N_{ig}}\sum_{r\in \bar{s}_i}\sum_{p=1}^{N_{ir}} E\left[(y_{igk}^{MMSE}(\theta) - y_{igk}^{MMSE}(\hat{\theta}))(y_{irp}^{MMSE}(\theta) - y_{irp}^{MMSE}(\hat{\theta}))\right]$$

$$= \frac{1}{N_i^2}\left[H_1 + 2H_2 + H_3\right]. \tag{C.1}$$

The corresponding approximations of those terms in right-hand side of (C.1) are derived below.
Note that,

$$(y_{ijk}^{MMSE}(\theta) - y_{ijk}^{MMSE}(\hat{\theta}))(y_{irp}^{MMSE}(\theta) - y_{irp}^{MMSE}(\hat{\theta})) =$$
$$y_{ijk}^{MMSE}(\theta)y_{irp}^{MMSE}(\theta) - y_{ijk}^{MMSE}(\theta)y_{irp}^{MMSE}(\hat{\theta}) - y_{ijk}^{MMSE}(\hat{\theta})y_{irp}^{MMSE}(\theta) + y_{ijk}^{MMSE}(\hat{\theta})y_{irp}^{MMSE}(\hat{\theta}). \tag{C.2}$$

Now we need to find the approximation of $y_{ijk}^{MMSE}(\hat{\theta})$. By (A.8) we have

$$\begin{aligned} \tilde{y}_{ijk}^{*}(\hat{\theta}) =& \exp\{x_{ij}^{T}\hat{\beta} + 1^{T}\mu_{1j}(\hat{\theta}) + \frac{1}{2}1^{T}\Sigma_{1j}(\hat{\theta})1 + \frac{1}{2}\hat{\sigma_e^2}\} \\ =& \exp\{\Delta_1(\hat{\theta}) + \Omega_1(\hat{\theta})\}, \end{aligned} \tag{C.3}$$

where

$$\begin{aligned} \Delta_1(\hat{\theta}) =& 1^{T}\mu_{1j}(\hat{\theta}) \\ \Omega_1(\hat{\theta}) =& x_{ij}^{T}\hat{\beta} + \frac{1}{2}(1^{T}\Sigma_{1j}(\hat{\theta})1 + \hat{\sigma_e^2}). \end{aligned}$$

As Δ_1 and Ω_1 are functions of $\theta = (\beta, \sigma_v^2, \sigma_u^2, \sigma_e^2)^T$. When its estimate $\hat{\theta}$ is used, $\Delta_1(\hat{\theta})$ and $\Omega_1(\hat{\theta})$ can be expanded respectively around $\Delta_1(\theta)$ and $\Omega_1(\theta)$, by a Taylor series as

$$\begin{aligned} \Delta_1(\hat{\theta}) \approx& \Delta_1(\theta) + (\hat{\theta} - \theta)^{T}\frac{\partial\Delta_1(\hat{\theta})}{\partial\hat{\theta}}\Big|_{\hat{\theta}=\theta} = \Delta_1(\theta) + (\hat{\theta} - \theta)^{T}\Delta_1^{*}(\theta) \\ \Omega_1(\hat{\theta}) \approx& \Omega_1(\theta) + (\hat{\theta} - \theta)^{T}\frac{\partial\Omega_1(\hat{\theta})}{\partial\hat{\theta}}\Big|_{\hat{\theta}=\theta} = \Omega_1(\theta) + (\hat{\theta} - \theta)^{T}\Omega_1^{*}(\theta), \end{aligned} \tag{C.4}$$

where the expressions in $\frac{\partial\Delta_1(\hat{\theta})}{\partial\hat{\theta}}\Big|_{\hat{\theta}=\theta}$ are calculated as follows

$$\Delta_1^{*}(\theta) = \frac{\partial\Delta_1(\hat{\theta})}{\partial\hat{\theta}}\Big|_{\hat{\theta}=\theta} = \begin{bmatrix} \frac{\partial\Delta_1(\hat{\theta})}{\partial\hat{\beta}}(\theta) \\ \frac{\partial\Delta_1(\hat{\theta})}{\partial\hat{\sigma_v^2}}(\theta) \\ \frac{\partial\Delta_1(\hat{\theta})}{\partial\hat{\sigma_u^2}}(\theta) \\ \frac{\partial\Delta_1(\hat{\theta})}{\partial\hat{\sigma_e^2}}(\theta) \end{bmatrix} = 1^{T}\begin{bmatrix} \frac{\partial\mu_{1j}(\hat{\theta})}{\partial\hat{\beta}}(\theta) \\ \frac{\partial\mu_{1j}(\hat{\theta})}{\partial\hat{\sigma_v^2}}(\theta) \\ \frac{\partial\mu_{1j}(\hat{\theta})}{\partial\hat{\sigma_u^2}}(\theta) \\ \frac{\partial\mu_{1j}(\hat{\theta})}{\partial\hat{\sigma_e^2}}(\theta) \end{bmatrix}$$

with

$$
\begin{aligned}
\frac{\partial \mu_{1j}(\hat{\theta})}{\partial \hat{\beta}}(\theta) &= -\begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} x_i, \\
\frac{\partial \mu_{1j}(\hat{\theta})}{\partial \hat{\sigma_v^2}}(\theta) &= \begin{bmatrix} 0^T \\ 1^T \end{bmatrix} \bar{V}^{-1}(\bar{l}_i - x_i\beta) - \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} J \bar{V}^{-1}(\bar{l}_i - x_i\beta) \\
&= \left(\begin{bmatrix} 0^T \\ 1^T \end{bmatrix} - \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} J \right) \bar{V}^{-1}(\bar{l}_i - x_i\beta), \\
\frac{\partial \mu_{1j}(\hat{\theta})}{\partial \hat{\sigma_u^2}}(\theta) &= \begin{bmatrix} 1_j^T \\ \\ 0^T \end{bmatrix} \bar{V}^{-1}(\bar{l}_i - x_i\beta) - \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} I \bar{V}^{-1}(\bar{l}_i - x_i\beta) \\
&= \left(\begin{bmatrix} 1_j^T \\ \\ 0^T \end{bmatrix} - \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} \right) \bar{V}^{-1}(\bar{l}_i - x_i\beta), \\
\frac{\partial \mu_{1j}(\hat{\theta})}{\partial \hat{\sigma_e^2}}(\theta) &= -\begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} D \bar{V}^{-1}(\bar{l}_i - x_i\beta),
\end{aligned}
$$

where

$$
1_j = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{(m_i \times 1)} \quad ; \quad J = \begin{bmatrix} 1 & \dots & 1 \\ \vdots & & \\ 1 & \dots & 1 \end{bmatrix}_{(m_i \times m_i)} \quad ; \quad D = \begin{bmatrix} \frac{1}{n_{i1}} & & & \\ & \frac{1}{n_{i2}} & & \\ & & \ddots & \\ & & & \frac{1}{n_{im_i}} \end{bmatrix}_{(m_i \times m_i)},
$$

and the expressions in $\frac{\partial \Omega_1(\hat{\theta})}{\partial \hat{\theta}}\Big|_{\hat{\theta}=\theta}$ are obtained as follows

$$
\Omega_1^*(\theta) = \frac{\partial \Omega_1(\hat{\theta})}{\partial \hat{\theta}}\Big|_{\hat{\theta}=\theta} = \begin{bmatrix} \frac{\partial \Omega_1(\hat{\theta})}{\partial \hat{\beta}}(\theta) \\ \frac{\partial \Omega_1(\hat{\theta})}{\partial \hat{\sigma_v^2}}(\theta) \\ \frac{\partial \Omega_1(\hat{\theta})}{\partial \hat{\sigma_u^2}}(\theta) \\ \frac{\partial \Omega_1(\hat{\theta})}{\partial \hat{\sigma_e^2}}(\theta) \end{bmatrix},
$$

with

$$\begin{aligned}
\frac{\partial \Omega_1(\hat{\theta})}{\partial \hat{\beta}}(\theta) =& x_{ij}^T, \\
\frac{\partial \Omega_1(\hat{\theta})}{\partial \hat{\sigma_v^2}}(\theta) =& \frac{1}{2} 1^T \frac{\partial \Sigma_{1j}(\hat{\theta})}{\partial \hat{\sigma_v^2}}(\theta) 1 = \frac{1}{2} 1^T \Sigma_{1j}^{(v)} 1, \\
\frac{\partial \Omega_1(\hat{\theta})}{\partial \hat{\sigma_u^2}}(\theta) =& \frac{1}{2} 1^T \frac{\partial \Sigma_{1j}(\hat{\theta})}{\partial \hat{\sigma_u^2}}(\theta) 1 = \frac{1}{2} 1^T \Sigma_{1j}^{(u)} 1, \\
\frac{\partial \Omega_1(\hat{\theta})}{\partial \hat{\sigma_e^2}}(\theta) =& \frac{1}{2} 1^T \frac{\partial \Sigma_{1j}(\hat{\theta})}{\partial \hat{\sigma_e^2}}(\theta) 1 + \frac{1}{2} = \frac{1}{2} 1^T \Sigma_{1j}^{(e)} 1 + \frac{1}{2},
\end{aligned}$$

where

$$\begin{aligned}
\Sigma_{1j}^{(v)} &= \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0^T \\ 1^T \end{bmatrix} \bar{V}^{-1} \begin{bmatrix} \alpha^{(j)} & \gamma \end{bmatrix} + \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} J \bar{V}^{-1} \begin{bmatrix} \alpha^{(j)} & \gamma \end{bmatrix} - \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} \begin{bmatrix} 0 & 1 \end{bmatrix}, \\
\Sigma_{1j}^{(u)} &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} 1_j^T \\ \\ 0^T \end{bmatrix} \bar{V}^{-1} \begin{bmatrix} \alpha^{(j)} & \gamma \end{bmatrix} + \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} I \bar{V}^{-1} \begin{bmatrix} \alpha^{(j)} & \gamma \end{bmatrix} - \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} \begin{bmatrix} 1_j & 0 \end{bmatrix}, \\
\Sigma_{1j}^{(e)} &= \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} D \bar{V}^{-1} \begin{bmatrix} \alpha^{(j)} & \gamma \end{bmatrix}.
\end{aligned}$$

From (C.3) and (C.4) we have

$$\begin{aligned}
\tilde{y}_{ijk}^*(\hat{\theta}) =& \exp\{x_{ij}^T \hat{\beta} + 1^T \mu_{1j}(\hat{\theta}) + \frac{1}{2} 1^T \Sigma_{1j}(\hat{\theta}) 1 + \frac{1}{2} \hat{\sigma_e^2}\} \\
\approx& \exp\{\Delta_1(\theta) + (\hat{\theta} - \theta)^T \Delta_1^*(\theta) + \Omega_1(\theta) + (\hat{\theta} - \theta)^T \Omega_1^*(\theta)\}. \qquad \text{(C.5)}
\end{aligned}$$

Then by (A.10)

$$\begin{aligned}
\tilde{y}_{ijk}^{**} =& \exp\{x_{ij}^T \hat{\beta} + 1^T \mu_2(\hat{\theta}) + \frac{1}{2} 1^T \Sigma_2(\hat{\theta}) 1 + \frac{1}{2} \hat{\sigma_e^2}\} \\
=& \exp\{\Delta_2(\hat{\theta}) + \Omega_2(\hat{\theta})\}, \qquad \text{(C.6)}
\end{aligned}$$

where

$$\begin{aligned}
\Delta_2(\hat{\theta}) =& 1^T \mu_2(\hat{\theta}), \\
\Omega_2(\hat{\theta}) =& x_{ij}^T \hat{\beta} + \frac{1}{2}(1^T \Sigma_2(\hat{\theta}) 1 + \hat{\sigma_e^2}).
\end{aligned}$$

Taking into account that Δ_2 and Ω_2 are functions of $\theta = (\beta, \sigma_v^2, \sigma_u^2, \sigma_e^2)^T$. When its estimate $\hat{\theta}$ is used, $\Delta_2(\hat{\theta})$ and $\Omega_2(\hat{\theta})$ can be expanded respectively around $\Delta_2(\theta)$ and $\Omega_2(\theta)$, by a Taylor series as

$$\begin{aligned}
\Delta_2(\hat{\theta}) \approx& \Delta_2(\theta) + (\hat{\theta}-\theta)^T \frac{\partial \Delta_2(\hat{\theta})}{\partial \hat{\theta}}\Big|_{\hat{\theta}=\theta} = \Delta_2(\theta) + (\hat{\theta}-\theta)^T \Delta_2^*(\theta), \\
\Omega_2(\hat{\theta}) \approx& \Omega_2(\theta) + (\hat{\theta}-\theta)^T \frac{\partial \Omega_2(\hat{\theta})}{\partial \hat{\theta}}\Big|_{\hat{\theta}=\theta} = \Omega_2(\theta) + (\hat{\theta}-\theta)^T \Omega_2^*(\theta),
\end{aligned} \tag{C.7}$$

where the expressions in $\frac{\partial \Delta_2(\hat{\theta})}{\partial \hat{\theta}}\Big|_{\hat{\theta}=\theta}$ are calculated as follows

$$\Delta_2^*(\theta) = \frac{\partial \Delta_2(\hat{\theta})}{\partial \hat{\theta}}\Big|_{\hat{\theta}=\theta} = \begin{bmatrix} \frac{\partial \Delta_2(\hat{\theta})}{\partial \hat{\beta}}(\theta) \\ \frac{\partial \Delta_2(\hat{\theta})}{\partial \hat{\sigma_v^2}}(\theta) \\ \frac{\partial \Delta_2(\hat{\theta})}{\partial \hat{\sigma_u^2}}(\theta) \\ \frac{\partial \Delta_2(\hat{\theta})}{\partial \hat{\sigma_e^2}}(\theta) \end{bmatrix} = 1^T \begin{bmatrix} \frac{\partial \mu_2(\hat{\theta})}{\partial \hat{\beta}}(\theta) \\ \frac{\partial \mu_2(\hat{\theta})}{\partial \hat{\sigma_v^2}}(\theta) \\ \frac{\partial \mu_2(\hat{\theta})}{\partial \hat{\sigma_u^2}}(\theta) \\ \frac{\partial \mu_2(\hat{\theta})}{\partial \hat{\sigma_e^2}}(\theta) \end{bmatrix},$$

with

$$\begin{aligned}
\frac{\partial \mu_2(\hat{\theta})}{\partial \hat{\beta}}(\theta) =& -\begin{bmatrix} 0^T \\ 1^T \end{bmatrix} \bar{V}^{-1} x_i, \\
\frac{\partial \mu_2(\hat{\theta})}{\partial \hat{\sigma_v^2}}(\theta) =& \begin{bmatrix} 0^T \\ 1^T \end{bmatrix} \bar{V}^{-1}(\bar{l}_i - x_i\beta) - \begin{bmatrix} 0^T \\ \gamma^T \end{bmatrix} \bar{V}^{-1} J \bar{V}^{-1}(\bar{l}_i - x_i\beta) \\
=& \Big(\begin{bmatrix} 0^T \\ 1^T \end{bmatrix} - \begin{bmatrix} 0^T \\ \gamma^T \end{bmatrix} \bar{V}^{-1} J \Big) \bar{V}^{-1}(\bar{l}_i - x_i\beta), \\
\frac{\partial \mu_2(\hat{\theta})}{\partial \hat{\sigma_u^2}}(\theta) =& -\begin{bmatrix} 0^T \\ \gamma^T \end{bmatrix} \bar{V}^{-1} I \bar{V}^{-1}(\bar{l}_i - x_i\beta), \\
\frac{\partial \mu_2(\hat{\theta})}{\partial \hat{\sigma_e^2}}(\theta) =& -\begin{bmatrix} 0^T \\ \gamma^T \end{bmatrix} \bar{V}^{-1} D \bar{V}^{-1}(\bar{l}_i - x_i\beta),
\end{aligned}$$

and the expressions in $\frac{\partial \Omega_2(\hat{\theta})}{\partial \hat{\theta}}\Big|_{\hat{\theta}=\theta}$ are obtained as follows

$$\Omega_2^*(\theta) = \frac{\partial \Omega_1(\hat{\theta})}{\partial \hat{\theta}}\Big|_{\hat{\theta}=\theta} = \begin{bmatrix} \frac{\partial \Omega_2(\hat{\theta})}{\partial \hat{\beta}}(\theta) \\ \frac{\partial \Omega_2(\hat{\theta})}{\partial \hat{\sigma_v^2}}(\theta) \\ \frac{\partial \Omega_2(\hat{\theta})}{\partial \hat{\sigma_u^2}}(\theta) \\ \frac{\partial \Omega_2(\hat{\theta})}{\partial \hat{\sigma_e^2}}(\theta) \end{bmatrix},$$

with

$$\begin{aligned}
\frac{\partial\Omega_2(\hat{\theta})}{\partial\hat{\beta}}(\theta) =& x_{ij}^T \\
\frac{\partial\Omega_2(\hat{\theta})}{\partial\hat{\sigma_v^2}}(\theta) =& \frac{1}{2}1^T\frac{\partial\Sigma_2(\hat{\theta})}{\partial\hat{\sigma_v^2}}(\theta)1 = \frac{1}{2}1^T\Sigma_2^{(v)}1, \\
\frac{\partial\Omega_2(\hat{\theta})}{\partial\hat{\sigma_u^2}}(\theta) =& \frac{1}{2}1^T\frac{\partial\Sigma_2(\hat{\theta})}{\partial\hat{\sigma_u^2}}(\theta)1 = \frac{1}{2}1^T\Sigma_2^{(u)}1, \\
\frac{\partial\Omega_2(\hat{\theta})}{\partial\hat{\sigma_e^2}}(\theta) =& \frac{1}{2}1^T\frac{\partial\Sigma_2(\hat{\theta})}{\partial\hat{\sigma_e^2}}(\theta)1 + \frac{1}{2} = \frac{1}{2}1^T\Sigma_2^{(e)}1 + \frac{1}{2},
\end{aligned}$$

where

$$\begin{aligned}
\Sigma_2^{(v)} &= \begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix} - \begin{bmatrix}0^T\\ 1^T\end{bmatrix}\bar{V}^{-1}\begin{bmatrix}0 & \gamma\end{bmatrix} + \begin{bmatrix}0^T\\ \gamma^T\end{bmatrix}\bar{V}^{-1}J\bar{V}^{-1}\begin{bmatrix}0 & \gamma\end{bmatrix} - \begin{bmatrix}0^T\\ \gamma^T\end{bmatrix}\bar{V}^{-1}\begin{bmatrix}0 & 1\end{bmatrix}, \\
\Sigma_2^{(u)} &= \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} + \begin{bmatrix}0^T\\ \gamma^T\end{bmatrix}\bar{V}^{-1}I\bar{V}^{-1}\begin{bmatrix}0 & \gamma\end{bmatrix}, \\
\Sigma_2^{(e)} &= \begin{bmatrix}0^T\\ \gamma^T\end{bmatrix}\bar{V}^{-1}D\bar{V}^{-1}\begin{bmatrix}0 & \gamma\end{bmatrix}.
\end{aligned}$$

From (C.6) and (C.7) we have

$$\begin{aligned}
\tilde{y}_{ijk}^{**}(\hat{\theta}) =& \exp\{x_{ij}^T\hat{\beta} + 1^T\mu_2(\hat{\theta}) + \frac{1}{2}1^T\Sigma_2(\hat{\theta})1 + \frac{1}{2}\hat{\sigma_e^2}\} \\
\approx& \exp\{\Delta_2(\theta) + (\hat{\theta}-\theta)^T\Delta_2^*(\theta) + \Omega_2(\theta) + (\hat{\theta}-\theta)^T\Omega_2^*(\theta)\}. \qquad \text{(C.8)}
\end{aligned}$$

In the both above cases, the approximation can be represented as follows

$$\begin{aligned}
\Delta(\hat{\theta}) + \Omega(\hat{\theta}) \approx& \Delta(\theta) + \Omega(\theta) + \frac{\partial\Delta^T}{\partial\hat{\theta}}(\theta)(\hat{\theta}-\theta) + \frac{\partial\omega^T}{\partial\hat{\theta}}(\theta)(\hat{\theta}-\theta) \\
=& \Delta(\theta) + \Omega(\theta) + \Big(\frac{\partial\Delta^T}{\partial\hat{\beta}}(\beta) + \frac{\partial\Omega^T}{\partial\hat{\beta}}(\beta)\Big)(\hat{\beta}-\beta) + \Big(\frac{\partial\Delta^T}{\partial\hat{\sigma}}(\sigma) + \frac{\partial\Omega^T}{\partial\hat{\sigma}}(\sigma)\Big)(\hat{\sigma}-\sigma), \\
& \qquad \text{(C.9)}
\end{aligned}$$

where $\sigma = (\sigma_v^2, \sigma_u^2, \sigma_e^2)^T$.
Considering the expressions from (C.2) and (C.9), we need to calculate the following expressions that correspond to the terms defined in (C.1):

- For H_1:

$$\begin{aligned}\tilde{y}^*_{ijk}(\theta)\tilde{y}^*_{iqp}(\hat{\theta}) &= \exp\{\Delta_1^{(j)}(\theta)+\Omega_1^{(j)}(\theta)+\Delta_1^{(q)}(\hat{\theta})+\Omega_1^{(q)}(\hat{\theta})\}\\ \tilde{y}^*_{ijk}(\hat{\theta})\tilde{y}^*_{iqp}(\hat{\theta}) &= \exp\{\Delta_1^{(j)}(\hat{\theta})+\Omega_1^{(j)}(\hat{\theta})+\Delta_1^{(q)}(\hat{\theta})+\Omega_1^{(q)}(\hat{\theta})\}.\end{aligned}$$

- For H_2:

$$\begin{aligned}\tilde{y}^*_{ijk}(\theta)\tilde{y}^{**}_{irp}(\hat{\theta}) &= \exp\{\Delta_1(\theta)+\Omega_1(\theta)+\Delta_2(\hat{\theta})+\Omega_2(\hat{\theta})\}\\ \tilde{y}^*_{ijk}(\hat{\theta})\tilde{y}^{**}_{irp}(\hat{\theta}) &= \exp\{\Delta_1(\hat{\theta})+\Omega_1(\hat{\theta})+\Delta_2(\hat{\theta})+\Omega_2(\hat{\theta})\}.\end{aligned}$$

- For H_3:

$$\begin{aligned}\tilde{y}^{**}_{igk}(\theta)\tilde{y}^{**}_{irp}(\hat{\theta}) &= \exp\{\Delta_2^{(g)}(\theta)+\Omega_2^{(g)}(\theta)+\Delta_2^{(r)}(\hat{\theta})+\Omega_2^{(r)}(\hat{\theta})\}\\ \tilde{y}^{**}_{igk}(\hat{\theta})\tilde{y}^{**}_{irp}(\hat{\theta}) &= \exp\{\Delta_2^{(g)}(\hat{\theta})+\Omega_2^{(g)}(\hat{\theta})+\Delta_2^{(r)}(\hat{\theta})+\Omega_2^{(r)}(\hat{\theta})\}.\end{aligned}$$

In continuation we discuss different scenarios for each case

- Case 1:

  1 Expressions in H_1, i.e $j, q \in s_i$ and $j = q$

$$\begin{aligned}\tilde{y}^*_{ijk}(\theta)\tilde{y}^*_{iqp}(\theta) &= \exp\{2(x_{ij}^T\beta + 1^T\mu_{1j} + \frac{1}{2}1^T\Sigma_{1j}1 + \frac{1}{2}\sigma_e^2)\}\\ &= \exp\{2(\Delta_1(\theta)+\Omega_1(\theta))\},\\ \tilde{y}^*_{ijk}(\theta)\tilde{y}^*_{iqp}(\hat{\theta}) &= \exp\{\Delta_1^{(j)}(\theta)+\Omega_1^{(j)}(\theta)+\Delta_1^{(q)}(\hat{\theta})+\Omega_1^{(q)}(\hat{\theta})\}\\ &\approx \exp\{2(\Delta_1(\theta)+\Omega_1(\theta)) + (\hat{\theta}-\theta)^T\Delta_1^*(\theta) + (\hat{\theta}-\theta)^T\Omega_1^*(\theta)\}\\ &= \exp\{2(\Delta_1(\theta)+\Omega_1(\theta)) + (\hat{\beta}-\beta)^T(\Delta_1^*(\beta)+\Omega_1^*(\beta)) +\\ &\quad (\hat{\sigma}-\sigma)^T(\Delta_1^*(\sigma)+\Omega_1^*(\sigma))\},\\ \tilde{y}^*_{ijk}(\hat{\theta})\tilde{y}^*_{iqp}(\hat{\theta}) &= \exp\{\Delta_1(\hat{\theta})+\Omega_1(\hat{\theta})+\Delta_1(\hat{\theta})+\Omega_1(\hat{\theta})\}\\ &\approx \exp\{2(\Delta_1(\theta)+\Omega_1(\theta) + (\hat{\theta}-\theta)^T\Delta_1^*(\theta) + (\hat{\theta}-\theta)^T\Omega_1^*(\theta))\}\\ &= \exp\{2(\Delta_1(\theta)+\Omega_1(\theta) + (\hat{\beta}-\beta)^T(\Delta_1^*(\beta)+\Omega_1^*(\beta)) +\\ &\quad (\hat{\sigma}-\sigma)^T(\Delta_1^*(\sigma)+\Omega_1^*(\sigma)))\}. \qquad \text{(C.10)}\end{aligned}$$

2 Expressions in H_1, i.e $j, q \in s_i$ and $j \neq q$

$$\begin{aligned}
\tilde{y}^*_{ijk}(\theta)\tilde{y}^*_{iqp}(\theta) &= \exp\{x_{ij}^T\beta + 1^T\mu_{1j} + \frac{1}{2}1^T\Sigma_{1j}1 + \frac{1}{2}\sigma_e^2\}\exp\{x_{iq}^T\beta + 1^T\mu_{1q} + \frac{1}{2}1^T\Sigma_{1q}1 + \frac{1}{2}\sigma_e^2\} \\
&= \exp\{\Delta_1^{(j)}(\theta) + \Omega_1^{(j)}(\theta) + \Delta_1^{(q)}(\theta) + \Omega_1^{(q)}(\theta)\}, \\
\tilde{y}^*_{ijk}(\theta)\tilde{y}^*_{iqp}(\hat{\theta}) &= \exp\{\Delta_1^{(j)}(\theta) + \Omega_1^{(j)}(\theta) + \Delta_1^{(q)}(\hat{\theta}) + \Omega_1^{(q)}(\hat{\theta})\} \\
&\approx \exp\{\Delta_1^{(j)}(\theta) + \Omega_1^{(j)}(\theta) + \Delta_1^{(q)}(\theta) + \Omega_1^{(q)}(\theta) + (\hat{\theta}-\theta)^T\Delta_1^{(q)*}(\theta) + \\
&\quad (\hat{\theta}-\theta)^T\Omega_1^{(q)*}(\theta)\} \\
&= \exp\{\Delta_1^{(j)}(\theta) + \Omega_1^{(j)}(\theta) + \Delta_1^{(q)}(\theta) + \Omega_1^{(q)}(\theta) + \\
&\quad (\hat{\beta}-\beta)^T(\Delta_1^{(q)*}(\beta) + \Omega_1^{(q)*}(\beta)) + \\
&\quad (\hat{\sigma}-\sigma)^T(\Delta_1^{(q)*}(\sigma) + \Omega_1^{(q)*}(\sigma))\}, \\
\tilde{y}^*_{ijk}(\hat{\theta})\tilde{y}^*_{iqp}(\hat{\theta}) &= \exp\{\Delta_1^{(j)}(\hat{\theta}) + \Omega_1^{(j)}(\hat{\theta}) + \Delta_1^{(q)}(\hat{\theta}) + \Omega_1^{(q)}(\hat{\theta})\} \\
&\approx \exp\{\Delta_1^{(j)}(\theta) + \Omega_1^{(j)}(\theta) + (\hat{\theta}-\theta)^T\Delta_1^{(j)*}(\theta) + (\hat{\theta}-\theta)^T \\
&\quad \Omega_1^{(j)*}(\theta) + \Delta_1^{(q)}(\theta) + \Omega_1^{(q)}(\theta) + (\hat{\theta}-\theta)^T\Delta_1^{(q)*}(\theta) + (\hat{\theta}-\theta)^T\Omega_1^{(q)*}(\theta)\} \\
&= \exp\{\Delta_1^{(j)}(\theta) + \Omega_1^{(j)}(\theta) + (\hat{\beta}-\beta)^T(\Delta_1^{(j)*}(\beta) + \Omega_1^{(j)*}(\beta)) + \\
&\quad (\hat{\sigma}-\sigma)^T(\Delta_1^{(j)*}(\sigma)) + \\
&\quad \Omega_1^{(j)*}(\sigma)) + \Delta_1^{(q)}(\theta) + \Omega_1^{(q)}(\theta) + (\hat{\beta}-\beta)^T(\Delta_1^{(r)*}(\beta) + \Omega_1^{(r)*}(\beta)) + \\
&\quad (\hat{\sigma}-\sigma)^T(\Delta_1^{(q)*}(\sigma) + \Omega_1^{(q)*}(\sigma))\}. \qquad \text{(C.11)}
\end{aligned}$$

- Case 2:

1 Expressions in H_2, i.e $j \in s_i$ and $r \in \bar{s}_i$

$$\begin{aligned}
\tilde{y}^*_{ijk}(\theta)\tilde{y}^{**}_{irp}(\theta) &= \exp\{x_{ij}^T\beta + 1^T\mu_{1j}\frac{1}{2}1^T\Sigma_{1j}1 + \frac{1}{2}\sigma_e^2\}\exp\{x_{ir}^T\beta + 1^T\mu_2\frac{1}{2}1^T\Sigma_2 1 + \frac{1}{2}\sigma_e^2\} \\
&= \exp\{\Delta_1(\theta) + \Omega_1(\theta) + \Delta_2(\theta) + \Omega_2(\theta)\}, \\
\tilde{y}^*_{ijk}(\theta)\tilde{y}^{**}_{irp}(\hat{\theta}) &= \exp\{\Delta_1(\theta) + \Omega_1(\theta) + \Delta_2(\hat{\theta}) + \Omega_2(\hat{\theta})\} \\
&\approx \exp\{\Delta_1(\theta) + \Omega_1(\theta) + \Delta_2(\theta) + \Omega_2(\theta) + (\hat{\theta}-\theta)^T\Delta_2^*(\theta) + (\hat{\theta}-\theta)^T\Omega_2^*(\theta)\} \\
&= \exp\{\Delta_1(\theta) + \Omega_1(\theta) + \Delta_2(\theta) + \Omega_2(\theta) + (\hat{\beta}-\beta)^T(\Delta_2^*(\beta) + \Omega_2^*(\beta)) + \\
&\quad (\hat{\sigma}-\sigma)^T(\Delta_2^*(\sigma) + \Omega_2^*(\sigma))\},
\end{aligned}$$

$$
\begin{aligned}
\tilde{y}^{*}_{ijk}(\hat{\theta})\tilde{y}^{**}_{irp}(\hat{\theta}) &= \exp\{\Delta_1(\hat{\theta}) + \Omega_1(\hat{\theta}) + \Delta_2(\hat{\theta}) + \Omega_2(\hat{\theta})\} \\
&\approx \exp\{\Delta_1(\theta) + \Omega_1(\theta) + (\hat{\theta}-\theta)^T\Delta_1^*(\theta) + (\hat{\theta}-\theta)^T\Omega_1^*(\theta) + \Delta_2(\theta) + \Omega_2(\theta) + \\
&\quad (\hat{\theta}-\theta)^T\Delta_2^*(\theta) + (\hat{\theta}-\theta)^T\Omega_2^*(\theta)\} \\
&= \exp\{\Delta_1(\theta) + \Omega_1(\theta) + (\hat{\beta}-\beta)^T(\Delta_1^*(\beta) + \Omega_1^*(\beta)) + \\
&\quad (\hat{\sigma}-\sigma)^T(\Delta_1^*(\sigma) + \Omega_1^*(\sigma)) + \\
&\quad \Delta_2(\theta) + \Omega_2(\theta) + (\hat{\beta}-\beta)^T(\Delta_2^*(\beta) + \Omega_2^*(\beta)) + (\hat{\sigma}-\sigma)^T(\Delta_2^*(\sigma) + \Omega_2^*(\sigma))\}.
\end{aligned} \tag{C.12}
$$

- Case 3:

1 Expressions in H_3, i.e $g, r \in \bar{s}_i$ and $g = r$

$$
\begin{aligned}
\tilde{y}^{**}_{igk}(\theta)\tilde{y}^{**}_{irp}(\theta) &= \exp\{2(x_{ig}^T\beta + 1^T\mu_2\frac{1}{2}1^T\Sigma_2 1 + \frac{1}{2}\sigma_e^2)\} \\
&= \exp\{2(\Delta_2(\theta) + \Omega_2(\theta))\}, \\
\tilde{y}^{**}_{igk}(\theta)\tilde{y}^{**}_{irp}(\hat{\theta}) &= \exp\{\Delta_2^{(g)}(\theta) + \Omega_2^{(g)}(\theta) + \Delta_2^{(r)}(\hat{\theta}) + \Omega_2^{(r)}(\hat{\theta})\} \\
&\approx \exp\{2(\Delta_2(\theta) + \Omega_2(\theta)) + (\hat{\theta}-\theta)^T\Delta_2^*(\theta) + (\hat{\theta}-\theta)^T\Omega_2^*(\theta)\} \\
&= \exp\{2(\Delta_2(\theta) + \Omega_2(\theta)) + (\hat{\beta}-\beta)^T(\Delta_2^*(\beta) + \Omega_2^*(\beta)) + \\
&\quad (\hat{\sigma}-\sigma)^T(\Delta_2^*(\sigma) + \Omega_2^*(\sigma))\}, \\
\tilde{y}^{**}_{igk}(\hat{\theta})\tilde{y}^{**}_{irp}(\hat{\theta}) &= \exp\{\Delta_2(\hat{\theta}) + \Omega_2(\hat{\theta}) + \Delta_2(\hat{\theta}) + \Omega_2(\hat{\theta})\} \\
&\approx \exp\{2(\Delta_2(\theta) + \Omega_2(\theta) + (\hat{\theta}-\theta)^T\Delta_2^*(\theta) + (\hat{\theta}-\theta)^T\Omega_2^*(\theta))\} \\
&= \exp\{2(\Delta_2(\theta) + \Omega_2(\theta) + (\hat{\beta}-\beta)^T(\Delta_2^*(\beta) + \Omega_2^*(\beta)) + \\
&\quad (\hat{\sigma}-\sigma)^T(\Delta_2^*(\sigma) + \Omega_2^*(\sigma)))\}.
\end{aligned} \tag{C.13}
$$

2 Expressions in H_3, i.e $g, r \in \bar{s}_i$ and $g \neq r$

$$
\begin{aligned}
\tilde{y}^{**}_{igk}(\theta)\tilde{y}^{**}_{irp}(\theta) &= \exp\{x_{ig}^T\beta + 1^T\mu_2\frac{1}{2}1^T\Sigma_2 1 + \frac{1}{2}\sigma_e^2\}\exp\{x_{ir}^T\beta + 1^T\mu_2\frac{1}{2}1^T\Sigma_2 1 + \frac{1}{2}\sigma_e^2\} \\
&= \exp\{\Delta_2^{(g)}(\theta) + \Omega_2^{(g)}(\theta) + \Delta_2^{(r)}(\theta) + \Omega_2^{(r)}(\theta)\},
\end{aligned}
$$

$$\begin{aligned}
\tilde{y}^{**}_{igk}(\theta)\tilde{y}^{**}_{irp}(\hat{\theta}) =& \exp\{\Delta_2^{(g)}(\theta) + \Omega_2^{(g)}(\theta) + \Delta_2^{(r)}(\hat{\theta}) + \Omega_2^{(r)}(\hat{\theta})\} \\
\approx& \exp\{\Delta_2^{(g)}(\theta) + \Omega_2^{(g)}(\theta) + \Delta_2^{(r)}(\theta) + \Omega_2^{(r)}(\theta) + (\hat{\theta}-\theta)^T\Delta_2^{(r)*}(\theta) + \\
& (\hat{\theta}-\theta)^T\Omega_2^{(r)*}(\theta)\} \\
=& \exp\{\Delta_2^{(g)}(\theta) + \Omega_2^{(g)}(\theta) + \Delta_2^{(r)}(\theta) + \Omega_2^{(r)}(\theta) + (\hat{\beta}-\beta)^T(\Delta_2^{(r)*}(\beta) + \Omega_2^{(r)*}(\beta)) \\
& + (\hat{\sigma}-\sigma)^T(\Delta_2^{(r)*}(\sigma) + \Omega_2^{(r)*}(\sigma))\}, \\
\tilde{y}^{**}_{igk}(\hat{\theta})\tilde{y}^{**}_{irp}(\hat{\theta}) =& \exp\{\Delta_2^{(g)}(\hat{\theta}) + \Omega_2^{(g)}(\hat{\theta}) + \Delta_2^{(r)}(\hat{\theta}) + \Omega_2^{(r)}(\hat{\theta})\} \\
\approx& \exp\{\Delta_2^{(g)}(\theta) + \Omega_2^{(g)}(\theta) + (\hat{\theta}-\theta)^T\Delta_2^{(g)*}(\theta) + (\hat{\theta}-\theta)^T\Omega_2^{(g)*}(\theta) + \\
& \Delta_2^{(r)}(\theta) + \Omega_2^{(r)}(\theta) + (\hat{\theta}-\theta)^T\Delta_2^{(r)*}(\theta) + (\hat{\theta}-\theta)^T\Omega_2^{(r)*}(\theta)\} \\
=& \exp\{\Delta_2^{(g)}(\theta) + \Omega_2^{(g)}(\theta) + (\hat{\beta}-\beta)^T(\Delta_2^{(g)*}(\beta) + \Omega_2^{(g)*}(\beta)) + \\
& (\hat{\sigma}-\sigma)^T(\Delta_2^{(g)*}(\sigma)) + \Omega_2^{(g)*}(\sigma) + \Delta_2^{(r)}(\theta) + \Omega_2^{(r)}(\theta) + \\
& (\hat{\beta}-\beta)^T(\Delta_2^{(r)*}(\beta) + \Omega_2^{(r)*}(\beta)) + (\hat{\sigma}-\sigma)^T(\Delta_2^{(r)*}(\sigma) + \Omega_2^{(r)*}(\sigma))\}.
\end{aligned} \tag{C.14}$$

The following step is to calculate the expected value for the above expressions. Let

$$\begin{aligned}
\delta_{1j} \equiv& \Delta_1^*(\beta) + \Omega_1^*(\beta) = x_{ij}^T - 1^T \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} x_i, \\
\rho_{1j} \equiv& \Delta_1^*(\sigma) + \Omega_1^*(\sigma) = \begin{bmatrix} a_{1j} & b_{1j} & c_{1j} \end{bmatrix}^T,
\end{aligned}$$

where

$$\begin{aligned}
a_{1j} =& 1^T\Big(\begin{bmatrix} 0^T \\ 1^T \end{bmatrix} - \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} J\Big)\bar{V}^{-1}(\bar{l}_i - x_i\beta) + \frac{1}{2}1^T\Sigma_{1j}^{(v)}1, \\
b_{1j} =& 1^T\Big(\begin{bmatrix} 1_j^T \\ 0^T \end{bmatrix} - \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1}\Big)\bar{V}^{-1}(\bar{l}_i - x_i\beta) + \frac{1}{2}1^T\Sigma_{1j}^{(u)}1, \\
c_{1j} =& -1^T \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} D \bar{V}^{-1}(\bar{l}_i - x_i\beta) + \frac{1}{2}1^T\Sigma_{1j}^{(e)}1 + \frac{1}{2}.
\end{aligned}$$

Assuming that $\hat{\beta}$ and $\hat{\sigma}$ are unbiased estimators of β and σ respectively (Rao, 2003, chap.6), it follows that

$$\begin{aligned}
E[\delta_{1j}^T(\hat{\beta}-\beta)] =& 0 \\
E[\rho_{1j}^T(\hat{\sigma}-\sigma)] =& 0,
\end{aligned}$$

and

$$\begin{aligned}
E[\Delta_1(\hat{\theta}) + \Omega_1(\hat{\theta})] =& \Omega_1(\theta) \\
\Phi_1(\theta) \equiv& var[\Delta_1(\hat{\theta}) + \Omega_1(\hat{\theta})] \\
=& var(\Delta_1(\theta)) + \delta_{1j}^T var(\hat{\beta})\delta_{1j} + E[\rho_{1j}(\hat{\sigma} - \sigma)]^2 \\
=& 1^T V_{\mu_{1j}} 1 + \lambda_{1j}
\end{aligned}$$

where

$$\begin{aligned}
V_{\mu_{1j}} =& 1^T \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} \begin{bmatrix} \alpha^{(j)} & \gamma \end{bmatrix} 1 \\
\lambda_{1j} =& \delta_{1j}^T var(\hat{\beta})\delta_{1j} + trace[E(\rho_{1j}\rho_{1j}^T) var(\hat{\sigma})].
\end{aligned}$$

The last term in λ_{1j} is calculated using $E(w^t u)^2 = trace[E(ww^T)E(u^T u)]$, where w and u are random vectors. $var(\hat{\beta})$ and $var(\hat{\sigma})$ are the asymptotic covariance matrices of the estimators, which are obtained from the inverse of the Fisher Information matrix under the REML procedure. then we have

$$\Delta_1(\hat{\theta}) + \Omega_1(\hat{\theta}) \sim N(\Omega_1(\theta), \Phi_1(\theta)). \tag{C.15}$$

Now,

- From case 1

  1 Expressions in H_1, i.e $j, q \in s_i$ and $j = q$

$$\begin{aligned}
E[\tilde{y}^*_{ijk}(\theta)\tilde{y}^*_{iqp}(\theta)] &= \exp\{2\Omega_1(\theta) + \frac{1}{2}4var(\Delta_1(\theta))\} \\
&= \exp\{2(\Omega_1(\theta) + 1^T V_{\mu_{1j}} 1)\} \\
E[\tilde{y}^*_{ijk}(\theta)\tilde{y}^*_{iqp}(\hat{\theta})] &= \exp\{2\Omega_1(\theta) + \frac{1}{2}\varphi_1(\theta)\} \\
E[\tilde{y}^*_{ijk}(\hat{\theta})\tilde{y}^*_{iqp}(\hat{\theta})] &= \exp\{2(\Omega_1(\theta) + \Phi_1(\theta))\},
\end{aligned} \tag{C.16}$$

  where

$$\varphi_1(\theta) = 4var(\Delta_1(\theta)) + \lambda_{1j}.$$

  Then

$$E[\tilde{y}^*_{ijk}(\theta)\tilde{y}^*_{iqp}(\theta)] - 2E[\tilde{y}^*_{ijk}(\theta)\tilde{y}^*_{iqp}(\hat{\theta})] + E[\tilde{y}^*_{ijk}(\hat{\theta})\tilde{y}^*_{iqp}(\hat{\theta})] =$$

$$\exp\{2(\Omega_1(\theta) + 1^T V_{\mu_{1j}} 1)\} - 2\exp\{2\Omega_1(\theta) + \frac{1}{2}\varphi_1(\theta)\} + \exp\{2(\Omega_1(\theta) + \Phi_1(\theta))\} \equiv h_{1ij}. \tag{C.17}$$

2 Expressions in H_1, i.e $j, q \in s_i$ and $j \neq q$

$$\begin{aligned}
E[\tilde{y}^*_{ijk}(\theta)\tilde{y}^*_{iqp}(\theta)] &= \exp\{\Omega_1^{(j)}(\theta) + \Omega_1^{(q)}(\theta) + \frac{1}{2}1^T(V_{\mu_{1j}} + V_{\mu_{1q}})1\} \\
E[\tilde{y}^*_{ijk}(\theta)\tilde{y}^*_{iqp}(\hat{\theta})] &= \exp\{\Omega_1^{(j)}(\theta) + \Omega_1^{(q)}(\theta) + \frac{1}{2}(1^T V_{\mu_{1j}}1 + \Phi_1^{(q)}(\theta))\} \\
E[\tilde{y}^*_{ijk}(\hat{\theta})\tilde{y}^*_{iqp}(\hat{\theta})] &= \exp\{\Omega_1^{(j)}(\theta) + \Omega_1^{(q)}(\theta) + \frac{1}{2}(\Phi_1^{(j)}(\theta) + \Phi_1^{(q)}(\theta))\}. \qquad (C.18)
\end{aligned}$$

Then

$$\begin{aligned}
&E[\tilde{y}^*_{ijk}(\theta)\tilde{y}^*_{iqp}(\theta)] - E[\tilde{y}^*_{ijk}(\theta)\tilde{y}^*_{iqp}(\hat{\theta})] - E[\tilde{y}^*_{iqp}(\theta)\tilde{y}^*_{ijk}(\hat{\theta})] + E[\tilde{y}^*_{ijk}(\hat{\theta})\tilde{y}^*_{iqp}(\hat{\theta})] = \\
&\exp\{\Omega_1^{(j)}(\theta) + \Omega_1^{(q)}(\theta) + \frac{1}{2}1^T(V_{\mu_{1j}} + V_{\mu_{1q}})1\} - \\
&\exp\{\Omega_1^{(j)}(\theta) + \Omega_1^{(q)}(\theta) + \frac{1}{2}(1^T V_{\mu_{1j}}1 + \Phi_1^{(q)}(\theta))\} - \\
&\exp\{\Omega_1^{(j)}(\theta) + \Omega_1^{(q)}(\theta) + \frac{1}{2}(1^T V_{\mu_{1q}}1 + \Phi_1^{(j)}(\theta))\} + \\
&\exp\{\Omega_1^{(j)}(\theta) + \Omega_1^{(q)}(\theta) + \frac{1}{2}(\Phi_1^{(j)}(\theta) + \Phi_1^{(q)}(\theta))\} \\
= &\exp\{\Omega_1^{(j)}(\theta) + \Omega_1^{(q)}(\theta)\}[\exp\{\frac{1}{2}1^T(V_{\mu_{1j}} + V_{\mu_{1q}})1\} - \exp\{\frac{1}{2}(1^T V_{\mu_{1j}}1 + \Phi_1^{(q)}(\theta))\} - \\
&\exp\{\frac{1}{2}(1^T V_{\mu_{1q}}1 + \Phi_1^{(j)}(\theta))\} + \exp\{\frac{1}{2}(\Phi_1^{(j)}(\theta) + \Phi_1^{(q)}(\theta))\}] \equiv h_{1ijq}. \qquad (C.19)
\end{aligned}$$

Then from (C.17) and (C.19) it follows that

$$H_1 = \sum_{j \in s_i}(N_{ij} - n_{ij})^2 h_{1ij} + \sum_{j \in s_i}\sum_{q \in s_i}(N_{ij} - n_{ij})(N_{iq} - n_{iq})h_{1ijq}I(q \neq j). \qquad (C.20)$$

- From case 2

1 Expressions in H_2, i.e $j \in s_i$ and $r \in \bar{s}_i$

$$\begin{aligned}
E[\tilde{y}^*_{ijk}(\theta)\tilde{y}^{**}_{irp}(\theta)] &= \exp\{\Omega_1(\theta) + \Omega_2(\theta) + \frac{1}{2}1^T(V_{\mu_{1j}} + V_{\mu_2})1\} \\
E[\tilde{y}^*_{ijk}(\theta)\tilde{y}^{**}_{irp}(\hat{\theta})] &= \exp\{\Omega_1(\theta) + \Omega_2(\theta) + \frac{1}{2}(1^T V_{\mu_{1j}}1 + \Phi_2(\theta))\} \\
E[\tilde{y}^*_{ijk}(\hat{\theta})\tilde{y}^{**}_{irp}(\hat{\theta})] &= \exp\{\Omega_1(\theta) + \Omega_2(\theta) + \frac{1}{2}(\Phi_1(\theta) + \Phi_2(\theta))\}. \qquad (C.21)
\end{aligned}$$

Then

$$
\begin{aligned}
& E[\tilde{y}^{*}_{ijk}(\theta)\tilde{y}^{**}_{irp}(\theta)] - E[\tilde{y}^{*}_{ijk}(\theta)\tilde{y}^{**}_{irp}(\hat{\theta})] - E[\tilde{y}^{*}_{ijk}(\theta)\tilde{y}^{**}_{irp}(\hat{\theta})] + E[\tilde{y}^{*}_{ijk}(\hat{\theta})\tilde{y}^{**}_{irp}(\hat{\theta})] = \\
& \exp\{\Omega_1(\theta) + \Omega_2(\theta) + \frac{1}{2}1^T(V_{\mu_{1j}} + V_{\mu_2})1\} - \exp\{\Omega_1(\theta) + \Omega_2(\theta) + \frac{1}{2}(1^T V_{\mu_{1j}}1 + \Phi_2(\theta))\} - \\
& \exp\{\Omega_1(\theta) + \Omega_2(\theta) + \frac{1}{2}(1^T V_{\mu_2}1 + \Phi_1(\theta))\} + \exp\{\Omega_1(\theta) + \Omega_2(\theta) + \frac{1}{2}(\Phi_1(\theta) + \Phi_2(\theta))\} \\
= & \exp\{\Omega_1(\theta) + \Omega_2(\theta)\}[\exp\{\frac{1}{2}1^T(V_{\mu_{1j}} + V_{\mu_2})1\} - \exp\{\frac{1}{2}(1^T V_{\mu_{1j}}1 + \Phi_2(\theta))\} - \\
& \exp\{\frac{1}{2}(1^T V_{\mu_2}1 + \Phi_1(\theta))\} + \exp\{\frac{1}{2}(\Phi_1(\theta) + \Phi_2(\theta))\}] \equiv h_{2ijr}. \qquad \text{(C.22)}
\end{aligned}
$$

Then from (C.22) it follows

$$
H_2 = \sum_{j \in s_i} \sum_{r \in \bar{s}_i} (N_{ij-n_{ij}}) N_{ir} h_{2ijr}. \qquad \text{(C.23)}
$$

- From case 3

  1 Expressions in H_3, i.e $g, r \in \bar{s}_i$ and $g = r$

$$
\begin{aligned}
E[\tilde{y}^{**}_{igk}(\theta)\tilde{y}^{**}_{irp}(\theta)] &= \exp\{2(\Omega_2(\theta) + 1^T V_{\mu_2}1)\} \\
E[\tilde{y}^{**}_{igk}(\theta)\tilde{y}^{**}_{irp}(\hat{\theta})] &= \exp\{2\Omega_2(\theta) + \frac{1}{2}\varphi_2(\theta)\} \\
E[\tilde{y}^{**}_{igk}(\hat{\theta})\tilde{y}^{**}_{irp}(\hat{\theta})] &= \exp\{2(\Omega_2(\theta) + \Phi_2(\theta))\}, \qquad \text{(C.24)}
\end{aligned}
$$

where

$$
\begin{aligned}
\varphi_2(\theta) &= 4var(\Delta_2(\theta)) + \lambda_2 \\
\lambda_2 &= \delta_2^T var(\hat{\beta})\delta_2 + trace[E(\rho_2\rho_2^T)var(\hat{\sigma})] \\
\delta_2^T &= \Delta_2^*(\beta) + \Omega_2^*(\beta) \\
\rho_2^T &= \Delta_2^*(\sigma) + \Omega_2^*(\sigma) = \begin{bmatrix} a_2 & b_2 & c_2 \end{bmatrix}^T \\
E[\Delta_2(\hat{\theta}) + \Omega_2(\hat{\theta})] &= \Omega_2(\theta) \\
\Phi_2(\theta) &\equiv var[\Delta_2(\hat{\theta}) + \Omega_2(\hat{\theta})] \\
&= var(\Delta_2(\hat{\theta})) + \delta_2^T var(\hat{\beta})\delta_2 + E[\rho_2(\hat{\sigma} - \sigma)]^2 \\
&= 1^T V_{\mu_2}1 + \lambda_2,
\end{aligned}
$$

with

$$\begin{aligned}
a_2 =& 1^T\Big(\begin{bmatrix}0^T\\1^T\end{bmatrix} - \begin{bmatrix}0^T\\\gamma^T\end{bmatrix}\bar{V}^{-1}J\Big)\bar{V}^{-1}(\bar{l}_i - x_i\beta) + \frac{1}{2}1^T\Sigma_2^{(v)}1\\
b_2 =& -1^T\begin{bmatrix}0^T\\\gamma^T\end{bmatrix}\bar{V}^{-1}I\bar{V}^{-1}(\bar{l}_i - x_i\beta) + \frac{1}{2}1^T\Sigma_2^{(u)}1\\
c_2 =& -1^T\begin{bmatrix}0^T\\\gamma^T\end{bmatrix}\bar{V}^{-1}D\bar{V}^{-1}(\bar{l}_i - x_i\beta) + \frac{1}{2}1^T\Sigma_2^{(e)}1 + \frac{1}{2}.
\end{aligned}$$

Then

$$E[\tilde{y}^{**}_{igk}(\theta)\tilde{y}^{**}_{irp}(\theta)] - 2E[\tilde{y}^{**}_{igk}(\theta)\tilde{y}^{**}_{irp}(\hat{\theta})] + E[\tilde{y}^{**}_{igk}(\hat{\theta})\tilde{y}^{**}_{irp}(\hat{\theta})] =$$
$$\exp\{2(\Omega_2(\theta) + 1^TV_{\mu_2}1)\} - 2\exp\{2\Omega_2(\theta) + \frac{1}{2}\varphi_2(\theta)\} + \exp\{2(\Omega_2(\theta) + \Phi_2(\theta))\} \equiv h_{3ig}. \tag{C.25}$$

2 Expressions in H_3, i.e $g, r \in \bar{s}_i$ and $g \neq r$

$$\begin{aligned}
E[\tilde{y}^{**}_{igk}(\theta)\tilde{y}^{**}_{irp}(\theta)] =& \exp\{\Omega_2^{(g)}(\theta) + \Omega_2^{(r)}(\theta) + \frac{1}{2}1^T(V_{\mu_{2g}} + V_{\mu_{2r}})1\}\\
E[\tilde{y}^{**}_{igk}(\theta)\tilde{y}^{**}_{irp}(\hat{\theta})] =& \exp\{\Omega_2^{(g)}(\theta) + \Omega_2^{(r)}(\theta) + \frac{1}{2}(1^TV_{\mu_{2g}}1 + \Phi_2^{(r)}(\theta))\}\\
E[\tilde{y}^{**}_{igk}(\hat{\theta})\tilde{y}^{**}_{irp}(\hat{\theta})] =& \exp\{\Omega_2^{(g)}(\theta) + \Omega_2^{(r)}(\theta) + \frac{1}{2}(\Phi_2^{(g)}(\theta) + \Phi_2^{(r)}(\theta))\}.
\end{aligned} \tag{C.26}$$

Then

$$\begin{aligned}
&E[\tilde{y}^{**}_{igk}(\theta)\tilde{y}^{**}_{irp}(\theta)] - E[\tilde{y}^{**}_{igk}(\theta)\tilde{y}^{**}_{irp}(\hat{\theta})] - E[\tilde{y}^{**}_{irp}(\theta)\tilde{y}^{**}_{igk}(\hat{\theta})] + E[\tilde{y}^{**}_{igk}(\hat{\theta})\tilde{y}^{**}_{irp}(\hat{\theta})] =\\
&\exp\{\Omega_2^{(g)}(\theta) + \Omega_2^{(r)}(\theta) + \frac{1}{2}1^T(V_{\mu_{2g}} + V_{\mu_{2r}})1\} -\\
&\exp\{\Omega_2^{(g)}(\theta) + \Omega_2^{(r)}(\theta) + \frac{1}{2}(1^TV_{\mu_{2g}}1 + \Phi_2^{(r)}(\theta))\} -\\
&\exp\{\Omega_2^{(g)}(\theta) + \Omega_2^{(r)}(\theta) + \frac{1}{2}(1^TV_{\mu_{2r}}1 + \Phi_2^{(g)}(\theta))\} +\\
&\exp\{\Omega_2^{(g)}(\theta) + \Omega_2^{(r)}(\theta) + \frac{1}{2}(\Phi_2^{(g)}(\theta) + \Phi_2^{(r)}(\theta))\}\\
=& \exp\{\Omega_2^{(g)}(\theta) + \Omega_2^{(r)}(\theta)\}[\exp\{\frac{1}{2}1^T(V_{\mu_{2g}} + V_{\mu_{2r}})1\} - \exp\{\frac{1}{2}(1^TV_{\mu_{2g}}1 + \Phi_2^{(r)}(\theta))\} -\\
&\exp\{\frac{1}{2}(1^TV_{\mu_{2r}}1 + \Phi_2^{(g)}(\theta))\} + \exp\{\frac{1}{2}(\Phi_2^{(g)}(\theta) + \Phi_2^{(r)}(\theta))\}] \equiv h_{3igr}.
\end{aligned} \tag{C.27}$$

Then from (C.25) and (C.27) it follows

$$H_3 = \sum_{g \in \bar{s}_i} N_{ig}^2 h_{3ig} + \sum_{g \in \bar{s}_i} \sum_{r \in \bar{s}_i} N_{ig} N_{ir} h_{3igr} I(r \neq g). \tag{C.28}$$

Then substituting the expressions (C.20), (C.23) and (C.28) into (C.1) we get

$$\begin{aligned} M_{2i} \approx & \frac{1}{N_i^2} \Bigg[\sum_{j \in s_i} (N_{ij} - n_{ij})^2 h_{1ij} + \sum_{j \in s_i} \sum_{q \in s_i} (N_{ij} - n_{ij})(N_{iq} - n_{iq}) h_{1ijq} I(q \neq j) + \\ & 2 \sum_{j \in s_i} \sum_{r \in \bar{s}_i} (N_{ij - n_{ij}}) N_{ir} h_{2ijr} + \sum_{g \in \bar{s}_i} N_{ig}^2 h_{3ig} + \sum_{g \in \bar{s}_i} \sum_{r \in \bar{s}_i} N_{ig} N_{ir} h_{3igr} I(r \neq g) \Bigg]. \end{aligned} \tag{C.29}$$

Appendix D

Double bootstrap for bias correction

Hall and Maiti (2006) proves that $2B_{1i} - B_{2i}$ is an unbiased estimator for $MSE(\hat{\theta}_i)$, with $\hat{\theta}_i$ being a EBLUP under the traditional small area models. In this settings, we show that this estimator is a second-order unbiased for the estimator $\bar{Y}_i^{EB.BCP}$ under the proposed model. Given the data $Y = (Y_1, \ldots, Y_m)$ and the vector parameter θ, the MSE of the predictor is given by

$$M_i(\theta) = MSE(\bar{Y}_i^{EB.BCP}) = M_{i1}(\theta) + M_{i2}(\theta), \tag{D.1}$$

where
$M_{1i}(\theta) = E_\theta[(\bar{Y}_i - \hat{\bar{Y}}_i^{MMSE})^2]$ is the leading term and $M_{2i}(\theta) = E_\theta[(\hat{\bar{Y}}_i^{MMSE} - \hat{\bar{Y}}_i^{EB})^2]$.

Stage 1 : After estimating θ by $\hat{\theta}$ from the data, we use $\hat{\theta}$ and the model to generate $Y^{(b_1)}$, $b_1 = 1, \ldots, B_1$. Then, we fit the model with data $\{(X, Y^{b_1})\}$ and we compute $\bar{Y}_i^{*(b_1)}$, $\hat{\theta}^{*(b_1)}$ and $\bar{Y}_i^{EB.BCP*(b_1)}$.
Now following (D.1), we have

$$M_i(\hat{\theta}) = E_{\hat{\theta}}[(\bar{Y}_i^{(b_1)} - \bar{Y}_i^{EB.BCP(b_1)})^2] = M_{i1}(\hat{\theta}) + M_{i2}(\hat{\theta}).$$

As Hall and Maiti (2006) has pointed out,

$$E_\theta(M_{i1}(\hat{\theta})) = M_{i1}(\theta) + b_i(\theta, \tag{D.2}$$

and the bias, b_i, to estimate is expected to be of the same order as M_{i2}.

Defining

$$B_{i1} = \frac{1}{B_1}\sum_{b_1=1}^{B_1}[(\bar{Y}_i^{*(b_1)} - \bar{Y}_i^{EB.BCP*(b_1)})^2],$$

it gives us

$$\begin{aligned} E_{\hat{\theta}}(B_{i1}) =& E_{\hat{\theta}}[(\bar{Y}_i^{(b_1)} - \bar{Y}_i^{EB.BCP(b_1)})^2] = M_i(\hat{\theta}) \\ =& M_{i1}(\hat{\theta}) + M_{i2}(\hat{\theta}). \end{aligned} \tag{D.3}$$

Stage 2 : Now, from the model and $\hat{\theta}^{(b_1)}$ we generate $Y^{b_2(b_1)}$, $b_2 = 1, \ldots, B_2$. We fit the model with data $\{(X, Y^{b_2(b_1)})\}$ and we compute $\bar{Y}_i^{**b_2(b_1)}$, $\hat{\theta}^{b_2(b_1)}$ and $\bar{Y}_i^{EB.BCP**(b2(b_1))}$.

Then, we define

$$B_{i2} = \frac{1}{B_1}\sum_{b_1=1}^{B_1}\frac{1}{B_2}\sum_{b_2=1}^{B_2}[(\bar{Y}_i^{**b_2(b_1)} - \bar{Y}_i^{EB.BCP**(b_2(b_1))})^2].$$

It follows that

$$\begin{aligned} E_{\hat{\theta}}(B_{i2}) =& E_{\hat{\theta}}\left[\frac{1}{B_1}\sum_{b_1=1}^{B_1}\frac{1}{B_2}\sum_{b_2=1}^{B_2}[(\bar{Y}_i^{b_2(b_1)} - \bar{Y}_i^{EB.BCP(b_2(b_1))})^2]\right] \\ =& E_{\hat{\theta}}\left[\frac{1}{B_1}\sum_{b_1=1}^{B_1}E_{\hat{\theta}^{(b_1)}}[(\bar{Y}_i^{b_2(b_1)} - \bar{Y}_i^{EB.BCP(b_2(b_1))})^2]\right], \qquad (\textbf{by D.3}) \\ =& E_{\hat{\theta}}\left[\frac{1}{B_1}\sum_{b_1=1}^{B_1}M_i(\hat{\theta}^{(b_1)})\right] \\ =& E_{\hat{\theta}}[M_i(\hat{\theta}^{(b_1)})], \qquad (\textbf{by taking the expected value on the mean}) \\ =& E_{\hat{\theta}}[M_{i1}(\hat{\theta}^{(b_1)}) + M_{i2}(\hat{\theta}^{(b_1)})] \\ =& M_{i1}(\hat{\theta}) + b_i(\hat{\theta}) + M_{i2}(\hat{\theta}), \qquad (\textbf{by D.2}). \end{aligned}$$

Therefore,

$$\begin{aligned} E_\theta(2B_{i1}) =& 2(M_{i1}(\theta) + b_i(\theta) + M_{i2}(\theta)) \\ E_\theta(B_{i2}) =& (M_{i1}(\theta) + b_i(\theta)) + b_i(\theta) + M_{i2}(\theta), \end{aligned}$$

and it follows that

$$E_\theta(2B_{i1} - B_{i2}) = M_{i1}(\theta) + M_{i2}(\theta) = M_i(\theta).$$

Which proves that $2B_{i1} - B_{i2}$ is a second-order unbiased estimator of $MSE(\bar{Y}_i^{EB.BCP})$.

Appendix E

Numerical tables corresponding to the case $\sigma_v^2 = 0.05$

This section presents the tables with the numerical values corresponding to the realization of the simulation experiments in the case $\sigma_v^2 = 0.05$ and varying the values of σ_u^2. The emphasize was put on the behavior of EB predictors with respect to the population quantities.

	y.pop	y.pred	mse.naive	mse.1bbc	mse.1bc
1	199.48	187.58	41.79	30.84	12.42
2	186.47	206.96	41.97	30.97	13.67
3	215.49	223.70	42.34	31.24	12.24
4	167.17	174.73	41.87	30.90	12.66
5	257.33	256.27	40.54	29.95	12.40
6	136.41	132.40	40.86	30.18	12.29
7	222.78	219.66	41.97	30.97	13.12
8	293.21	278.07	41.32	30.51	13.09
9	189.09	172.09	43.07	31.76	13.70
10	220.77	213.70	40.94	30.23	13.03
11	275.77	256.81	43.86	32.33	12.81
12	209.07	207.44	41.27	30.47	12.86
13	228.27	229.47	41.74	30.81	14.24
14	233.48	228.61	41.92	30.93	12.63
15	162.07	177.93	42.16	31.10	12.60
16	185.38	168.60	42.58	31.39	12.57
17	201.60	173.37	41.71	30.77	12.37
18	188.80	201.02	41.02	30.27	11.94
19	213.57	224.88	42.70	31.47	13.06
20	153.97	149.04	41.94	30.92	12.74
21	277.73	255.78	42.22	31.12	12.39
22	164.10	172.40	41.64	30.69	13.51
23	156.03	159.32	40.31	29.74	12.35
24	223.94	196.03	42.02	30.97	12.75
25	281.37	238.18	39.91	29.45	12.02
26	220.43	224.46	40.60	29.93	13.23
27	134.45	149.25	43.35	31.89	13.60
28	184.22	179.44	41.92	30.85	12.50
29	254.94	233.55	41.42	30.50	12.72
30	203.11	207.39	42.01	30.90	12.72

Table E.1: Population, predictor, and MSE quantities for $\sigma_v^2 = 0.05$ and $\sigma_u^2 = 0.05$

	y.pop	y.pred	mse.naive	mse.1bbc	mse.1bc
1	200.96	185.37	42.00	28.48	17.40
2	189.57	218.03	42.14	28.59	19.00
3	222.82	231.58	42.47	28.79	16.92
4	167.59	180.67	42.02	28.52	17.68
5	267.34	261.09	40.74	27.73	17.40
6	138.59	136.50	41.03	27.87	17.34
7	234.99	229.24	42.08	28.56	18.55
8	302.19	278.23	41.44	28.17	18.52
9	195.81	175.66	43.08	29.21	18.90
10	223.26	212.37	41.06	27.93	18.08
11	281.51	254.09	43.81	29.68	17.83
12	211.47	209.05	41.34	28.09	18.04
13	237.09	237.33	41.78	28.38	19.86
14	240.73	234.41	41.94	28.47	17.78
15	164.55	187.15	42.15	28.61	17.43
16	190.62	171.06	42.53	28.84	17.66
17	206.13	169.47	41.69	28.31	17.16
18	191.55	210.45	41.02	27.88	16.62
19	222.95	236.12	42.58	28.86	18.07
20	159.71	155.26	41.85	28.39	17.71
21	285.63	251.80	42.13	28.57	17.31
22	166.52	179.75	41.54	28.17	18.75
23	161.60	166.84	40.28	27.35	17.21
24	225.89	189.46	41.90	28.41	17.52
25	290.13	227.86	39.90	27.13	16.65
26	225.92	231.23	40.52	27.50	18.46
27	137.82	161.40	43.09	29.13	19.01
28	191.40	185.03	41.71	28.20	17.60
29	263.37	232.14	41.25	27.92	17.67
30	208.95	217.58	41.78	28.23	17.59

Table E.2: Population, predictor, and MSE quantities for $\sigma_v^2 = 0.05$ and $\sigma_u^2 = 0.1$

	y.pop	y.pred	mse.naive	mse.1bbc	mse.1bc
1	203.99	185.93	42.77	27.23	21.37
2	193.30	227.35	42.90	27.36	23.14
3	230.14	238.16	43.18	27.49	20.63
4	168.96	186.77	42.76	27.30	21.62
5	276.43	265.29	41.50	26.61	21.37
6	141.02	141.43	41.75	26.65	21.39
7	246.09	237.41	42.78	27.30	22.87
8	311.04	279.46	42.14	26.98	22.86
9	202.38	180.16	43.71	27.87	22.95
10	226.39	213.06	41.77	26.76	21.97
11	287.91	253.94	44.41	28.28	21.80
12	214.15	211.55	41.99	26.86	22.10
13	244.95	243.68	42.41	27.14	24.23
14	247.37	239.58	42.56	27.20	21.88
15	167.89	195.29	42.75	27.30	21.22
16	195.72	174.76	43.10	27.49	21.69
17	211.29	168.99	42.29	27.04	20.89
18	194.85	218.73	41.63	26.66	20.28
19	232.04	245.19	43.12	27.50	21.96
20	165.33	161.70	42.40	27.06	21.58
21	293.75	250.82	42.67	27.25	21.18
22	169.29	186.47	42.07	26.83	22.81
23	166.85	173.81	40.84	26.10	21.06
24	228.81	187.12	42.42	27.06	21.21
25	298.91	222.62	40.48	25.95	20.28
26	231.48	236.89	41.06	26.24	22.54
27	141.46	172.23	43.53	27.68	23.20
28	198.47	190.52	42.16	26.78	21.65
29	272.43	232.74	41.72	26.54	21.54
30	214.70	226.26	42.21	26.79	21.41

Table E.3: Population, predictor, and MSE quantities for $\sigma_v^2 = 0.05$ and $\sigma_u^2 = 0.15$

	y.pop	y.pred	mse.naive	mse.1bbc	mse.1bc
1	207.88	187.82	43.89	26.71	24.87
2	197.42	235.84	44.02	26.87	26.72
3	237.57	244.21	44.26	26.92	23.91
4	170.87	192.92	43.86	26.81	25.07
5	285.12	269.54	42.61	26.20	24.89
6	143.58	146.63	42.82	26.14	24.96
7	256.74	245.01	43.85	26.77	26.65
8	319.93	281.72	43.20	26.52	26.66
9	208.96	185.02	44.73	27.31	26.46
10	229.86	214.86	42.83	26.34	25.31
11	294.74	255.29	45.41	27.68	25.26
12	216.98	214.56	43.01	26.37	25.64
13	252.34	249.42	43.43	26.65	28.02
14	253.71	244.61	43.56	26.67	25.49
15	171.74	202.80	43.74	26.75	24.52
16	200.80	178.97	44.07	26.92	25.23
17	216.90	170.12	43.27	26.51	24.12
18	198.50	226.50	42.61	26.19	23.46
19	241.10	253.27	44.05	26.92	25.32
20	170.99	168.11	43.33	26.49	24.93
21	302.10	251.67	43.62	26.71	24.54
22	172.27	192.79	42.98	26.25	26.29
23	171.99	180.39	41.77	25.58	24.44
24	232.24	186.83	43.34	26.50	24.38
25	307.82	220.08	41.44	25.51	23.45
26	237.11	242.17	41.98	25.71	26.07
27	145.33	182.22	44.40	27.06	26.80
28	205.62	195.92	43.02	26.12	25.20
29	282.04	234.54	42.58	25.94	24.89
30	220.46	234.25	43.05	26.13	24.72

Table E.4: Population, predictor, and MSE quantities for $\sigma_v^2 = 0.05$ and $\sigma_u^2 = 0.2$

Appendix F

Simulation results

In this appendix we present table with the numerical values corresponding to the realization of the simulation results described in section 5.8.2

	pop.c	pred.c	syth.c	msq.naive	msq.1bc	msq.dbc	msq.syth	msq.2bt
1	209.19	208.08	185.97	7.94	7.76	7.45	8.55	5.83
2	233.18	259.96	189.13	7.97	7.79	7.48	8.58	5.59
3	168.18	180.38	184.88	7.92	7.73	7.41	8.82	6.40
4	239.08	219.52	190.00	8.00	7.80	7.47	9.10	6.67
5	201.82	222.64	182.65	7.88	7.68	7.36	8.39	6.06
6	211.86	214.57	184.77	7.94	7.75	7.41	8.71	5.98
7	235.04	240.75	186.56	7.97	7.76	7.38	9.39	6.40
8	222.30	198.27	183.77	7.94	7.74	7.37	8.63	6.07
9	179.93	172.02	181.45	7.84	7.64	7.31	7.63	5.42
10	253.82	258.58	182.80	7.88	7.68	7.32	8.07	4.98
11	220.81	201.85	182.34	7.86	7.64	7.25	8.45	5.61
12	230.68	227.46	184.97	7.92	7.71	7.33	8.70	6.79
13	198.44	195.93	181.67	7.87	7.65	7.27	7.89	5.45
14	202.41	205.43	185.18	7.95	7.73	7.31	8.74	6.22
15	191.09	200.50	182.51	7.84	7.62	7.22	8.17	5.08
16	210.28	214.94	179.19	7.79	7.56	7.14	7.77	6.12
17	225.22	215.58	180.14	7.81	7.59	7.19	8.20	6.04
18	253.73	241.24	177.56	7.76	7.52	7.09	7.80	5.12
19	193.95	214.86	182.76	7.84	7.60	7.15	9.00	6.30
20	177.18	173.70	181.78	7.82	7.59	7.15	8.24	5.90
21	230.82	240.50	179.12	7.77	7.52	7.07	7.29	5.83
22	211.06	219.68	178.79	7.80	7.55	7.07	7.89	5.21
23	214.24	189.86	183.18	7.85	7.60	7.14	8.34	5.54
24	224.22	227.68	180.36	7.81	7.56	7.08	9.15	6.59
25	224.13	226.66	179.04	7.76	7.51	7.04	8.15	5.74
26	200.58	200.90	182.43	7.84	7.57	7.07	7.94	5.37
27	180.10	177.61	179.98	7.80	7.54	7.04	8.08	5.58
28	238.29	257.83	182.32	7.85	7.58	7.06	9.52	7.35
29	189.68	190.83	177.66	7.76	7.49	6.99	8.12	5.29
30	196.11	189.99	178.39	7.78	7.50	6.99	8.29	6.15

Table F.1: Population, EB predictor, Synthetic estimator and MSE quantities for $\sigma_v^2 = 0.05$ and $\sigma_u^2 = 0.2$

Bibliography

[1] Barnett, V., Lewis, T., (1994). *Outliers in Statistical Data*, John Wiley and Sons, 3edition.

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

[3] Berg, E., Chandra, H. (2014). Small area prediction for unit-level lognormal model. *Computational Statistics and Data Analysis*, 78, 159-175.

[4] Butar, F. B., Lahiri, P. (2003). On measures of uncertainty of empirical bayes small-area estimators, *Journal of Statistical Planning and Inference*, 112, 1, 6376.

[5] Chandra, H., Chambers, R. (2011). Small area estimation under transformation to linearity. *Survey Mothodology*, 37, 39-51.

[6] Chang, J., Hall, P. (2015). Double-bootstrap methods that use a single double-bootstrap simulation, *Biometrica*, 102, 1, 203-214.

[7] Datta, G. S., Ghosh, M. (1991). Bayesian prediction in linear models: Applications to small area estimation. *Annals of Statistics*, 19, 1748-1770.

[8] Davison, A. C., Hinkley, D. V., (1997). *Bootstrap methods and their application*, Cambridge University Press, United Kingdom.

[9] Diggle, P., Heagerty, P., Liang, K., and Zeger, S. (2002). *Analysis of Longitudinal Data*, 2nd ed., Oxford University Press, New York.

[10] Efron, B. and Morris, C. (1975). Data analysis using steinś estimator and and its generalizations. *Journal of the American Statistical Association*, 70(350), 311-313.

[11] Esteban, M. D., Morales.D., Prez, A., and Santamaría, L. (2012). Small area estimation of poverty proportions under area-level time models. *Computational Statistics and Data Analysis*, 56, 2840-2855.

[12] Fay, R. E., Herriot, R. A. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 85, 398-409.

[13] Fuller, W. A., Battese, G. E. (1973). Transformations for estimation of linear models with nested-error structure. *Journal of the American Statistical Association*, 68, 626-32.

[14] Ghosh, M., Maiti, T. (2004). Small-area estimation based on natural exponential family quadratic variance function models and survey weights. *Biometrika*, 91, 95-112.

[15] Ghosh, M. and Rao, J. N. K. (1994). Small area estimation: An appraisal. *Statistical Science*, 9(1), 55-76.

[16] Gonzalez, M. E. (1973). Use and evaluation of synthetic estimates. In Proceedings of the Social Statistics Section. *American Statistical Association*. USA, pp. 33-36.

[17] Hall P., Maiti T., (2006). On parametric bootstrap methods for small area prediction, *textJ.R. Statist. Soc. B*, 68, 2, 221-238.

[18] Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72, 320-340.

[19] Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 75, 423-447.

[20] Henderson, C. R. (1953). Estimation of variance and covariance components. *Biometrics*, 9, 226-252.

[21] Kackar, R. N., Harville, D. A. (1984). Approximations for Standard Errors of Estimators of Fixed and Random Effects in Mixed Linear Models. *Journal of the American Statistical Association*, 79, 853-862.

[22] Karlberg, F. (2014). Small area estimation for skewed data in the presence of zeroes. *Statistics in transition new series and Survey methodology*, 16, 4, 541-562.

[23] Kubokawa,T., Nagashima, B., (2012). Parametric bootstrap methods for bias correction in linear mixed models, *Journal of Multivariate Analysis*, 106, 116.

[24] McCulloch, C. E., Searle, S. R. (2001). *Generalized linear mixed models*, New York: Wiley.

[25] Molina, I., Rao, J. N. K. (2010). Small area estimation of poverty indicators. *Canadian Journal of Statistics*,38, 369-385.

[26] Pereira, L.N. and Coelho, P.S. (2012). Small area estimation using a spatio-temporal linear mixed model. *Statistical Journal*, 10,3, 285-308.

[27] Petrucci, A. and Pratesi, M. (2014). Spatial models in small area estimation in the context of official statistics. *Italian Journal of Applied Statistcs*, 24, 1.

[28] Petrucci, A. and Salvati, N. (2006). Small area estimation for spatial correlation in watershed erosion assessment.*Journal of Agricultural, Biological, and Environmental Statistics*, 11, 2, 169-182.

[29] Pfeffermann, D. (2013). New Important Developments in Small Area Estimation. *Statistical Science*, 28(1), 40-68.

[30] Pfeffermann, D., Barnard, C.H. (1991). Some new estimators for small area meanswith application to the assessment of farmland values. *Journal of Business and Economic Statistics*, 73-84.

[31] Pinheiro, J. C., Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*, Springer.

[32] Prasad, N. G. N., Rao, J. N. K. (1990). The Estimation of the Mean Squared Error of Small-Area Estimators. *Journal of the American Statistical Association*, 85, 409, 163-171.

[33] Pratesi, M. and Salvati, N. (2008). Small area estimation: the EBLUP estimator based on spatially correlated random area effects,*Statistical Methods and Applications*, 17, 1, 113141.

[34] Rachida O. (2011). Computationally efficient approximation for the double bootstrap mean bias correction. *Economics Bulletin*, 31, 3, 2388-2403.

[35] Rao, J. N. K. (2003). *Small Area Estimation*. New Jersey, John Wiley and Sons, Inc.

[36] Rao, J. N. K. (1999). Some recent advances in model-based small area estimation. *Survey Methodology*, 25, 175-186.

[37] Rao, J. N. K., Molina, I. (2015). *Small Area Estimation.* New Jersey, John Wiley and Sons, Inc.

[38] Robinson, G. K. (1991). That BLUP is a good thing: The estimation of random effects. *Statistical Science*, 6, 15-32.

[39] Royall, R. M. (1976). The linear least-squares prediction approach to two-stage sampling. *Journal of the American Statistical Association*, 71, 355, 657-664.

[40] Saei, A., Chambers, R. (2003). Small area estimates: A review of methods based on the application of mixed models. Technical Report Southampton Statistical Sciences Research Institute, Methodology Working Paper, M03/16.

[41] Sarndal, C. E., Swensson, B., and Wretman, J. (1992). *Model assisted survey sampling.* New York, Springer-Verlag.

[42] Searle, S. R., Casella, G., and McCulloch, C. E. (1992). *Variance Components.* John Wiley and Sons, New York.

[43] Slud, E. V., Maiti, T. (2006). Mean-squared error estimation in transformed Fay-Herriot models. *Journal of the Royal Statistical Society: Sec.B*,68,239-257.

[44] Sugasawa, S., Kubokawa, T., (2017). Transforming response values in small area prediction, *Journal of Computational Statistics and Data Analysis*, 114, 47-60.

[45] Stukel, D. M., Rao, J. N. K. (1999). On Small area estimation under two-fold nested error regression models. *Journal of Statistical Planning and Inference*, 78, 131-147.

[46] Valliant, R., Dortman, A. H., and Royall, R. M. (2000). *Finite Population Sampling and Inference.* John Wiley and Sons, New York.

[47] Verbeke, G., Molenberghs, G. (2000). *Linear mixed models for longitudinal data*, Springer-Verlag, Berlin.