Topics in small area estimation with applications to the National Resources Inventory

by

Junyuan Wang

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Major Professors: Wayne A. Fuller and F. Jay Breidt

Iowa State University

Ames, Iowa

2000

UMI Number: 9990499

Graduate College
Iowa State University

This is to certify that the Doctoral dissertation of

Junyuan Wang

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Co-major Professor

Signature was redacted for privacy.

Co-major Professor

Signature was redacted for privacy.

For the Major Program

Signature was redacted for privacy.

For the Graduate College

To my parents

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 OVERVIEW OF SMALL AREA ESTIMATION

1.1 Introduction

Sample surveys are a more cost-effective way of obtaining information than complete enumerations or censuses for most purposes. The surveys are usually designed to ensure that reliable estimators of totals and means for the population, pre-specified domains of interest, or major subpopulations can be derived from the survey data. There are also many situations in which it is desirable to derive reliable estimators for additional domains of interest, especially small geographical areas or small subpopulations, from existing survey data. The terms "small area" and "local area" are commonly used to denote a small geographical area, such as a county, a municipality or a census division and the term "small domain" is used to denote a small subpopulation, such as a specific age-sex-race group of people. In this discussion, these terms will not be differentiated and only "small area" will be used.

For example, suppose there is a state-wide survey about the average household income. Many cities in that state want to know the average household income for individual cities without conducting their own survey. Naturally, they turn to the state-wide survey. However, sample sizes for these cities (small areas) are typically small due to the size of cities. Therefore, the usual direct survey estimators of the average household income for such small areas, based on data only from the sample units in the area, are likely to yield unacceptably large standard errors (compared to the interesting statistic, the average household income) due to the small sample size in the area. This makes it necessary to "borrow strength" from data related to the city of interest using some form of model-dependent estimator to find more accurate estimates for the given area or, simultaneously, for several areas. The potential sources for the related data sets can be divided into three groups:

- Administrative data known at the small area level or sample element level
- Data measured for the same characteristics in other 'similar' areas
- Data measured for the same characteristics in the same area in a previous sample or census.

Classical survey sampling practitioners and Bayesians frequently disagree on the philosophical basis for estimation. But small area estimation is one area where these groups of statistician have a consensus on the need for *model-dependent* estimation. The idea is to relate similar small areas via supplementary data (e.g. census and administrative data) through explicit or implicit models. Ghosh and Rao (1994) provide an excellent review of many of the models found in the current literature and evaluate them in the light of practical considerations.

As pointed out by Marker (1999), most of the small area estimation models can be identified as special cases of generalized linear regression. This paper will focus on discussion of multiple linear regression models that include both fixed effects and random area-specific effects. There are two types of such models according to the level of aggregation at which the models are developed. Following is a review of these two models and of the estimation procedures in the literature.

1.2 Random effects model

The random effects model was originally proposed for small area estimation by Fay and Herriot (1979). The basic model is an old one. See, for example, page 260 of Snedecor and Cochran (1956). For this type of model, area-level auxiliary data $\boldsymbol{x}_i = (x_{i1}, ..., x_{ip})^T$

are available. The small area parameters (e.g. small area totals and means) of interest. y_i. are assumed to be related to $\boldsymbol{x}_i$ through the linear model:

$$y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + z_i b_i, \qquad i = 1, \ldots, m. \tag{1.1}$$

where $\boldsymbol{x}_i^T$ are known constants and the z_i are known positive constants, $\boldsymbol{\beta}$ is the vector of regression parameters and the b_i's are independent and identically distributed (iid) random variables with $E(b_i) = 0, V(b_i) = \sigma_b^2$. In addition, normality of the random effects, b_i, is often assumed.

We can not observe y_i directly due to the constraint of budget. The direct estimators, Y_i, are available through the sample and

$$Y_i = y_i + e_i, \qquad i = 1, \ldots, m, \tag{1.2}$$

where the e_i's are sampling errors with $E(e_i|y_i) = 0$ and $V(e_i|y_i) = \sigma_{ei}^2$. The Y_i is a design unbiased estimator of y_i with unacceptably large variance, σ_{ei}^2. It is customary in the survey literature to assume that σ_{ei}^2 are known. These assumptions may be quite restrictive in some applications. One such example is discussed in Chapter 2. We will discuss ways to relax these assumptions in later chapters.

Combining (1.1) and (1.2), we obtain the model

$$Y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + z_i b_i + e_i, \qquad i = 1, \ldots, m. \tag{1.3}$$

which is a special case of the general mixed linear model. It is worthwhile to point out that e_i's in (1.3) are design-induced random variables and b_i's are model-based random variables which are not observable.

There are many extensions of model (1.3) in the literature. Some of the extensions will be discussed in later chapters. In the next three sections, we will limit ourselves to model (1.3) and present estimation approaches for point estimation of y_i and for the measurement of uncertainty associated with the estimators. The three major approaches to estimation for random effects models are empirical (estimated) best linear unbiased predictor (EBLUP), empirical Bayes and hierarchical Bayes. The EBLUP approach is also called the variance components approach in some literature.

1.2.1 EBLUP Approach

Recall that the random effects model (1.3) is a special case of the general mixed linear model

$$\mathbf{Y} = \boldsymbol{X\beta} + \mathbf{Zb} + \boldsymbol{e}, \tag{1.4}$$

where $\boldsymbol{X}$ and $\mathbf{Z}$ are known matrices, $\boldsymbol{\beta}$ is a vector of p unknown parameters that have fixed values (fixed effects), and $\boldsymbol{b}$ and $\boldsymbol{e}$ are vectors of dimension q and n, respectively. The $\boldsymbol{b}$ and $\boldsymbol{e}$ are unobservable random variables (random effects) such that $E(\mathbf{b}) = \mathbf{0}$, $E(\boldsymbol{e}) = \mathbf{0}$ and

$$Var\begin{bmatrix} \mathbf{b} \\ \boldsymbol{e} \end{bmatrix} = \begin{bmatrix} \boldsymbol{G} & 0 \\ 0 & \boldsymbol{R} \end{bmatrix}.$$

Usually, $\boldsymbol{G}$ and $\boldsymbol{R}$ depend on some parameter $\boldsymbol{\theta}$.

Given model (1.4), we wish to estimate some linear function of $\boldsymbol{\beta}$ and predict some linear function of $\boldsymbol{b}$. In other words, we want to find the BLUP of $\boldsymbol{k}'\boldsymbol{\beta} + \boldsymbol{m}'\boldsymbol{b}$. The acronym BLUP stands for "Best Linear Unbiased Predictor". The BLUP of $\boldsymbol{k}'\boldsymbol{\beta} + \boldsymbol{m}'\boldsymbol{b}$ is *linear* in the sense that it is a linear function of the data $\mathbf{Y}$; *unbiased* in the sense that the average value of the predictor is equal to the average value of the quantity being predicted; *best* in the sense that it has minimum mean square error (MSE) within the class of linear unbiased predictors. *Predictor* is used to distinguish the statistic from estimators of fixed effects. There is some discussion about the appropriateness of the terminology "prediction", but it has become common practice to "estimate" fixed effects and to "predict" random effects. Robinson (1991) gives an excellent summary of BLUP theory and examples of its application.

To study BLUP, we first assume that the model parameter $\boldsymbol{\theta}$ is known. We also assume that $\boldsymbol{X}$ and $\mathbf{Z}$ are full rank matrices. Henderson (1950) showed that the predictor of $\boldsymbol{k}'\boldsymbol{\beta} + \boldsymbol{m}'\boldsymbol{b}$ is $\boldsymbol{k}'\widehat{\boldsymbol{\beta}} + \boldsymbol{m}'\widehat{\boldsymbol{b}}$, where $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{b}}$ are any solutions to

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\boldsymbol{R}^{-1}\boldsymbol{X} & \mathbf{Z}'\boldsymbol{R}^{-1}\mathbf{Z} + \boldsymbol{G}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\mathbf{Y} \\ \mathbf{Z}'\boldsymbol{R}^{-1}\mathbf{Y} \end{bmatrix}. \tag{1.5}$$

The equation (1.5) is known as the "mixed model equation", $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{b}}$ are called "mixed model solutions". Henderson *et al.* (1959) proved that any solution $\widehat{\boldsymbol{\beta}}$ to

$$\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}'\boldsymbol{V}^{-1}\mathbf{Y}, \tag{1.6}$$

where $\boldsymbol{V} = \boldsymbol{R} + \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}'$, is also a solution for $\boldsymbol{\beta}$ of (1.5). Henderson (1963) further proved that $\widehat{\mathrm{b}} = \boldsymbol{G}\boldsymbol{Z}'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})$ is a solution for $\boldsymbol{b}$ of (1.5). In short, the BLUP estimator of $\boldsymbol{k}'\boldsymbol{\beta} + \boldsymbol{m}'\boldsymbol{b}$ is given by

$$\boldsymbol{k}'\widehat{\boldsymbol{\beta}} + \boldsymbol{m}'\boldsymbol{G}\boldsymbol{Z}'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}). \tag{1.7}$$

where $\widehat{\boldsymbol{\beta}}$ is defined in (1.6). Note that these results do not depend on normality, a result that is similar to that for the best linear unbiased estimators (BLUEs) of fixed parameters.

Under model (1.3), we want to predict the small area parameter $y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + z_ib_i$, which is a linear combination of fixed and random effects. For the random effects model (1.3),

$$\boldsymbol{V} = diag(z_1^2\sigma_b^2 + \sigma_{e1}^2, \ldots, z_m^2\sigma_b^2 + \sigma_{em}^2),$$

$\boldsymbol{G} = diag(\sigma_b^2, \ldots, \sigma_b^2)$, and $\boldsymbol{Z} = diag(z_1, \ldots, z_m)$. Taking $\boldsymbol{k} = \boldsymbol{x}_i$ and $\boldsymbol{m} = (0, \ldots, 0, z_i, 0, \ldots, 0)^T$ with z_i in the i-th position, the BLUP of $y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + z_ib_i$ is

$$\widetilde{y}_i^H = \boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}} + \gamma_i(Y_i - x_i^T\widehat{\boldsymbol{\beta}}), \tag{1.8}$$

where the superscript H stands for Henderson,

$$\widehat{\beta} = \left[\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right]^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\mathbf{Y} = \left[\sum_{i=1}^{m} \boldsymbol{x}_i\boldsymbol{x}_i^T(z_i^2\sigma_b^2 + \sigma_{ei}^2)^{-1}\right]^{-1} \left[\sum_{i=1}^{m} \boldsymbol{x}_iY_i(z_i^2\sigma_b^2 + \sigma_{ei}^2)^{-1}\right].$$

and

$$\gamma_i = z_i^2\sigma_b^2(z_i^2\sigma_b^2 + \sigma_{ei}^2)^{-1}. \tag{1.9}$$

Rewriting (1.8) as

$$\widetilde{y}_i^H = \gamma_i Y_i + (1-\gamma_i)\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}}, \tag{1.10}$$

we see that $\widetilde{y}_i^H$ is a weighted average of the direct estimator Y_i and the regression estimator $\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}}$. Thus, the BLUP takes proper account of between-area variation relative to the precision of the direct estimator. The BLUP is also design consistent in the sense that $\gamma_i \to 1$ when the sampling variance $\sigma_{ei}^2 \to 0$, i.e., $\widetilde{y}_i^H \to Y_i = y_i$ if $\sigma_{ei}^2 \to 0$.

Henderson (1975) gave a general result for $var(\widetilde{\boldsymbol{y}}^H - \boldsymbol{y}) = MSE(\widetilde{\boldsymbol{y}}^H - \boldsymbol{y})$. However, because Henderson's simplification process involves lots of matrix algebra, we use direct calculation to obtain $M_1(\boldsymbol{\theta}) \equiv MSE(\widetilde{y}_i^H) = E(\widetilde{y}_i^H - y_i)^2$. Note that

$$\begin{aligned}
\widetilde{y}_i^H - y_i &= \gamma_i Y_i + (1-\gamma_i)\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}} - y_i \\
&= \gamma_i Y_i + (1-\gamma_i)\boldsymbol{x}_i^T\boldsymbol{\beta} - y_i + (1-\gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\
&= \gamma_i(\boldsymbol{x}_i^T\boldsymbol{\beta} + z_i b_i + e_i) + (1-\gamma_i)\boldsymbol{x}_i^T\boldsymbol{\beta} - (\boldsymbol{x}_i^T\boldsymbol{\beta} + z_i b_i)(1-\gamma_i)\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}} \\
&= \gamma_i e_i + (\gamma_i - 1)z_i b_i + (1-\gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}),
\end{aligned} \tag{1.11}$$

and

$$\begin{aligned}
&cov\{\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}), \gamma_i e_i + (\gamma_i - 1)z_i b_i\} \\
&= cov\{\boldsymbol{x}_i^T\left[\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X}\right]^{-1}\sum_{i=1}^{m}\boldsymbol{x}_i Y_i(z_i^2\sigma_b^2 + \sigma_{ei}^2)^{-1}, \gamma_i e_i - (1-\gamma_i)z_i b_i\} \\
&= \left\{\gamma_i\boldsymbol{x}_i^T\left[\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X}\right]^{-1}\boldsymbol{x}_i\sigma_{ei}^2 - (1-\gamma_i)\boldsymbol{x}_i^T\left[\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X}\right]^{-1}\boldsymbol{x}_i z_i^2\sigma_b^2\right\}(z_i^2\sigma_b^2 + \sigma_{ei}^2)^{-1} \\
&= 0.
\end{aligned} \tag{1.12}$$

Following the notation of Prasad and Rao (1990) and by using (1.11) and (1.12), we have

$$M_1(\boldsymbol{\theta}) \equiv E(\widetilde{y}_i^H - y_i)^2 = g_{1i}(\boldsymbol{\theta}) + g_{2i}(\boldsymbol{\theta}), \tag{1.13}$$

where

$$g_{1i}(\boldsymbol{\theta}) = \gamma_i \sigma_{ei}^2 = E\left[\gamma_i Y_i + (1-\gamma_i)\boldsymbol{x}_i^T \boldsymbol{\beta} - y_i\right]^2 \tag{1.14}$$

and

$$g_{2i}(\boldsymbol{\theta}) = (1-\gamma_i)^2 \boldsymbol{x}_i^T \left[\sum_{i=1}^{m} \boldsymbol{x}_i \boldsymbol{x}_i^T (z_i^2 \sigma_b^2 + \sigma_{ei}^2)^{-1}\right]^{-1} \boldsymbol{x}_i. \tag{1.15}$$

The term $g_{2i}(\boldsymbol{\theta})$ is due to the estimation of $\boldsymbol{\beta}$.

The BLUP $\widetilde{y}_i^H$ in (1.8) depends on the variance components σ_b^2 and $\sigma_{ei}^2, i = 1, \ldots, m$. In the survey literature, the sampling error variances $\sigma_{ei}^2, i = 1, \ldots, m$ are customarily assumed to be known or to depend on some parameter $\boldsymbol{\theta}_1$. Also, σ_b^2 is unknown in practical applications. Various methods of estimating the unknown variance components $\boldsymbol{\theta} = (\sigma_b^2, \boldsymbol{\theta}_1')'$ are available. Several of these methods such as Henderson's method 3, maximum likelihood (ML), and restricted maximum likelihood (REML), are derived for more general settings and yield asymptotically consistent estimators under reasonable regularity conditions.

Replacing $\boldsymbol{\theta}$ with an asymptotically consistent estimator $\widehat{\boldsymbol{\theta}}$ in (1.9), we get the estimator

$$\widehat{y}_i^H = \widehat{\gamma}_i Y_i + (1-\widehat{\gamma}_i)\boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}}, \tag{1.16}$$

where $\widehat{y}_i^H$ is called the empirical BLUP or EBLUP estimator. Kackar and Harville (1984) showed that $E(\widehat{y}_i^H) = E(y_i)$ if $E(\widetilde{y}_i^H)$ is finite, the elements of $\widehat{\boldsymbol{\theta}}$ are even functions of $\boldsymbol{y}$ and are translation-invariant (standard methods of estimating $\boldsymbol{\theta}$ satisfy these requirements), and the distributions of $\boldsymbol{b}$ and $\boldsymbol{e}$ are symmetric. The assumption of symmetric distribution is critical and is not always true in practice.

Though the EBLUP $\widehat{y}_i^H$ is unbiased under these regularity conditions, $M_1(\boldsymbol{\theta})$, the MSE of $\widetilde{y}_i^H - y_i$ in (1.13), underestimates (sometimes severely) the $MSE(\widehat{y}_i^H - y_i)$ due to the estimation of $\boldsymbol{\theta}$. Kackar and Harville (1984) showed that

$$\begin{aligned} M_2(\boldsymbol{\theta}) &\equiv E(\widehat{y}_i^H - y_i)^2 = E(\widetilde{y}_i^H - y_i)^2 + E(\widehat{y}_i^H - \widetilde{y}_i^H)^2 \\ &= M_1(\boldsymbol{\theta}) + E(\widehat{y}_i^H - \widetilde{y}_i^H)^2 \end{aligned} \tag{1.17}$$

if $\boldsymbol{b}$ and $\boldsymbol{e}$ are normal and $\widehat{\boldsymbol{\theta}}$ is translation-invariant. The second term in (1.17) is not tractable in general. By expanding the second term in a Taylor series in $\widehat{\boldsymbol{\theta}}$ about $\boldsymbol{\theta}$. Kackar and Harville (1984) obtained a second-order approximation

$$(\widehat{y}_i^H - \widetilde{y}_i^H)^2 = [\widetilde{y}_i^H(\widehat{\boldsymbol{\theta}}) - \widetilde{y}_i^H(\boldsymbol{\theta})]^2 \doteq [d(\boldsymbol{\theta})'(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})]^2,$$

where $d(\boldsymbol{\theta})' = \partial\widetilde{y}_i^H(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. They then proposed that

$$E[d(\boldsymbol{\theta})'(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})]^2 \doteq tr[A(\boldsymbol{\theta})E(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})'], \tag{1.18}$$

where $A(\boldsymbol{\theta})$ is the covariance matrix of $d(\boldsymbol{\theta})$.

Prasad and Rao (1990) proposed a further approximation to g_{3i} given by

$$g_{3i} = tr[A(\boldsymbol{\theta})E(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})'] \doteq tr[(\boldsymbol{B}'\boldsymbol{V}\boldsymbol{B}E(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})'], \tag{1.19}$$

where $\boldsymbol{B} = [(\partial(\boldsymbol{m}'\boldsymbol{G}\boldsymbol{Z}'\boldsymbol{V}^{-1})/\partial\theta_1)', \ldots, (\partial(\boldsymbol{m}'\boldsymbol{G}\boldsymbol{Z}'\boldsymbol{V}^{-1})/\partial\theta_p)']$ and $\boldsymbol{m}'\boldsymbol{G}\boldsymbol{Z}'\boldsymbol{V}^{-1}$ is defined in (1.7). They also evaluated (1.19) for three small area models. As an example, the $g_{3i}(\boldsymbol{\theta})$ for model (1.3) when $\boldsymbol{\theta} = \sigma_b^2$ is

$$g_{3i}(\sigma_b^2) = \sigma_{ei}^4 z_i^4(\sigma_b^2 z_i^2 + \sigma_{ei}^2)^{-3} V(\widehat{\sigma}_b^2). \tag{1.20}$$

Under some regularity conditions, the order of the neglected terms in the approximation (1.18) and (1.19) is $o(m^{-1})$ for large m. Also, an estimator of $g_{1i}(\boldsymbol{\theta})$ is obtained by adjusting $g_{1i}(\widehat{\boldsymbol{\theta}})$ for its bias to $O(m^{-1})$. After considerable algebraic simplification. Prasad and Rao (1990) obtained

$$E(g_{1i}(\widehat{\boldsymbol{\theta}})) = g_{1i}(\boldsymbol{\theta}) + g_{3i}(\boldsymbol{\theta}) + o(m^{-1}), \tag{1.21}$$

where $g_{1i}(\widehat{\boldsymbol{\theta}})$ is defined in (1.14) and $g_{3i}(\boldsymbol{\theta})$ is defined in (1.20). Therefore, an approximately unbiased estimator of $MSE(\widehat{y}_i^H - y_i)$ with expectation correct to $o(m^{-1})$ is

$$MSE(\widehat{y}_i^H) = g_{1i}(\widehat{\boldsymbol{\theta}}) + g_{2i}(\widehat{\boldsymbol{\theta}}) + 2g_{3i}(\widehat{\boldsymbol{\theta}}). \tag{1.22}$$

The bias of (1.22) is of lower order than m^{-1}. Lahiri and Rao (1995) show that the approximation (1.22) is robust to departures from the assumption of normality of b_i under model (1.3).

1.2.2 Empirical Bayes approach

From a Bayesian perspective, model (1.1) with normal errors gives the distribution of $\boldsymbol{y} = (y_1, \ldots, y_m)^T$ conditional on $\boldsymbol{\beta}$ and σ_b^2, and model (1.2) with normal errors gives the distribution of $\boldsymbol{Y} = (Y_1, \ldots, Y_m)^T$ conditional on $\boldsymbol{y}$ and $\boldsymbol{\theta}_1$. Under these assumptions, we have conditional distributions

$$f(\boldsymbol{y}|\boldsymbol{\beta}, \sigma_b^2) \propto \exp\left[\sum_{i=1}^{m}(2z_i^2)^{-1}\sigma_b^2(y_i - x_i^T\boldsymbol{\beta})^2\right] \tag{1.23}$$

and

$$f(\boldsymbol{Y}|\boldsymbol{y}, \boldsymbol{\theta}_1) \propto \exp\left[\sum_{i=1}^{m}(2\sigma_{ei})^{-1}(Y_i - y_i)^2\right], \tag{1.24}$$

where $\propto$ means that the density has the specified form up to a constant of proportionality. If both $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are known, we derive from (1.23) and (1.24) the posterior distribution of the unknown $\boldsymbol{y}$

$$f(\boldsymbol{y}|\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{\theta}) \propto \exp\left[\sum_{i=1}^{m}(2g_{1i})^{-1}(y_i - y_i^B)^2\right], \tag{1.25}$$

where

$$y_i^B = \boldsymbol{x}_i^T\boldsymbol{\beta} + \gamma_i(Y_i - x_i^T\boldsymbol{\beta}), \quad i = 1, \ldots, m, \tag{1.26}$$

and γ_i and g_{1i} are defined in (1.9) and (1.14) respectively. Therefore, the posterior distributions of $y_i, i = 1, \ldots, m$, are independent normal distributions with mean y_i^B and variance g_{1i}. Under quadratic loss, y_i^B is the Bayes estimator of y_i.

When $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are unknown, we can estimate them from the marginal distribution of $\boldsymbol{Y}$. That is, we estimate $\boldsymbol{\beta}$ by using (1.6) and estimate $\boldsymbol{\theta}$ by using ML, REML, or other procedures. We then substitute the estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$ in (1.26) and obtain the empirical Bayes estimator $\widehat{y}_i^{EB}$ of y_i. Note that $\widehat{y}_i^{EB}$ is equal to the EBLUP estimator $\widehat{y}_i^H$ in (1.12) if the same estimator for $\boldsymbol{\theta}$ is used.

From a Bayesian perspective, the inference statements are conditional on the observed value $\boldsymbol{Y}$. In other words, we draw conclusions about $\boldsymbol{y}$ in terms of $f(\boldsymbol{y}|\boldsymbol{Y})$. Under quadratic loss. $E(y_i|\boldsymbol{Y})$ and $Var(y_i|\boldsymbol{Y})$ are the most important statistics. When $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are known, we can use (1.25) to obtain $E(y_i|\boldsymbol{Y}) = y_i^B$ and $Var(y_i|\boldsymbol{Y}) = g_{1i}$ since $f(\boldsymbol{y}|\boldsymbol{Y}) = f(\boldsymbol{y}|\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{\theta})$. When $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are unknown, $\widehat{y}_i^{EB} = E(y_i|\boldsymbol{Y}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}})$ and $g_{1i}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}) = Var(y_i|\boldsymbol{Y}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}})$. Note that

$$E(y_i|\boldsymbol{Y}) = E_{\boldsymbol{\beta},\boldsymbol{\theta}}[E(y_i|\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{\theta})] \tag{1.27}$$

and

$$Var(y_i|\boldsymbol{Y}) = E_{\boldsymbol{\beta},\boldsymbol{\theta}}[V(y_i|\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{\theta})] + Var_{\boldsymbol{\beta},\boldsymbol{\theta}}[E(y_i|\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{\theta})], \tag{1.28}$$

where $E_{\boldsymbol{\beta},\boldsymbol{\theta}}$ and $Var_{\boldsymbol{\beta},\boldsymbol{\theta}}$ denote the expectation and variance with respect to the posterior distribution of $\boldsymbol{\beta}, \boldsymbol{\theta}$ given $\boldsymbol{Y}$. We can see that $\widehat{y}_i^{EB}$ is a reasonable approximation to the true posterior mean $E(y_i|\boldsymbol{Y})$. However, the estimated posterior variance $g_{1i}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}})$ is only a good approximation to the first term of the right side of (1.28) and could severely underestimate the posterior variance $Var(y_i|\boldsymbol{Y})$. Therefore, $g_{1i}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}})$ is not a good measure of $Var(y_i|\boldsymbol{Y})$, the true variability of $E(y_i|\boldsymbol{Y})$, because it fails to account for the uncertainty about the unknown parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$.

The posterior distribution of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ given $\boldsymbol{Y}$ is not available because the prior distribution of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ is not specified in the empirical Bayes approach. Therefore, we can not use (1.28) to evaluate the posterior variance $Var(y_i|\boldsymbol{Y})$. Laird and Louis (1987) proposed a bootstrap approach to get around this. Kass and Steffy (1989) proposed an asymptotic approximation to $Var(y_i|\boldsymbol{Y})$ by adding a positive correction term to account for the underestimation. This term depends on the observed information matrix and the partial derivatives of y_i^B with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, evaluated at the ML estimates $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$. In a more recent result, Singh *et al* (1998) use Monte Carlo approximations for some of the error measures.

If we want to evaluate the empirical Bayes estimates in the frequentist framework, MSE is a natural measure of uncertainty. This makes empirical Bayes and EBLUP essentially equivalent. Therefore, the empirical Bayes approach is only discussed briefly in this paper. Morris (1983) offered a good introduction to the empirical Bayes approach with many applications. Ghosh and Rao (1994) provide a summary of results prior to 1994 and Ghosh (1999) collects and unifies many recent results.

1.2.3 Hierarchical Bayes approach

In addition to the empirical Bayes model assumptions (1.23) and (1.24), the hierarchical Bayes procedure models the uncertainty of hyper-parameters by assigning a prior distribution to the model parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ and derives the posterior distribution $f(\boldsymbol{y}|\boldsymbol{Y})$.

To demonstrate the hierarchical Bayes approach under model (1.23) and (1.24), we assume that the prior distribution of $\boldsymbol{\beta}$ is uniform over R^p and that $\boldsymbol{\theta}$ is known. Without loss of generality, we assume that $\boldsymbol{X}$ is full rank. The joint (improper) probability density function (pdf) of $\boldsymbol{Y}, \boldsymbol{y}$, and $\boldsymbol{\beta}$ is

$$f(\boldsymbol{Y}, \boldsymbol{y}, \boldsymbol{\beta}) \propto \exp\left[-\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{y})^T\boldsymbol{R}^{-1}(\boldsymbol{Y}-\boldsymbol{y})\right]\exp\left[-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{X\beta})^T\boldsymbol{G}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta})\right], \tag{1.29}$$

where $\boldsymbol{R} = diag(\sigma_{e1}^2, \ldots, \sigma_{em}^2)$ and $\boldsymbol{G} = diag(z_1^2\sigma_b^2, \ldots, z_m^2\sigma_b^2)$. Using the fact that

$$\begin{aligned}(\boldsymbol{y}-\boldsymbol{X\beta})^T\boldsymbol{G}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta}) &= \boldsymbol{y}^T\left[\boldsymbol{G}^{-1}-\boldsymbol{G}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{G}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{G}^{-1}\right]\boldsymbol{y}\\ &+\left[\boldsymbol{\beta}-(\boldsymbol{X}^T\boldsymbol{G}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{G}^{-1}\boldsymbol{y}\right]^T(\boldsymbol{X}^T\boldsymbol{G}^{-1}\boldsymbol{X})\left[\boldsymbol{\beta}-(\boldsymbol{X}^T\boldsymbol{G}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{G}^{-1}\boldsymbol{y}\right]\end{aligned} \tag{1.30}$$

and integrating with respect to $\boldsymbol{\beta}$ in (1.29), one finds the joint (improper) pdf of $\boldsymbol{Y}$ and $\boldsymbol{y}$ to be

$$\begin{aligned}f(\boldsymbol{Y}, \boldsymbol{y}) &\propto \exp\left[-\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{y})^T\boldsymbol{R}^{-1}(\boldsymbol{Y}-\boldsymbol{y})\right]\\ &\quad\times\exp\left\{-\frac{1}{2}\boldsymbol{y}^T\left[\boldsymbol{G}^{-1}-\boldsymbol{G}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{G}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{G}^{-1}\right]\boldsymbol{y}\right\}.\end{aligned} \tag{1.31}$$

Letting $\boldsymbol{S} = \boldsymbol{R}^{-1} + \left[\boldsymbol{G}^{-1}-\boldsymbol{G}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{G}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{G}^{-1}\right]$, we have

$$(\boldsymbol{Y}-\boldsymbol{y})^T\boldsymbol{R}^{-1}(\boldsymbol{Y}-\boldsymbol{y}) + \boldsymbol{y}^T\left[\boldsymbol{G}^{-1}-\boldsymbol{G}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{G}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{G}^{-1}\right]\boldsymbol{y}$$

$$= \boldsymbol{y}^T\boldsymbol{S}\boldsymbol{y} - 2\boldsymbol{y}^T\boldsymbol{R}^{-1}\boldsymbol{Y} + \boldsymbol{Y}^T\boldsymbol{R}^{-1}\boldsymbol{Y}$$

$$= (\boldsymbol{y} - \boldsymbol{S}^{-1}\boldsymbol{R}^{-1}\boldsymbol{Y})^T\boldsymbol{S}(\boldsymbol{y} - \boldsymbol{S}^{-1}\boldsymbol{R}^{-1}\boldsymbol{Y}) + \boldsymbol{Y}^T(\boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{S}^{-1}\boldsymbol{R}^{-1})\boldsymbol{Y}. \quad (1.32)$$

From (1.31) and (1.32), it follows that

$$f(\boldsymbol{y}|\boldsymbol{Y}) \propto \exp\left[-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{S}^{-1}\boldsymbol{R}^{-1}\boldsymbol{Y})^T\boldsymbol{S}(\boldsymbol{y} - \boldsymbol{S}^{-1}\boldsymbol{R}^{-1}\boldsymbol{Y})\right] \quad (1.33)$$

with

$$E(\boldsymbol{y}|\boldsymbol{Y}) = \boldsymbol{S}^{-1}\boldsymbol{R}^{-1}\boldsymbol{Y}; \quad Var(\boldsymbol{y}|\boldsymbol{Y}) = \boldsymbol{S}^{-1}. \quad (1.34)$$

Write $\boldsymbol{S} = (\boldsymbol{R}^{-1} + \boldsymbol{G}^{-1}) - \boldsymbol{G}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{G}^{-1}\ X)^{-1}\boldsymbol{X}^T\boldsymbol{G}^{-1}$ and note that

$$\left[\boldsymbol{\Sigma}_{ww}^{-1} - \boldsymbol{H}^T\boldsymbol{\Sigma}_{uu}^{-1}\boldsymbol{H}\right]^{-1} = \boldsymbol{\Sigma}_{ww} - \boldsymbol{\Sigma}_{ww}\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{\Sigma}_{ww}\boldsymbol{H}^T - \boldsymbol{\Sigma}_{uu})^{-1}\boldsymbol{H}\boldsymbol{\Sigma}_{ww}. \quad (1.35)$$

Letting $\boldsymbol{\Sigma}_{ww} = (\boldsymbol{R}^{-1} + \boldsymbol{G}^{-1})^{-1} = (\boldsymbol{I} - \boldsymbol{\Gamma})\boldsymbol{G}, \boldsymbol{H} = \boldsymbol{X}^T\boldsymbol{G}^{-1}, \boldsymbol{\Sigma}_{uu} = \boldsymbol{X}^T\boldsymbol{G}^{-1}\boldsymbol{X}$, we have

$$\boldsymbol{S}^{-1} = (\boldsymbol{I} - \boldsymbol{\Gamma})\boldsymbol{G} + (\boldsymbol{I} - \boldsymbol{\Gamma})\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{I} - \boldsymbol{\Gamma}) \quad (1.36)$$

and

$$\boldsymbol{S}^{-1}\boldsymbol{R}^{-1} = \boldsymbol{\Gamma} + (\boldsymbol{I} - \boldsymbol{\Gamma})\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}, \quad (1.37)$$

where $\boldsymbol{\Gamma} = diag(\gamma_1, \ldots, \gamma_m)$ and $\boldsymbol{V} = diag(z_1^2\sigma_b^2 + \sigma_{e1}^2, \ldots, z_m^2\sigma_b^2 + \sigma_{em}^2)$. Combining (1.34) and (1.37), we have

$$E(\boldsymbol{y}|\boldsymbol{Y}) = \boldsymbol{S}^{-1}\boldsymbol{R}^{-1}\boldsymbol{Y} = \boldsymbol{\Gamma}\boldsymbol{Y} + (\boldsymbol{I} - \boldsymbol{\Gamma})\boldsymbol{X}\widehat{\boldsymbol{\beta}}, \quad (1.38)$$

By (1.36) and (1.38), $E(y_i|\boldsymbol{Y}) = \widetilde{y}_i^H$ and $Var(y_i|\boldsymbol{Y}) = M_{1i} = MSE(\widetilde{y}_i^H - y_i)$ when $\boldsymbol{\beta}$ is unknown with non-informative prior and $\boldsymbol{\theta}$ is known. Under quadratic loss, the hierarchical Bayes approach and the BLUP approach lead to the same estimation.

When σ_b^2 is unknown, we should write $f(\boldsymbol{y}|\boldsymbol{Y})$ in (1.33) as $f(\boldsymbol{y}|\boldsymbol{Y}, \sigma_b^2)$, $E(\boldsymbol{y}|\boldsymbol{Y})$ in (1.34) as $E(\boldsymbol{y}|\boldsymbol{Y}, \sigma_b^2)$, and $Var(\boldsymbol{y}|\boldsymbol{Y})$ in (1.34) as $Var(\boldsymbol{y}|\boldsymbol{Y}, \sigma_b^2)$. Ghosh (1992) derived the closed form for $f(\sigma_b^2|\boldsymbol{Y})$ when σ_b^2 has a non-informative prior on $(0, \infty)$ and $z_i =$

$1, i = 1, \ldots, m$. By applying one-dimensional integration to (1.34), we have the posterior distribution

$$f(\boldsymbol{y}|\boldsymbol{Y}) = \int f(\boldsymbol{y}|\boldsymbol{Y}, \sigma_b^2) f(\sigma_b^2|\boldsymbol{Y}) d\sigma_b^2. \tag{1.39}$$

When $\boldsymbol{\theta}$ is unknown, we have the general form of (1.39)

$$f(\boldsymbol{y}|\boldsymbol{Y}) = \int f(\boldsymbol{y}|\boldsymbol{Y}, \boldsymbol{\theta}) f(\boldsymbol{\theta}|\boldsymbol{Y}) d\boldsymbol{\theta}. \tag{1.40}$$

where $f(\boldsymbol{y}|\boldsymbol{Y}, \boldsymbol{\theta})$ is $f(\boldsymbol{y}|\boldsymbol{Y})$ defined in (1.33). The posterior mean and variance of $\boldsymbol{y}$ are then given by

$$E(y_i|\boldsymbol{Y}) = E_{\boldsymbol{\theta}}[E(y_i|\boldsymbol{Y}, \boldsymbol{\theta})] \tag{1.41}$$

and

$$Var(y_i|\boldsymbol{Y}) = E_{\boldsymbol{\theta}}[V(y_i|\boldsymbol{Y}, \boldsymbol{\theta})] + Var_{\boldsymbol{\theta}}[E(y_i|\boldsymbol{Y}, \boldsymbol{\theta})], \tag{1.42}$$

where $E(\boldsymbol{y}|\boldsymbol{Y}, \boldsymbol{\theta})$ is $E(\boldsymbol{y}|\boldsymbol{Y})$ defined in (1.34) and $V(\boldsymbol{y}|\boldsymbol{Y}, \boldsymbol{\theta})$ is $V(\boldsymbol{y}|\boldsymbol{Y})$ defined in (1.34).

The empirical Bayes procedure attempts to estimate the unknown model parameters (usually called hyper-parameters) from the marginal distributions of the observations by classical methods and substitutes the estimates into the expressions for $E(\boldsymbol{y}|\boldsymbol{Y}, \boldsymbol{\theta})$. The empirical Bayes and hierarchical Bayes approaches often lead to comparable results in the context of point estimation. When it comes to measuring the standard errors associated with these estimators, the hierarchical Bayes method has a clear edge over a naive empirical Bayes method.

The idea of the hierarchical Bayes approach is straightforward. But the posterior distribution $f(\boldsymbol{\theta}|\boldsymbol{Y})$ in (1.40) usually does not have a closed form. Therefore, numerical evaluation is necessary and is computationaly intensive in many cases. There are many other ways to evaluate $f(\boldsymbol{y}|\boldsymbol{Y})$. One method is the Gibbs sampler advocated by Gelfand and Smith (1990). The Gibbs sampler is a Markov Chain Monte Carlo (MCMC) method that proceeds as follows. Suppose the random vector $\boldsymbol{U} = (U_1, \ldots, U_k)^T$ has been divided into k components or sub-vectors. Starting from an arbitrary set of values

$(U_1^{(0)}, \ldots, U_k^{(0)})$, we draw $U_1^{(1)} \sim f(U_1|U_2^{(0)}, \ldots, U_k^{(0)})$, $U_2^{(1)} \sim f(U_2|U_1^{(1)}, U_3^{(0)}, \ldots, U_k^{(0)})$, $U_3^{(1)} \sim f(U_3|U_1^{(1)}, U_2^{(1)}, U_4^{(0)}, \ldots, U_k^{(0)})$, and so on, up to $U_k^{(1)} \sim f(U_k|U_1^{(1)}, \ldots, U_{k-1}^{(1)})$. Thus, each component U_i is updated conditional on the latest value of $\boldsymbol{U}$ for the other components. After n such iterations, we arrive at $(U_1^{(n)}, \ldots, U_k^{(n)})$. Under some mild conditions, Geman and Geman (1984) showed that $(U_1^{(n)}, \ldots, U_k^{(n)}) \xrightarrow{d} f(U_1, \ldots, U_k)$. To draw a sample of size J from the joint distribution $\boldsymbol{U} = (U_1, \ldots, U_k)^T$, we perform a large number of cycles, say n, until convergence. Then we discard the first n samples and treat $\{(U_1^{(n+j)}, \ldots, U_k^{(n+j)}), j = 1, \ldots, J\}$ as J simulated samples from the joint distribution of $\boldsymbol{U} = (U_1, \ldots, U_k)^T$.

To illustrate the application of the Gibbs sampler in obtaining $f(\boldsymbol{y}|\boldsymbol{Y})$ in (1.39), we use model (1.23) and (1.24). We further assume that $\boldsymbol{\theta}_1$ is known, the prior on $\boldsymbol{\beta}$ is noninformative, and the prior on σ_b^2 is the inverse-gamma(a, b) distribution with parameters $\alpha > 0$ and $\nu > 0$, denoted by $f(\sigma_b^2) \propto \exp[-\alpha(\sigma_b^2)^{-1}](\sigma_b^2)^{-(\nu+1)}$. Therefore,

$$\begin{aligned} f(\boldsymbol{y}, \boldsymbol{\beta}, \sigma_b^2|\boldsymbol{Y}) &\propto f(\boldsymbol{Y}|\boldsymbol{y}) f(\boldsymbol{y}|\boldsymbol{\beta}, \sigma_b^2) f(\sigma_b^2) \\ &\propto \exp\left[-\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{y})^T \boldsymbol{R}^{-1}(\boldsymbol{Y}-\boldsymbol{y})\right] \exp\left[-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{X\beta})^T \boldsymbol{G}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta})\right] \\ &\quad \times \exp[-\alpha(\sigma_b^2)^{-1}](\sigma_b^2)^{-(\nu+1)}. \end{aligned} \tag{1.43}$$

From (1.43), we can derive

$$y_i \,|\, \boldsymbol{\beta}, \sigma_b^2, \boldsymbol{Y} \sim N(\gamma_i Y_i + (1-\gamma_i) x_i^T \boldsymbol{\beta}, g_{1i}) \quad i = 1, \ldots, m, \tag{1.44}$$

and

$$\boldsymbol{\beta} \,|\, \boldsymbol{y}, \sigma_b^2, \boldsymbol{Y} \sim N\left((\boldsymbol{X}^T \boldsymbol{G}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{G}^{-1} \boldsymbol{y}, (\boldsymbol{X}^T \boldsymbol{G}^{-1} \boldsymbol{X})^{-1}\right), \tag{1.45}$$

where $\boldsymbol{G} = diag(z_1^2 \sigma_b^2, \ldots, z_m^2 \sigma_b^2)$, and

$$\begin{aligned} \sigma_b^2 \,|\, \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{Y} \;\sim\;& \text{Inverse-gamma}(\alpha_1, \nu_1) \text{ with parameters} \\ & \alpha_1 = \frac{1}{2}\sum_{i=1}^m z_i^{-2}(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta})^2 + \alpha \text{ and } \nu_1 = \frac{m}{2} + \nu. \end{aligned} \tag{1.46}$$

Following the Gibbs sampler algorithm, we can draw a sample of size J from $f(\boldsymbol{y}, \boldsymbol{\beta}, \sigma_b^2|\boldsymbol{Y})$. If we simply discard the information about $\boldsymbol{\beta}$ and σ_b^2 from the sample, we obtain a sample of size J from the marginal distribution $f(\boldsymbol{y}|\boldsymbol{Y})$. The posterior mean and posterior variance of $\boldsymbol{y}$ are estimated by using the J simulated samples.

1.3 Nested error regression model

The nested error regression model for small area estimation was originally considered by Battese et al (1988). The model assumes that auxiliary information $\boldsymbol{x}_{ij} = (x_{ij1}, \ldots, x_{ijp})^T$ is available for every element in the *population* (or at least for every element in the sample) and that the population area totals of these variables are known. The variable of interest, y_{ij}, is assumed to be related to $\boldsymbol{x}_{ij}$ through a nested error linear model:

$$y_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + b_i + e_{ij}, \qquad j = 1, \ldots N_i, i = 1, \ldots, m, \tag{1.47}$$

where N_i is the number of the elements in the i-th small area. Here e_{ij} is independent of b_k for all i, j, and k, and $E(e_{ij}) = 0, V(e_{ij}) = \sigma_{e_{ij}}^2 = k_{ij}\sigma_e^2$ with k_{ij} being known constants. Normality of the b_i's and e_{ij}'s is often assumed. The parameters of interest are usually the small area totals $y_{i.}$, or the means $\bar{y}_i = y_{i.}N_i^{-1}$.

A sample of size n_i is taken from the i-th area to make inference about $y_{i.}$ or $\bar{y}_{i.}$. This model does not incorporate *any* sample design features. It is not appropriate for some complex sampling designs which have selection bias. However, it is possible to extend this model to account for such features.

The matrix form of model (1.47) is

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + b_i\mathbf{1}_i + \boldsymbol{e}_i, \tag{1.48}$$

where $\boldsymbol{X}_i$ is $N_i \times p$, $\boldsymbol{y}_i, \boldsymbol{e}_i$ and $\mathbf{1}_i$ are $N_i \times 1$ and $\mathbf{1}_i = (1, \ldots, 1)^T$. Using superscript * to denote the non-sampled elements and superscript s to denote the sampled elements, we can partition (1.48) as

$$\boldsymbol{y}_i = \begin{bmatrix} \boldsymbol{y}_i^s \\ \boldsymbol{y}_i^* \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_i^s \\ \boldsymbol{X}_i^* \end{bmatrix} \boldsymbol{\beta} + b_i \begin{bmatrix} \mathbf{1}_i^s \\ \mathbf{1}_i^* \end{bmatrix} + \begin{bmatrix} \boldsymbol{e}_i^s \\ \boldsymbol{e}_i^* \end{bmatrix}. \tag{1.49}$$

Under model (1.49), suppose we are interested in estimating the small area means $\overline{y}_i$. Rewrite $\overline{y}_i$ as

$$\overline{y}_i = f_i \overline{y}_i^s + (1 - f_i)\overline{y}_i^*, \tag{1.50}$$

where $f_i = n_i N_i^{-1}$, $\overline{y}_i^s$ and $\overline{y}_i^*$ are the means for sampled and non-sampled elements respectively. Therefore, the estimation of $\overline{y}_i$ is equivalent to prediction of $\overline{y}_i^*$ given the sampled data $\boldsymbol{y}_i^s$ and auxiliary information $\boldsymbol{X}_i$. However, the problem is not posed this way in most of the literature because the sample fractions f_i's are usually negligible. The estimated means of interest are defined to be

$$\overline{y}_i = \overline{\boldsymbol{X}}_i \boldsymbol{\beta} + b_i. \tag{1.51}$$

Parallel to the random effects model, there are three major estimation approaches to the nested error model. There are also many extensions of model (1.49) in the literature. We will limit ourselves to model (1.49) and present approaches for point estimation of $\overline{y}_i$ and measurement of uncertainty associated with the estimators.

First, we obtain the BLUP of $\overline{\boldsymbol{X}}_i \boldsymbol{\beta} + b_i$ based on sampled elements for the nested-error model (1.49). For simplification, we assume that $k_{ij} = 1$ for all i and j. Thus, $V(\epsilon_{ij}) = \sigma_\epsilon^2$. Note that $\boldsymbol{V} = diag(\boldsymbol{V}_1, \ldots, \boldsymbol{V}_m)$ with $\boldsymbol{V}_i = \sigma_e^2 \boldsymbol{I}_{n_i} + z_i^2 \sigma_b^2 \mathbf{1}_{n_i} \mathbf{1}_{n_i}^T$, $\boldsymbol{G} = diag(\sigma_b^2, \ldots, \sigma_b^2)$ for model (1.49) and $\boldsymbol{Z} = diag(\mathbf{1}_{n_1}, \ldots, \mathbf{1}_{n_m})_{(\sum n_i) \times m}$. We use the fact that $\boldsymbol{V}_i^{-1} = (\sigma_e^2)^{-1} \boldsymbol{I}_{n_i} - \gamma_i (n_i \sigma_e^2)^{-1} \mathbf{1}_{n_i} \mathbf{1}_{n_i}^T$, where

$$\gamma_i = z_i^2 \sigma_b^2 (z_i^2 \sigma_b^2 + n_i^{-1} \sigma_e^2)^{-1}. \tag{1.52}$$

Substituting $\boldsymbol{V}^{-1}$ into (1.6), we get $\widehat{\boldsymbol{\beta}}$. Taking $\boldsymbol{k} = \overline{\boldsymbol{X}}_i$ and $\boldsymbol{m} = (0, \ldots, 0, 1, 0, \ldots, 0)^T$ with 1 in the i-th position, we get the BLUP of $\overline{\boldsymbol{X}}_i^T \boldsymbol{\beta} + b_i$ by using Henderson's general result (1.7):

$$\widetilde{(\overline{y}_i)}^H = \overline{\boldsymbol{X}}_i\widehat{\boldsymbol{\beta}} + \gamma_i[\overline{y}_i^s - \overline{\boldsymbol{X}}_i^s\widehat{\boldsymbol{\beta}}]. \tag{1.53}$$

where $\overline{\boldsymbol{X}}_i^s$ is the mean of $\boldsymbol{x}_{ij}$ for the sampled elements.

Rewrite (1.53) as

$$\widetilde{(\overline{y}_i)}^H = \gamma_i[\overline{y}_i^s + (\overline{\boldsymbol{X}}_i - \overline{\boldsymbol{X}}_i^s)\widehat{\boldsymbol{\beta}}] + (1-\gamma_i)\overline{\boldsymbol{X}}_i\widehat{\boldsymbol{\beta}}. \tag{1.54}$$

Therefore, $\widetilde{(\overline{y}_i)}^H$ is a weighted average of the synthetic estimator $\overline{\boldsymbol{X}}_i\widehat{\boldsymbol{\beta}}$ and the "survey regression" estimator $\overline{y}_i^s + (\overline{\boldsymbol{X}}_i - \overline{\boldsymbol{X}}_i^s)\widehat{\boldsymbol{\beta}}$.

The EBLUP $\widehat{(\overline{y}_i)}^H$ is obtained by replacing σ_b^2 and σ_e^2 with consistent estimators $\widehat{\sigma}_b^2$ and $\widehat{\sigma}_e^2$ in (1.54). Similar to (1.22) for estimator (1.16), the MSE for $\widehat{(\overline{y}_i)}^H$ is

$$MSE\left\{\widehat{(\overline{y}_i)}^H - y_i\right\} = g_{1i} + g_{2i} + 2g_{3i} + o(m^{-1}), \tag{1.55}$$

where

$$g_{1i} = \gamma_i\sigma_e^2 n_i^{-1}, \tag{1.56}$$

$$g_{2i} = (\overline{\boldsymbol{X}}_i - \gamma_i\overline{\boldsymbol{X}}_i^s)^T(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}(\overline{\boldsymbol{X}}_i - \gamma_i\overline{\boldsymbol{X}}_i^s), \tag{1.57}$$

and

$$g_{3i} = n_i^{-2}z_i^4(z_i^2\sigma_b^2 + n_i^{-1}\sigma_e^2)^{-3}Var(\widehat{\sigma}_e^2\sigma_b^2 - \widehat{\sigma}_b^2\sigma_e^2). \tag{1.58}$$

For Bayes approaches, we assume that b_i and e_{ij} are normal errors for all i and j. The empirical Bayes estimator $\widehat{(\overline{y}_i)}^{EB}$ for model (1.49) is the same as the EBLUP estimator $\widehat{(\overline{y}_i)}^H$ though the naive variance estimator for $\widehat{(\overline{y}_i)}^{EB}$ may severely underestimate the true posterior variance $Var(\overline{y}_i|\boldsymbol{Y})$.

When both σ_b^2 and σ_e^2 are known, the hierarchical Bayes approach and the BLUP approach lead to the same estimator. When σ_b^2 and σ_e^2 are unknown, priors are assumed for σ_b^2 and σ_e^2 in the hierarchical Bayes approach. Datta and Ghosh (1991) derive the closed form for $E(\overline{y}_i|\boldsymbol{Y},\lambda)$, $Var(\overline{y}_i|\boldsymbol{Y},\lambda)$, and $f(\lambda|\boldsymbol{Y})$, where $\lambda = \sigma_e^2\sigma_b^{-2}$. Therefore, the posterior mean and variance of $\overline{y}_i$ given $\boldsymbol{Y}$ is available. Datta and Ghosh (1991) also compare EBLUP, empirical Bayes, and hierarchical Bayes estimators and the corresponding MSE estimators by using the data from Battese, Harter and Fuller (1988).

1.4 Some extensions and recent development

1.4.1 Multi-level model

Holt and Moura (1993) extended model (1.49) to a multi-level model with random regression coefficients. Let $\boldsymbol{Y}^s = vec(Y_{ij}), i = 1, \dots, m; j = 1, \dots, n_i$ be the sampled data. The model is specified as

$$\boldsymbol{\beta}_i = \boldsymbol{Z}_i^T \boldsymbol{\zeta} + \boldsymbol{b}_i, \tag{1.59}$$

where $\boldsymbol{b}_i \sim N_p(\mathbf{0}, \boldsymbol{\Phi})$ for $i = 1, \dots, m$ with $\boldsymbol{\Phi}$ unknown and

$$Y_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_i + e_{ij} \quad i = 1, \dots, m; j = 1, \dots, n_i, \tag{1.60}$$

where e_{ij} are independent identical distributed $N(0, \sigma_e^2)$ and $\boldsymbol{b}_i$ and e_{ij} are independent for all i and j. We are interested in finding the estimates for $\boldsymbol{\beta}_i$ given the sampled data $\boldsymbol{Y}^s$. We are also interested in estimating the small area means $\overline{\boldsymbol{X}}_i^T \boldsymbol{\beta}_i$ where $\overline{\boldsymbol{X}}_i^T$ is the $p \times 1$ vector of population means of the auxiliary variables for the i-th small area.

Note that model (1.59) and (1.60) incorporate both the unit level auxiliary information $\boldsymbol{x}_{ij}$ and the area-level auxillary information $\boldsymbol{Z}_i$ into a single model. Holt and Moura (1993) obtain the EBLUP of $\overline{\boldsymbol{X}}_i^T \boldsymbol{\beta}_i$ and a second-order approximation to the MSE of the EBLUP for the multi-level model (1.59) and (1.60). You and Rao (1999) study the multi-level model in a hierarchical Bayes framework and extend the model to more general multi-level models with unequal error variances.

1.4.2 Models with discrete measurements

When the measurements Y_{ij} are categorical or discrete and the small area quantities of interest are proportions or counts, the mixed linear models considered in Section 1.2

and Section 1.3 no longer apply. An example is the model proposed by Jiang and Lahiri (1999):

$$Pr(Y_{ij} = 1|p_{ij}) = p_{ij} \quad Pr(Y_{ij} = 0|p_{ij}) = 1 - p_{ij}. \tag{1.61}$$

and

$$logit(p_{ij}) = \log(p_{ij}(1 - p_{ij})^{-1}) = \boldsymbol{X}_{ij}\boldsymbol{\beta} + b_i, \tag{1.62}$$

where $b_i \sim N(0, \sigma^2)$.

Let $\boldsymbol{Y} = (y_{ij})_{1\le i\le m, 1\le j\le n_i}$, $\boldsymbol{Y}_i = (y_{ij})_{1\le j\le n_i}$, and $Y_{i.} = \sum_{j=1}^{n_i} Y_{ij}$. Let $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0^T, \sigma_0)^T$ be the vector of true parameters. When $\boldsymbol{\theta}_0$ is known, the best predictor (BP) of b_i, in terms of MSE, is given by

$$\begin{aligned} E(b_i|\boldsymbol{Y}) &= E(b_i|\boldsymbol{Y}_i) \\ &= \sigma_0 \left\{ \int \exp[\phi_i - (2\sigma_0^2)^{-1}u_i^2]du_i \right\}^{-1} \left\{ \int u_i \exp[\phi_i - (2\sigma_0^2)^{-1}u_i^2]du_i \right\} \\ &\equiv \psi(Y_{i.}, \boldsymbol{\theta}_0), \end{aligned} \tag{1.63}$$

where $\phi_i = Y_{i.}u_i - \sum_{j=1}^{n_i} log[1 + \exp(\boldsymbol{X}_{ij}\boldsymbol{\beta} + u_i)]$ and $u_i \sim N(0, \sigma_0^2)$. Therefore, the best predictor for $logit(p_{ij})$ is $\boldsymbol{X}_{ij}\boldsymbol{\beta}_0 + E(b_i|\boldsymbol{Y}_i)$.

An empirical best predictor (EBP) of b_i is obtained by replacing the unknown parameter $\boldsymbol{\theta}_0$ by a consistent estimator $\widehat{\boldsymbol{\theta}}$, i.e.

$$\widehat{b}_i = \psi(Y_{i.}, \widehat{\boldsymbol{\theta}}). \tag{1.64}$$

Jiang and Lahiri (1999) obtain an approximation to the MSE of the EBP with bias of order $o(m^{-1})$. They also discus the estimation of functions of fixed and random effects for this model.

1.4.3 Time series model

Rao and Yu (1992, 1994) proposed an extension of the random effects model to time series and cross-sectional data. The model is of the form

$$Y_{it} = y_{it} + e_{it} \quad i = 1, \ldots, m; t = 1, \ldots, T, \tag{1.65}$$

where e_{it} is sampling error. Let $\boldsymbol{e}_i = (e_{i1}, \ldots, e_{iT})'$. We assume that $\boldsymbol{e}_i \sim N_T(\mathbf{0}, \boldsymbol{\Phi}_i)$.

$$y_{it} = \boldsymbol{x}_{it}^T \boldsymbol{\beta} + v_i + u_{it}, \tag{1.66}$$

where $v_i \overset{i.i.d}{\sim} N(0, \sigma_v^2)$, and u_{it}'s are assumed to follow a common first order autoregressive process ($AR(1)$) for each i, i.e.,

$$u_{it} = \rho u_{i,t-1} + \epsilon_{it}, \quad |\rho| < 1, \tag{1.67}$$

where $\epsilon_{it} \overset{i.i.d}{\sim} N(0, \sigma^2)$.

Model (1.66) and (1.67) are used extensivly in the econometric literature. Rao and Yu (1992, 1994) obtain the EBLUP and hierarchical Bayes estimators and their standard errors under (1.65) and (1.66) and an $AR(1)$ process u_{it}. More complex models on the u_{it} can be formulated by assuming an autoregressive moving average (ARMA) process. But the resulting efficiency gains relative to use $AR(1)$ process are unlikely to be significant. There are some alternative time series models for small area estimation. See e.g., Pfeffermann and Burck (1990) and Ghosh et al (1996).

1.5 Dissertation organization

This dissertation is organized as follows. In Chapter 2, a practical application of small area estimation in the National Resources Inventory (NRI) is presented. Several issues raised by this application are discussed in the following chapters. Chapter 3 derives an estimator for the MSE of the EBLUP when the small area sampling variances are directly estimated. In Chapter 4, small area estimation under a restriction is discussed. The bias of the EBLUP are assessed in Chapter 5 when the sampling errors are not normally distrbuted. Finally, the summarization of our results and the comparisons to other results in the literature are presented in Chapter 6.

2 AN EXAMPLE OF SMALL AREA ESTIMATION IN THE NATIONAL RESOURCES INVENTORY

2.1 The U.S. National Resources Inventory

The Iowa State Statistical Laboratory cooperates with the U.S. Natural Resources Conservation Service on a large survey of land use in the United States. The survey is a panel survey and was conducted in 1982, 1987, 1992, and 1997.

The survey collects data on soil characteristics, land use, land cover, wind erosion, water erosion, and conservation practices. The data are collected by employees of the Natural Resources Conservation Service. Iowa State University has responsibility for sample design and for estimation. See Nusser and Goebel (1997) for a complete description of the survey.

The sample is a stratified sample of all states and Puerto Rico. The sampling units are areas of land called *segments*. The segments vary in size, from 40 acres to 640 acres. Data are collected for the entire segment on items such as urban land and water area. Detailed data on soil properties and land use are collected at a random sample of points within the segment. Generally, there are three points per segment, but 40-acre segments contain two points and the samples in two states contain one point per segment. Some data, such as total land area, federally-owned land and area in large water bodies, are collected on a census basis external to the sample survey. The current sample contains about 300,000 segments and about 800,000 points.

The sample size is such that direct estimates have acceptable variances for subdivisions of the surface area called *hydrologic units*. Hydrologic units are, essentially, drainage areas for major streams. There are about 200 hydrologic units in the United States. The estimation procedure is designed to reproduce the correct acreage for counties where counties are important political subdivisions. There are about 3,100 counties in the United States. Because the sample must provide consistent acreage estimates for both counties and hydrologic units, the basic tabulation unit is the portion of a hydrologic unit within a county. This unit is called a *HUCCO*. There are about 5,000 HUCCOs. Some HUCCOs are relatively small and may contain only one segment. Some are relatively large and contain more than 100 segments.

2.2 Small area estimation

In the National Resources Inventory, small area estimation is used in the estimation of area in roads and in the estimation of change in acres for urban and built-up areas. Urban land is divided into two categories on the basis of the size of the tract. We present the analysis for the change in the sum of the two categories of urban land acreage from 1992 to 1997.

2.2.1 Small area estimation model for urban change

The urban area was determined for each year of the NRI survey. Let $U92_{kl}$ and $U97_{kl}$ denote urban area in HUCCO l of county k in 1992 and in 1997 respectively. The quantity used in small area estimation for urban change is $D_{kl} = U97_{kl} - U92_{kl}$, the direct estimated change from 1992 to 1997. The auxiliary information is the 1992 population and 1997 population. Population data are only available for counties, not for HUCCOs. Therefore, we defined two variables

$$
\begin{aligned}
z_{kl,1} &= (U92_k)^{-1} U92_{kl} P92_k \\
z_{kl,2} &= (U92_k)^{-1} U92_{kl} (P97_k - P92_k),
\end{aligned}
$$

where $U92_k$ is the urban acres in county k in 1992, $P92_k$ and $P97_k$ are the populations of county k in 1992 and in 1997, respectively. We expect both variables to be positively related to the change in urban acres. Because a reduction in urban area is extremely rare, we set the population change variable to a small positive number if the population change is negative.

Some heavily urbanized areas have a large population and very little area in non-urban uses that can be converted to urban use. Therefore the actual regression variables used in the analysis were constructed to recognize the availability of potentially convertible land. The variables are

$$
\begin{aligned}
z_{kl,1} &= min(R_1\tilde{z}_{kl,1}, 0.5C_{kl}) \\
z_{kl,2} &= min(R_2\tilde{z}_{kl,2}, 0.5C_{kl}),
\end{aligned}
$$

where R_j is the ratio of the total change in urban for the state to the sum of $\tilde{z}_{kl,i}$ for the state, and C_{kl} is the total area available for conversion to urban use in HUCCO kl.

In the sequel, we use the single subscript i in place of the double subscript kl as the index for the HUCCO. Our goal is to predict d_i, the unobservable true value of change in urban area. The distribution of D_i is highly skewed. Also, the empirical results presented later show that the sampling variance $V(D_i)$ is approximately proportional to $[E(D_i)]^{1.25}$. This relationship is also shown in Figure 2.2. Therefore, a power transformation $Y_i = D_i^{0.375}$ is used in the estimation to stablize the variance components estimation. By choosing the power index equal to 0.375, we will show later that $V(Y_i)$ is not a function of $E(Y_i)$ anymore. A model for small area estimation is

$$
\begin{aligned}
y_i &= x_{1i}\beta_1 + x_{2i}\beta_2 + n_i^{0.375}b_i & (2.1)\\
Y_i &= y_i + e_i, & (2.2)
\end{aligned}
$$

where $y_i = d_i^{0.375}, x_{1i} = z_{1i}^{0.375}, x_{2i} = z_{2i}^{0.375}$, n_i is the number of sample points in the i-th HUCCO, b_i is the area effect, e_i is the sampling error, and $(b_i, e_i), i = 1, 2, \ldots, m$, are independent vectors with a normal distribution.

$$\begin{pmatrix} b_i \\ \epsilon_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_b^2 & 0 \\ 0 & \sigma_{ei}^2 \end{pmatrix} \right). \tag{2.3}$$

We designed the small area estimation procedure to be fully automated because estimates were to be constructed for about 5,000 small areas using more than fifty analysis units, where the typical analysis unit is a state. Therefore, we used relatively simple procedures in place of complicated procedures that might produce marginal gains in efficiency. One could design an iterative estimation procedure for β_1 and β_2 in which the estimated between-area component of variance is used to estimate the covariance matrix of $u_i = n_i^{0.375} b_i$. We used a simple weighted least squares procedure where the weights were a function of n_i. The estimator of $(\beta_1, \beta_2)'$ is

$$\begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{pmatrix} = \mathbf{H}^{-1} \begin{pmatrix} \sum_i n_i^{-0.25} x_{1i} Y_i \\ \sum_i n_i^{-0.25} x_{2i} Y_i \end{pmatrix}, \tag{2.4}$$

where

$$\mathbf{H} = \begin{pmatrix} \sum_i n_i^{-0.25} x_{1i}^2 & \sum_i n_i^{-0.25} x_{1i} x_{2i} \\ \sum_i n_i^{-0.25} x_{1i} x_{2i} & \sum_i n_i^{-0.25} x_{2i}^2 \end{pmatrix}.$$

Because the variance of Y_i is closely related to $n_i^{-0.25}$, the procedure is close to estimated generalized least squares. Our model estimator of y_i is

$$\widehat{y}_i = x_{1i}\widehat{\beta}_1 + x_{2i}\widehat{\beta}_2. \tag{2.5}$$

Let $V(D_i) = V(Y_i^\eta | d_i^\eta) = V(\epsilon_i^\eta)$, where $\eta = 0.375^{-1}$ is the within HUCCO sample error variance. The sample variance $V(D_i)$ can be estimated directly from the sample data. However, the sample size is relatively small in some HUCCOs and, hence, the variance of the estimated variance is large. Therefore, a model was developed for $V(D_i)$ to provide an improved sample variance estimator for small HUCCOs.

If we had a simple random sample of points with a zero-one indicator for urban change, the sample variance of D_i would be

$$V(D_i) = V(\epsilon_i^{\eta}) = V(N\widehat{p}) = N^2 n_i^{-1} PQ = N n_i^{-1}(NPQ) \doteq N n_i^{-1}(NP) \ . \tag{2.6}$$

where N is the population number of sample units, P is the proportion of the units that change into urban, and $\widehat{p}$ is the corresponding sample proportion. In our notation, $D_i = N\widehat{p}$. Now $n_i^{-1}N$ is constant under proportional sampling. Thus, $V(D_i)$ should be proportional to $E(D_i) = NP$. If P were nearly constant, then $n_i^{-1}E(D_i) = n_i^{-1}NP$ would be nearly constant too. This means that $E(D_i)$ would be almost proportional to n_i. Therefore, it seems reasonable to approximate the variance of D_i with a function of both n_i and $E(D_i)$. The empirical results show that

$$V(D_i) \doteq C_1 \, n_i^{-0.25} [E(D_i)]^{1.25} \, . \tag{2.7}$$

where C_1 is a constant to be determined, is a good approximation to the sample variance $V(D_i)$. A plot of the direct variance estimator against the model variance estimator in (2.7) is also shown in Figure 2.2 of the next section. By Taylor approximation, the variance of e_i is

$$V(e_i) = V(D_i^{\lambda}) = [E(D_i)]^{2(\lambda-1)} \lambda^2 V(D_i) \, , \tag{2.8}$$

where $\lambda = \eta^{-1}$. For $\lambda = 0.375$, we have

$$V(\epsilon_i) \doteq \sigma_u^2 n_i^{-0.25} . \tag{2.9}$$

where σ_u^2 is a constant to be determined. Therefore, $V(\epsilon_i)$ does not depend on $E(Y_i)$.

Let $\widehat{V}(D_i)$ be the directly estimated variance of D_i from the sample data. The direct estimated variance of e_i is

$$\widehat{V}(\epsilon_i) = 0.96 (D_i^{*})^{2(\lambda-1)} \lambda^2 \widehat{V}(D_i) \tag{2.10}$$

where 0.96 is an empirical adjustment and D_i^* is an estimator of $E\{D_i\}$. An estimator of the within-area component of variance is

$$\widehat{\sigma}_w^2 = \left(\sum_i \delta_i\right)^{-1} \sum_i n_i^{0.25} \widehat{V}\{e_i\}\, \delta_i\,. \tag{2.11}$$

where

$$\delta_i = \begin{cases} 1 & \text{if } n_i > 2 \\ 0 & \text{otherwise.} \end{cases}$$

An estimator of the between-area component of variance is

$$\widehat{\sigma}_b^2 = \left(\sum_i n_i \delta_i\right)^{-1} \sum_i \ddot{\sigma}_{b,i}^2 n_i \delta_i. \tag{2.12}$$

where $\ddot{\sigma}_{b,i}^2 = (n_i^{-0.375})^2 \left[(\widehat{Y}_i - Y_i)^2 - \widehat{V}(e_i)\right]$.

The predictor of y_i for the i-th HUCCO is

$$\widetilde{y}_i = \widehat{y}_i + \widehat{\gamma}_i\,(Y_i - \widehat{y}_i) = \widehat{\gamma}_i Y_i + (1 - \widehat{\gamma}_i)\,\widehat{y}_i\,, \tag{2.13}$$

where

$$\widehat{\gamma}_i = \left(n_i^{0.75}\widehat{\sigma}_b^2 + n_i^{-0.25}\widehat{\sigma}_w^2\right)^{-1} n_i^{0.75}\widehat{\sigma}_b^2 = \left(\widehat{\sigma}_b^2 + n_i^{-1}\widehat{\sigma}_w^2\right)^{-1} \widehat{\sigma}_b^2\,. \tag{2.14}$$

Under the model, the error in $\widetilde{y}_i$ as an estimator of y_i is

$$\widetilde{y}_i - y_i = \widehat{\gamma}\,(u_i + e_i) - u_i + (1 - \widehat{\gamma}_i)\left[x_{1i}(\widehat{\beta}_1 - \beta_1) + x_{2i}(\widehat{\beta}_2 - \beta_2)\right]. \tag{2.15}$$

where $u_i = n_i^{0.375} b_i$. If $\widehat{\gamma}_i$ is treated as a fixed quantity and the possible covariance between (e_i, u_i) and $(\widehat{\beta}_1, \widehat{\beta}_2)$ is ignored, then

$$\widehat{V}\{\widetilde{y}_i - y_i\} = n_i^{0.75}\left[(1 - \widehat{\gamma}_i)^2\widehat{\sigma}_b^2 + \widehat{\gamma}_i^2 n_i^{-1}\widehat{\sigma}_w^2\right] + (1 - \widehat{\gamma}_i)^2(x_{1i}, x_{2i})\widehat{V}\{(\widehat{\beta}_1, \widehat{\beta}_2)\}(x_{1i}, x_{2i})'. \tag{2.16}$$

The estimator (2.16) does not contain a contribution to the variance from estimating the variance components.

The predictor of change in total urban from 1992 to 1997 for HUCCO i is

$$\widetilde{d}_i = \widetilde{y}_i^{\eta}. \tag{2.17}$$

The corresponding estimated variance of $\widetilde{d}_i$ is

$$\widehat{V}\left\{\widetilde{d}_i - d_i\right\} \doteq (0.96)^{-1}(D_i^{*})^{1.25}(0.375)^{-2}\widehat{V}\left\{\widetilde{y}_i - y_i\right\}. \tag{2.18}$$

For confidence limits of $\widetilde{d}_i$, it is preferable to set limits for $\widetilde{y}_i$ and then exponentiate those limits.

The predictor $\widetilde{d}_i$ performs well overall, but poorly for a few small areas that are not consistent with the small area model. In other words, the predictor $\widetilde{d}_i$ performs poorly for the small areas where the regression model relating Y_i and the auxiliary information $\boldsymbol{x}_i$ does not fit well. To avoid this problem, Efron and Morris (1972) and Fay and Harriot (1979) suggested a compromise estimator which limits the difference between the small area estimator and the direct estimator by some multiple of the standard error of the direct estimator. Using this idea, we used the compromise estimator

$$\widehat{d}_i = \begin{cases} \widetilde{d}_i, & \text{if } -\phi_i \geq \widetilde{d}_i - D_i \leq \phi_i \\ D_i + \phi_i, & \text{if } \widetilde{d}_i - D_i > \phi_i \\ D_i - \phi_i, & \text{if } \widetilde{d}_i - D_i < -\phi_i, \end{cases} \tag{2.19}$$

where $\phi_i = \max\left(2000, \left[\widehat{V}\left(D_i\right)\right]^{0.5}\right)$, as the small area estimator.

The direct survey estimator for urban change acres D_i has a large variance as an estimator of the true urban change acres d_i in HUCCO i. Therefore, the small area estimator $\widehat{d}_i$ is used to predict d_i. However, the direct survey estimator for state urban change acres $\sum_i D_i$ is design unbiased for the true urban state change acres, and it has relatively small variance. Hence, the survey estimator $\sum_i D_i$ is adequate for the state urban change acres. It then becomes desirable to modify the individual HUCCO predictors $\widehat{d}_i$ so that $\sum_i \widehat{d}_i$ is equal to the unbiased survey estimator $\sum_i D_i$. As a rough approach, a ratio adjustment is used to serve this purpose.

The predictor $\widetilde{y}_i$ is unbiased for y_i under normal errors. However, $E(\widetilde{y}_i^{\eta})$ is not necessarily equal to $E(y_i^{\eta}) = E(d_i)$ by Jensen's inequality. The ratio adjustment described before can also partially adjust for this bias. Therefore, we do not make additional adjustments to correct the bias. We will discuss more about non-normal errors and ratio adjustment in Chapter 5.

2.2.2 Results and model checks

We present some plots and statistics for a set of HUCCOs in the state of Illinois. The analyses are for a preliminary data set. There are 163 HUCCOs in the state. The HUCCOs are divided into ten groups on the basis of the model predicted change in acres. The first group is the group of HUCCOs with $(x_{1i}, x_{2i}) = (0, 0)$. This is the group of HUCCOs with zero estimated change. There are either 15 or 16 HUCCOs in each of the remaining nine groups.

Table 2.1 contains summary statistics for the grouped data. While the groups were ordered on the predicted change, the number of segments is generally larger for HUCCOs

Table 2.1 Summary statistics for mean model

	HUCCOs in group	Mean sample size	Group mean		t-stat
			Y_i	$\widehat{y}_i$	
1	21	10.2	0.9	0.0	1.36
2	15	20.3	5.4	4.0	1.25
3	16	41.2	7.6	6.1	1.37
4	16	65.8	5.3	7.8	-2.14
5	16	74.1	8.2	9.1	-0.65
6	16	73.3	10.1	10.3	-0.14
7	16	78.2	10.2	11.5	-1.22
8	16	64.4	14.5	13.7	0.53
9	16	75.9	18.1	17.2	0.43
10	15	109.9	30.1	30.2	0.10

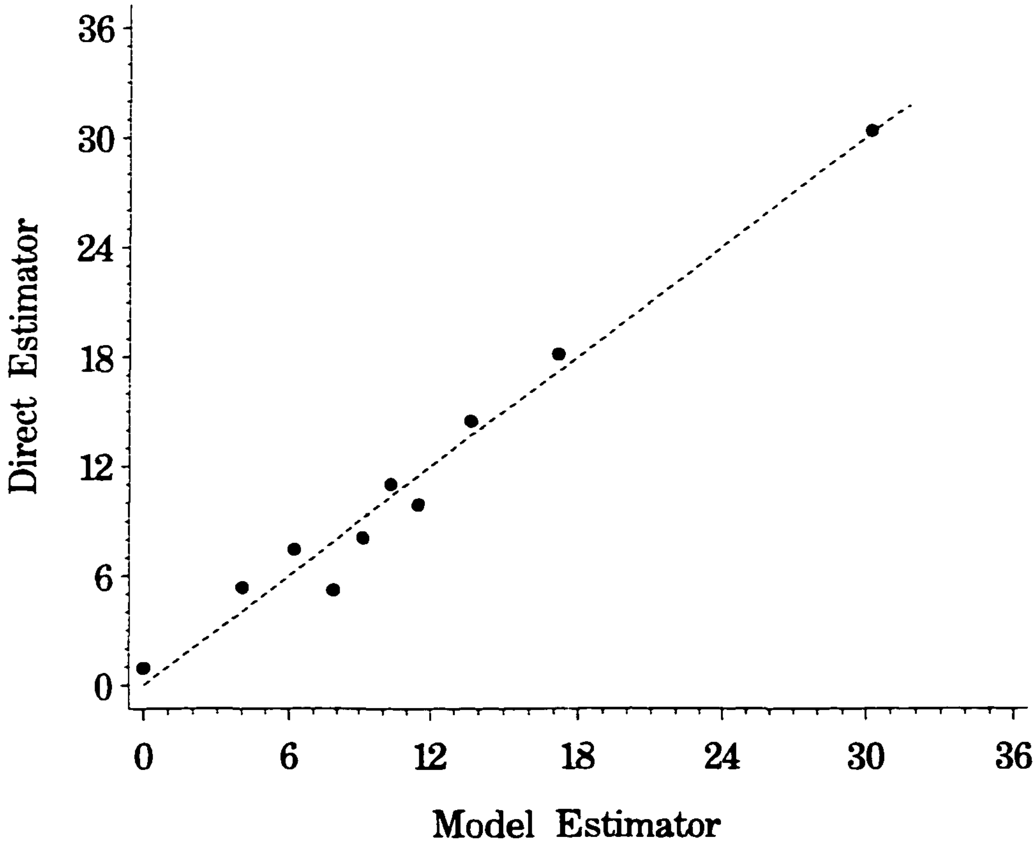

Figure 2.1 Mean Model

with larger changes. The fact that the mean model (2.1) fits quite well is demonstrated by agreement between the two columns Y_i and $\widehat{y}_i$, where $\widehat{y}_i$ is defined in (2.5). A plot of the two variables is shown in Figure 2.1.

The t-statistics in the column "t-stat" in Table 2.1 were calculated as the difference between the group means divided by the standard error of the difference. The standard error was calculated using the mean square of the individual differences. The sum of squares of the ten t-statistics is 12.27. This value can be compared to the 5% tabular value of 18.31 for the chi-square distribution with 10 degrees of freedom. The model is easily accepted on the basis of this test. The R^2 computed for the regression of Y_i on (x_{1i}, x_{2i}) is 0.79.

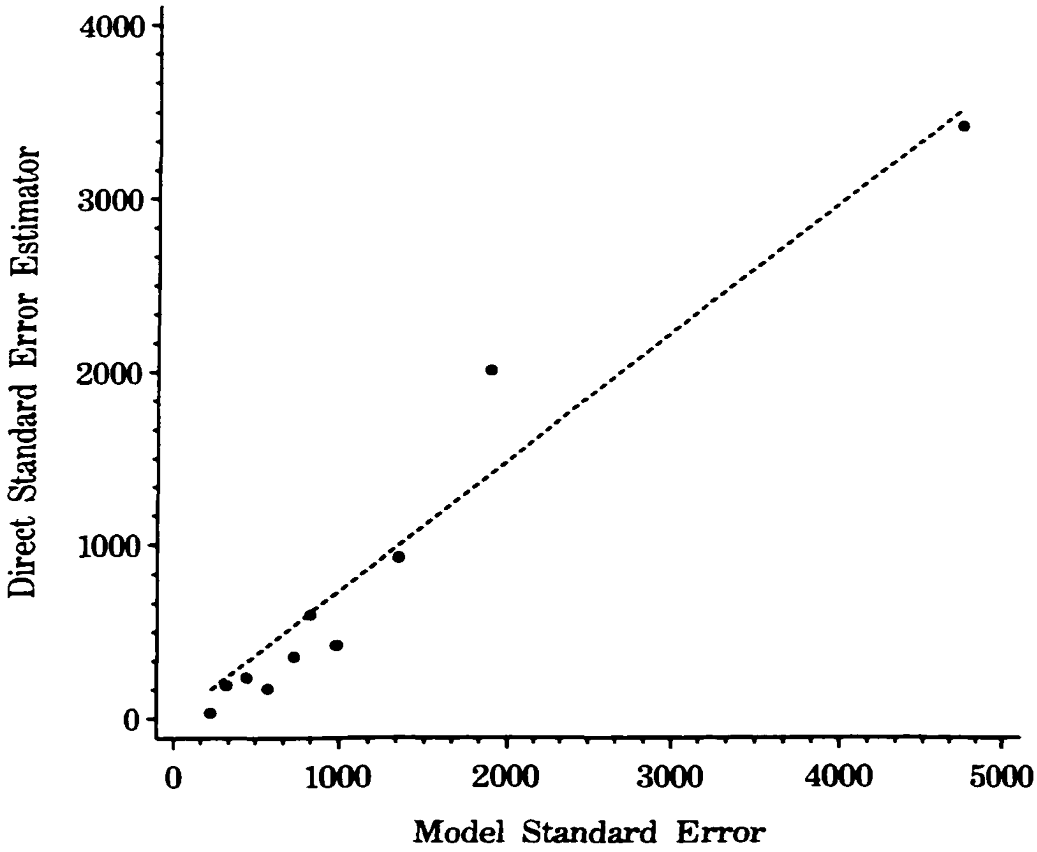

Figure 2.2 Variance Model

Figure 2.2 is the plot of grouped means of $[\widehat{V}(D_i]^{0.5}$ against the grouped means of $[(D_i^*)^{1.25}n_i^{-0.25}]^{0.5}$. The plot indicates that (2.7) is a reasonable approximation to the sample variance $V(D_i) = \widehat{V}(e_i^n)$.

Table 2.2 contains statistics for sample standard errors using the groups of Table 2.1. The column headed $[\widehat{V}(e_i)]^{0.5}$ contains the group averages of the square root of $\widehat{V}(e_i)$, where $\widehat{V}(e_i)$ is defined in (2.10). The averages of $(n_i^{-0.25}\widehat{\sigma}_w^2)^{0.5}$, which is our modeled estimator of $V(e_i)$ is put in the adjacent column. The $(n_i^{-0.25}\widehat{\sigma}_w^2)^{0.5}$ decrease as one moves down the column because the average sample size increases. The entries in the column for $(n_i^{-0.25}\widehat{\sigma}_w^2)^{0.5}$ were ratio adjusted so that the sum is equal to the sum of $[\widehat{V}(e_i)]^{0.5}$.

The t-statistics in the column "t-stat" in Table 2.2 were calculated as the difference

Table 2.2 Summary statistics for variance model

Group	HUCCOs in group	Mean sample size	Group mean $[\widehat{V}(e_i)]^{0.5}$	Group mean $(n_i^{-0.25}\widehat{\sigma}_w^2)^{0.5}$	t-stat	Mean of $[\widehat{V}\{\widetilde{y}_i - y_i\}]^{0.5}$	Efficiency
1	21	10.2	1.16	4.02	-	0.50	-
2	15	20.3	3.79	3.76	0.02	0.84	4.48
3	16	41.2	3.66	3.35	0.37	1.38	2.42
4	16	65.8	1.76	3.16	-2.04	1.79	1.77
5	16	74.1	2.98	3.10	-0.14	1.94	1.60
6	16	73.3	3.63	3.14	0.33	1.93	1.62
7	16	78.2	2.59	3.04	-0.86	2.06	1.48
8	16	64.4	3.86	3.14	0.90	1.83	1.72
9	16	75.9	5.79	3.07	1.34	2.06	1.49
10	15	109.9	3.89	2.92	2.38	2.45	1.19

between the means of $[\widehat{V}(\epsilon_i)]^{0.5}$ and means of $(n_i^{-0.25}\widehat{\sigma}_w^2)^{0.5}$ divided by the standard error of the difference. There are many direct estimates of zero in the first group. so that a comparison of $[\widehat{V}(\epsilon_i)]^{0.5}$ and $(n_i^{-0.25}\widehat{\sigma}_w^2)^{0.5}$ for that group says little about model adequacy. The sum of squares for the other nine t-statistics associated with groups two through ten is 13.44. considerably less than the tabular value of 16.92 for chi-square with nine degrees of freedom. Thus. there is no reason to reject the variance model.

The root transformation has a strong variance stabilizing effect. The standard errors of the original estimates of change for the largest group are about 3386 and the corresponding standard errors for the second group are about 110.

In this particular state. the estimator of the between-HUCCO component of variance was negative. In the production version of the program we impose a lower bound of 0.008 on the ratio of $\widehat{\sigma}_b^2$ to $\widehat{\sigma}_w^2$ in the computation of $\widehat{\gamma}$. The lower bound means that the direct estimator in a HUCCO with 125 segments receives a weight of 0.5.

Because the estimated value for $\widehat{\sigma}_b^2$ is zero. the estimated variance of the prediction error for $\widetilde{y}_i$ was computed as

$$\begin{aligned}\widehat{V}\{\widetilde{y}_i - y_i\} &= \widehat{\gamma}_i^2 n_i^{-0.25}\widehat{\sigma}_w^2 \\ &+ (1-\widehat{\gamma}_i)^2(x_{1i}, x_{2i})\widehat{V}\{(\widehat{\beta}_1, \widehat{\beta}_2)\}(x_{1i}, x_{2i})'. \qquad (2.20)\end{aligned}$$

The seventh column of Table 2.2 contains the mean of the prediction standard errors. where the standard error is the square root of the model variance of the predictor computed with equation (2.20). The last column of Table 2.2 is the ratio of the fifth column to the seventh column. There are large estimated gains in efficiency from using the small area model. particularly for HUCCOs with a small number of segments.

3 MSE OF EBLUP WHEN SAMPLING ERROR VARIANCES ARE ESTIMATED

3.1 Introduction

The MSE estimator of $\tilde{y}_i$ in (2.16), which is derived from (1.13), contains a component for the estimation of $\boldsymbol{\beta}$, but contains no contribution for the variability due to the estimation of the variance components. The refined MSE estimator defined in (1.22) contains a component for the estimation of the between-area and within-area variance components when the variance components depend on some parameter $\boldsymbol{\theta}$. To use (1.22), we need to derive $g_{3i}(\widetilde{\boldsymbol{\theta}})$. We can not simply use $g_{3i}(\widetilde{\boldsymbol{\theta}})$ defined in (1.20) because (1.20) is true only when $\boldsymbol{\theta} = \sigma_b^2$. In the example of Chapter 2, we model the sampling error variance σ_{ei}^2 (within-area variance components) in (2.9) as a function of the parameter σ_w^2. Also, the between-area variance component depends on the parameter σ_b^2. Therefore, the form of $g_{3i}(\widetilde{\boldsymbol{\theta}})$ is more complex than (1.20) because $\boldsymbol{\theta} = (\sigma_b^2, \sigma_w^2)$. We can use (1.19) to derive $g_{3i}(\widetilde{\boldsymbol{\theta}})$ for the example of Chapter 2.

Instead of assuming that the variance components depend on some parameter $\boldsymbol{\theta}$, Arora, Lahiri and Mukherjee (1997) modeled σ_{ei}^2 in a different way. For the nested error model, they assume that the σ_{ei}^2 are independent identically distributed with distribution $G(\xi, \nu)$, where $G(\xi, \nu)$ represents a gamma distribution with mean ξ and variance ν. They derived the empirical Bayes estimator and assessed the asymptotic properties of the estimator under the gamma model.

In many practical applications, it is difficult to find a good model for the sampling

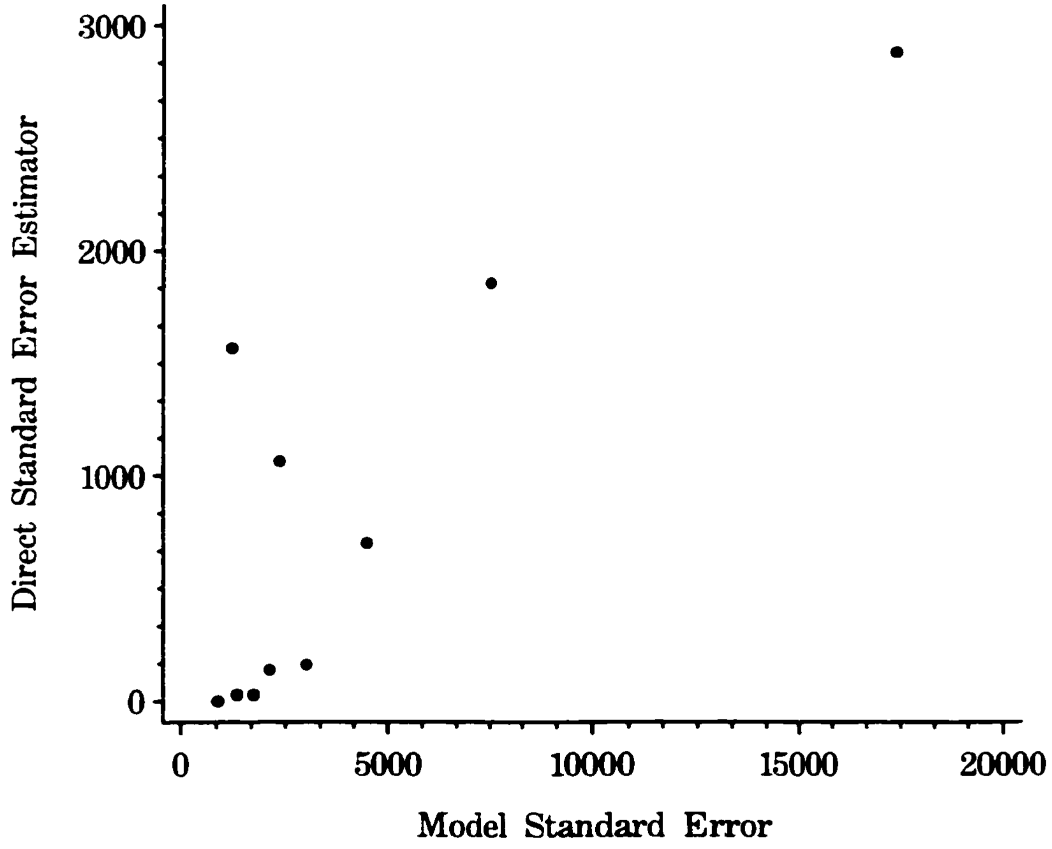

Figure 3.1 Variance Model for Colorado (Grouped)

error variances σ^2_{ei}. Considering the sampling variance model for urban change described in Chapter 2, there are some states where the variance model does not work well. One such example is the state of Colorado. We group the data in the same way as in Chapter 2: the first group is the group of HUCCOs with zero estimated change, other HUCCOs are divided evenly into nine groups. Then we calculate the t-statistics in the same way as in Table 2.2 for the variance model. The sum of squares for the nine t-statistics associated with groups two through ten is 404.87 for the state of Colorado. This number is considerably larger than the tabular value of 16.92 for chi-square with nine degrees of freedom. Thus, there is compelling reason to reject the variance model. To get a better idea about how different the direct variance estimator and modeled variance estimator

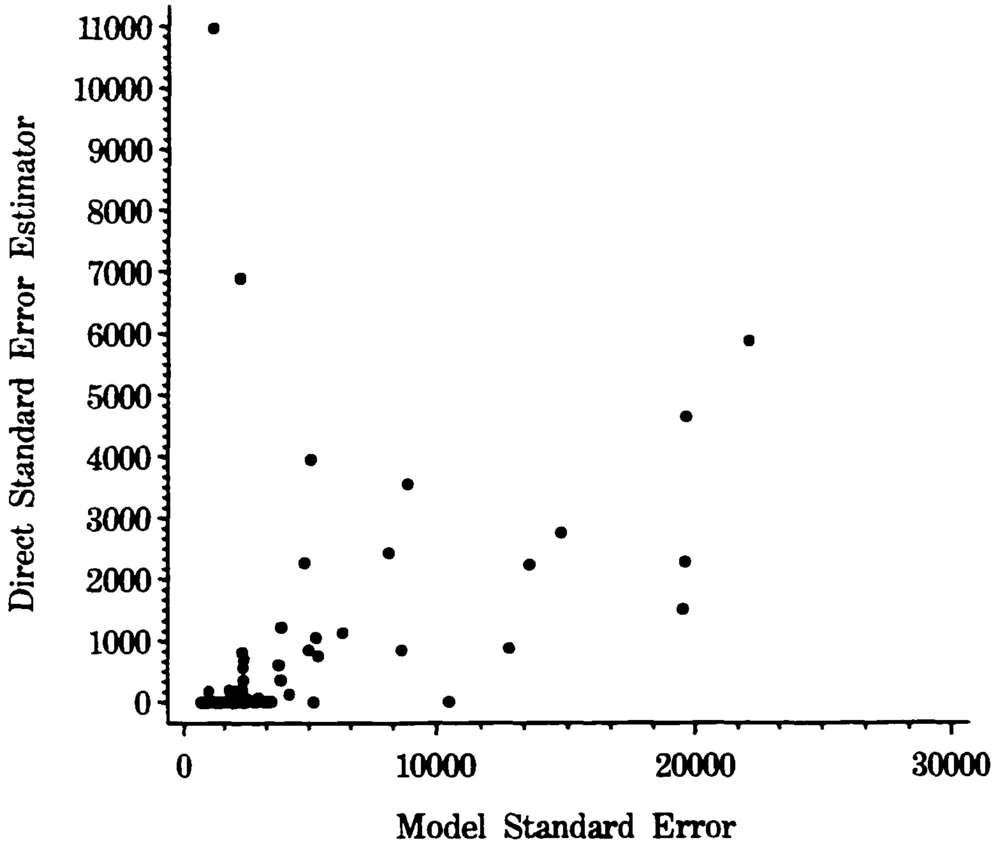

Figure 3.2 Variance Model for Colorado (Individual HUCCO)

are, a plot that corresponds to Figure 2.2 for Colorado is shown in Figure 3.1. The plot on the HUCCO basis is Figure 3.2.

In this chapter, we consider the EBLUP for y_i obtained by replacing σ_{ei}^2 in (2.14) with the individual direct variance estimator $\widehat{\sigma}_{ei}^2$ and assess the impact on the predictors due to estimation of the within-area variances with $\widehat{\sigma}_{ei}^2$. The approach outlined in Prasad and Rao (1990) is not appropriate for this problem because their approach involves proving that

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} = O_p(m^{-0.5})$$

for a fixed dimension parameter $\boldsymbol{\theta}$ and proving

$$E[d(\boldsymbol{\theta})'(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})]^2 = tr[(\boldsymbol{B}'\boldsymbol{V}\boldsymbol{B}E(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})'] + o(m^{-1}), \tag{3.1}$$

where

$$\boldsymbol{B} = [\partial(\boldsymbol{m}'\boldsymbol{G}\boldsymbol{Z}'\boldsymbol{V}^{-1})'/\partial\theta_1, \dots, \partial(\boldsymbol{m}'\boldsymbol{G}\boldsymbol{Z}'\boldsymbol{V}^{-1})'/\partial\theta_p]$$

and $\boldsymbol{m}'\boldsymbol{G}\boldsymbol{Z}'\boldsymbol{V}^{-1}$ is defined in (1.7). If $\widehat{\sigma}^2_{ei}$ is estimated individually, the dimension of the variance component parameter $\boldsymbol{\theta} = (\sigma^2_b, \sigma^2_{e1}, \dots, \sigma^2_{em})^T$ is $m+1$, which is not fixed, and $(\widehat{\sigma}^2_{ei} - \sigma^2_{ei})^2$ is not $O_p(m^{-0.5})$. Therefore, (3.1) does not hold.

We will use a Taylor expansion to obtain an approximate MSE estimator of $\tilde{y}_i$. Before we derive the MSE of the EBLUP estimators, we state two theorems that are frequently used in this chapter.

Theorem 3.1 *Let $\overline{\boldsymbol{x}}_n = (\overline{x}_{1n}, \dots, \overline{x}_{kn})^T$ be the mean of a random sample of n vector random variables selected from a distribution function with mean vector zero and finite Bth moment. Consider the sequence $\{\overline{\boldsymbol{x}}_n\}_{n=1}^{\infty}$, and let $p_1, p_2, \dots, p_k$ be nonnegative integers such that $s = \sum_{i=1}^{k} p_i$. Then*

$$E\{\overline{x}^{p_1}_{1n}\overline{x}^{p_2}_{2n}\dots\overline{x}^{p_k}_{kn}\} = \begin{cases} O(n^{-0.5s}) & \text{if } s \text{ is even}, \\ O(n^{-0.5(s+1)}), & \text{if } s \text{ is odd}. \end{cases}$$

Proof: See page 242 of Fuller (1996). ■

Theorem 3.2 *Let $\{\boldsymbol{X}_n\}$ be a sequence of k-dimensional random variables with corresponding distribution functions $\{\boldsymbol{F}_n(\boldsymbol{x})\}$, and let $\{f_n(\boldsymbol{x})\}$ be a sequence of functions mapping $\Re^k$ into $\Re$. Let $\delta \in (0, \infty)$, and define $\alpha = \delta^{-1}(1+\delta)$. Assume that for some positive integers s and N_0:*

(i) $\int |\boldsymbol{x} - \boldsymbol{\mu}|^{\alpha s} dF_n(\boldsymbol{x}) = r_n^{\alpha s}$, *where* $r_n \to 0$ *as* $n \to \infty$.

(ii) $\int |f_n(\boldsymbol{x})|^{1+\delta} dF_n(\boldsymbol{x}) = O(1)$.

(iii) $f_n^{(i_1,\ldots,i_n)}(\boldsymbol{x})$ *is continuous in* $\boldsymbol{x}$ *over a closed and bounded sphere* S *for all* n *greater than* N_0, *where*

$$f_n^{(i_1,\ldots,i_n)}(\boldsymbol{x}_0) = \frac{\partial^p}{\partial x_{i_1}\ldots\partial x_{i_p}} f_n(\boldsymbol{x})\,|_{\boldsymbol{x}=\boldsymbol{x}_0}\,.$$

(iv) $\boldsymbol{\mu}$ *is an interior point of* S.

(v) There is a finite number K *such that, for* $n > N_0$,

$$|f_n^{(i_1,\ldots,i_n)}(\boldsymbol{x})| \le K \quad \textit{for all } \boldsymbol{x} \in S,$$

$$|f_n^{(i_1,\ldots,i_n)}(\boldsymbol{\mu})| \le K \quad \textit{for } p = 1, 2, \ldots, s-1,$$

and

$$|f_n(\boldsymbol{\mu})| \le K.$$

Then

$$\int f_n(\boldsymbol{x})dF_n(\boldsymbol{x}) = f_n(\boldsymbol{\mu}) + \sum_{j=1}^{s-1}\frac{1}{j!}\int D^j f_n(\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^j dF_n(\boldsymbol{x}) + O(r_n^s),$$

where

$$D^p f_n(\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^p = \sum_{i_1=1}^{k}\sum_{i_2=1}^{k}\cdots\sum_{i_p=1}^{k}\frac{\partial^p}{\partial x_{i_1}\ldots\partial x_{i_p}} f_n(\boldsymbol{x})\,|_{\boldsymbol{x}=\boldsymbol{\mu}}\prod_{j=1}^{p}(x_{i_j}-\mu_{i_j})$$

and, for $s = 1$, *it is understood that*

$$\int f_n(\boldsymbol{x})dF_n(\boldsymbol{x}) = f_n(\boldsymbol{\mu}) + O(r_n).$$

The results also holds if we replace (ii) with the condition that the $f_n(\boldsymbol{x})$ *are uniformly bounded for* n *sufficiently large and assume that (i) and (iii) hold for* $\alpha = 1$.

Proof: See page 245-246 of Fuller (1996). ■

3.2 MSE of EBLUP under alternative states of knowledge

Consider the random effects model

$$Y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + z_i b_i + e_i, \qquad i = 1, \ldots, m,$$

defined in (1.3) with normal errors and unknown sampling error σ_{ei}^2. If we use the individual direct estimators $\widehat{\sigma}_{ei}^2, i = 1, \dots, m$, to obtain the EBLUP $\widehat{y}_i$, we need an approximation for the MSE of $\widehat{y}_i - y_i$. For simplicity, we will consider the case with $z_i = 1, i = 1, \dots, m$, where $Var(b_i) = z_i^2\sigma_b^2$, throughout this section. The same asymptotic results hold for general z_i.

We first consider an ideal case: $\boldsymbol{\beta}$ and σ_b^2 are known. The EBLUP of y_i is

$$\widehat{y}_i^H = \widehat{\gamma}_i Y_i + (1 - \widehat{\gamma}_i)\boldsymbol{x}_i^T\boldsymbol{\beta}, \tag{3.2}$$

where $\widehat{\gamma}_i = (\sigma_b^2 + \widehat{\sigma}_{ei}^2)^{-1}\sigma_b^2$. We wish to derive

$$\begin{aligned} MSE(\widehat{y}_i^H - y_i) &= E\left\{[\widehat{\gamma}_i Y_i + (1 - \widehat{\gamma}_i)\boldsymbol{x}_i^T\boldsymbol{\beta} - y_i]^2\right\} \\ &= E\left\{[(\widehat{\gamma}_i - 1)b_i + \widehat{\gamma}_i e_i]^2\right\} \\ &= E\left\{(1 - \widehat{\gamma}_i)^2 b_i^2\right\} + 2E\left\{(\widehat{\gamma}_i - 1)\widehat{\gamma}_i b_i e_i\right\} + E\left\{\widehat{\gamma}_i^2 e_i\right\} \\ &= \sigma_b^2 E\left\{(1 - \widehat{\gamma}_i)^2\right\} + \sigma_{ei}^2 E\left\{\widehat{\gamma}_i^2\right\}. \end{aligned} \tag{3.3}$$

The cross-product term $E\left\{(\widehat{\gamma}_i - 1)\widehat{\gamma}_i b_i e_i\right\}$ is equal to zero because b_i and e_i are independent of $\widehat{\gamma}_i$ with mean zero.

We assume that the σ_{ei}^2, $i = 1, 2, \dots,$ form a fixed sequence and $\widehat{\sigma}_{ei}^2, i = 1, 2, \dots,$ are unbiased estimators of σ_{ei}^2 and are independent of $\boldsymbol{Y}$. We also assume that $d_i\widehat{\sigma}_{ei}^2 \sim \chi_{d_i}^2\sigma_{ei}^2$. By the variance property of the χ^2-distribution, we have

$$E(\widehat{\sigma}_{ei}^2 - \sigma_{ei}^2)^2 = 2\sigma_{ei}^4 d_i^{-1}. \tag{3.4}$$

Since any moment of the χ^2-distribution exists, by Theorem 3.1, we have

$$E(\widehat{\sigma}_{ei}^2 - \sigma_{ei}^2)^s = \begin{cases} O(d^{-0.5s}) & \text{if } s \text{ is even,} \\ O(d^{-0.5(s+1)}), & \text{if } s \text{ is odd.} \end{cases} \tag{3.5}$$

We consider a sequence of estimators in which every d_i increases at a common rate d, where $d \to \infty$. We treat $\widehat{\gamma}_i$ as function of $\widehat{\sigma}_{ei}^2$. Note that $0 \le \widehat{\gamma}_i \le 1$. Using Theorem 3.2

for the uniformly bounded case and a Taylor expansion to the third order.

$$\begin{aligned} E(\widehat{\gamma}_i) &= E[\gamma_i - (\sigma_b^2 + \sigma_{ei}^2)^{-2}\sigma_b^2(\widehat{\sigma}_{ei}^2 - \sigma_{ei}^2) \\ &\quad +(\sigma_b^2 + \sigma_{ei}^2)^{-3}\sigma_b^2(\widehat{\sigma}_{ei}^2 - \sigma_{ei}^2)^2] + O(d^{-2}) \\ &= \gamma_i + [2\sigma_b^2\sigma_{ei}^4(\sigma_b^2 + \sigma_{ei}^2)^{-3}]d_i^{-1} + O(d^{-2}) \end{aligned} \tag{3.6}$$

and

$$\begin{aligned} E\{\widehat{\gamma}_i^2\} &= E[\gamma_i^2 - 2(\sigma_b^2 + \sigma_{ei}^2)^{-3}\sigma_b^4(\widehat{\sigma}_{ei}^2 - \sigma_{ei}^2) \\ &\quad +3(\sigma_b^2 + \sigma_{ei}^2)^{-4}\sigma_b^4(\widehat{\sigma}_{ei}^2 - \sigma_{ei}^2)^2] + O(d^{-1.5}) \\ &= \gamma_i^2 + [6\sigma_b^4\sigma_{ei}^4(\sigma_b^2 + \sigma_{ei}^2)^{-4}]d_i^{-1} + O(d^{-2}) \end{aligned} \tag{3.7}$$

by the results in (3.4) and (3.5).

Combining (3.6) and (3.7), we have

$$\begin{aligned} E\{(1 - \widehat{\gamma}_i)^2\} &= E[1 - 2\widehat{\gamma}_i + \widehat{\gamma}_i^2] \\ &= (1 - \gamma_i)^2 + [2\sigma_b^2\sigma_{ei}^4(\sigma_b^2 - 2\sigma_{ei}^2)(\sigma_b^2 + \sigma_{ei}^2)^{-4}]d_i^{-1} + O(d^{-2}), \end{aligned} \tag{3.8}$$

and

$$Var\{\widehat{\gamma}_i\} = E(\widehat{\gamma}_i^2) - 2\gamma_i E(\widehat{\gamma}_i) + \gamma_i^2 = 2\sigma_b^2\sigma_{ei}^4(\sigma_b^2 + \sigma_{ei}^2)^{-4}d_i^{-1} + O(d^{-2}). \tag{3.9}$$

By (3.7), (3.8), and (3.9), we have

$$\begin{aligned} MSE(\widehat{y}_i^H - y_i) &= \sigma_{ei}^2\gamma_i + [2\sigma_b^4\sigma_{ei}^4(\sigma_b^2 + \sigma_{ei}^2)^{-3}]d_i^{-1} + O(d^{-1.5}) \\ &= \sigma_{ei}^2\gamma_i + (\sigma_b^2 + \sigma_{ei}^2)Var\{\widehat{\gamma}_i\} + O(d^{-2}). \end{aligned} \tag{3.10}$$

Comparing (3.10) with (1.14), we see that the second term of (3.6) is due to the estimation of σ_{ei}^2.

Now, we further assume that σ_b^2 is unknown, but continue to assume $\boldsymbol{\beta}$ is known. An unbiased estimator of σ_b^2 is

$$\widetilde{\sigma}_{b0,m}^2 = \sum_{i=1}^m c_{im}[(Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2 - \widehat{\sigma}_{ei}^2], \tag{3.11}$$

where $c_{im}, i = 1, \ldots, m$ are positive fixed values such that $\sum_{i=1}^m c_{im} = 1$.

Lemma 3.1 *Consider the sequence* $\{\widetilde{\sigma}^2_{b0,m}\}_{m=1}^{\infty}$, *where* $\widetilde{\sigma}^2_{b0,m}$ *is defined in (3.11) and* m *is the number of small areas. Assume the* σ^2_{ei}, $i = 1, 2, \ldots$ *form a fixed sequence, assume* $\widehat{\sigma}^2_{ei}, i = 1, 2, \ldots$ *are unbiased estimators of* σ^2_{ei}, *and assume* $\widehat{\sigma}^2_{ei}$ *are independent of* $\boldsymbol{Y}$. *Assume that:*

(i) $\sigma^2_{ei} < C_1$ *for all* i.

(ii) $c_{im}, i = 1, \ldots, m$, *are positive fixed values such that* $\sum_{i=1}^{m} c_{im} = 1$ *and* $c_{im} = O(m^{-1})$.

(iii) $d_i\widehat{\sigma}^2_{ei} \sim \chi^2_{d_i}\sigma^2_{ei}$.

Then

$$E\left\{|\widetilde{\sigma}^2_{b0,m} - \sigma^2_b|^s\right\} = O(m^{-0.5s}) \tag{3.12}$$

for any integer $s \geq 1$.

Proof: Note that

$$(a_1 + \ldots + a_q)^s \leq q^{s-1}(a_1^s + \ldots + a_q^s) \text{ for any } a_i \geq 0, i = 1, \ldots, q, \text{ and } s \geq 1. \tag{3.13}$$

By assumption (ii), there exists C_2 such that $0 < c_{im} \leq m^{-1}C_2$. Using inequality (3.13) for $q = 2$, we have

$$\begin{aligned} |\widetilde{\sigma}^2_{b0,m} - \sigma^2_b|^s &= \left|\sum_{i=1}^{m} c_{im}[(Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2 - \widehat{\sigma}^2_{ei} - \sigma^2_b]\right|^s \\ &\leq 2^{s-1}C_2^s\left(\left|m^{-1}\sum_{i=1}^{m}[(Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2 - \sigma^2_b - \sigma^2_{ei}]\right|^s + \left|m^{-1}\sum_{i=1}^{m}(\widehat{\sigma}^2_{ei} - \sigma^2_{ei})\right|^s\right). \end{aligned} \tag{3.14}$$

By condition (iii), $d_i\widehat{\sigma}^2_{ei} \sim \chi^2_{d_i}\sigma^2_{ei}$. Therefore, all moments of $(\widehat{\sigma}^2_{ei} - \sigma^2_{ei})$ exist. The moments are multiples of the powers of σ^2_{ei} and d_i^{-1}. Given any s fixed, there is a constant C_{3s} such that $E\{(\widehat{\sigma}^2_{ei} - \sigma^2_{ei})^s\} < C_{3s}$ for any i because $\sigma^2_{ei} < C_1$ for all i. If s is even, by following the same argument as used in the proof of Theorem 3.1, we have

$$E\left\{\left|m^{-1}\sum_{i=1}^{m}(\widehat{\sigma}^2_{ei} - \sigma^2_{ei})\right|^s\right\} = E\left\{\left[m^{-1}\sum_{i=1}^{m}(\widehat{\sigma}^2_{ei} - \sigma^2_{ei})\right]^s\right\} = O(m^{-0.5s}), \tag{3.15}$$

and

$$E\left\{\left|m^{-1}\sum_{i=1}^{m}[(Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2 - \sigma^2_{ei} - \sigma^2_b]\right|^s\right\}$$

$$= E\left\{\left[m^{-1}\sum_{i=1}^{m}[(Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2 - \sigma_{ei}^2 - \sigma_b^2]\right]^s\right\} = O(m^{-0.5s}). \qquad (3.16)$$

By (3.14), (3.15), and (3.16), we have

$$E\left\{|\widetilde{\sigma}_{b0,m}^2 - \sigma_b^2|^s\right\} = O(m^{-0.5s}).$$

if s is even. If s is odd, then $s+1$ is even. By Holder's inequality,

$$E\left\{|\widetilde{\sigma}_{b0,m}^2 - \sigma_b^2|^s\right\} \leq \left(E\left\{|\widetilde{\sigma}_{b0,m}^2 - \sigma_b^2|^{s+1}\right\}\right)^{(s+1)^{-1}s}\left(E\left\{1\right\}\right)^{(s+1)^{-1}} = O(m^{-0.5s}). \quad \blacksquare$$

It is possible for $\widetilde{\sigma}_{b0,m}^2$ to take negative values. We define $\widehat{\sigma}_{b0,m}^2 = max(\widetilde{\sigma}_{b0,m}^2, 0)$ in practical use.

Lemma 3.2 *In addition to the conditions of Lemma 3.1, assume that* $\sigma_b^2 > 0$. *We have*

$$E\left\{|\widehat{\sigma}_{b0,m}^2 - \sigma_b^2|^s\right\} = O(m^{-0.5s}). \qquad (3.17)$$

for any integer $s \geq 1$.

Proof: We have

$$\begin{aligned} P(\widetilde{\sigma}_{b0,m}^2 \leq 0) &= P(\widetilde{\sigma}_{b0,m}^2 - \sigma_b^2 \leq -\sigma_b^2) \leq P(|\widetilde{\sigma}_{b0,m}^2 - \sigma_b^2| \geq \sigma_b^2) \\ &\leq (\sigma_b^2)^{-s} E\left\{|\widetilde{\sigma}_{b0,m}^2 - \sigma_b^2|^s\right\} \qquad (3.18) \end{aligned}$$

by Markov's inequality. Therefore, $P(\widetilde{\sigma}_{b0,m}^2 \leq 0) = O(m^{-s})$ for $s \geq 1$, by Lemma 3.1. Now

$$\begin{aligned} E\left\{|\widehat{\sigma}_{b0,m}^2 - \sigma_b^2|^s\right\} &= E[|\widehat{\sigma}_{b0,m}^2 - \sigma_b^2|^s|\widetilde{\sigma}_{b0,m}^2 \leq 0]P(\widetilde{\sigma}_{b0,m}^2 \leq 0) \\ &\quad + E[|\widehat{\sigma}_{b0,m}^2 - \sigma_b^2|^s|\widetilde{\sigma}_{b0,m}^2 > 0]P(\widetilde{\sigma}_{b0,m}^2 > 0) \\ &\leq (\sigma_b^2)^s P(\widetilde{\sigma}_{b0,m}^2 \leq 0) + E\left\{|\widetilde{\sigma}_{b0,m}^2 - \sigma_b^2|^s\right\} = O(m^{-0.5s}). \end{aligned}$$

by using Lemma 3.1 again. $\blacksquare$

Now an EBLUP of y_i assuming $\boldsymbol{\beta}$ is known is $\widehat{y}_i^H = \widehat{\gamma}_i Y_i + (1 - \widehat{\gamma}_i)\boldsymbol{x}_i^T\boldsymbol{\beta}$, where

$$\widehat{\gamma}_i = (\widehat{\sigma}_{b0,m}^2 + \widehat{\sigma}_{ei}^2)^{-1}\widehat{\sigma}_{b0,m}^2. \qquad (3.19)$$

We need to derive an estimator of

$$MSE(\widehat{y}_i^H - y_i) = E\left\{[(\widehat{\gamma}_i - 1)b_i + \widehat{\gamma}_i e_i]^2\right\}$$

$$= E\left\{[(\gamma_i - 1)b_i + \gamma_i e_i + (\widehat{\gamma}_i - \gamma_i)(b_i + e_i)]^2\right\}$$

$$= E\left\{[(\gamma_i - 1)b_i + \gamma_i e_i]^2\right\} + E\left\{[(\widehat{\gamma}_i - \gamma_i)(b_i + e_i)]^2\right\}$$

$$+2E\left\{[(\gamma_i - 1)b_i + \gamma_i e_i][(\widehat{\gamma}_i - \gamma_i)(b_i + e_i)]\right\}$$

$$= \gamma_i \sigma_{ei}^2 + E\left\{[(\widehat{\gamma}_i - \gamma_i)(b_i + e_i)]^2\right\} + 2E\left\{[(\gamma_i - 1)b_i + \gamma_i e_i][(\widehat{\gamma}_i - \gamma_i)(b_i + e_i)]\right\}. \quad (3.20)$$

We will evaluate the last two terms. To do this, we need $E\left\{(\widehat{\gamma}_i - \gamma_i)b_i^2\right\}$, $E\left\{(\widehat{\gamma}_i - \gamma_i)^2 b_i^2\right\}$, $E\left\{(\widehat{\gamma}_i - \gamma_i)b_i e_i\right\}$, $E\left\{(\widehat{\gamma}_i - \gamma_i)^2 b_i e_i\right\}$, $E\left\{(\widehat{\gamma}_i - \gamma_i)e_i^2\right\}$, and $E\left\{(\widehat{\gamma}_i - \gamma_i)^2 e_i^2\right\}$.

We will ignore the bias of $\widehat{\sigma}_{b0,m}^2 = \max(0, \widetilde{\sigma}_{b0,m}^2)$ because $P(\widetilde{\sigma}_{b0,m}^2 \leq 0) = O(m^{-s})$ for any $s \geq 1$ and $\sigma_b^2 > 0$ by Lemma 3.2. Let $r_m = \max(m^{-1.5}, m^{-1}d^{-1}, d^{-2})$. We have

$$E\left\{(\widehat{\sigma}_{b0,m}^2 - \sigma_b^2)b_i^2\right\} = 2\sigma_b^4 c_{im}, \quad (3.21)$$

$$E\left\{(\widehat{\sigma}_{b0,m}^2 - \sigma_b^2)(\widehat{\sigma}_{ei}^2 - \sigma_{ei}^2)b_i^2\right\} = 2\sigma_b^2\sigma_{ei}^4(c_{im}d_i^{-1}) = O(r_m), \quad (3.22)$$

and

$$E\left\{(\widehat{\sigma}_{b0,m}^2 - \sigma_b^2)^2\right\} = 2\sum_{i=1}^{m} c_{im}^2(\sigma_b^2 + \sigma_{ei}^2)^2 + O(r_m).$$

From now on, we will let $V(\widehat{\sigma}_{b0,m}^2)$ denote $2\sum_{i=1}^{m} c_{im}^2(\sigma_b^2 + \sigma_{ei}^2)^2$. We also have that

$$E\left\{(\widehat{\sigma}_{b0,m}^2 - \sigma_b^2)^2 b_i^2\right\} = \sigma_b^2 V(\widehat{\sigma}_{b0,m}^2) + O(r_m). \quad (3.23)$$

We consider a sequence of estimators in which every d_i increases at a common rate d, where $d \to \infty$ and the number of small areas $m \to \infty$. We treat $\widehat{\gamma}_i$ defined in (3.19) as a function of $\widehat{\sigma}_{b0,m}^2$ and $\widehat{\sigma}_{ei}^2$. Note that $0 \leq \widehat{\gamma}_i \leq 1$. Using Theorem 3.2 for the uniformly bounded case and Taylor expansions to the third order, and by the results in (3.4), (3.5), (3.21), (3.22), (3.23), and Lemma 3.2, we have

$$E\left\{(\widehat{\gamma}_i - \gamma_i)b_i^2\right\} = E\left\{(\sigma_b^2 + \sigma_{ei}^2)^{-2}[\sigma_{ei}^2(\widehat{\sigma}_{b0,m}^2 - \sigma_b^2) - \sigma_b^2(\widehat{\sigma}_{ei}^2 - \sigma_{ei}^2)]b_i^2\right.$$

$$+(\sigma_b^2+\sigma_{ei}^2)^{-3}[-\sigma_{ei}^2(\widehat{\sigma}_{b0,m}^2-\sigma_b^2)^2+\sigma_b^2(\widehat{\sigma}_{ei}^2-\sigma_{ei}^2)^2$$

$$+(\sigma_b^2-\sigma_{ei}^2)(\widehat{\sigma}_{b0,m}^2-\sigma_b^2)(\widehat{\sigma}_{ei}^2-\sigma_{ei}^2)]b_i^2\}+O(\max(d^{-2},m^{-1.5}))$$

$$=(\sigma_b^2+\sigma_{ei}^2)^{-2}2\sigma_b^4\sigma_{ei}^2c_{im}+(\sigma_b^2+\sigma_{ei}^2)^{-3}[2\sigma_b^4\sigma_{ei}^4d_i^{-1}-\sigma_b^2\sigma_{ei}^2V(\widehat{\sigma}_{b0,m}^2)]+O(r_m)$$

$$=(\sigma_b^2+\sigma_{ei}^2)^{-3}2\sigma_b^2\sigma_{ei}^2\left\{\sigma_b^2\sigma_{ei}^2d_i^{-1}+c_{im}(\sigma_b^4+\sigma_b^2\sigma_{ei}^2)-\frac{1}{2}V(\widehat{\sigma}_{b0,m}^2)\right\}+O(r_m). \quad (3.24)$$

and

$$E\left\{(\widehat{\gamma}_i-\gamma_i)^2b_i^2\right\}=E\left\{(\sigma_b^2+\sigma_{ei}^2)^{-4}[\sigma_{ei}^4(\widehat{\sigma}_{b0,m}^2-\sigma_b^2)^2\right.$$

$$\left.+\sigma_b^4(\widehat{\sigma}_{ei}^2-\sigma_{ei}^2)^2-\sigma_b^2\sigma_{ei}^2(\widehat{\sigma}_{b0,m}^2-\sigma_b^2)(\widehat{\sigma}_{ei}^2-\sigma_{ei}^2)]b_i^2\right\}+O(\max(d^{-2},m^{-1.5}))$$

$$=(\sigma_b^2+\sigma_{ei}^2)^{-4}2\sigma_b^2\sigma_{ei}^4\left\{\sigma_b^4d_i^{-1}+\frac{1}{2}V(\widehat{\sigma}_{b0,m}^2)\right\}+O(r_m). \quad (3.25)$$

Similarly, we have

$$E\left\{(\widehat{\gamma}_i-\gamma_i)e_i^2\right\}=(\sigma_b^2+\sigma_{ei}^2)^{-3}2\sigma_{ei}^4\left\{\sigma_b^2\sigma_{ei}^2d_i^{-1}+c_{im}(\sigma_b^2\sigma_{ei}^2+\sigma_{ei}^4)-\frac{1}{2}V(\widehat{\sigma}_{b0,m}^2)\right\}+O(r_m). \quad (3.26)$$

$$E\left\{(\widehat{\gamma}_i-\gamma_i)^2e_i^2\right\}=(\sigma_b^2+\sigma_{ei}^2)^{-4}2\sigma_{ei}^6\left\{\sigma_b^4d_i^{-1}+\frac{1}{2}V(\widehat{\sigma}_{b0,m}^2)\right\}+O(r_m). \quad (3.27)$$

We also have

$$E\left\{(\widehat{\gamma}_i-\gamma_i)b_ie_i\right\}=(\sigma_b^2+\sigma_{ei}^2)^{-2}2c_{im}\sigma_b^2\sigma_{ei}^4+O(r_m), \quad (3.28)$$

and

$$E\left\{(\widehat{\gamma}_i-\gamma_i)^2b_ie_i\right\}=O(r_m). \quad (3.29)$$

Combining (3.25), (3.27), and (3.29), we obtain

$$E\left\{[(\widehat{\gamma}_i-\gamma_i)(b_i+e_i)]^2\right\}=[2\sigma_b^4\sigma_{ei}^4(\sigma_b^2+\sigma_{ei}^2)^{-3}]d_i^{-1}+\sigma_{ei}^4(\sigma_b^2+\sigma_{ei}^2)^{-3}V(\widehat{\sigma}_{b0,m}^2)+O(r_m). \quad (3.30)$$

Combining (3.24), (3.26), and (3.28), we have

$$E\left\{[(\gamma_i-1)b_i+\gamma_ie_i][(\widehat{\gamma}_i-\gamma_i)(b_i+e_i)]\right\}$$

$$=(\gamma_i-1)E\left\{(\widehat{\gamma}_i-\gamma_i)b_i^2\right\}+\gamma_iE\left\{(\widehat{\gamma}_i-\gamma_i)e_i^2\right\}+(2\gamma_i-1)E\left\{(\widehat{\gamma}_i-\gamma_i)b_ie_i\right\}$$

$$= (\sigma_b^2+\sigma_{ei}^2)^{-4}2\sigma_b^2\sigma_{ei}^4[-\sigma_b^2\sigma_{ei}^2 d_i^{-1} - (\sigma_b^4+\sigma_b^2\sigma_{ei}^2)c_{im} + \frac{1}{2}V(\widehat{\sigma}_{b0,m}^2) + \sigma_b^2\sigma_{ei}^2 d_i^{-1}$$
$$+(\sigma_b^2\sigma_{ei}^2+\sigma_{ei}^4)c_{im} - \frac{1}{2}V(\widehat{\sigma}_{b0,m}^2)] + (\sigma_b^2+\sigma_{ei}^2)^{-3}2\sigma_b^2\sigma_{ei}^4(\sigma_b^2-\sigma_{ei}^2)c_{im} + O(r_m)$$
$$= (\sigma_b^2+\sigma_{ei}^2)^{-3}2\sigma_b^2\sigma_{ei}^4(\sigma_{ei}^2-\sigma_b^2)c_{im} + (\sigma_b^2+\sigma_{ei}^2)^{-3}2\sigma_b^2\sigma_{ei}^4(\sigma_b^2-\sigma_{ei}^2)c_{im} + O(r_m)$$
$$= O(r_m). \tag{3.31}$$

By (3.20), (3.30), and (3.31), we have

$$MSE(\widehat{y}_i^H - y_i) = \gamma_i\sigma_{ei}^2 + [2\sigma_b^4\sigma_{ei}^4(\sigma_b^2+\sigma_{ei}^2)^{-3}]d_i^{-1} + \sigma_{ei}^4(\sigma_b^2+\sigma_{ei}^2)^{-3}V(\widehat{\sigma}_{b0,m}^2) + O(r_m). \tag{3.32}$$

where $V(\widehat{\sigma}_{b0,m}^2) = 2\sum_{i=1}^m c_{im}^2(\sigma_b^2+\sigma_{ei}^2)^2$. Comparing (3.32) with (3.10), we can see that the second term of (3.32) is due to the estimation of σ_{ei}^2 and the third term of (3.32) is due to the estimation of σ_b^2.

Note that

$$\begin{aligned} E\{(\widehat{\gamma}_i-\gamma_i)^2\} &= E\{(\sigma_b^2+\sigma_{ei}^2)^{-4}[\sigma_{ei}^4(\widehat{\sigma}_{b0,m}^2-\sigma_b^2)^2 \\ &\quad + \sigma_b^4(\widehat{\sigma}_{ei}^2-\sigma_{ei}^2)^2 - \sigma_b^2\sigma_{ei}^2(\widehat{\sigma}_{b0,m}^2-\sigma_b^2)(\widehat{\sigma}_{ei}^2-\sigma_{ei}^2)]\} + O(r_m) \\ &= (\sigma_b^2+\sigma_{ei}^2)^{-4}2\sigma_{ei}^4\left\{\sigma_b^4 d_i^{-1} + \frac{1}{2}V(\widehat{\sigma}_{b0,m}^2)\right\} + O(r_m). \end{aligned} \tag{3.33}$$

We write $E\{(\widehat{\gamma}_i-\gamma_i)^2\}$ as $Var\{\widehat{\gamma}_i\}$ by ignoring the order $O(\max(d^{-1}, m^{-1}))$ bias of $\widehat{\gamma}_i$. By (3.32), we have

$$MSE(\widehat{y}_i - y_i) = \gamma_i\sigma_{ei}^2 + (\sigma_b^2+\sigma_{ei}^2)Var\{\widehat{\gamma}_i\} + O(r_m). \tag{3.34}$$

Finally, we assume that σ_b^2, $\sigma_{ei}^2, i = 1,\ldots,m$, and $\boldsymbol{\beta}$ are all unknown. To simplify the computation, we do not use the empirical generalized least squares estimator

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\widehat{\boldsymbol{V}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\widehat{\boldsymbol{V}}^{-1}\boldsymbol{Y}. \tag{3.35}$$

Instead, we use the weighted least squares (WLS) estimator

$$\widehat{\boldsymbol{\beta}}_O = (\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{Y}, \tag{3.36}$$

where $\boldsymbol{W}_m$ is a known fixed diagonal matrix and is chosen to be close to V^{-1}. An estimator of σ_b^2 is

$$\widetilde{\sigma}_{b,m}^2 = \sum_{i=1}^{m} c_{im}[(Y_i - \boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}}_O)^2 - \widehat{\sigma}_{ei}^2]. \tag{3.37}$$

where $c_{im}, i = 1, \ldots, m$ are positive fixed value such that $\sum_{i=1}^{m} c_{im} = 1$.

Lemma 3.3 *Consider the sequence* $\{\widetilde{\sigma}_{b,m}^2\}_{m=1}^{\infty}$, *where* $\widetilde{\sigma}_{b,m}^2$ *is defined in (3.37) and* m *is the number of small areas. Assume the* σ_{ei}^2, $i = 1, 2, \ldots$, *form a fixed sequence, assume* $\widehat{\sigma}_{ei}^2, i = 1, 2, \ldots$ *are unbiased estimators of* σ_{ei}^2, *and assume* $\widehat{\sigma}_{ei}^2$ *are independent of* $\boldsymbol{Y}$. *Assume that:*

(i) $\sigma_{ei}^2 < C_1$ *for all* i.

(ii) $c_{im}, i = 1, \ldots, m$, *are positive fixed values such that* $\sum_{i=1}^{m} c_{im} = 1$ *and* $c_{im} = O(m^{-1})$.

(iii) $d_i \widehat{\sigma}_{ei}^2 \sim \chi_{d_i}^2 \sigma_{ei}^2$.

(iv) $\lim_{m\to\infty} (\boldsymbol{X}^T \boldsymbol{W}_m \boldsymbol{X})^{-1} = m^{-1}\boldsymbol{A}_1$, *and* $\lim_{m\to\infty} m^{-1}\boldsymbol{X}^T \boldsymbol{W}_m \boldsymbol{V} \boldsymbol{W}_m \boldsymbol{X} = \boldsymbol{A}_2$ *for some positive definite matrices* $\boldsymbol{A}_1$ *and* $\boldsymbol{A}_2$.

Then

$$E\left\{|\widetilde{\sigma}_{b,m}^2 - \sigma_b^2|^s\right\} = O(m^{-0.5s}), \tag{3.38}$$

for any integer $s \geq 1$.

Proof: By inequality (3.13),

$$\begin{aligned}
\left|\widetilde{\sigma}_{b,m}^2 - \sigma_b^2\right|^s &= \left|\sum_{i=1}^{m} c_{im}[(Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2 - \widehat{\sigma}_{ei}^2] - \sigma_b^2 + \sum_{i=1}^{m} c_{im}[\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})]^2 \right. \\
&\quad \left. -2\sum_{i=1}^{m} c_{im}(Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})\right|^s \\
&\leq 3^{s-1}\left|\widetilde{\sigma}_{b0,m}^2 - \sigma_b^2\right|^s + 3^{s-1}\left|\sum_{i=1}^{m} c_{im}[\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})]^2\right|^s \\
&\quad +3^{s-1}\left|-2\sum_{i=1}^{m} c_{im}(Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})\right|^s.
\end{aligned} \tag{3.39}$$

Since $\boldsymbol{b}$ and $\boldsymbol{e}$ are normal errors, $\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta} \sim N_p(\boldsymbol{0}, V(\widehat{\boldsymbol{\beta}}_O))$. By condition (iv),

$$V(\widehat{\boldsymbol{\beta}}_O) = (\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{V}\boldsymbol{W}_m\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1} = O(m^{-1}).$$

Note that $c_{im} = O(m^{-1})$. If s is even, by following the same argument as used in the proof of Theorem 3.1, we have

$$E\left\{\left|\sum_{i=1}^{m} c_{im}[\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})]^2\right|^s\right\} = O(m^{-s}). \tag{3.40}$$

and

$$\begin{aligned} &E\left\{\left|-2\sum_{i=1}^{m} c_{im}(Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})\right|^s\right\} \\ &= E\left\{\left|-2\sum_{i=1}^{m} c_{im}(b_i + e_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})\right|^s\right\} = O(m^{-0.5s}). \end{aligned} \tag{3.41}$$

Combining Lemma 3.1, (3.40), and (3.41), the result follows (3.39). If s is odd, then $s+1$ is even. By Holder's inequality,

$$E\left\{|\widetilde{\sigma}^2_{b,m} - \sigma^2_b|^s\right\} \le \left(E\left\{|\widetilde{\sigma}^2_{b,m} - \sigma^2_b|^{s+1}\right\}\right)^{(s+1)^{-1}s}(E\{1\})^{(s+1)^{-1}} = O(m^{-0.5s}). \quad \blacksquare$$

Remark: For any estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, Lemma 3.3 still holds if $E\left\{\boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right\} = O(m^{-1})$.

As previously mentioned, it is possible for $\widetilde{\sigma}^2_{b,m}$ to take negative values and we define $\widehat{\sigma}^2_{b,m} = \max(\widetilde{\sigma}^2_{b,m}, 0)$ in practical use.

Lemma 3.4 *Under the conditions of Lemma 3.3 and $\sigma^2_b > 0$, we have*

$$P(\widetilde{\sigma}^2_{b,m} \le 0) = O(m^{-s})$$

and

$$E\left\{|\widehat{\sigma}^2_{b,m} - \sigma^2_b|^s\right\} = O(m^{-0.5s}) \tag{3.42}$$

for any $s \ge 1$.

Proof: Follows the same argument used in the proof of Lemma 3.2. $\blacksquare$

Consider the random effects model

$$Y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + b_i + e_i, \qquad i = 1, \ldots, m,$$

where $b_i \sim N(0, \sigma_b^2)$, $e_i \sim N(0, \sigma_{ei}^2)$. Let the estimator of $\boldsymbol{\beta}$ be $\widehat{\boldsymbol{\beta}}_O$ defined in (3.36), the estimator of σ_b^2 be $\widetilde{\sigma}_{b,m}^2$ defined in (3.37), and the estimator of σ_{ei}^2 be $\widehat{\sigma}_{ei}^2$. Now we consider the estimator

$$\widehat{y}_i = \widehat{\gamma}_i Y_i + (1 - \widehat{\gamma}_i)\boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}}_O, \tag{3.43}$$

where

$$\widehat{\gamma}_i = (\widetilde{\sigma}_{b,m}^2 + \widehat{\sigma}_{ei}^2)^{-1} \widetilde{\sigma}_{b,m}^2 \tag{3.44}$$

and $\widetilde{\sigma}_{b,m}^2$ is defined in (3.37). The only difference between $\widehat{y}_i$ and the EBLUP $\widehat{y}_i^H$ is the estimator of $\boldsymbol{\beta}$ used. The $\widehat{y}_i$ uses $\widehat{\boldsymbol{\beta}}_O = (\boldsymbol{X}^T \boldsymbol{W}_m \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{W}_m \boldsymbol{Y}$ as the estimator of $\boldsymbol{\beta}$. The EBLUP $\widehat{y}_i^H$ uses $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T \widehat{\boldsymbol{V}}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^T \widehat{\boldsymbol{V}}^{-1} \boldsymbol{Y}$ as the estimator of $\boldsymbol{\beta}$. If $\boldsymbol{W}_m$ is chosen to be close to $\boldsymbol{V}^{-1}$, $\boldsymbol{W}_m$ should be reasonablely close to $\widehat{\boldsymbol{V}}^{-1}$. Then $\widehat{\boldsymbol{\beta}}_O$ should be close to $\widehat{\boldsymbol{\beta}}$ and $\widehat{y}_i$ is close to $\widehat{y}_i^H$. For $\widehat{y}_i$ defined in (3.43), we have the major theorem of this chapter.

Theorem 3.3 *Let there be a sequence of small areas $i = 1, \ldots, m$, where the number of small areas $m \to \infty$. Assume $\sigma_{ei}^2, i = 1, 2, \ldots,$ and $\boldsymbol{x}_i^T, i = 1, 2, \ldots,$ are fixed sequences. Assume $\widehat{\sigma}_{ei}^2, i = 1, 2, \ldots,$ with $d_i \widehat{\sigma}_{ei}^2 \sim \chi_{d_i}^2 \sigma_{ei}^2$ are unbiased estimators of σ_{ei}^2. Assume $\{\widehat{\sigma}_{ei}^2\}_{i=1}^m$ are independent of $\boldsymbol{Y} = (Y_1, \ldots, Y_m)^T$ for any m. Assume that:*

(i) every d_i increases at a common rate d, where $d \to \infty$.

(ii) $\sigma_b^2 > 0$, $\sigma_{ei}^2 < C_1$ for all i, and $x_{ij} < C_x$ for all i and j.

(iii) $w_{ii,m} < C_w$, where $w_{ii,m}$ are the diagonal elements of $\boldsymbol{W}_m$, for all i and m.

(iv) $c_{im}, i = 1, \ldots, m$, which are the weights in (3.37) such that $\sum_{i=1}^m c_{im} = 1$, are positive fixed values and $c_{im} = O(m^{-1})$.

(v) $\lim_{m\to\infty} (\boldsymbol{X}^T \boldsymbol{W}_m \boldsymbol{X})^{-1} = m^{-1}\boldsymbol{A}_1$, and $\lim_{m\to\infty} m^{-1} \boldsymbol{X}^T \boldsymbol{W}_m \boldsymbol{V} \boldsymbol{W}_m \boldsymbol{X} = \boldsymbol{A}_2$ for some positive definite matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$.

Let $\widehat{\gamma}_i$ be defined by (3.44). Then

$$E\left\{(\widehat{\gamma}_i - \gamma_i)^2\right\} = (\sigma_b^2 + \sigma_{ei}^2)^{-4} 2\sigma_{ei}^4 \left\{\sigma_b^4 d_i^{-1} + \frac{1}{2} V(\widetilde{\sigma}_{b,m}^2)\right\} + O(r_m), \tag{3.45}$$

where $V(\widehat{\sigma}^2_{b,m}) = 2\sum_{i=1}^{m} c^2_{im}(\sigma^2_b + \sigma^2_{ei})^2$ *and* $r_m = \max(m^{-1.5}, m^{-1}d^{-1}, d^{-2})$. *Let* $\widehat{y}_i$ *be defined by (3.43). Then*

$$\begin{aligned} MSE(\widehat{y}_i - y_i) = E\left\{(\widehat{y}_i - y_i)^2\right\} &= \sigma^2_{ei}\gamma_i + (1-\gamma_i)^2 \boldsymbol{x}_i^T V(\widehat{\boldsymbol{\beta}}_O)\boldsymbol{x}_i \\ &+ (\sigma^2_b + \sigma^2_{ei})V(\widehat{\gamma}_i) + O(r_m). \end{aligned} \tag{3.46}$$

where $V(\widehat{\boldsymbol{\beta}}_O) = (\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{V}\boldsymbol{W}_m\boldsymbol{X})(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}$ *and*

$$V(\widehat{\gamma}_i) = (\sigma^2_b + \sigma^2_{ei})^{-4} 2\sigma^4_{ei}\left\{\sigma^4_b d_i^{-1} + \frac{1}{2}V(\widehat{\sigma}^2_{b,m})\right\}.$$

If $i \neq j$, *we have*

$$E\left\{(\widehat{y}_i - y_i)(\widehat{y}_j - y_j)\right\} = (1-\gamma_i)(1-\gamma_j)\boldsymbol{x}_i^T V(\widehat{\boldsymbol{\beta}}_O)\boldsymbol{x}_j + O(r_m). \tag{3.47}$$

Proof: By (3.4) and (3.5), we have

$$E(\widehat{\sigma}^2_{ei} - \sigma^2_{ei})^2 = 2\sigma^4_{ei} d_i^{-1}.$$

and

$$E(\widehat{\sigma}^2_{ei} - \sigma^2_{ei})^s = \begin{cases} O(d^{-0.5s}) & \text{if } s \text{ is even,} \\ O(d^{-0.5(s+1)}), & \text{if } s \text{ is odd.} \end{cases}$$

By Lemma 3.4, we have

$$E\left\{|\widehat{\sigma}^2_{b,m} - \sigma^2_b|^s\right\} = O(m^{-0.5s}).$$

We will ignore the bias of $\widehat{\sigma}^2_{b,m}$ because $P(\widetilde{\sigma}^2_{b,m} \leq 0) = O(m^{-s})$ for any $s \geq 1$ and $\sigma^2_b > 0$. Let $r_m = max(m^{-1.5}, m^{-1}d^{-1}, d^{-2})$.

$$E\left\{(\widehat{\sigma}^2_{b0,m} - \sigma^2_b)(\widehat{\sigma}^2_{ei} - \sigma^2_{ei})\right\} = 2\sigma^4_{ei}(c_{im}d_i^{-1}) = O(r_m),$$

and

$$E\left\{(\widehat{\sigma}^2_{b,m} - \sigma^2_b)^2\right\} = V(\widehat{\sigma}^2_{b,m}) + O(r_m),$$

where $V(\widehat{\sigma}^2_{b,m}) = 2\sum_{i=1}^{m} c^2_{im}(\sigma^2_b + \sigma^2_{ei})^2$.

We treat $\widehat{\gamma}_i$ defined in (3.44) as a function of $\widehat{\sigma}^2_{b,m}$ and $\widehat{\sigma}^2_{ei}$. Note that $0 \le \widehat{\gamma}_i \le 1$. We use Theorem 3.2 to obtain the asymptotic results and do a Taylor expansion to the third order. Using this idea, we have

$$\begin{aligned} E\left\{(\widehat{\gamma}_i - \gamma_i)^2\right\} &= E\left\{(\sigma_b^2 + \sigma_{ei}^2)^{-4}[\sigma_{ei}^4(\widehat{\sigma}^2_{b0,m} - \sigma_b^2)^2 \right. \\ &\quad \left. + \sigma_b^4(\widehat{\sigma}^2_{ei} - \sigma^2_{ei})^2 - \sigma_b^2\sigma_{ei}^2(\widehat{\sigma}^2_{b,m} - \sigma_b^2)(\widehat{\sigma}^2_{ei} - \sigma^2_{ei})]\right\} + O(r_m) \\ &= (\sigma_b^2 + \sigma_{ei}^2)^{-4} 2\sigma_{ei}^4 \left\{\sigma_b^4 d_i^{-1} + \frac{1}{2} V(\widehat{\sigma}^2_{b,m})\right\} + O(r_m). \end{aligned}$$

Therefore, (3.45) is true.

Note that

$$\begin{aligned} \widehat{y}_i - y_i &= (1-\gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta}) - (\widehat{\gamma}_i - \gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta}) + [(\widehat{\gamma}_i - 1)b_i + \widehat{\gamma}_i e_i] \\ &= (\gamma_i - 1)b_i + \gamma_i e_i + (\widehat{\gamma}_i - \gamma_i)(b_i + e_i) + (1-\gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta}) \\ &\quad + O_p\left(\max(m^{-1}, m^{-0.5}d^{-0.5})\right). \end{aligned} \tag{3.48}$$

and

$$\begin{aligned} (\widehat{y}_i - y_i)^2 &= [(1-\gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})]^2 + 2(1-\gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})[(\widehat{\gamma}_i - 1)b_i + \widehat{\gamma}_i e_i] \\ &\quad + [(\widehat{\gamma}_i - \gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})]^2 - 2(1-\gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})(\widehat{\gamma}_i - \gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta}) \\ &\quad - 2(\widehat{\gamma}_i - \gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})[(\widehat{\gamma}_i - 1)b_i + \widehat{\gamma}_i e_i] + [(\widehat{\gamma}_i - 1)b_i + \widehat{\gamma}_i e_i]^2. \end{aligned} \tag{3.49}$$

For the first term in (3.49), we have

$$E\left\{[(1-\gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})]^2\right\} = (1-\gamma_i)^2\boldsymbol{x}_i^T V(\widehat{\boldsymbol{\beta}}_O)\boldsymbol{x}_i. \tag{3.50}$$

For the last term in (3.49), we have

$$\begin{aligned} E\left\{[(\widehat{\gamma}_i - 1)b_i + \widehat{\gamma}_i e_i]^2\right\} &= E\left\{[(\gamma_i - 1)b_i + \gamma_i e_i + (\widehat{\gamma}_i - \gamma_i)(b_i + e_i)]^2\right\} \\ &= \gamma_i\sigma_{ei}^2 + E\left\{[(\widehat{\gamma}_i - \gamma_i)(b_i + e_i)]^2\right\} \\ &\quad + 2E\left\{[(\gamma_i - 1)b_i + \gamma_i e_i][(\widehat{\gamma}_i - \gamma_i)(b_i + e_i)]\right\}. \end{aligned} \tag{3.51}$$

The $\widehat{\gamma}_i$ defined in (3.44) is not the same as the $\widehat{\gamma}_i$ defined in (3.19) because the estimators for σ_b^2 are $\widehat{\sigma}^2_{b,m}$ and $\widehat{\sigma}^2_{b0,m}$, respectively. However, by the method used in deriving (3.30) and (3.31), we can show that equations (3.30) and (3.31) still hold if $\widehat{\sigma}^2_{b0,m}$ is replaced with $\widehat{\sigma}^2_{b,m}$. That is

$$\begin{aligned} E\left\{[(\widehat{\gamma}_i-\gamma_i)(b_i+e_i)]^2\right\} &= [2\sigma_b^4\sigma_{ei}^4(\sigma_b^2+\sigma_{ei}^2)^{-3}]d_i^{-1}+\sigma_{ei}^4(\sigma_b^2+\sigma_{ei}^2)^{-3}V(\widehat{\sigma}^2_{b,m})+O(r_m) \\ &= (\sigma_b^2+\sigma_{ei}^2)V(\widehat{\gamma}_i)+O(r_m), \end{aligned} \tag{3.52}$$

and

$$E\left\{[(\gamma_i-1)b_i+\gamma_i e_i][(\widehat{\gamma}_i-\gamma_i)(b_i+e_i)]\right\}=O(r_m).$$

Therefore, the expectation of the last term of (3.49) is

$$E\left\{[(\widehat{\gamma}_i-1)b_i+\widehat{\gamma}_i e_i]^2\right\}=\gamma_i\sigma_{ei}^2+(\sigma_b^2+\sigma_{ei}^2)V(\widehat{\gamma}_i)+O(r_m). \tag{3.53}$$

To prove (3.46) is true, we need to show the expectations of the other terms in (3.49) are of order $O(r_m)$. We will show that $E\left\{[\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O-\boldsymbol{\beta})][(\widehat{\gamma}_i-1)b_i+\widehat{\gamma}_i e_i]\right\}=O(r_m)$ as an example. By Taylor expansion, we have

$$cov\left\{(\widehat{\gamma}_i-\gamma_i)(b_i+e_i),(b_j+e_j)\right\}=\begin{cases} O(r_m) & \text{if } i\neq j, \\ O(d^{-1}), & \text{if } i=j. \end{cases} \tag{3.54}$$

By condition (iii) and (v), for sufficiently large m,

$$\boldsymbol{x}_j^T(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}\boldsymbol{x}_k w_{kk,m}\leq m^{-1}C_x^2(\sum_{i=1}^p\sum_{j=1}^p a_{1ij})w_{kk,m}=m^{-1}C_4, \tag{3.55}$$

where a_{1ij} is the ij-th element of matrix $\boldsymbol{A}_1$.

By the order results of (3.54) and (3.55), we have

$$\begin{aligned} &\left|E\left\{[(\widehat{\gamma}_i-\gamma_i)(b_i+e_i)][\boldsymbol{x}_j^T(\widehat{\boldsymbol{\beta}}_O-\boldsymbol{\beta})]\right\}\right| \\ &=\left|E[(\widehat{\gamma}_i-\gamma_i)(b_i+e_i)\sum_{k=1}^m\boldsymbol{x}_j^T(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}\boldsymbol{x}_k w_{kk,m}(b_k+e_k)]\right| \\ &\leq m^{-1}C_4E\left\{(\widehat{\gamma}_i-\gamma_i)(b_i+e_i)^2\right\}+m^{-1}C_4\sum_{k\neq i}E\left\{(\widehat{\gamma}_i-\gamma_i)(b_k+e_k)(b_i+e_i)\right\} \\ &=m^{-1}C_4O(d^{-1})+m^{-1}(m-1)O(r_m)=O(r_m) \end{aligned} \tag{3.56}$$

for any i and j. Note that

$$(\widehat{\gamma}_i - 1)b_i + \widehat{\gamma}_i e_i = [(\gamma_i - 1)b_i + \gamma_i e_i] + (\widehat{\gamma}_i - \gamma_i)(b_i + e_i).$$

Also

$$E\left\{[\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})][(\gamma_i - 1)b_i + \gamma_i e_i]\right\} = 0.$$

because

$$E\left\{(b_i + e_i)[(\gamma_i - 1)b_i + \gamma_i e_i]\right\} = 0.$$

Therefore,

$$\left|E\left\{\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})[(\widehat{\gamma}_i - 1)b_i + \widehat{\gamma}_i e_i]\right\}\right| = \left|E\left\{\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})[(\widehat{\gamma}_i - \gamma_i)(b_i + e_i)]\right\}\right| = O(r_m). \tag{3.57}$$

Similarly, we can show

$$E\left\{[(\widehat{\gamma}_i - \gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})]^2\right\} = O(r_m), \tag{3.58}$$

$$E\left\{[(\widehat{\gamma}_i - \gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})][(\widehat{\gamma}_i - 1)b_i + \widehat{\gamma}_i e_i]\right\} = O(r_m), \tag{3.59}$$

and

$$E\left\{[\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})(\widehat{\gamma}_i - \gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})]\right\} = O(r_m). \tag{3.60}$$

Result (3.46) follows (3.49), (3.50), (3.57), (3.58), (3.59), (3.60), and (3.53).

To prove (3.47), we expand $\widehat{y}_i - y_i$ in the form of (3.48) and get the expression for $(\widehat{y}_i - y_i)(\widehat{y}_j - y_j)$. One of the terms is

$$(1 - \gamma_i)(1 - \gamma_j)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})^T\boldsymbol{x}_j.$$

We have

$$E\left\{(1 - \gamma_i)(1 - \gamma_j)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})^T\boldsymbol{x}_j\right\} = (1 - \gamma_i)(1 - \gamma_j)\boldsymbol{x}_i^T V(\widehat{\boldsymbol{\beta}}_O)\boldsymbol{x}_j.$$

So we only need to show that the expecatation of all the remaining terms are of order $O(r_m)$. After routine and lengthy calculation of the same type as we used to prove (3.46), we can show that this claim is indeed true. Therefore, (3.47) is true. ■

Remark: If $d^{-2} \leq O(m^{-1.5})$, then $O(r_m) = O\{max(m^{-1.5}, m^{-1}d^{-1}, d^{-2})\} = O(m^{-1.5})$. The convergence rate is fast under this situation. Even though $d^{-2} > O(m^{-1.5})$, $O(r_m) = O\{max(m^{-1.5}, m^{-1}d^{-1}, d^{-2})\} = o(m^{-1})$ as long as $d^{-2} = o(m^{-1})$. This means that (3.46) is the order $o(m^{-1})$ asymptotic approximation to $MSE(\widehat{y}_i - y_i)$. Last, if $d^{-2} = O(m^{-1})$, then we have $O(r_m) = O(m^{-1})$. The convergence rate is slower under this situation.

Since we do not know σ_b^2, $\sigma_{ei}^2, i = 1, \ldots, m$, and $\boldsymbol{\beta}$, it is desirable to estimate $MSE(\widehat{y}_i - y_i)$ in (3.46). Simply replacing σ_b^2 and σ_{ei}^2 in $\gamma_i\sigma_{ei}^2$ with $\widehat{\sigma}_{b,m}^2$, and $\widehat{\sigma}_{ei}^2$ yields an estimator of $\gamma_i\sigma_{ei}^2$ that underestimates $\gamma_i\sigma_{ei}^2$ because

$$\begin{aligned} E\left\{\widehat{\gamma}_i\widehat{\sigma}_{ei}^2\right\} &= \gamma_i\sigma_{ei}^2 - (\sigma_b^2 + \sigma_{ei}^2)^{-3}2\sigma_{ei}^4\left[\sigma_b^4 d_i^{-1} + \frac{1}{2}V(\widehat{\sigma}_{b,m}^2)\right] + O(r_m) \\ &= \gamma_i\sigma_{ei}^2 - (\sigma_b^2 + \sigma_{ei}^2)V(\widehat{\gamma}_i) + O(r_m). \end{aligned} \tag{3.61}$$

Therefore, an estimator of $MSE(\widehat{y}_i - y_i)$ in (3.46) with bias of order $O(r_m)$ is

$$\widehat{MSE}(\widehat{y}_i - y_i) = \widehat{\gamma}_i\widehat{\sigma}_{ei}^2 + (1 - \widehat{\gamma}_i)^2\boldsymbol{x}_i^T\widehat{V}(\widehat{\boldsymbol{\beta}}_0)\boldsymbol{x}_i + 2(\widehat{\sigma}_{b,m}^2 + \widehat{\sigma}_{ei}^2)\widehat{Var}\{\widehat{\gamma}_i\}, \tag{3.62}$$

where

$$\widehat{Var}\{\widehat{\gamma}_i\} = 2\widehat{\sigma}_{ei}^4(\widehat{\sigma}_{b,m}^2 + \widehat{\sigma}_{ei}^2)^{-4}\left\{\widehat{\sigma}_{b,m}^4 d_i^{-1} + \sum_{i=1}^m c_{im}^2(\widehat{\sigma}_b^2 + \widehat{\sigma}_{ei}^2)^2\right\}. \tag{3.63}$$

3.3 Simulation study

A small Monte Carlo study was conducted on the univariate model to investigate the properties of our approximations to $MSE(\widehat{y}_i - y_i)$. Each Monte Carlo sample was constructed in the following way. First, observations were generated to satisfy the model

$$Y_{ij} = \mu + b_i + e_{ij}, \qquad j = 1, \ldots, d+1, i = 1, \ldots, m, \tag{3.64}$$

where $b_i \sim N(0, \sigma_b^2)$, $e_{ij} \sim N(0, (d+1)\sigma_{ei}^2)$. The random effects model for the small area mean is

$$Y_i = \mu + b_i + e_i, \qquad i = 1, \dots, m, \tag{3.65}$$

where

$$Y_i = \overline{Y}_{i.} = (d+1)^{-1} \sum_{j=1}^{d+1} Y_{ij},$$

$$e_i = \overline{e}_{i.} = (d+1)^{-1} \sum_{j=1}^{d+1} e_{ij},$$

and $e_i \sim N(0, \sigma_{ei}^2)$. The quantities we estimate are $y_i = \mu + b_i, i = 1, \dots, m$.

To ease our programing effort, we set all d_i equal to d, one third of σ_{ei}^2 equal to σ_e^2, one third of σ_{ei}^2 equal to $1.5\sigma_e^2$, and one third of σ_{ei}^2 equal to $2\sigma_e^2$ for a given σ_e^2. We estimate the σ_{ei}^2 individually. We also set $\boldsymbol{W}_m = \boldsymbol{I}_m$. This leads to the ordinary least squares (OLS) estimator of $\boldsymbol{\beta}$.

The unbiased estimator of σ_{ei}^2 is

$$\widehat{\sigma}_{ei}^2 = (d+1)^{-1} d^{-1} \sum_{j=1}^{d+1} (Y_{ij} - Y_i)^2,$$

where $d\widehat{\sigma}_{ei}^2 \sim \chi_d^2 \sigma_{ei}^2$. The OLS estimator of μ is

$$\widehat{\mu} = m^{-1} \sum_{i=1}^{m} Y_i = m^{-1}(d+1)^{-1} \sum_{i=1}^{m} \sum_{j=1}^{d+1} Y_{ij}.$$

We set $c_{im} = m^{-1}$. The estimator for σ_b^2 is

$$\widehat{\sigma}_{b,m}^2 = \max\left(0, \; m^{-1} \sum_{i=1}^{m} [m(m-1)^{-1}(Y_i - \widehat{\mu})^2 - \widehat{\sigma}_{ei}^2]\right).$$

The predictor for y_i is

$$\widehat{y}_i = \widehat{\gamma}_i Y_i + (1 - \widehat{\gamma}_i)\widehat{\mu}, \tag{3.66}$$

where

$$\widehat{\gamma}_i = (\widehat{\sigma}_{b,m}^2 + \widehat{\sigma}_{ei}^2)^{-1} \widehat{\sigma}_{b,m}^2.$$

By (3.46), an approximation to $MSE(\widehat{y}_i - y_i)$ with bias of order $O(r_m)$, where $r_m = \max(m^{-1.5}, m^{-1}d^{-1}, d^{-2})$, is

$$MSE(\widehat{y}_i - y_i) = \gamma_i\sigma_{ei}^2 + (1-\gamma_i)^2(\sigma_b^2 + \overline{\sigma_{ei}^2})m^{-1} + (\sigma_b^2 + \sigma_{ei}^2)^{-3}2\sigma_b^4\sigma_{ei}^4 d^{-1}$$
$$+(\sigma_b^2 + \sigma_{ei}^2)^{-3}2\sigma_{ei}^4(\sigma_b^4 + \overline{\sigma_{ei}^4} + 2\sigma_b^2\overline{\sigma_{ei}^2})m^{-1}, \tag{3.67}$$

where $\gamma_i = (\sigma_b^2 + \sigma_{ei}^2)^{-1}\sigma_b^2$, $\overline{\sigma_{ei}^2} = m^{-1}\sum_{i=1}^{m}\sigma_{ei}^2$, and $\overline{\sigma_{ei}^4} = m^{-1}\sum_{i=1}^{m}\sigma_{ei}^4$.
By (3.62), an estimator of $\widehat{MSE}(\widehat{y}_i - y_i)$ with bias of order $O(r_m)$ is

$$\widehat{MSE}_1(\widehat{y}_i - y_i) = \widehat{\gamma}_i\widehat{\sigma}_{ei}^2 + (1-\widehat{\gamma}_i)^2(\widehat{\sigma}_{b,m}^2 + \overline{\widehat{\sigma}_{ei}^2})m^{-1} + (\widehat{\sigma}_{b,m}^2 + \widehat{\sigma}_{ei}^2)^{-3}4\widehat{\sigma}_{ei}^4\widehat{\sigma}_{b,m}^4 d^{-1}$$
$$+(\widehat{\sigma}_{b,m}^2 + \widehat{\sigma}_{ei}^2)^{-3}4\widehat{\sigma}_{ei}^4(\widehat{\sigma}_{b,m}^4 + \overline{\widehat{\sigma}_{ei}^4} + 2\widehat{\sigma}_{b,m}^2\overline{\widehat{\sigma}_{ei}^2})m^{-1}. \tag{3.68}$$

In our simulation, we set $\mu = 10$, $\sigma_e^2 = 1$, and $\sigma_b^2 = 0.01, 0.1, 0.25, \frac{3}{7}, \frac{2}{3}, 1, 1.5, \frac{7}{3}$. For each of the parameter settings, we generated 1000 samples for (d, m)=(5, 36), (9, 36), (9, 99), (14, 99), and (14, 225). The results are reported in Table 3.1 through Table 3.5.

Every table has the same format. The first column of the table contains the parameter σ_b^2. We have three distinct γ_i, which are listed in the second column of the table. We divide the data into three groups according to the values of γ_i. For each sample, the mean of the simulated $MSE(\widehat{y}_i - y_i)$ for the G-th group is

$$MSE(\widehat{y}_i - y_i)_G = 3m^{-1}\sum_{i\in G}(\widehat{y}_i - y_i)^2.$$

We calculate the mean of the 1000 simulated $MSE(\widehat{y}_i - y_i)_G$ for each group and the result is reported in the third column of the table. The corresponding theoretical $MSE(\widehat{y}_i - y_i)_G$ value defined in (3.67) is in the fourth column.

For each sample, we compute $\widehat{MSE}_1(\widehat{y}_i - y_i)$ defined in (3.68) and the corresponding mean for the G-th group, which is

$$\widehat{MSE}_1(\widehat{y}_i - y_i)_G = 3m^{-1}\sum_{i\in G}\widehat{MSE}_1(\widehat{y}_i - y_i).$$

The mean of the 1000 simulated $\widehat{MSE}_1(\widehat{y}_i - y_i)_G$ for each group is reported in the fifth column of the table.

From these tables, we can see that the theoretical approximation to $MSE(\widehat{y}_i - y_i)$ and simulated $MSE(\widehat{y}_i - y_i)$ are very close to each other when $\sigma_{ei}^{-2}\sigma_b^2$ is not close to zero, even when d and m are small. When d and m are large, the theoretical approximation and simulated $MSE(\widehat{y}_i - y_i)$ are almost identical. This indicates that (3.65) provides a reasonable approximation to the true $MSE(\widehat{y}_i - y_i)$ when $\sigma_{ei}^{-2}\sigma_b^2$ is not close to zero. When $\sigma_{ei}^{-2}\sigma_b^2$ is close to zero, the theoretical approximation to $MSE(\widehat{y}_i - y_i)$ is much larger than the simulated $MSE(\widehat{y}_i - y_i)$. Note that when $\sigma_{ei}^{-2}\sigma_b^2$ is close to zero, σ_b^2 is very small compared to σ_{ei}^2 and $\widehat{\sigma}_{b,m}^2 = 0$ with probability close to 0.5. If $\widehat{\sigma}_{b,m}^2 = 0$, the estimator for y_i is $\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \widehat{\mu}$. Even if $\widehat{\sigma}_{b,m}^2 > 0$ but small, the estimator for y_i is still close to $\widehat{\mu}$. Therefore, the true MSE of $\widehat{y}_i - y_i$ should be close to $\sigma_b^2 + (\sigma_b^2 + \overline{\sigma_{ei}^2})m^{-1}$. However, the theoretical approximation to $MSE(\widehat{y}_i - y_i)$ is close to $\sigma_b^2 + (\sigma_b^2 + \overline{\sigma_{ei}^2} + 2\sigma_{ei}^{-2}\overline{\sigma_{ei}^4})m^{-1}$. In other words, the impact on $MSE(\widehat{y}_i - y_i)$ due to estimation of σ_b^2 and σ_{ei}^2 is not as large as that given by the Taylor expansion for small d, small m and small $\sigma_{ei}^{-2}\sigma_b^2$.

Comparing the third column with the fifth column, we can see that $\widehat{MSE}_1(\widehat{y}_i - y_i)$ is close to the simulated $MSE(\widehat{y}_i - y_i)$ when $\sigma_{ei}^{-2}\sigma_b^2$ is large if d and m are small and that $\widehat{MSE}_1(\widehat{y}_i - y_i)$ is close to the simulated $MSE(\widehat{y}_i - y_i)$ when σ_b^2 is moderately large or large if d and m are large, though $\widehat{MSE}_1(\widehat{y}_i - y_i)$ slightly underestimates the simulated $MSE(\widehat{y}_i - y_i)$ in these cases. This indicates that $\widehat{MSE}_1(\widehat{y}_i - y_i)$ is a reasonable estimator of the true $MSE(\widehat{y}_i - y_i)$ when $\sigma_{ei}^{-2}\sigma_b^2$ is large.

3.4 Improved estimator for $MSE(\widehat{y}_i - y_i)$

The estimator $\widehat{MSE}_1(\widehat{y}_i - y_i)$ yields an unsatisfactory approximation to the true $MSE(\widehat{y}_i - y_i)$ when $\sigma_{ei}^{-2}\sigma_b^2$ is small. Three factors contribute to the severe overestimation. First, $\widehat{MSE}_1(\widehat{y}_i - y_i)$ defined in (3.68) estimates the theoretical $MSE(\widehat{y}_i - y_i)$ defined in

(3.67), which overestimates the true $MSE(\widehat{y}_i - y_i)$ when $\sigma_{ei}^{-2}\sigma_b^2$ is small. We cannot do much about the overestimation of the theoretical approximation. Second, $\widehat{\sigma}_{b,m}^2$ severely overestimates σ_b^2 due to truncation when σ_b^2 is small. For example, the average of the 1000 $\widehat{\sigma}_{b,m}^2$ is 0.12 when $\sigma_b^2 = 0.01$. This overestimation of σ_b^2 in turn leads to overestimation of $\widehat{MSE}_1(\widehat{y}_i - y_i)$. Third, $\widehat{\gamma}_i\widehat{\sigma}_{ei}^2$ does not underestimate $\gamma_i\sigma_{ei}^2$ when $\sigma_{ei}^{-2}\sigma_b^2$ is small. Therefore, the adjustment made in (3.61) inflates $\widehat{MSE}_1(\widehat{y}_i - y_i)$. For real data, we do not know σ_b^2 and σ_{ei}^2. Therefore, we cannot determine when $\widehat{\sigma}_{b,m}^2$ overestimates σ_b^2 and $\widehat{\gamma}_i\widehat{\sigma}_{ei}^2$ overestimates $\gamma_i\sigma_{ei}^2$. We need to tackle the overestimation from another angle.

Note that $\widehat{MSE}_1(\widehat{y}_i - y_i)$ defined in (3.68) is a special case of (3.62). That is

$$\widehat{MSE}_1(\widehat{y}_i - y_i) = \widehat{\gamma}_i\widehat{\sigma}_{ei}^2 + (1-\widehat{\gamma}_i)^2\boldsymbol{x}_i^T\widehat{V}(\widehat{\boldsymbol{\beta}}_0)\boldsymbol{x}_i + 2(\widehat{\sigma}_{b,m}^2 + \widehat{\sigma}_{ei}^2)\widehat{Var}\{\widehat{\gamma}_i\}.$$

To improve the estimator defined in (3.68), we need to deal with two issues. First, $\widehat{\sigma}_{b,m}^2$ severely overestimates σ_b^2 due to truncation, which in turn leads to $(\widehat{\sigma}_{b,m}^2 + \widehat{\sigma}_{ei}^2)\widehat{Var}\{\widehat{\gamma}_i\}$ overestimating $(\sigma_b^2 + \sigma_{ei}^2)Var\{\widehat{\gamma}_i\}$, when $\sigma_{ei}^{-2}\sigma_b^2$ is small. We need to find a better estimator to $(\sigma_b^2 + \sigma_{ei}^2)Var\{\widehat{\gamma}_i\}$, the contribution due to the varibility of $\widehat{\gamma}_i$. Second, $\widehat{\gamma}_i\widehat{\sigma}_{ei}^2$ overestimates $\gamma_i\sigma_{ei}^2$ when $\sigma_{ei}^{-2}\sigma_b^2$ is small and underestimates $\gamma_i\sigma_{ei}^2$ when $\sigma_{ei}^{-2}\sigma_b^2$ is large. It is not appropriate to simply use $\widehat{\gamma}_i\widehat{\sigma}_{ei}^2 + (\widehat{\sigma}_{b,m}^2 + \widehat{\sigma}_{ei}^2)\widehat{Var}\{\widehat{\gamma}_i\}$ to estimate $\gamma_i\sigma_{ei}^2$. We need to derive a better estimator to $\gamma_i\sigma_{ei}^2$.

By (3.52), we have

$$E\left\{(\widehat{\gamma}_i - \gamma_i)^2(b_i + e_i)^2\right\} = (\sigma_b^2 + \sigma_{ei}^2)Var\{\widehat{\gamma}_i\} + O(r_m). \tag{3.69}$$

Therefore, the $(\sigma_b^2 + \sigma_{ei}^2)Var\{\widehat{\gamma}_i\}$ term in (3.46) is due to the approximation to $E\left\{(\widehat{\gamma}_i - \gamma_i)^2(b_i + e_i)^2\right\}$. We will use the approximation employed by Dodd (1999) to estimate $E\left\{(\widehat{\gamma}_i - \gamma_i)^2(b_i + e_i)^2\right\}$ directly instead of estimating $(\sigma_b^2 + \sigma_{ei}^2)Var\{\widehat{\gamma}_i\}$.

The idea is to approximate the standard normal distribution by a set of points $\{(c_k, w_k)\}_{k=1}^9$, where $(c_1, \ldots, c_9)^T = (-2.1, -1.3, -0.8, -0.5, 0, 0.5, 0.8, 1.3, 2.1)^T$ and $(w_1, \ldots, w_9)^T = (0.063345,\ 0.080255,\ 0.070458,\ 0.159698,\ 0.252489,\ 0.159698,\ 0.070458,$

$0.080255, 0.063345)^T$. Note that

$$\sum_{k=1}^{9} w_k = 1, \quad \sum_{l=k}^{9} w_k c_k^r = 0 \text{ for } r = 1, 3, 5,$$

and

$$\sum_{k=1}^{9} w_k c_k^2 = 1, \quad \sum_{k=1}^{9} w_k c_k^4 = 3.$$

Let $\widetilde{\sigma}^2_{b,m}$ be the non-truncated estimator of σ_b^2 and let $\widehat{\sigma}^2_{b,m}$ be the truncated estimator of σ_b^2. The $\widetilde{\sigma}^2_{b,m}$ is approximately normal distributed in large samples. Thus we can approximate the distribution of $\widehat{\sigma}^2_{b,m} = \max(0, \widetilde{\sigma}^2_{b,m})$ with a truncated normal distribution. Let

$$\ddot{\sigma}^2_{bk} = \max(0, \widetilde{\sigma}^2_{b,m} + c_k s_{bm}),$$

where $s_{bm} = \sqrt{\widehat{V}(\widehat{\sigma}^2_{b,m})}$ is the estimated standard error of $\widehat{\sigma}^2_{b,m}$. Then $\{(\ddot{\sigma}^2_{bk}, w_k)\}_{k=1}^9$ would be a reasonable approximation to the distribution of $\widehat{\sigma}^2_{b,m}$. Suppose $U \sim \chi^2_d$, then

$$(d^{-1}U)^{\frac{1}{3}} \dot{\sim} N(\mu_d, s_d^2), \tag{3.70}$$

where $(\mu_5, s_5) = (0.953, 0.2047)$, $(\mu_9, s_9) = (0.970, 0.170)$, and $(\mu_{14}, s_{14}) = (0.983, 0.123)$. Note that $d_i \sigma_{ei}^{-2} \widehat{\sigma}^2_{ei} \sim \chi^2_{d_i}$. Let

$$\ddot{\sigma}^2_{eik} = \max\left(0, (\mu_{d_i} + c_k s_{d_i})^3 \widehat{\sigma}^2_{ei}\right),$$

where μ_{d_i}, and s_{d_i} are defined in (3.70). Then $\{(\ddot{\sigma}^2_{eik}, w_k)\}_{k=1}^9$ would be a reasonable approximation to the distribution of $\widehat{\sigma}^2_{ei}$. The covariance between $\widehat{\sigma}^2_{b,m}$ and $\widehat{\sigma}^2_{ei}$ is $O(r_m)$. Therefore, we can approximate the joint distribution of $(\widehat{\sigma}^2_{b,m}, \widehat{\sigma}^2_{ei})$ by these two independent quantities.

Note that

$$cov\left\{(b_i + e_i)^2, \widehat{\sigma}^2_{ei}\right\} = 0,$$

and

$$cov\left\{(b_i + e_i)^2, \widehat{\sigma}^2_{b,m}\right\} = 2m^{-1}(\sigma_b^2 + \sigma_{ei}^2)^2.$$

Therefore,

$$E\left\{(b_i+e_i)^2|\widehat{\sigma}^2_{b.m}=\ddot{\sigma}^2_{bk},\widehat{\sigma}^2_{ei}=\ddot{\sigma}^2_{eik}\right\}=\widehat{\sigma}^2_{b.m}+\alpha_b(\ddot{\sigma}^2_{bk}-\widehat{\sigma}^2_{b.m})+\widehat{\sigma}^2_{ei}, \tag{3.71}$$

where $\alpha_b=[\widehat{V}(\widehat{\sigma}^2_{b,m})]^{-1}[2m^{-1}(\widehat{\sigma}^2_{b,m}+\widehat{\sigma}^2_{ei})^2]$. Since

$$E\left\{[(\widehat{\gamma}_i-\gamma_i)(b_i+e_i)]^2\right\}=E\left\{E\{[(\widehat{\gamma}_i-\gamma_i)(b_i+e_i)]^2|\widehat{\sigma}^2_{b,m},\widehat{\sigma}^2_{ei}\}\right\},$$

an approximation to $E\{[(\widehat{\gamma}_i-\gamma_i)(b_i+e_i)]^2\}$ is

$$\widehat{E}\left\{[(\widehat{\gamma}_i-\gamma_i)(b_i+e_i)]^2\right\}=\sum_{j=1}^{9}\sum_{k=1}^{9}w_jw_k(\ddot{\gamma}_{ijk}-\widetilde{\gamma}_i)^2[\widehat{\sigma}^2_{b,m}+\alpha_b(\ddot{\sigma}^2_{bk}-\widehat{\sigma}^2_{b,m})+\widehat{\sigma}^2_{ei}]. \tag{3.72}$$

where

$$\ddot{\gamma}_{ijk}=(\ddot{\sigma}^2_{bj}+\ddot{\sigma}^2_{eik})^{-1}\ddot{\sigma}^2_{bj}$$

and

$$\widetilde{\gamma}_i=\sum_{j=1}^{9}\sum_{k=1}^{9}w_jw_k\ddot{\gamma}_{ijk}.$$

Now we derive a better estimator to $\gamma_i\sigma^2_{ei}$. Let

$$\widehat{V}^*(\widehat{\sigma}^2_{b,m})=\widehat{V}(\widehat{\sigma}^2_{b,m})\max\left\{0,\min(1,0.6[\widehat{V}(\widehat{\sigma}^2_{b,m})]^{-0.5}\widehat{\sigma}^2_{b,m})\right\}. \tag{3.73}$$

The $\widehat{V}^*(\widehat{\sigma}^2_{b,m})$ is an estimator to the variance of $\widehat{\sigma}^2_{b,m})$ that is zero if $\widehat{\sigma}^2_{b,m}=0$ and is $\widehat{V}(\widehat{\sigma}^2_{b,m})$ if $\widehat{\sigma}^2_{b,m}$ is greater than $0.6^{-1}[\widehat{V}(\widehat{\sigma}^2_{b,m})]^{0.5}$. Thus it is a better approximation to the variance of $\widehat{\sigma}^2_{b,m}$ when σ^2_b is small.

Assume $E(\overline{x})=\mu_x$ and $E(\overline{y})=\mu_y$. Then $[(\overline{y})^2+var(\overline{y})]^{-1}[(\overline{x})(\overline{y})+cov(\overline{x},\overline{y})]$ is a better estimator of $\mu_y^{-1}\mu_x$ than $(\overline{y})^{-1}\overline{x}$. Using this idea, an improved estimator of $\gamma_i\sigma^2_{ei}$ is

$$\begin{aligned}\widehat{\gamma_i\sigma^2_{ei}} &= [(\widehat{\sigma}^2_{b,m}+\widehat{\sigma}^2_{ei})^2+\widehat{V}^*(\widehat{\sigma}^2_{b,m})+2d^{-1}(1-2m^{-1})\widehat{\sigma}^4_{ei}]^{-1}\\ &\quad\times[(\widehat{\sigma}^2_{b,m}+\widehat{\sigma}^2_{ei})\widehat{\sigma}^2_{b,m}\widehat{\sigma}^2_{ei}+\widehat{\sigma}^2_{ei}\widehat{V}^*(\widehat{\sigma}^2_{b,m})+\widehat{\sigma}^2_{b,m}2d^{-1}\widehat{\sigma}^4_{ei}], \end{aligned} \tag{3.74}$$

where $\widehat{V}^*(\widehat{\sigma}^2_{b,m})$ is defined in (3.73). This leads to the estimator

$$\begin{aligned}\widehat{MSE}_2(\widehat{y}_i-y_i) &= \widehat{\gamma_i\sigma^2_{ei}}+(1-\widehat{\gamma}_i)^2\boldsymbol{x}_i^T\widehat{V}(\widehat{\boldsymbol{\beta}}_0)\boldsymbol{x}_i\\ &\quad+\widehat{E}\left\{[(\widehat{\gamma}_i-\gamma_i)(b_i+e_i)]^2\right\}, \end{aligned} \tag{3.75}$$

where $\widehat{\gamma_i\sigma_{ei}^2}$ is defined in (3.74) and $\widehat{E}\{[(\widehat{\gamma}_i - \gamma_i)(b_i + e_i)]^2\}$ is defined in (3.72).

For our simulation study, $\boldsymbol{x}_i^T\widehat{V}(\widehat{\boldsymbol{\beta}}_0)\boldsymbol{x}_i = (\widehat{\sigma}_{b,m}^2 + \overline{\widehat{\sigma}_{ei}^2})m^{-1}$ because we use the simple mean as the estimator of μ. For each sample, we compute $\widehat{MSE}_2(\widehat{y}_i - y_i)$ defined in (3.75) and the corresponding mean for the G-th group, which is $\widehat{MSE}_2(\widehat{y}_i - y_i)_G = 3m^{-1}\sum_{i\in G}\widehat{MSE}_2(\widehat{y}_i - y_i)$. The mean of the 1000 simulated $\widehat{MSE}_2(\widehat{y}_i - y_i)_G$ for each group is reported in the sixth columns of Table 3.1 through Table 3.5. From these tables, we can see that $\widehat{MSE}_2(\widehat{y}_i - y_i)$ approximates both the simulated $MSE(\widehat{y}_i - y_i)$ and theoretical $MSE(\widehat{y}_i - y_i)$ much better than $\widehat{MSE}_1(\widehat{y}_i - y_i)$. While the overestimation at $\sigma_b^2 \doteq 0$ is still very large with $\widehat{MSE}_2(\widehat{y}_i - y_i)$, it is much smaller than the overestimation of $\widehat{MSE}_1(\widehat{y}_i - y_i)$. For large values of σ_b^2, the two procedures yield very similar estimates.

Table 3.1 Simulation study of $MSE(\widehat{y}_i - y_i)$ $(d = 5, m = 36)$

σ_b^2	γ_i	Simulated $MSE(\widehat{y}_i - y_i)$	Theoretical $MSE(\widehat{y}_i - y_i)$	Estimated $\widehat{MSE}_1(\widehat{y}_i - y_i)$	Estimated $\widehat{MSE}_2(\widehat{y}_i - y_i)$
0.010	0.0099	0.104	0.183	0.615	0.222
	0.0066	0.114	0.140	0.489	0.243
	0.0050	0.126	0.118	0.419	0.257
0.100	0.0909	0.193	0.244	0.598	0.273
	0.0625	0.208	0.218	0.507	0.304
	0.0476	0.223	0.203	0.453	0.323
0.250	0.2	0.313	0.336	0.594	0.341
	0.143	0.342	0.336	0.547	0.388
	0.111	0.372	0.332	0.511	0.415
0.429	0.3	0.409	0.426	0.594	0.404
	0.222	0.465	0.457	0.576	0.464
	0.176	0.497	0.470	0.569	0.504
0.667	0.4	0.510	0.518	0.605	0.490
	0.308	0.594	0.589	0.637	0.575
	0.25	0.657	0.628	0.656	0.630
1.000	0.5	0.609	0.612	0.648	0.604
	0.4	0.753	0.734	0.744	0.730
	0.333	0.823	0.810	0.819	0.832
1.500	0.6	0.700	0.704	0.706	0.706
	0.5	0.873	0.888	0.880	0.891
	0.429	1.020	1.020	1.010	1.040
2.333	0.7	0.757	0.791	0.792	0.795
	0.609	1.030	1.050	1.030	1.040
	0.538	1.230	1.250	1.210	1.230

Table 3.2 Simulation study of $MSE(\widehat{y}_i - y_i)$ $(d = 9, m = 36)$

σ_b^2	γ_i	Simulated $MSE(\widehat{y}_i - y_i)$	Theoretical $MSE(\widehat{y}_i - y_i)$	Estimated $\widehat{MSE}_1(\widehat{y}_i - y_i)$	Estimated $\widehat{MSE}_2(\widehat{y}_i - y_i)$
0.010	0.0099	0.099	0.183	0.487	0.214
	0.0066	0.104	0.140	0.405	0.231
	0.0050	0.108	0.118	0.361	0.242
0.100	0.0909	0.180	0.243	0.486	0.260
	0.0625	0.189	0.217	0.428	0.284
	0.0476	0.209	0.202	0.398	0.299
0.250	0.2	0.289	0.330	0.503	0.310
	0.143	0.317	0.331	0.465	0.344
	0.111	0.323	0.329	0.445	0.365
0.429	0.3	0.404	0.414	0.524	0.386
	0.222	0.448	0.446	0.524	0.438
	0.176	0.476	0.461	0.530	0.474
0.667	0.4	0.485	0.501	0.562	0.478
	0.308	0.559	0.572	0.606	0.555
	0.25	0.611	0.612	0.634	0.607
1.000	0.5	0.582	0.590	0.615	0.586
	0.4	0.704	0.708	0.721	0.704
	0.333	0.775	0.783	0.790	0.787
1.500	0.6	0.675	0.678	0.687	0.691
	0.5	0.848	0.855	0.856	0.862
	0.429	0.972	0.979	0.974	0.988
2.333	0.7	0.765	0.765	0.771	0.781
	0.609	1.010	1.010	1.000	1.010
	0.538	1.190	1.200	1.180	1.200

Table 3.3 Simulation study of $MSE(\widehat{y}_i - y_i)$ $(d = 9, m = 99)$

σ_b^2	γ_i	Simulated $MSE(\widehat{y}_i - y_i)$	Theoretical $MSE(\widehat{y}_i - y_i)$	Estimated $\widehat{MSE}_1(\widehat{y}_i - y_i)$	Estimated $\widehat{MSE}_2(\widehat{y}_i - y_i)$
0.010	0.0099	0.049	0.073	0.218	0.129
	0.0066	0.049	0.057	0.189	0.135
	0.0050	0.048	0.049	0.173	0.138
0.100	0.0909	0.136	0.147	0.246	0.177
	0.0625	0.139	0.139	0.228	0.189
	0.0476	0.143	0.135	0.219	0.196
0.250	0.2	0.248	0.252	0.301	0.261
	0.143	0.263	0.260	0.304	0.285
	0.111	0.271	0.264	0.305	0.299
0.429	0.3	0.350	0.350	0.366	0.352
	0.222	0.389	0.383	0.396	0.395
	0.176	0.409	0.399	0.412	0.421
0.667	0.4	0.456	0.450	0.454	0.458
	0.308	0.523	0.516	0.520	0.533
	0.25	0.561	0.554	0.558	0.580
1.000	0.5	0.551	0.550	0.550	0.559
	0.4	0.658	0.660	0.659	0.677
	0.333	0.749	0.730	0.733	0.761
1.500	0.6	0.644	0.649	0.645	0.653
	0.5	0.810	0.815	0.810	0.827
	0.429	0.941	0.931	0.924	0.952
2.333	0.7	0.737	0.744	0.739	0.744
	0.609	0.972	0.979	0.967	0.980
	0.538	1.150	1.160	1.140	1.160

Table 3.4 Simulation study of $MSE(\widehat{y}_i - y_i)$ $(d = 14, m = 99)$

σ_b^2	γ_i	Simulated $MSE(\widehat{y}_i - y_i)$	Theoretical $MSE(\widehat{y}_i - y_i)$	Estimated $\widehat{MSE}_1(\widehat{y}_i - y_i)$	Estimated $\widehat{MSE}_2(\widehat{y}_i - y_i)$
0.010	0.0099	0.047	0.073	0.201	0.121
	0.0066	0.045	0.057	0.175	0.125
	0.0050	0.044	0.049	0.162	0.126
0.100	0.0909	0.129	0.147	0.233	0.168
	0.0625	0.131	0.139	0.217	0.177
	0.0476	0.135	0.134	0.209	0.181
0.250	0.2	0.246	0.249	0.286	0.246
	0.143	0.261	0.258	0.290	0.265
	0.111	0.266	0.262	0.292	0.276
0.429	0.3	0.350	0.345	0.357	0.336
	0.222	0.380	0.378	0.386	0.374
	0.176	0.397	0.395	0.402	0.396
0.667	0.4	0.448	0.443	0.444	0.440
	0.308	0.511	0.508	0.508	0.507
	0.25	0.556	0.546	0.547	0.550
1.000	0.5	0.542	0.540	0.533	0.534
	0.4	0.648	0.648	0.637	0.640
	0.333	0.727	0.718	0.705	0.712
1.500	0.6	0.643	0.637	0.637	0.638
	0.5	0.801	0.800	0.798	0.801
	0.429	0.912	0.914	0.912	0.918
2.333	0.7	0.735	0.733	0.731	0.730
	0.609	0.954	0.962	0.960	0.960
	0.538	1.130	1.140	1.130	1.130

Table 3.5 Simulation study of $MSE(\widehat{y}_i - y_i)$ $(d = 14, m = 225)$

σ_b^2	γ_i	Simulated $MSE(\widehat{y}_i - y_i)$	Theoretical $MSE(\widehat{y}_i - y_i)$	Estimated $\widehat{MSE}_1(\widehat{y}_i - y_i)$	Estimated $\widehat{MSE}_2(\widehat{y}_i - y_i)$
0.010	0.0099	0.027	0.038	0.111	0.077
	0.0066	0.026	0.031	0.099	0.078
	0.0050	0.025	0.027	0.093	0.078
0.100	0.0909	0.110	0.116	0.150	0.126
	0.0625	0.112	0.114	0.144	0.130
	0.0476	0.112	0.113	0.141	0.132
0.250	0.2	0.225	0.224	0.238	0.230
	0.143	0.239	0.236	0.249	0.246
	0.111	0.243	0.242	0.254	0.255
0.429	0.3	0.327	0.325	0.327	0.328
	0.222	0.362	0.358	0.360	0.364
	0.176	0.378	0.376	0.379	0.385
0.667	0.4	0.431	0.427	0.425	0.427
	0.308	0.501	0.490	0.488	0.494
	0.25	0.536	0.528	0.527	0.535
1.000	0.5	0.527	0.528	0.528	0.530
	0.4	0.632	0.633	0.633	0.638
	0.333	0.704	0.701	0.703	0.711
1.500	0.6	0.625	0.628	0.622	0.623
	0.5	0.792	0.787	0.776	0.780
	0.429	0.897	0.899	0.886	0.893
2.333	0.7	0.723	0.726	0.724	0.724
	0.609	0.947	0.952	0.946	0.948
	0.538	1.120	1.130	1.120	1.120

4 ESTIMATION UNDER A RESTRICTION

4.1 Introduction

We are considering estimation in the situation where the direct survey estimator Y_i for small area i has a large variance as an estimator of y_i due to small sample size. However, the direct survey estimator of the total across small areas is often satisfactory. In the example of Chapter 2, the direct survey estimator for urban change acres D_i has a large variance as an estimator of the true urban change acres d_i in HUCCO i. However, the direct survey estimator for state urban change acres $\sum_i D_i$ is design unbiased for the true urban state change acres, and it has relatively small variance.

Therefore, it is desirable to put a restriction on the weighted total of the small area estimators such that the weighted total of the small area estimators is equal to the the weighted total of the direct survey estimators. Equivalently, we can put the restriction on the weighted mean of the small area estimators. Thus, we want to adjust the small area estimators such that

$$\sum_{i=1}^{m} \omega_i \widehat{y}_i^M = \sum_{i=1}^{m} \omega_i Y_i, \tag{4.1}$$

where $\omega_i > 0, i = 1, \cdots, m$, are the weights, $\sum_{i=1}^{m} \omega_i = 1$, and $\widehat{y}_i^M$ is the adjusted small area estimator. Usually the ω_i are the sampling weights such that $\sum_{i=1}^{m} \omega_i Y_i$ is an unbiased estimator of the population mean. One heuristic approach is to make a ratio adjustment

$$\widehat{y}_i^M = \left(\sum_{j=1}^{m} \omega_j \widehat{y}_j^H\right)^{-1} \left(\sum_{j=1}^{m} \omega_j Y_j\right) \widehat{y}_i^H, \tag{4.2}$$

where $\widehat{y}_i^H$ is the EBLUP of y_i. This adjustment was made in Chapter 2. The disadvantage of this approach is that it is hard to assess the bias and variance of $\widehat{y}_i^M$.

Pfeffermann and Barnard (1991) proposed an alternative approach. The mixed model equation for the random effects model defined in (1.3) is

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{\Sigma}_e^{-1}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{\Sigma}_e^{-1}\mathbf{Z} \\ \mathbf{Z}'\boldsymbol{\Sigma}_e^{-1}\boldsymbol{X} & \mathbf{Z}'\boldsymbol{\Sigma}_e^{-1}\mathbf{Z} + \boldsymbol{\Sigma}_b^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{\Sigma}_e^{-1}\mathbf{Y} \\ \mathbf{Z}'\boldsymbol{\Sigma}_e^{-1}\mathbf{Y} \end{bmatrix}. \tag{4.3}$$

where $\boldsymbol{\Sigma}_e = diag(\sigma_{e1}^2, \cdots, \sigma_{em}^2)$ and $\boldsymbol{\Sigma}_b = diag(z_1^2\sigma_b^2, \cdots, z_m^2\sigma_b^2)$. Let $\widetilde{\boldsymbol{y}}^H = (\widetilde{y}_1^H, \cdots, \widetilde{y}_m^H)^T$ denote the BLUP estimator of $\boldsymbol{y} = (y_1, \cdots, y_m)^T$. We have

$$\widetilde{\boldsymbol{y}}^H = \boldsymbol{X}\widehat{\boldsymbol{\beta}} + \mathbf{Z}\widehat{\boldsymbol{b}}, \tag{4.4}$$

where $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{b}}$ are any solutions to the mixed model equation (4.3). Note that finding a solution to the mixed model equation (4.3) is equivalent to finding a solution to the minimization problem

$$\min_{\boldsymbol{\beta}, \boldsymbol{b}} \left\{ (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{b})^T \boldsymbol{\Sigma}_e^{-1} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{b}) + \boldsymbol{b}^T \boldsymbol{\Sigma}_b^{-1} \boldsymbol{b} \right\}. \tag{4.5}$$

To make (4.1) hold, Pfeffermann and Barnard (1991) proposed the modified estimator

$$\widehat{\boldsymbol{y}}^M = \boldsymbol{X}\widehat{\boldsymbol{\beta}}^M + \mathbf{Z}\widehat{\boldsymbol{b}}^M, \tag{4.6}$$

where $\widehat{\boldsymbol{\beta}}^M$ and $\widehat{\boldsymbol{b}}^M$ are any solutions to the minimization problem (4.5) with $\boldsymbol{\beta}$ and $\boldsymbol{b}$ subject to the constraint

$$\sum_{i=1}^m \omega_i(\boldsymbol{x}_i^T\boldsymbol{\beta} + z_i b_i) = \sum_{i=1}^m \omega_i Y_i. \tag{4.7}$$

This leads to the estimator

$$\widehat{y}_i^M = \widetilde{y}_i^H + cov(\widetilde{y}_i^H, \widetilde{y}_.)[Var(\widetilde{y}_.)]^{-1}[\sum_{j=1}^m \omega_j Y_j - \widetilde{y}_.], \tag{4.8}$$

where $\widetilde{y}_. = \sum_{i=1}^m \omega_i \widetilde{y}_i^H$. They did not give the expression for $cov(\widetilde{y}_i^H, \widetilde{y}_.)$ and $Var(\widetilde{y}_.)$ in their paper. They gave the formula for $E\{(\boldsymbol{\beta}', \boldsymbol{b}')'(\boldsymbol{\beta}', \boldsymbol{b}')\}$ when the variance components parameters are known. They did not give a formula for $MSE(\widehat{y}_i^M - y_i)$ in their paper.

The Pfeffermann-Barnard (1991) approach is a natural way to make the estimator $\widehat{y}_i^M$ satisfy (4.1). The derivation of (4.8) relies on the fact that $\boldsymbol{b}$ can be treated as a fixed parameter. When there is no restriction, Henderson (1950) showed that we can solve the mixed model equation (4.3) and estimate $\boldsymbol{b}$ as a fixed parameter. However, (4.7) puts a constraint on the random vector $\boldsymbol{b}$. The constraint (4.7) makes the distribution of $\boldsymbol{b}$ a degenerate one. Thus, the variance structure of $\boldsymbol{b}$ is changed and $Var(\boldsymbol{b}) \neq diag(z_1^2\sigma_b^2, \ldots, z_m^2\sigma_b^2)$. The underlying model assumptions about the random effects model (1.3) have been changed by the constraint. It is not clear that the estimation of $\boldsymbol{b}$ as a fixed parameter is still justified.

In light of this fact, we consider restriction (4.1) as an adjustment problem instead of a constraint problem proposed by Pfeffermann and Barnard. Suppose we have the small area estimator $\widehat{y}_i$. To make the final estimator $\widehat{y}_i^M$ satisfy (4.1), we allocate the difference $\sum_{j=1}^{m} \omega_j Y_j - \sum_{j=1}^{m} \omega_j \widehat{y}_j$ to the small area estimators $\widehat{y}_i$, $i = 1, \cdots, m$, according to a rule. We define the modified (adjusted) estimator

$$\widehat{y}_i^M = \widehat{y}_i + a_i \left[\sum_{j=1}^{m} \omega_j Y_j - \sum_{j=1}^{m} \omega_j \widehat{y}_j \right], \tag{4.9}$$

where $\sum_{i=1}^{m} \omega_i a_i = 1$. Clearly, the modified estimator satisfies (4.1).

The estimator defined in (4.2) is an estimator of the form in (4.9) since

$$\begin{aligned} \widehat{y}_i^M &= \left(\sum_{j=1}^{m} \omega_j \widehat{y}_j^H \right)^{-1} \left(\sum_{j=1}^{m} \omega_j Y_j \right) \widehat{y}_i^H \\ &= \widehat{y}_i^H + \left(\sum_{j=1}^{m} \omega_j \widehat{y}_j^H \right)^{-1} \widehat{y}_i^H \left(\sum_{j=1}^{m} \omega_j Y_j - \sum_{j=1}^{m} \omega_j \widehat{y}_j^H \right). \end{aligned} \tag{4.10}$$

The estimator defined in (4.8) is also of the form (4.9) though the estimator is derived from a constraint minimization problem. An estimator similar to (4.8) proposed by Battese, Harter and Fuller (1988) is

$$\widehat{y}_i^M = \widehat{y}_i^H + \left[\sum_{j=1}^{m} \omega_j^2 \widehat{Var}(\widehat{y}_j^H) \right]^{-1} \omega_i \widehat{Var}(\widehat{y}_i^H) \left(\sum_{j=1}^{m} \omega_j Y_j - \sum_{j=1}^{m} \omega_j \widehat{y}_j^H \right). \tag{4.11}$$

Isaki, Tsay and Fuller (1999) imposed the restriction by a procedure that, approximately, constructed the best predictors of $m-1$ quantities that are estimated to be uncorrelated with $\sum_{i=1}^{m} \omega_i Y_i$. Let $\boldsymbol{w} = (w_1, \cdots, w_m)^T$, and

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_e = Var(\boldsymbol{Y}). \tag{4.12}$$

Let $\widehat{\boldsymbol{\Sigma}}$ be the estimator of $\boldsymbol{\Sigma}$ and let $\ddot{\boldsymbol{C}} = \ddot{\boldsymbol{A}}\boldsymbol{T}$, where

$$\boldsymbol{T} = \begin{pmatrix} & \boldsymbol{\omega}' \\ \boldsymbol{0}_{m-1} & \boldsymbol{I}_{m-1} \end{pmatrix},$$

$$\ddot{\boldsymbol{A}} = \begin{pmatrix} 1 & \boldsymbol{0}'_{m-1} \\ -\ddot{\boldsymbol{a}}_{m-1} & \boldsymbol{I}_{m-1} \end{pmatrix},$$

$$\ddot{\boldsymbol{a}}_{m-1} = \left(\boldsymbol{\omega}^T \widehat{\boldsymbol{\Sigma}} \boldsymbol{\omega}\right)^{-1} \begin{pmatrix} \boldsymbol{0}_{m-1} & \boldsymbol{I}_{m-1} \end{pmatrix} \widehat{\boldsymbol{\Sigma}} \boldsymbol{\omega}.$$

The modified estimator of $\boldsymbol{y}$ is

$$\widehat{\boldsymbol{y}}^M = \boldsymbol{Y} - \ddot{\boldsymbol{C}}^{-1} \boldsymbol{B} \ddot{\boldsymbol{C}} (\boldsymbol{I}_m - \widehat{\boldsymbol{\Gamma}})(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}), \tag{4.13}$$

where $\widehat{\boldsymbol{\Gamma}} = \widehat{\boldsymbol{\Sigma}}_b \widehat{\boldsymbol{\Sigma}}^{-1}$, and

$$\boldsymbol{B} = \begin{pmatrix} 0 & \boldsymbol{0}'_{m-1} \\ \boldsymbol{0}_{m-1} & \boldsymbol{I}_{m-1} \end{pmatrix}. \tag{4.14}$$

Isaki, Tsay and Fuller argued that the estimator in (4.13) gave the BLUP of the $m-1$ quantities orthogonal to $\sum_{i=1}^{m} \omega_i Y_i$ when variance components parameters are known, but no other theoretical justification was provided for the particular choice of $\ddot{\boldsymbol{a}}_{m-1}$.

Let $\ddot{\boldsymbol{H}} = \ddot{\boldsymbol{C}}^{-1} \boldsymbol{B} \ddot{\boldsymbol{C}} (\boldsymbol{I}_m - \widehat{\boldsymbol{\Gamma}})$, $\widehat{\boldsymbol{V}}_\beta$ be an estimator of the contribution to the variance due to estimating $\boldsymbol{\beta}$, $\widehat{\boldsymbol{\Omega}}_{33}$ be an estimator of the contribution to the variance due to using $\widehat{\boldsymbol{\Sigma}}_b$ to estimate $\boldsymbol{\Sigma}_b$, and $\widehat{\boldsymbol{\Omega}}_{44}$ be an estimator of the contribution to the variance due to using $\widehat{\boldsymbol{\Sigma}}_e$ to estimate $\boldsymbol{\Sigma}_e$. Isaki et. al. also proposed an estimator of $var\left\{\widehat{\boldsymbol{y}}^M - \boldsymbol{y}\right\}$ as

$$(\boldsymbol{I}_m - \ddot{\boldsymbol{H}}) \widehat{\boldsymbol{\Sigma}}_e (\boldsymbol{I}_m - \ddot{\boldsymbol{H}})' + \ddot{\boldsymbol{H}} \widehat{\boldsymbol{\Sigma}}_b \ddot{\boldsymbol{H}}' + \ddot{\boldsymbol{C}}^{-1} \boldsymbol{B} \ddot{\boldsymbol{C}} (\ddot{\boldsymbol{H}} \boldsymbol{X} \widehat{\boldsymbol{V}}_\beta \boldsymbol{X}' \ddot{\boldsymbol{H}}' + \widehat{\boldsymbol{\Omega}}_{33} + \widehat{\boldsymbol{\Omega}}_{44}) \ddot{\boldsymbol{C}} \boldsymbol{B} \ddot{\boldsymbol{C}}'^{-1}.$$

For the random effects model defined in (1.3), $\widehat{\boldsymbol{\Sigma}} = diag(z_1^2\widehat{\sigma}_b^2 + \widehat{\sigma}_{e1}^2, \cdots, z_m^2\widehat{\sigma}_b^2 + \widehat{\sigma}_{em}^2)$, $\widehat{\boldsymbol{\Gamma}} = diag(\widehat{\gamma}_1, \cdots, \widehat{\gamma}_m)$, and $\widehat{\gamma}_i = (z_i^2\widehat{\sigma}_b^2 + \widehat{\sigma}_{ei}^2)^{-1} z_i^2\widehat{\sigma}_b^2$. After some matrix operation, we can rewrite (4.13) in component form, rather than matrix form, as

$$\begin{aligned}
\widehat{y}_i^M &= \widetilde{y}_i^H + \left[\sum_{j=1}^{m} \omega_j^2 (z_j^2\widehat{\sigma}_b^2 + \widehat{\sigma}_{ej}^2)\right]^{-1} \omega_i (z_i^2\widehat{\sigma}_b^2 + \widehat{\sigma}_{ei}^2) \left[\sum_{j=1}^{m} \omega_j Y_j - \sum_{j=1}^{m} \omega_j \widetilde{y}_j^H\right] \\
&= \widetilde{y}_i^H + \left[\sum_{j=1}^{m} \omega_j^2 \widehat{Var}(Y_j)\right]^{-1} \omega_i \widehat{Var}(Y_i) \left(\sum_{j=1}^{m} \omega_j Y_j - \sum_{j=1}^{m} \omega_j \widetilde{y}_j^H\right). \qquad (4.15)
\end{aligned}$$

Therefore, the modified estimator in (4.13) also has the form of (4.9).

4.2 Best linear unbiased estimator under a restriction

We want to find the "best" linear unbiased estimator for $\boldsymbol{y}$ that satisfies restriction (4.1). Just as in the derivation of BLUP, we first assume the parameters for the variance components are known. Let $R(\widehat{\boldsymbol{y}})$ denote the collection of all linear unbiased estimators that satisfies (4.1). Suppose the BLUP of $\boldsymbol{y} = (y_1, \cdots, y_m)^T$ is $\widetilde{\boldsymbol{y}}^H = (\widetilde{y}_1^H, \cdots, \widetilde{y}_m^H)^T$ and $\widetilde{\boldsymbol{y}}^H \notin R(\widehat{\boldsymbol{y}})$.

First of all, we need to define the meaning of "best". We cannot obtain the BLUP for all $y_i, i = 1, \ldots, m$ anymore. Consider a family of estimators $\widetilde{\boldsymbol{y}}^{(j)}$, where

$$\widetilde{y}_i^{(j)} = \begin{cases} \widetilde{y}_i^H & \text{if } i \neq j, \\ \widetilde{y}_i^H + \omega_i^{-1} \left[\sum_{j=1}^{m} \omega_j Y_j - \sum_{j=1}^{m} \omega_j \widetilde{y}_j^H\right] & \text{if } i = j. \end{cases} \qquad (4.16)$$

In other words, $\widetilde{\boldsymbol{y}}^{(j)}$ is the estimator in which the i-th component is the BLUP of y_i but the j-th component is the BLUP of y_j plus ω_i^{-1} multiplied by the difference $\left[\sum_{j=1}^{m} \omega_j Y_j - \sum_{j=1}^{m} \omega_j \widetilde{y}_j^H\right]$. It is easy to see that $\widetilde{\boldsymbol{y}}^{(j)} \in R(\widehat{\boldsymbol{y}})$. For any $\widehat{\boldsymbol{y}}^M \in R(\widehat{\boldsymbol{y}})$, there is at least one component $\widehat{y}_k^M \neq \widetilde{y}_k^H$ since $\widetilde{\boldsymbol{y}}^H \notin R(\widehat{\boldsymbol{y}})$. Therefore,

$$Var(\widehat{y}_k - y_k) > Var(\widetilde{y}_k^{(j)} - y_k) = Var(\widetilde{y}_k^H - y_k) \qquad (4.17)$$

for any $j \neq k$ because $\tilde{y}_k^H$ is the BLUP of y_k. This indicates that no estimator can be found in $R(\hat{\boldsymbol{y}})$ with smallest prediction variance for every component. There is always another estimator in $R(\hat{\boldsymbol{y}})$ with smaller prediction variance for at least one component.

Since it is impossible to compare estimators in $R(\hat{\boldsymbol{y}})$ component-by-component to find the best estimator, some kind of overall criterion is desirable. A natural choice is to find $\hat{\boldsymbol{y}}^M \in R(\hat{\boldsymbol{y}})$ that minimizes

$$Q(\hat{\boldsymbol{y}}^M) = \sum_{i=1}^{m} \varphi_i E(\hat{y}_i^M - y_i)^2, \tag{4.18}$$

where the $\varphi_i, i = 1, \ldots, m$ are positive weights. Usually, φ_i depends on the variance components. To obtain a general result, we do not specify φ_i now. We will discuss the choice of φ_i later. We give the major theorem of this chapter.

Theorem 4.1 *Assume the random effects model*

$$Y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + z_i b_i + e_i, \qquad i = 1, \ldots, m,$$

where the b_i have independent identical distributions with mean zero and variance σ_b^2, the e_i have independent distributions with mean zero and variance σ_{ei}^2, and $\boldsymbol{b} = (b_1, \ldots, b_m)^T$ is independent of $\boldsymbol{e} = (e_1, \ldots, e_m)^T$. Assume z_i, σ_b^2, and σ_{ei}^2 are known and $\boldsymbol{\beta}$ is unknown. Let $\tilde{y}_i^H$ be the BLUP of y_i defined in (1.8). Let

$$\hat{y}_i^M = \tilde{y}_i^H + \breve{a}_i \left(\sum_{j=1}^{m} \omega_j Y_j - \sum_{j=1}^{m} \omega_j \tilde{y}_j^H \right), \tag{4.19}$$

where $\breve{a}_i = \left(\sum_{i=1}^{m} \varphi_i^{-1} \omega_i^2 \right)^{-1} \varphi_i^{-1} \omega_i$ and ω_i are the fixed weights of (4.1). Then $\hat{\boldsymbol{y}}^M = (\hat{y}_1^M, \ldots, \hat{y}_m^M)^T$ is the **unique** *estimator among all linear unbiased estimators that satisfies (4.1) and minimizes criterion (4.18).*

Proof: Let $R^0(\tilde{\boldsymbol{y}}_{\boldsymbol{a}})$ denote the collection of all estimators that have the form

$$\tilde{y}_{i\boldsymbol{a}}^M = \tilde{y}_i^H + a_i \left(\sum_{j=1}^{m} \omega_j Y_j - \sum_{j=1}^{m} \omega_j \tilde{y}_j^H \right), \tag{4.20}$$

where $\sum_{i=1}^{m} \omega_i a_i = 1$. Clearly, $R^0(\widetilde{\boldsymbol{y}}_{\boldsymbol{a}})$ is a subset of $R(\widehat{\boldsymbol{y}})$. We first find the best estimator in $R^0(\widetilde{\boldsymbol{y}}_{\boldsymbol{a}})$. For any $\widetilde{\boldsymbol{y}}_{\boldsymbol{a}}^M \in R^0(\widetilde{\boldsymbol{y}}_{\boldsymbol{a}})$, let

$$
\begin{aligned}
f(a_1, \ldots, a_m) &= Q(\widetilde{\boldsymbol{y}}_{\boldsymbol{a}}^M) = \sum_{i=1}^{m} \varphi_i E(\widetilde{y}_i^M - y_i)^2 \\
&= \sum_{i=1}^{m} \varphi_i E\left\{\left[(\widetilde{y}_i^H - y_i) + a_i(\sum_{j=1}^{m} \omega_j Y_j - \sum_{j=1}^{m} \omega_j \widetilde{y}_j^H)\right]^2\right\}.
\end{aligned}
$$

Note that

$$
\begin{aligned}
&E\left[(\widetilde{y}_i^H - y_i)(Y_j - \widetilde{y}_j^H)\right] \\
&= cov\left\{(\gamma_i - 1)b_i + \gamma_i e_i + (1-\gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}), \omega_j(1-\gamma_j)[(b_j + e_j) - \boldsymbol{x}_j^T(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})]\right\},
\end{aligned}
$$

where γ_i is defined in (1.9). We have

$$
\begin{aligned}
&cov\left\{(1-\gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}), \omega_j(1-\gamma_j)[(b_j + e_j) - \boldsymbol{x}_j^T(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})]\right\} \\
&= \omega_j(1-\gamma_j)(1-\gamma_i) cov\left\{\boldsymbol{x}_j^T(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}(\boldsymbol{b} + \boldsymbol{e}), (b_j + e_j)\right\} \\
&\quad -\omega_j(1-\gamma_j)(1-\gamma_i)\boldsymbol{x}_i^T V(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\boldsymbol{x}_j^T \\
&= \omega_j(1-\gamma_j)(1-\gamma_i)\left\{\boldsymbol{x}_j^T(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{l}_i - \boldsymbol{x}_i^T(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{x}_j^T\right\} = 0,
\end{aligned}
$$

where $\boldsymbol{l}_i$ is the vector with the i-th element equal to one and others equal to zero. Also,

$$
\begin{aligned}
&cov\left\{(\gamma_i - 1)b_i + \gamma_i e_i, \omega_j(1-\gamma_j)[(b_j + e_j) - \boldsymbol{x}_j^T(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})]\right\} \\
&\qquad = cov\left\{(\gamma_i - 1)b_i + \gamma_i e_i, h_j(b_j + e_j)\right\} = 0,
\end{aligned}
$$

where $h_j = \omega_j(1-\gamma_j)\left[1 - (\sigma_b^2 + \sigma_{ei}^2)^{-1}\boldsymbol{x}_j^T(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{x}_i^T\right]$.

This leads to

$$
E\left[(\widetilde{y}_i^H - y_i)(Y_j - \widetilde{y}_j^H)\right] = 0. \tag{4.21}
$$

Therefore,

$$
f(a_1, \ldots, a_m) = \sum_{i=1}^{m} \varphi_i E(\widetilde{y}_i^H - y_i)^2 + V\left[\sum_{j=1}^{m} \omega_j(Y_j - \widetilde{y}_j^H)\right]\sum_{i=1}^{m} \varphi_i a_i^2. \tag{4.22}
$$

Using Lagrangian multiplier methods to minimize $f(a_1, \ldots, a_m)$ subject to the restriction $\sum_{i=1}^{m} \omega_i a_i = 1$, we obtain the system

$$2a_i\varphi_i V\left(\sum_{j=1}^{m}\omega_j Y_j - \sum_{j=1}^{m}\omega_j \widetilde{y}_j^H\right) + \lambda\omega_i = 0, \quad i = 1, \ldots, m.$$

The solution to the linear system subject to the restriction $\sum_{i=1}^{m}\omega_i a_i = 1$ is

$$\breve{a}_i = \left(\sum_{i=1}^{m}\varphi_i^{-1}\omega_i^2\right)^{-1}\varphi_i^{-1}\omega_i. \tag{4.23}$$

Therefore, $\widehat{\boldsymbol{y}}^M$ defined in (4.19) is an estimator of the form (4.20) and minimizes criterion (4.18).

Let $\widehat{\boldsymbol{y}}$ be any linear unbiased estimator of $\boldsymbol{y}$ that satisfies (4.1), i.e., $\widehat{\boldsymbol{y}} \in R(\widehat{\boldsymbol{y}})$. By standard results for BLUP (See, for example, Robinson (1991) and Harville (1976)), we have

$$cov(\widetilde{y}_i^H - y_i, \widehat{y}_i - \widetilde{y}_i^H) = 0.$$

This leads to

$$E(\widehat{y}_i - y_i)^2 = E(\widetilde{y}_i^H - y_i)^2 + E(\widehat{y}_i - \widetilde{y}_i^H)^2. \tag{4.24}$$

Therefore,

$$Q(\widehat{\boldsymbol{y}}) = \sum_{i=1}^{m}\varphi_i E(\widehat{y}_i - y_i)^2 = \sum_{i=1}^{m}\varphi_i E(\widetilde{y}_i^H - y_i)^2 + \sum_{i=1}^{m}\varphi_i E(\widehat{y}_i^M - \widetilde{y}_i^H)^2. \tag{4.25}$$

Since $\widehat{\boldsymbol{y}}$ satisfies (4.1), we have $\sum_{i=1}^{m}\omega_i\widehat{y}_i = \sum_{i=1}^{m}\omega_i Y_i$. For the $\widehat{\boldsymbol{y}}^M$ defined in (4.19), $\widehat{y}_i^M = \widetilde{y}_i^H + \breve{a}_i(\sum_{j=1}^{m}\omega_j Y_j - \sum_{j=1}^{m}\omega_j\widetilde{y}_j^H) = \widetilde{y}_i^H + a_i\left[\sum_{j=1}^{m}\omega_j(\widehat{y}_j - \widetilde{y}_j^H)\right]$. By (4.22), we have

$$\begin{aligned} Q(\widehat{\boldsymbol{y}}^M) &= \sum_{i=1}^{m}\varphi_i E(\widetilde{y}_i^H - y_i)^2 + V\left\{\sum_{j=1}^{m}\omega_j(Y_j - \widetilde{y}_j^H)\right\}\sum_{i=1}^{m}\varphi_i\breve{a}_i^2 \\ &= \sum_{i=1}^{m}\varphi_i E(\widetilde{y}_i^H - y_i)^2 + V\left\{\sum_{j=1}^{m}\omega_j(Y_j - \widetilde{y}_j^H)\right\}\sum_{i=1}^{m}\varphi_i\left[\left(\sum_{i=1}^{m}\varphi_i^{-1}\omega_i^2\right)^{-1}\varphi_i^{-1}\omega_i\right]^2 \end{aligned}$$

$$= \sum_{i=1}^{m} \varphi_i E(\widetilde{y}_i^H - y_i)^2 + V\left\{\sum_{j=1}^{m} \omega_j(\widehat{y}_j - \widetilde{y}_j^H)\right\}\left(\sum_{i=1}^{m} \varphi_i^{-1}\omega_i^2\right)^{-1}. \tag{4.26}$$

Note that

$$\begin{aligned} V\left\{\sum_{j=1}^{m} \omega_j(\widehat{y}_j - \widetilde{y}_j^H)\right\} &= \sum_{j=1}^{m}\sum_{k=1}^{m} \omega_j\omega_k cov\left\{(\widehat{y}_j - \widetilde{y}_j^H), (\widehat{y}_k - \widetilde{y}_k^H)\right\} \\ &\le \sum_{j=1}^{m}\sum_{k=1}^{m} \omega_j\omega_k g_j g_k = \left(\sum_{i=1}^{m} \omega_i g_i\right)^2, \end{aligned} \tag{4.27}$$

where $g_j = \sqrt{Var(\widehat{y}_j - \widetilde{y}_j^H)}$. By Cauchy's inequality,

$$\left(\sum_{j=1}^{m} \omega_j g_j\right)^2 \le \left(\sum_{i=1}^{m} \varphi_i^{-1}\omega_i^2\right)\left(\sum_{i=1}^{m} \varphi_i g_i^2\right) = \left(\sum_{i=1}^{m} \varphi_i^{-1}\omega_i^2\right)\left[\sum_{i=1}^{m} \varphi_i E(\widehat{y}_i^M - \widetilde{y}_i^H)^2\right]. \tag{4.28}$$

Combining (4.26), (4.27) and (4.28), we have

$$\left[\sum_{i=1}^{m} \varphi_i E(\widehat{y}_i^M - \widetilde{y}_i^H)^2\right] \ge V\left\{\sum_{j=1}^{m} \omega_j(\widehat{y}_j - \widetilde{y}_j^H)\right\}\left(\sum_{i=1}^{m} \varphi_i^{-1}\omega_i^2\right)^{-1}.$$

Therefore, we have shown that $Q(\widehat{\boldsymbol{y}}) \ge Q(\widehat{\boldsymbol{y}}^M)$.

To show the uniqueness of $\widehat{\boldsymbol{y}}^M$, we need to check when the inequalities (4.27) and (4.28) become equalities. Inequality (4.27) becomes an equality if and only if

$$\widehat{y}_j - \widetilde{y}_j^H = c_j^0 + c_j^1(\widehat{y}_1 - \widetilde{y}_1^H) \tag{4.29}$$

for some constants c_j^0 and $c_j^1, j = 2, \ldots, m$. Inequality (4.28) becomes an equality if and only if

$$\sqrt{\varphi_i^{-1}\omega_i^2}\sqrt{v_j g_j^2} - \sqrt{v_j^{-1}\omega_j^2}\sqrt{\varphi_i g_i^2} = 0,$$

or, equivalently,

$$\varphi_i^2\omega_i^{-2}Var(\widehat{y}_i - \widetilde{y}_i^H) = v_j^2\omega_j^{-2}Var(\widehat{y}_j - \widetilde{y}_j^H). \tag{4.30}$$

Also,

$$\sum_{i=1}^{m} \omega_i\widehat{y}_i = \sum_{i=1}^{m} \omega_i Y_i. \tag{4.31}$$

Combining (4.29), (4.30), and (4.31), we have that the equality holds if and only if $\widehat{y}_j = \widehat{y}_j^M$. Thus, we have shown that $\widehat{\boldsymbol{y}}^M$ is the unique linear unbiased estimator that satisfies (4.1) and minimizes criterion (4.18). ■

Remark 1. If the variance component parameters are unknown, we replace the variance components with the estimated variance components, just as in EBLUP. This leads to the modified estimator

$$\widehat{y}_i^M = \widehat{y}_i^H + \hat{a}_i\left(\sum_{j=1}^{m}\omega_j Y_j - \sum_{j=1}^{m}\omega_j \widehat{y}_j^H\right), \tag{4.32}$$

where $\widehat{y}_i^H$ is the EBLUP defined in (1.16).

Remark 2. With a different choice of φ_i, we have a different estimator that minimizes the criterion (4.18). If $\varphi_i = \omega_i Y_i^{-1}$, we have the ratio estimator defined in (4.10). Since Y_i could be less than zero and it is not reasonable to have negative φ_i, we can see that the ratio adjustment is not always a good choice. If $\varphi_i = \omega_i[cov(\widetilde{y}_i^H, \widetilde{y}_.)]^{-1}$, where $\widetilde{y}_. = \sum_{j=1}^{m}\omega_j\widetilde{y}_j^H$, we have the estimator (4.8) derived by Pfeffermann and Barnard (1991). The estimators in (4.11) and (4.15) are estimators when the variance components are unknown. When $\varphi_i = [\widehat{Var}(\widehat{y}_j^H)]^{-1}$, we have the Battese, Harter and Fuller estimator of (4.11). The Isaki, Tsay and Fuller estimator in (4.15) results from $\varphi_i = [\widehat{Var}(Y_i)]^{-1}$. Therefore, Theorem 4.1 provides a unified way to derive the estimators described in this chapter.

Since the "best" estimator depends on the φ_i used in the criterion (4.18), we want to find a reasonable choice of φ_i. To get an insight into the problem, we first assume all the variance component parameters to be known. We argue that $\varphi_i = [Var(Y_i)]^{-1}$ is the most reasonable choice by showing the properties of the corresponding modified estimator. Let $\dot{\boldsymbol{C}} = \dot{\boldsymbol{A}}\boldsymbol{T}$, where

$$\boldsymbol{T} = \begin{pmatrix} & \boldsymbol{\omega}' & \\ \boldsymbol{0}_{m-1} & & \boldsymbol{I}_{m-1} \end{pmatrix}, \tag{4.33}$$

$$\dot{\boldsymbol{A}} = \begin{pmatrix} 1 & \mathbf{0}'_{m-1} \\ -\dot{\boldsymbol{a}}_{m-1} & \boldsymbol{I}_{m-1} \end{pmatrix}.$$

$$\dot{\boldsymbol{a}}_{m-1} = (\dot{a}_2, \ldots, \dot{a}_m)' = \left(\boldsymbol{\omega}^T \boldsymbol{\Sigma} \boldsymbol{\omega}\right)^{-1} \begin{pmatrix} \mathbf{0}_{m-1} & \boldsymbol{I}_{m-1} \end{pmatrix} \boldsymbol{\Sigma}\boldsymbol{\omega}.$$

Let $\bar{Y} = \sum_{i=1}^{m} \omega_i Y_i$ and observe that $\dot{\boldsymbol{C}}\boldsymbol{Y} = (\bar{Y}, Y_2 - \dot{a}_2\bar{Y}, \ldots, Y_m - \dot{a}_m\bar{Y})'$. Note that $Y_i - \dot{a}_i\bar{Y}, i = 2, \ldots, m$ are uncorrelated with $\bar{Y}$, i.e., $(Y_2 - \dot{a}_2\bar{Y}, \ldots, Y_m - \dot{a}_m\bar{Y})'$ is a basis for the space that is orthogonal to $\bar{Y}$ in the space spanned by $\boldsymbol{Y}$. We want to estimate $\boldsymbol{y}$. Equivalently, we can estimate $\dot{\boldsymbol{C}}\boldsymbol{y} = (\bar{y}, y_2 - \dot{a}_2\bar{y}, \ldots, y_m - \dot{a}_m\bar{y})'$, where $\bar{y} = \sum_{i=1}^{m} \omega_i y_i$. When there is no restriction, the BLUP of $\dot{\boldsymbol{C}}\boldsymbol{y}$ is

$$\widetilde{\dot{\boldsymbol{C}}\boldsymbol{y}} = \dot{\boldsymbol{C}}\boldsymbol{Y} - \dot{\boldsymbol{C}}(\boldsymbol{I}_m - \boldsymbol{\Gamma})(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) = \dot{\boldsymbol{C}}\widetilde{\boldsymbol{y}}. \tag{4.34}$$

Let $\dot{\boldsymbol{c}}'_1$ be the first row of $\dot{\boldsymbol{C}}$, $\dot{\boldsymbol{c}}'_1\boldsymbol{y} = \bar{y} = \sum_{i=1}^{m} \omega_i y_i$. To impose the restriction (4.1) on $\dot{\boldsymbol{C}}\boldsymbol{y}$, we only need to replace $\widetilde{\dot{\boldsymbol{c}}'_1\boldsymbol{y}}$ with $\sum_{i=1}^{m} \omega_i Y_i$ and use BLUP to estimate the other $m-1$ quantities in $\dot{\boldsymbol{C}}\boldsymbol{y}$. This leads to the modified estimator

$$\dot{\boldsymbol{C}}\widetilde{\boldsymbol{y}}^M = \dot{\boldsymbol{C}}\boldsymbol{Y} - \boldsymbol{B}\dot{\boldsymbol{C}}(\boldsymbol{I}_m - \boldsymbol{\Gamma})(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}), \tag{4.35}$$

where $\boldsymbol{B}$ is defined in (4.14). To obtain the modified estimator for $\boldsymbol{y}$, we multiply equation (4.35) by $\dot{\boldsymbol{C}}^{-1}$ on both sides to obtain

$$\widetilde{\boldsymbol{y}}^M = \boldsymbol{Y} - \dot{\boldsymbol{C}}^{-1}\boldsymbol{B}\dot{\boldsymbol{C}}(\boldsymbol{I}_m - \boldsymbol{\Gamma})(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}). \tag{4.36}$$

This estimator does not depend on the choice of basis in the space that is orthogonal to $\bar{Y}$ in the space spanned by $\boldsymbol{Y}$. Suppose $(s_2, \ldots, s_m)'$ is another basis in the space that is orthogonal to Y_T in the space spanned by $\boldsymbol{Y}$. There is an invertible matrix $\boldsymbol{S}_1$ such that $(s_2, \ldots, s_m)' = \boldsymbol{S}_1(Y_2 - \dot{a}_2\bar{Y}, \ldots, Y_m - \dot{a}_m\bar{Y})'$. In other words, $(\bar{Y}, s_2, \ldots, s_m)' = \boldsymbol{S}\dot{\boldsymbol{C}}\boldsymbol{Y}$, where $\boldsymbol{S} = \begin{pmatrix} 1 & \mathbf{0}'_{m-1} \\ \mathbf{0}_{m-1} & \boldsymbol{S}_1 \end{pmatrix}$. Letting $\boldsymbol{C}_1 = \boldsymbol{S}\dot{\boldsymbol{C}}$ and following the same argument, we have

$$\begin{aligned} \tilde{\boldsymbol{y}}_s^M &= \boldsymbol{Y} - \boldsymbol{C}_1^{-1}\boldsymbol{B}\boldsymbol{C}_1(\boldsymbol{I}_m - \boldsymbol{\Gamma})(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \\ &= \boldsymbol{Y} - \dot{\boldsymbol{C}}\boldsymbol{B}\dot{\boldsymbol{C}}(\boldsymbol{I}_m - \boldsymbol{\Gamma})(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) = \tilde{\boldsymbol{y}}^M. \end{aligned} \tag{4.37}$$

because $\boldsymbol{C}_1^{-1}\boldsymbol{B}\boldsymbol{C}_1 = \dot{\boldsymbol{C}}^{-1}\boldsymbol{S}^{-1}\boldsymbol{B}\boldsymbol{S}^{-1}\dot{\boldsymbol{C}} = \dot{\boldsymbol{C}}^{-1}\boldsymbol{B}\dot{\boldsymbol{C}}$. Therefore, (4.36) is derived by applying the restriction and making estimation based on the information from the space that is orthogonal to $\bar{Y}$.

There are exactly $m-1$ linearly independent BLUPs in any $m-1$ dimensional subspace spanned by $\boldsymbol{Y}$ that does not contain $\bar{Y} = \sum_{i=1}^{m} \omega_i Y_i$. Let $\boldsymbol{C}_{m-1}\boldsymbol{Y}$ be the space spanned by the $m-1$ linearly independent BLUPs and let $\boldsymbol{C}_{\boldsymbol{a}} = (\boldsymbol{\omega}, \boldsymbol{C}'_{m-1})'$. Similar to the derivation of $\tilde{\boldsymbol{y}}^M$ in (4.36), we can impose the restriction (4.1) by using the linear combination of $\sum_{i=1}^{m} \omega_i Y_i$ and the $m-1$ BLUPs to construct the modified estimator $\tilde{\boldsymbol{y}}_{\boldsymbol{a}}^M$. First, we estimate $\boldsymbol{C}_{\boldsymbol{a}}\boldsymbol{y}$ by replacing $\boldsymbol{c}_1'\tilde{\boldsymbol{y}}$, the BLUP of the first quantity of $\boldsymbol{C}_{\boldsymbol{a}}\boldsymbol{y}$, with $\sum_{i=1}^{m} \omega_i Y_i$ and use BLUP to estimate the other $m-1$ quantities in $\boldsymbol{C}_{\boldsymbol{a}}\boldsymbol{y}$. This is represented by

$$\widetilde{\boldsymbol{C}_{\boldsymbol{a}}\boldsymbol{y}_{\boldsymbol{a}}}^M = \boldsymbol{C}_{\boldsymbol{a}}\boldsymbol{Y} - \boldsymbol{B}\boldsymbol{C}_{\boldsymbol{a}}(\boldsymbol{I}_m - \boldsymbol{\Gamma})(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}). \tag{4.38}$$

Then we multiply the estimator of $\boldsymbol{C}_{\boldsymbol{a}}\boldsymbol{y}$ by $\boldsymbol{C}_{\boldsymbol{a}}^{-1}$ to obtain the estimator of $\boldsymbol{y}$

$$\widetilde{\boldsymbol{y}_{\boldsymbol{a}}}^M = \boldsymbol{Y} - \boldsymbol{C}_{\boldsymbol{a}}^{-1}\boldsymbol{B}\boldsymbol{C}_{\boldsymbol{a}}(\boldsymbol{I}_m - \boldsymbol{\Gamma})(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}). \tag{4.39}$$

It is possible to derive some linear unbiased estimators for $\boldsymbol{y}$ by using less than $m-1$ linearly independent BLUPs in the space spanned $\boldsymbol{Y}$. Obviously, these estimators are less efficient than the estimators in the form of (4.39). Therefore, the question, "which estimator for $\boldsymbol{y}$ is the most reasonable estimator?" is equivalent to the question, "which $m-1$ BLUPs are the most reasonable choice for constructing the restricted estimator?" If we interpret restriction (4.1) as meaning that $\bar{Y} = \sum_{i=1}^{m} \omega_i Y_i$ is the best estimator for $\bar{y} = \sum_{i=1}^{m} \omega_i y_i$, we should choose to evaluate the $m-1$ components that are orthogonal to Y. This indicates that the estimator in (4.36) is the most reasonable estimator. Therefore. $\varphi_i = [Var(Y_i)]^{-1}$ is the most sensible choice for φ_i. When the variance components are unknown. we replace the variance components with the corresponding estimated value. This leads to the estimator in (4.15).

We want to say a few more words about the estimators in the form of (4.39). Let $\boldsymbol{a}$ be the first column of $\boldsymbol{C}_{\boldsymbol{a}}^{-1}$ and let

$$\widetilde{\boldsymbol{y}}^H = (\widehat{y}_1^H, \ldots, \widehat{y}_m^H)' = [\boldsymbol{Y} - (\boldsymbol{I}_m - \boldsymbol{\Gamma})(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})].$$

Then

$$\begin{aligned}
\widetilde{\boldsymbol{y}}_{\boldsymbol{a}}^M &= \boldsymbol{Y} - \boldsymbol{C}_{\boldsymbol{a}}^{-1}\boldsymbol{B}\boldsymbol{C}_{\boldsymbol{a}}(\boldsymbol{I}_m - \boldsymbol{\Gamma})(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \\
&= [\boldsymbol{Y} - (\boldsymbol{I}_m - \boldsymbol{\Gamma})(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})] - \boldsymbol{C}_{\boldsymbol{a}}^{-1}(\boldsymbol{I}_m - \boldsymbol{B})\boldsymbol{C}_{\boldsymbol{a}}(\boldsymbol{I}_m - \boldsymbol{\Gamma})(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \\
&= \widetilde{\boldsymbol{y}}^H - \boldsymbol{a}\boldsymbol{\omega}'(\boldsymbol{I}_m - \boldsymbol{\Gamma})(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \\
&= \widetilde{\boldsymbol{y}}^H - \boldsymbol{a}[\sum_{j=1}^m \omega_j Y_j - \sum_{j=1}^m \omega_j \widetilde{y}_j^H] \qquad (4.40)
\end{aligned}$$

because $\boldsymbol{C}_{\boldsymbol{a}}^{-1}(\boldsymbol{I}_m - \boldsymbol{B})\boldsymbol{C}_{\boldsymbol{a}} = [\boldsymbol{C}_{\boldsymbol{a}}^{-1}(\boldsymbol{I}_m - \boldsymbol{B})][(\boldsymbol{I}_m - \boldsymbol{B})\boldsymbol{C}_{\boldsymbol{a}}] = \boldsymbol{a}\boldsymbol{\omega}'$. In other words. any estimator of the form (4.39) can be written in the form of (4.20). On the other hand. any estimator of the form (4.20) can be written in the form of (4.39) by letting $\boldsymbol{C}_{\boldsymbol{a}} = \boldsymbol{A}_{\boldsymbol{a}}\boldsymbol{T}$. where

$$\boldsymbol{A}_{\boldsymbol{a}} = \begin{pmatrix} 1 & \boldsymbol{0}'_{m-1} \\ -\boldsymbol{a}_{m-1} & \boldsymbol{I}_{m-1} \end{pmatrix}.$$

$$\boldsymbol{a}_{m-1} = (a_2, \ldots, a_m)'.$$

Therefore. (4.20) and (4.39) are just different representations of the same family of estimators.

4.3 The MSE of the modified estimator

When the variance component parameters are known, we argued in Section 4.2 that the estimator defined in (4.36) is the most reasonable estimator satisfying restriction

(4.1). If the variance components are unknown, we replace the variance components with the corresponding estimated values. This leads to the modified estimator in (4.15). If we replace $\widehat{y}_i^H$ in (4.15) with

$$\widehat{y}_i = \widehat{\gamma}_i Y_i + (1 - \widehat{\gamma}_i)\boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}}_O,$$

which is defined in (3.46), and replace $\widehat{Var}(Y_i)$ with $Var(Y_i)$, we have

$$\widehat{y}_i^M = \widehat{y}_i + [\sum_{j=1}^m \omega_j^2 Var(Y_j)]^{-1} \omega_i Var(Y_i)[\sum_{j=1}^m \omega_j (Y_j - \widehat{y}_j)]. \tag{4.41}$$

For $MSE(\widehat{y}_i^M - y_i)$, we have the following theorem corresponding to the Theorem 3.3 in Chapter 3.

Theorem 4.2 *Let the random effects model*

$$Y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + b_i + e_i, \qquad i = 1, \ldots, m,$$

where $b_i \sim N(0, \sigma_b^2)$, $e_i \sim N(0, \sigma_{ei}^2)$, hold. Assume there is a sequence of small areas $i = 1, \ldots, m$, where the number of small areas $m \to \infty$. Assume $\sigma_{ei}^2, i = 1, 2, \ldots$ and $\boldsymbol{x}_i^T, i = 1, 2, \ldots$ are fixed sequences. Assume $\widehat{\sigma}_{ei}^2, i = 1, 2, \ldots$, with $d_i \widehat{\sigma}_{ei}^2 \sim \chi_{d_i}^2 \sigma_{ei}^2$, are unbiased estimators of σ_{ei}^2. Assume $\{\widehat{\sigma}_{ei}^2\}_{i=1}^m$ are independent of $\boldsymbol{Y} = (Y_1, \ldots, Y_m)^T$ for any m. Assume that:

(i) every d_i increases at a common rate d, where $d \to \infty$.

(ii) $\sigma_b^2 > 0$, $\sigma_{ei}^2 < C_1$ for all i, and $x_{ij} < C_x$ for all i and j.

(iii) $w_{ii,m} < C_w$ for all i and m, where $w_{ii,m}$ are the diagonal elements of $\boldsymbol{W}_m$.

(iv) $c_{im}, i = 1, \ldots, m$, the weights in (3.37), are such that $\sum_{i=1}^m c_{im} = 1$, are positive fixed values and $c_{im} = O(m^{-1})$.

(v) $\lim_{m\to\infty} (\boldsymbol{X}^T \boldsymbol{W}_m \boldsymbol{X})^{-1} = m^{-1}\boldsymbol{A}_1$, and $\lim_{m\to\infty} m^{-1}\boldsymbol{X}^T \boldsymbol{W}_m \boldsymbol{V} \boldsymbol{W}_m \boldsymbol{X} = \boldsymbol{A}_2$ for some positive definite matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$.

(vi) the $\omega_i, i = 1, \ldots, m$, which are the weights in (4.1), satisfy $\sum_{i=1}^m \omega_i = 1$.

Let $\dot{a}_i = [\sum_{j=1}^{m} \omega_j^2 Var(Y_j)]^{-1} \omega_i Var(Y_i),\ i = 1, \ldots, m.$ *Then*

$$MSE(\widehat{y}_i^M - y_i) = \left[\gamma_i \sigma_{ei}^2 + (1-\gamma_i)^2 \boldsymbol{x}_i^T V(\widehat{\boldsymbol{\beta}}_O)\boldsymbol{x}_i + (\sigma_b^2 + \sigma_{ei}^2)V(\widehat{\gamma}_i)\right]$$
$$+2\dot{a}_i(1-\gamma_i)E\left\{(\widehat{\gamma}_i - \gamma_i)(b_i + e_i)^2\right\}$$
$$+2\dot{a}_i(1-\gamma_i)\left[\boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{V}\boldsymbol{l}_w - \boldsymbol{x}_i^T V(\widehat{\boldsymbol{\beta}}_O)\boldsymbol{X}^T\boldsymbol{l}_w\right]$$
$$+\dot{a}_i^2 \sum_{j=1}^{m} \omega_j^2 (1-\gamma_j)^2 \left[(\sigma_b^2 + \sigma_{ej}^2) + (\sigma_b^2 + \sigma_{ej}^2)V(\widehat{\gamma}_j) + E\left\{(\widehat{\gamma}_j - \gamma_j)(b_j + e_j)^2\right\}\right]$$
$$+\dot{a}_i^2\left[\boldsymbol{l}_w^T\boldsymbol{X}V(\widehat{\boldsymbol{\beta}}_O)\boldsymbol{X}^T\boldsymbol{l}_w - 2\boldsymbol{l}_w^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{V}\boldsymbol{l}_w\right] + O(r_m), \qquad (4.42)$$

where $\boldsymbol{l}_w = [\omega_1(1-\gamma_1), \ldots, \omega_m(1-\gamma_m)]^T$, $r_m = \max(m^{-1.5}, m^{-1}d^{-1}, d{-}2)$, *and*

$$E\left\{(\widehat{\gamma}_i - \gamma_i)(b_i + e_i)^2\right\} = (\sigma_b^2+\sigma_{ei}^2)^{-2} 2\sigma_b^2\sigma_{ei}^4 d_i^{-1} - (\sigma_b^2+\sigma_{ei}^2)^{-2}\sigma_{ei}^2 V(\widehat{\sigma}_{b,m}^2) + 2c_{im}\sigma_{ei}^2. \quad (4.43)$$

If $i \neq j$, *we have*

$$E\left\{(\widehat{y}_i^M - y_i)(\widehat{y}_j^M - y_j)\right\} = (1-\gamma_i)(1-\gamma_j)\boldsymbol{x}_i^T V(\widehat{\boldsymbol{\beta}}_O)\boldsymbol{x}_j$$
$$+\dot{a}_i(1-\gamma_i)E\left\{(\widehat{\gamma}_i - \gamma_i)(b_i + e_i)^2\right\} + \dot{a}_j(1-\gamma_j)E\left\{(\widehat{\gamma}_j - \gamma_j)(b_j + e_j)^2\right\}$$
$$+\dot{a}_i(1-\gamma_i)\left[\boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{V}\boldsymbol{l}_w - \boldsymbol{x}_i^T V(\widehat{\boldsymbol{\beta}}_O)\boldsymbol{X}^T\boldsymbol{l}_w\right]$$
$$+\dot{a}_j(1-\gamma_j)\left[\boldsymbol{x}_j^T(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{V}\boldsymbol{l}_w - \boldsymbol{x}_j^T V(\widehat{\boldsymbol{\beta}}_O)\boldsymbol{X}^T\boldsymbol{l}_w\right]$$
$$+\dot{a}_i\dot{a}_j \sum_{j=1}^{m} \omega_j^2(1-\gamma_j)^2\left[(\sigma_b^2 + \sigma_{ej}^2) + (\sigma_b^2 + \sigma_{ej}^2)V(\widehat{\gamma}_j) + E\left\{(\widehat{\gamma}_j - \gamma_j)(b_j + e_j)^2\right\}\right]$$
$$+\dot{a}_i\dot{a}_j\left[\boldsymbol{l}_w^T\boldsymbol{X}V(\widehat{\boldsymbol{\beta}}_O)\boldsymbol{X}^T\boldsymbol{l}_w - 2\boldsymbol{l}_w^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{V}\boldsymbol{l}_w\right] + O(r_m). \qquad (4.44)$$

Proof: Note that

$$(\widehat{y}_i^M - y_i)^2 = \left[(\widehat{y}_i - y_i) + \dot{a}_i \sum_{j=1}^{m} \omega_j (Y_j - \widehat{y}_j)\right]^2$$
$$= (\widehat{y}_i - y_i)^2 + 2\dot{a}_i \sum_{j=1}^{m} \omega_j(\widehat{y}_i - y_i)(Y_j - \widehat{y}_j) + \dot{a}_i^2 \sum_{j=1}^{m}\sum_{k=1}^{m} \omega_j\omega_k(Y_j - \widehat{y}_j)(Y_k - \widehat{y}_k),$$

and because the moments are finite, we have

$$MSE(\widehat{y}_i^M - y_i) = MSE(\widehat{y}_i - y_i) + 2\dot{a}_i \sum_{j=1}^{m} \omega_j cov\left\{(\widehat{y}_i - y_i), (Y_j - \widehat{y}_j)\right\}$$

$$+\dot{a}_i^2 \sum_{j=1}^{m} \sum_{k=1}^{m} \omega_j \omega_k cov \left\{ (Y_j - \widehat{y}_j), (Y_k - \widehat{y}_k) \right\}. \tag{4.45}$$

By (3.46) in Theorem 3.3, we have

$$MSE(\widehat{y}_i - y_i) = \gamma_i \sigma_{ei}^2 + (1-\gamma_i)^2 \boldsymbol{x}_i^T V(\widehat{\boldsymbol{\beta}}_O) \boldsymbol{x}_i + (\sigma_b^2 + \sigma_{ei}^2) V(\widehat{\gamma}_i) + O(r_m), \tag{4.46}$$

where $r_m = max(m^{-1.5}, m^{-1}d^{-1}, d^{-2})$.

Because $\sum_{i=1}^m \omega_i = 1$ by condition (vi), there are at most a finite number of ω_i with $\omega_i = O(1)$ and $\omega_i^{-1} = O(1)$. All other $\omega_i = O(m^{-1})$. Therefore,

$$\sum_{j=1}^{m} \omega_j O(r_m) = O(r_m) \tag{4.47}$$

and

$$\sum_{j=1}^{m} \sum_{k=1}^{m} \omega_j \omega_k O(r_m) = O(r_m). \tag{4.48}$$

Because $\sigma_b^2 > 0$ and $\sigma_{ei}^2 < C_1$ for all i by condition (ii), we have

$$0 < \sigma_b^2 \leq Var(Y_j) = \sigma_b^2 + \sigma_{ej}^2 \leq \sigma_b^2 + C_1.$$

If there is some ω_j such that $\omega_i = O(1)$, then $\sum_{j=1}^{m} \omega_j^2 = O(1)$ and

$$|\dot{a}_i| = \left| \left[\sum_{j=1}^{m} \omega_j^2 Var(Y_j) \right]^{-1} \omega_i Var(Y_i) \right| \leq O(1).$$

If $\omega_i = O(m^{-1})$ for $i = 1, \ldots, m$, note that $\sum_{j=1}^{m} \omega_j^2 \geq m^{-1}$ because $\sum_{i=1}^m \omega_i = 1$. Hence, $\sum_{j=1}^{m} \omega_j^2 Var(Y_j) \geq \sigma_b^2 \sum_{j=1}^{m} \omega_j^2 \geq \sigma_b^2 m^{-1}$ and

$$|\dot{a}_i| = \left| \left[\sum_{j=1}^{m} \omega_j^2 Var(Y_j) \right]^{-1} \omega_i Var(Y_i) \right| \leq (\sigma_b^2 m^{-1})^{-1} O(m^{-1}) = O(1).$$

Therefore, $|\dot{a}_i| = O(1)$.

By (4.47), (4.48), and $|\dot{a}_i| = O(1)$, if we have an order $O(r_m)$ approximation to each component of the second and the third terms of (4.45), we have an order $O(r_m)$ approximation to $MSE(\widehat{y}_i^M - y_i)$. We first approximate $cov\left\{(\widehat{y}_i - y_i), (Y_j - \widehat{y}_j)\right\}$. Recall that

$$\begin{aligned}
\widehat{y}_i - y_i &= [(\widehat{\gamma}_i - 1)b_i + \widehat{\gamma}_i e_i] + (1-\gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta}) - (\widehat{\gamma}_i - \gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta}) \\
&= (\gamma_i - 1)b_i + \gamma_i e_i + (\widehat{\gamma}_i - \gamma_i)(b_i + e_i) + (1-\gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta}) \\
&\quad + O_p\left(\max(m^{-1}, m^{-0.5}d^{-0.5})\right). \qquad (4.49)
\end{aligned}$$

and

$$(Y_j - \widehat{y}_j) = (1-\gamma_j)[b_j + e_j - \boldsymbol{x}_j^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})] - (\widehat{\gamma}_j - \gamma_j)[b_j + e_j - \boldsymbol{x}_j^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})]. \qquad (4.50)$$

By Taylor expansion, we have

$$cov\left\{(\widehat{\gamma}_i - \gamma_i)(b_i + e_i), (b_j + e_j)\right\} = \begin{cases} O(r_m) & \text{if } i \neq j, \\ E\left\{(\widehat{\gamma}_i - \gamma_i)(b_i + e_i)^2\right\} + O(r_m), & \text{if } i = j. \end{cases} \qquad (4.51)$$

where $E\left\{(\widehat{\gamma}_i - \gamma_i)(b_i + e_i)^2\right\}$ is defined in (4.43). By (3.56), we have

$$E\left\{[(\widehat{\gamma}_i - \gamma_i)(b_i + e_i)][\boldsymbol{x}_j^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})]\right\} = O(r_m) \qquad (4.52)$$

for any i and j. Therefore,

$$\begin{aligned}
& cov\left\{[(\widehat{\gamma}_i - 1)b_i + \widehat{\gamma}_i e_i], (1-\gamma_j)[b_j + e_j - \boldsymbol{x}_j^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})]\right\} \\
&= cov\left\{[(\gamma_i - 1)b_i + \gamma_i e_i], (1-\gamma_j)[b_j + e_j - \boldsymbol{x}_j^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})]\right\} \\
&+ cov\left\{(\widehat{\gamma}_i - \gamma_i)(b_i + e_i), (1-\gamma_j)[b_j + e_j - \boldsymbol{x}_j^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})]\right\} \\
&= 0 + (1-\gamma_j)cov\left\{(\widehat{\gamma}_i - \gamma_i)(b_i + e_i), (b_j + e_j)\right\} \\
&- (1-\gamma_j)cov\left\{(\widehat{\gamma}_i - \gamma_i)(b_i + e_i), \boldsymbol{x}_j^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})\right\} \\
&= \begin{cases} O(r_m) & \text{if } i \neq j, \\ E\left\{(\widehat{\gamma}_i - \gamma_i)(b_i + e_i)^2\right\} + O(r_m) & \text{if } i = j. \end{cases} \qquad (4.53)
\end{aligned}$$

We also have

$$\begin{aligned} & cov\left\{\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta}), [b_j + e_j - \boldsymbol{x}_j^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})]\right\} \\ & = \left[\boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}\boldsymbol{x}_j w_{jj,m}(\sigma_b^2 + \sigma_{ej}^2) - \boldsymbol{x}_i^T V(\widehat{\boldsymbol{\beta}}_O)\boldsymbol{x}_j\right], \end{aligned} \tag{4.54}$$

where $V(\widehat{\boldsymbol{\beta}}_O) = (\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{V}\boldsymbol{W}_m\boldsymbol{X})(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}$. By using (3.55) and (3.56), we have

$$cov\left\{(\widehat{\gamma}_i - \gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta}), (1-\gamma_j)[b_j + e_j - \boldsymbol{x}_j^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})]\right\} = O(r_m). \tag{4.55}$$

The derivation of (4.55) is quite long. Heuristically, we can look at the problem in the following way. By condition (v) and (4.54),

$$cov\left\{\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta}), (1-\gamma_j)[b_j + e_j - \boldsymbol{x}_j^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})]\right\} = O(m^{-1}). \tag{4.56}$$

Since $E(\widehat{\gamma}_i - \gamma_i) = O(d^{-1})$, $cov\left\{(\widehat{\gamma}_i - \gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta}), (1-\gamma_j)[b_j + e_j - \boldsymbol{x}_j^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})]\right\}$ should be of order $O(m^{-1}d^{-1}) = O(r_m)$. Similar arguments (or direct calculation) lead to

$$cov\left\{[(\widehat{\gamma}_i - 1)b_i + \widehat{\gamma}_i e_i], (\widehat{\gamma}_j - \gamma_j)[b_j + e_j - \boldsymbol{x}_j^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})]\right\} = O(r_m), \tag{4.57}$$

$$cov\left\{(1-\gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta}), (\widehat{\gamma}_j - \gamma_j)[b_j + e_j - \boldsymbol{x}_j^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})]\right\} = O(r_m), \tag{4.58}$$

and

$$cov\left\{(\widehat{\gamma}_i - \gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta}), (\widehat{\gamma}_j - \gamma_j)[b_j + e_j - \boldsymbol{x}_j^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})]\right\} = O(r_m). \tag{4.59}$$

By (4.53), (4.54), (4.55), (4.57), (4.58), and (4.59), we have

$$\begin{aligned} cov\{(\widehat{y}_i - y_i), (Y_j - \widehat{y}_j)\} &= (1-\gamma_j)\left[\boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}\boldsymbol{x}_j w_{jj,m}(\sigma_b^2 + \sigma_{ej}^2)\right. \\ &\quad \left. -\boldsymbol{x}_i^T V(\widehat{\boldsymbol{\beta}}_O)\boldsymbol{x}_j\right] + O(r_m) \end{aligned} \tag{4.60}$$

if $i \neq j$ and

$$\begin{aligned} cov\{(\widehat{y}_i - y_i), (Y_i - \widehat{y}_i)\} &= (1-\gamma_i)\left[\boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}\boldsymbol{x}_i w_{ii,m}(\sigma_b^2 + \sigma_{ej}^2)\right. \\ &\quad \left. -\boldsymbol{x}_i^T V(\widehat{\boldsymbol{\beta}}_O)\boldsymbol{x}_i\right] + E\left\{(\widehat{\gamma}_i - \gamma_i)(b_i + e_i)^2\right\} + O(r_m) \end{aligned} \tag{4.61}$$

if $i = j$. By (4.47), (4.60), and (4.61),

$$\begin{aligned}
&\sum_{j=1}^{m} \omega_j cov\left\{(\widehat{y}_i - y_i), (Y_j - \widehat{y}_j)\right\} \\
&= (1-\gamma_i)\sum_{j=1}^{m} \omega_j(1-\gamma_j)\left[\boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}\boldsymbol{x}_j w_{jj,m}(\sigma_b^2+\sigma_{e_j}^2) - \boldsymbol{x}_i^T V(\widehat{\boldsymbol{\beta}}_O)\boldsymbol{x}_j\right] \\
&+E\left\{(\widehat{\gamma}_i-\gamma_i)(b_i+e_i)^2\right\} + O(r_m) \\
&= (1-\gamma_i)\left[\boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{V}\boldsymbol{l}_w - \boldsymbol{x}_i^T V(\widehat{\boldsymbol{\beta}}_O)\boldsymbol{X}^T\boldsymbol{l}_w\right] + O(r_m), \qquad (4.62)
\end{aligned}$$

where $\boldsymbol{l}_w = [\omega_1(1-\gamma_1), \ldots, \omega_m(1-\gamma_m)]^T$.

Now we approximate $cov\left\{(Y_j - \widehat{y}_j), (Y_k - \widehat{y}_k)\right\}$. By (4.50), we have

$$(Y_j - \widehat{y}_j) = (1-\gamma_j)[b_j + e_j - \boldsymbol{x}_j^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})] - (\widehat{\gamma}_j - \gamma_j)[b_j + e_j - \boldsymbol{x}_j^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})].$$

Note that

$$\begin{aligned}
&cov\left\{[b_j + e_j - \boldsymbol{x}_j^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})], [b_k + e_k - \boldsymbol{x}_k^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})]\right\} \\
&= \begin{cases}
\Big[\boldsymbol{x}_j^T V(\widehat{\boldsymbol{\beta}}_O)\boldsymbol{x}_k - \boldsymbol{x}_j^T(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}\boldsymbol{x}_k w_{kk,m}(\sigma_b^2+\sigma_{ek}^2) & \\
\qquad -\boldsymbol{x}_k^T(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}\boldsymbol{x}_j w_{jj,m}(\sigma_b^2+\sigma_{ej}^2)\Big], & \text{if } j \neq k \\
\Big[(\sigma_b^2+\sigma_{ej}^2) + \boldsymbol{x}_j^T V(\widehat{\boldsymbol{\beta}}_O)\boldsymbol{x}_j & \\
\qquad -2\boldsymbol{x}_j^T(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}\boldsymbol{x}_j w_{jj,m}(\sigma_b^2+\sigma_{ej}^2)\Big], & \text{if } j = k.
\end{cases} \qquad (4.63)
\end{aligned}$$

$$\begin{aligned}
&cov\left\{[b_j + e_j - \boldsymbol{x}_j^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})], (\widehat{\gamma}_k - \gamma_k)[b_k + e_k - \boldsymbol{x}_k^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})]\right\} \\
&\qquad = \begin{cases}
O(r_m) & \text{if } j \neq k, \\
E\left\{(\widehat{\gamma}_j - \gamma_j)(b_j+e_j)^2\right\} + O(r_m) & \text{if } j = k.
\end{cases} \qquad (4.64)
\end{aligned}$$

$$\begin{aligned}
&cov\left\{(\widehat{\gamma}_j - \gamma_j)[b_j + e_j - \boldsymbol{x}_j^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})], (\widehat{\gamma}_k - \gamma_k)[b_k + e_k - \boldsymbol{x}_k^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})]\right\} \\
&= \begin{cases}
O(r_m) & \text{if } j \neq k, \\
(\sigma_b^2 + \sigma_{ei}^2)V(\widehat{\gamma}_i) + O(r_m) & \text{if } j = k.
\end{cases} \qquad (4.65)
\end{aligned}$$

By (4.63), (4.64), and (4.65), we have

$$cov\left\{(Y_j-\widehat{y}_j),(Y_k-\widehat{y}_k)\right\}=(1-\gamma_j)(1-\gamma_k)\left[-\boldsymbol{x}_j^T(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}\boldsymbol{x}_k w_{kk,m}(\sigma_b^2+\sigma_{ek}^2)\right.$$
$$\left.-\boldsymbol{x}_k^T(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}\boldsymbol{x}_j w_{jj,m}(\sigma_b^2+\sigma_{ej}^2))+\boldsymbol{x}_j^T V(\widehat{\boldsymbol{\beta}}_O)\boldsymbol{x}_k\right]+O(r_m) \tag{4.66}$$

if $j \neq k$ and

$$cov\left\{(Y_j-\widehat{y}_j),(Y_j-\widehat{y}_j)\right\}=(1-\gamma_j)^2\left[(\sigma_b^2+\sigma_{ej}^2)+\boldsymbol{x}_j^T V(\widehat{\boldsymbol{\beta}}_O)\boldsymbol{x}_j\right.$$
$$-2\boldsymbol{x}_j^T(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}\boldsymbol{x}_j w_{jj,m}(\sigma_b^2+\sigma_{ej}^2)$$
$$\left.+(\sigma_b^2+\sigma_{ej}^2)V(\widehat{\gamma}_j)+E\left\{(\widehat{\gamma}_j-\gamma_j)(b_j+e_j)^2\right\}\right]+O(r_m) \tag{4.67}$$

if $i \neq j$. By (4.48), (4.66), and (4.67),

$$\sum_{j=1}^{m}\sum_{k=1}^{m}\omega_j\omega_k cov\left\{(Y_j-\widehat{y}_j),(Y_k-\widehat{y}_k)\right\}=\boldsymbol{l}_w^T\boldsymbol{X}V(\widehat{\boldsymbol{\beta}}_O)\boldsymbol{X}^T\boldsymbol{l}_w$$
$$-2\boldsymbol{l}_w^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{V}\boldsymbol{l}_w+\sum_{j=1}^{m}\omega_j^2(1-\gamma_j)^2\left[(\sigma_b^2+\sigma_{ej}^2)\right.$$
$$\left.+(\sigma_b^2+\sigma_{ej}^2)V(\widehat{\gamma}_j)+E\left\{(\widehat{\gamma}_j-\gamma_j)(b_j+e_j)^2\right\}\right]+O(r_m). \tag{4.68}$$

where $\boldsymbol{l}_w=[\omega_1(1-\gamma_1),\ldots,\omega_m(1-\gamma_m)]^T$. The $MSE(\widehat{y}_i^M-y_i)$ of (4.42) follows (4.45), (4.46), (4.62), and (4.68).

To prove (4.44), note that

$$\begin{aligned}E\left\{(\widehat{y}_i^M-y_i)(\widehat{y}_j^M-y_j)\right\} &= E\left\{(\widehat{y}_i-y_i)(\widehat{y}_j-y_j)\right\}+\dot{a}_i\sum_{k=1}^{m}\omega_k cov\left\{(\widehat{y}_i-y_i),(Y_k-\widehat{y}_k)\right\}\\ &\quad+\dot{a}_j\sum_{k=1}^{m}\omega_k cov\left\{(\widehat{y}_j-y_j),(Y_k-\widehat{y}_k)\right\}\\ &\quad+\dot{a}_i\dot{a}_j\sum_{k=1}^{m}\sum_{k_1=1}^{m}\omega_k\omega_{k_1} cov\left\{(Y_k-\widehat{y}_k),(Y_{k_1}-\widehat{y}_{k_1})\right\}.\end{aligned} \tag{4.69}$$

The $E\left\{(\widehat{y}_i^M-y_i)(\widehat{y}_j^M-y_j)\right\}$ of (4.44) follows (4.69), (4.62), (4.68), and (3.47) of Theorem 3.3. ■

Remark 1. It is easy to verify that the last four terms of (4.42) are all of order $O(m^{-1})$. Therefore, $\widehat{y}_i^M-\widehat{y}_i$ is of order $O_p(m^{-0.5})$.

Remark 2. If $\dot{a}_i = O(m^{-1})$, the second term of (4.42) is of order $O(r_m)$ because $E\left\{(\widehat{\gamma}_j - \gamma_j)(b_j + e_j)^2\right\}$ is of order $O(d^{-1})$.

Remark 3. Generally, the third term of (4.42)

$$2\dot{a}_i(1-\gamma_i)\left[\boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{V}\boldsymbol{l}_w - \boldsymbol{x}_i^T V(\widehat{\boldsymbol{\beta}}_O)\boldsymbol{X}^T\boldsymbol{l}_w\right] = O(m^{-1}).$$

However, it is negligible in many practical situations because $\boldsymbol{x}_i^T V(\widehat{\boldsymbol{\beta}}_O)\boldsymbol{X}^T\boldsymbol{l}_w$ is close to $\boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{V}\boldsymbol{l}_w$. We can see this by noting that

$$cov\left\{(1-\gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}), (1-\gamma_j)[(b_j+e_j)-\boldsymbol{x}_j^T(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})]\right\} = 0,$$

where $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{V}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}\boldsymbol{Y}$ is the GLS estimator. We also note that the third term of (4.42) is due to

$$cov\left\{(1-\gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O-\boldsymbol{\beta}), (1-\gamma_j)[b_j+e_j-\boldsymbol{x}_j^T(\widehat{\boldsymbol{\beta}}_O-\boldsymbol{\beta})]\right\}.$$

Therefore, the third term of (4.42) should be negligible if the WLS estimator $\widehat{\boldsymbol{\beta}}_O$ is close to the GLS estimator $\widehat{\boldsymbol{\beta}}$.

4.4 Simulation study

We will use the settings of the simulation study in Chapter 3 in the simulation for the restricted estimator. The restriction that we use is $m^{-1}\sum_{i=1}^{m}\widehat{y}_i^M = m^{-1}\sum_{i=1}^{m}Y_i$, which means $\omega_i = m^{-1}$ for $i = 1,\ldots,m$. We use $\ddot{a}_i = [\sum_{j=1}^{m}(\widehat{\sigma}_b^2+\widehat{\sigma}_{ei}^2)]^{-1}m(\widehat{\sigma}_b^2+\widehat{\sigma}_{ei}^2)$ to replace $\dot{a}_i = [\sum_{j=1}^{m}(\sigma_b^2+\sigma_{ei}^2)]^{-1}m(\sigma_b^2+\sigma_{ei}^2)$ of Theorem 4.2. We treat $\ddot{a}_i$ as fixed values so that we can use (4.42) of Theorem 4.2 to get the order $O(r_m)$ approximation to $MSE(\widehat{y}_i^M - y_i)$ for our simulation study. The second term of (4.42) is of order $O(r_m)$ in the simulation set up by Remark 2 of Theorem 4.2. We also have

$$\ddot{a}_i^2\sum_{j=1}^{m}\omega_j^2(1-\gamma_j)^2\left[(\sigma_b^2+\sigma_{ej}^2)V(\widehat{\gamma}_j) + E\left\{(\widehat{\gamma}_j-\gamma_j)(b_j+e_j)^2\right\}\right] = O(m^{-1}d^{-1})$$

because $\omega_i = m^{-1}$, $V(\widehat{\gamma}_j) = O(d^{-1})$, and $E\{(\widehat{\gamma}_j - \gamma_j)(b_j + e_j)^2\} = O(d^{-1})$. However, when d is small, $\sum_{j=1}^{m} \omega_j^2(1-\gamma_j)^2 \left[(\sigma_b^2 + \sigma_{ej}^2)V(\widehat{\gamma}_j) + E\left\{(\widehat{\gamma}_j - \gamma_j)(b_j + e_j)^2\right\}\right]$ is much larger than

$$\ddot{a}_i^2 \left[\sum_{j=1}^{m} \omega_j^2(1-\gamma_j)^2(\sigma_b^2 + \sigma_{ej}^2) + \boldsymbol{l}_w^T\boldsymbol{X}V(\widehat{\boldsymbol{\beta}}_O)\boldsymbol{X}^T\boldsymbol{l}_w - 2\boldsymbol{l}_w^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{V}\boldsymbol{l}_w\right].$$

which is the $O(m^{-1})$ part of the fourth and fifth terms of (4.42). Therefore, we do not simply eliminate

$$\sum_{j=1}^{m} \omega_j^2(1-\gamma_j)^2 \left[(\sigma_b^2 + \sigma_{ej}^2)V(\widehat{\gamma}_j) + E\left\{(\widehat{\gamma}_j - \gamma_j)(b_j + e_j)^2\right\}\right]$$

in $MSE(\widehat{y}_i^M - y_i)$ because it is an order $O(r_m)$ term in our simulation set up. We still include the order $O(m^{-1}d^{-1})$ part of this term, which is

$$2d^{-1}\sum_{j=1}^{m}(1-\gamma_j)^4(\sigma_b^2 + \gamma_j),$$

in $MSE(\widehat{y}_i^M - y_i)$. This leads to an order $O(r_m)$ theoretical approximation to $MSE(\widehat{y}_i^M - y_i)$

$$\begin{aligned} MSE(\widehat{y}_i^M - y_i) &= MSE(\widehat{y}_i - y_i) \\ &+2\ddot{a}_i(1-\gamma_i)\left[m^{-1}\sum_{j=1}^{m}\sigma_{ej}^2 - m^{-2}\sum_{j=1}^{m}(\sigma_b^2+\sigma_{ej}^2)\sum_{j=1}^{m}(1-\gamma_j)\right] \\ &+\ddot{a}_i^2\left\{\sum_{j=1}^{m}(1-\gamma_j)^2(\sigma_b^2+\sigma_{ej}^2) + m^{-2}\left[\sum_{j=1}^{m}(1-\gamma_j)\right]^2\sum_{j=1}^{m}(\sigma_b^2+\sigma_{ej}^2)\right. \\ &\left. -2m^{-1}\sum_{j=1}^{m}(1-\gamma_j)\sum_{j=1}^{m}\sigma_{ej}^2 + 2d^{-1}\sum_{j=1}^{m}(1-\gamma_j)^4(\sigma_b^2+\gamma_j)\right\}. \end{aligned} \tag{4.70}$$

An estimator to $MSE(\widehat{y}_i^M - y_i)$ is

$$\begin{aligned} \widehat{MSE}(\widehat{y}_i^M - y_i) &= \widehat{MSE}_2(\widehat{y}_i - y_i) \\ &+2\ddot{a}_i(1-\widehat{\gamma}_i)\left[m^{-1}\sum_{j=1}^{m}\widehat{\sigma}_{ej}^2 - m^{-2}\sum_{j=1}^{m}(\widehat{\sigma}_b^2+\widehat{\sigma}_{ej}^2)\sum_{j=1}^{m}(1-\widehat{\gamma}_j)\right] \end{aligned}$$

$$+\ddot{a}_i^2 \left\{ \sum_{j=1}^{m} (1-\widehat{\gamma}_j)^2 (\widehat{\sigma}_b^2 + \widehat{\sigma}_{ej}^2) + m^{-2} \left[\sum_{j=1}^{m} (1-\widehat{\gamma}_j) \right]^2 \sum_{j=1}^{m} (\widehat{\sigma}_b^2 + \widehat{\sigma}_{ej}^2) \right.$$
$$\left. -2m^{-1} \sum_{j=1}^{m} (1-\widehat{\gamma}_j) \sum_{j=1}^{m} \widehat{\sigma}_{ej}^2 + 2d^{-1} \sum_{j=1}^{m} (1-\widehat{\gamma}_j)^4 (\widehat{\sigma}_b^2 + \widehat{\gamma}_j) \right\}. \qquad (4.71)$$

where $\widehat{MSE}_2(\widehat{y}_i - y_i)$ is defined in (3.72).

The simulation results are reported in Table 4.1 through Table 4.5. These tables are similar to the tables in Chapter 3. For comparison purposes, both the unrestricted and restricted simulation MSE are reported in the tables.

From these tables, we can see that the simulated MSE of the restricted estimator is very close to the simulated MSE of the unrestricted estimator. The inflation in the MSE due to the modification is almost negligible (less than 0.005). For large d and m, there is no visible difference between the two MSE. The theoretical MSE approximates the simulated MSE very well except for very small σ_b^2. The estimated MSE also yields a very good estimator of the simulated and theoretical MSE except for very small σ_b^2. Considering the facts that the adjustments we made to the MSE due to the restriction are quite small, and that the theoretical and estimated MSE in Chapter 3 yield good approximations to the simulated MSE, these two observations are not surprising.

Table 4.1 Simulation study of $MSE(\widehat{y}_i^M - y_i)$ $(d = 5, m = 36)$

σ_b^2	γ_i	Unrestricted Simulation MSE	Restricted Simulation MSE	Theoretical MSE	Estimated MSE
0.010	0.0099	0.104	0.104	0.183	0.223
	0.0066	0.114	0.114	0.140	0.246
	0.0050	0.126	0.128	0.119	0.260
0.100	0.0909	0.193	0.193	0.245	0.275
	0.0625	0.208	0.209	0.220	0.308
	0.0476	0.223	0.225	0.205	0.328
0.250	0.2	0.313	0.313	0.337	0.344
	0.143	0.342	0.343	0.339	0.392
	0.111	0.372	0.375	0.337	0.421
0.429	0.3	0.409	0.409	0.427	0.407
	0.222	0.465	0.466	0.460	0.468
	0.176	0.497	0.499	0.475	0.511
0.667	0.4	0.510	0.510	0.520	0.493
	0.308	0.594	0.594	0.593	0.580
	0.25	0.657	0.658	0.633	0.638
1.000	0.5	0.609	0.608	0.614	0.608
	0.4	0.753	0.753	0.737	0.737
	0.333	0.823	0.825	0.814	0.841
1.500	0.6	0.700	0.700	0.705	0.710
	0.5	0.873	0.873	0.890	0.897
	0.429	1.02	1.020	1.020	1.050
2.333	0.7	0.757	0.757	0.792	0.798
	0.609	1.030	1.020	1.050	1.050
	0.538	1.230	1.230	1.250	1.240

Table 4.2 Simulation study of $MSE(\widehat{y}_i^M - y_i)$ $(d = 9, m = 36)$

σ_b^2	γ_i	Unrestricted Simulation MSE	Restricted Simulation MSE	Theoretical MSE	Estimated MSE
0.010	0.0099	0.099	0.0987	0.183	0.212
	0.00662	0.104	0.104	0.140	0.227
	0.00498	0.108	0.109	0.119	0.238
0.100	0.0909	0.180	0.180	0.244	0.257
	0.0625	0.189	0.191	0.218	0.280
	0.0476	0.209	0.212	0.204	0.295
0.250	0.2	0.289	0.289	0.331	0.307
	0.143	0.317	0.318	0.333	0.340
	0.111	0.323	0.325	0.331	0.361
0.429	0.3	0.404	0.405	0.415	0.384
	0.222	0.448	0.449	0.448	0.434
	0.176	0.476	0.477	0.464	0.469
0.667	0.4	0.485	0.485	0.503	0.475
	0.308	0.559	0.559	0.574	0.551
	0.25	0.611	0.612	0.615	0.602
1.000	0.5	0.582	0.581	0.591	0.583
	0.4	0.704	0.705	0.710	0.700
	0.333	0.775	0.776	0.786	0.781
1.500	0.6	0.675	0.675	0.679	0.689
	0.5	0.848	0.848	0.857	0.857
	0.429	0.972	0.972	0.981	0.981
2.333	0.7	0.765	0.765	0.765	0.778
	0.609	1.010	1.010	1.010	1.000
	0.538	1.190	1.190	1.20	1.190

Table 4.3 Simulation study of $MSE(\widehat{y}_i^M - y_i)$ $(d = 9, m = 99)$

σ_b^2	γ_i	Unrestricted Simulation MSE	Restricted Simulation MSE	Theoretical MSE	Estimated MSE
0.010	0.0099	0.049	0.049	0.073	0.127
	0.0066	0.049	0.049	0.057	0.132
	0.0050	0.048	0.049	0.050	0.135
0.100	0.0909	0.136	0.136	0.148	0.174
	0.0625	0.139	0.139	0.140	0.185
	0.0476	0.143	0.143	0.135	0.192
0.250	0.2	0.248	0.248	0.252	0.257
	0.143	0.263	0.263	0.261	0.280
	0.111	0.271	0.272	0.265	0.294
0.429	0.3	0.350	0.350	0.351	0.348
	0.222	0.389	0.389	0.383	0.389
	0.176	0.409	0.409	0.401	0.415
0.667	0.4	0.456	0.456	0.451	0.453
	0.308	0.523	0.524	0.516	0.526
	0.25	0.561	0.562	0.555	0.572
1.000	0.5	0.551	0.552	0.551	0.554
	0.4	0.658	0.658	0.660	0.669
	0.333	0.749	0.749	0.731	0.751
1.500	0.6	0.644	0.644	0.649	0.648
	0.5	0.810	0.810	0.815	0.818
	0.429	0.941	0.942	0.932	0.940
2.333	0.7	0.737	0.737	0.745	0.739
	0.609	0.972	0.972	0.980	0.971
	0.538	1.150	1.150	1.160	1.150

Table 4.4 Simulation study of $MSE(\widehat{y}_i^M - y_i)$ $(d = 14, m = 99)$

σ_b^2	γ_i	Unrestricted Simulation MSE	Restricted Simulation MSE	Theoretical MSE	Estimated MSE
0.010	0.0099	0.047	0.047	0.073	0.121
	0.0066	0.045	0.045	0.057	0.125
	0.0050	0.044	0.045	0.050	0.126
0.100	0.0909	0.129	0.129	0.147	0.169
	0.0625	0.131	0.131	0.139	0.177
	0.0476	0.135	0.135	0.135	0.182
0.250	0.2	0.246	0.246	0.250	0.246
	0.143	0.261	0.261	0.259	0.266
	0.111	0.266	0.266	0.263	0.277
0.429	0.3	0.350	0.350	0.346	0.337
	0.222	0.380	0.380	0.379	0.375
	0.176	0.397	0.397	0.396	0.398
0.667	0.4	0.448	0.447	0.443	0.440
	0.308	0.511	0.511	0.508	0.508
	0.25	0.556	0.556	0.547	0.551
1.000	0.5	0.542	0.543	0.541	0.535
	0.4	0.648	0.649	0.649	0.642
	0.333	0.727	0.728	0.719	0.713
1.500	0.6	0.643	0.643	0.638	0.639
	0.5	0.801	0.801	0.800	0.802
	0.429	0.912	0.912	0.915	0.919
2.333	0.7	0.735	0.735	0.733	0.731
	0.609	0.954	0.954	0.962	0.961
	0.538	1.130	1.130	1.140	1.130

Table 4.5 Simulation study of $MSE(\hat{y}_i^M - y_i)$ $(d = 14, m = 225)$

σ_b^2	γ_i	Unrestricted Simulation MSE	Restricted Simulation MSE	Theoretical MSE	Estimated MSE
0.010	0.0099	0.027	0.027	0.038	0.077
	0.0066	0.026	0.026	0.031	0.078
	0.0050	0.025	0.025	0.027	0.078
0.100	0.0909	0.110	0.110	0.116	0.126
	0.0625	0.112	0.112	0.114	0.130
	0.0476	0.112	0.112	0.113	0.132
0.250	0.2	0.225	0.225	0.224	0.230
	0.143	0.239	0.239	0.236	0.246
	0.111	0.243	0.244	0.242	0.255
0.429	0.3	0.327	0.327	0.325	0.328
	0.222	0.362	0.363	0.358	0.364
	0.176	0.378	0.378	0.376	0.385
0.667	0.4	0.431	0.431	0.427	0.428
	0.308	0.501	0.501	0.490	0.494
	0.25	0.536	0.537	0.528	0.536
1.000	0.5	0.527	0.527	0.528	0.530
	0.4	0.632	0.632	0.633	0.639
	0.333	0.704	0.704	0.702	0.712
1.500	0.6	0.625	0.625	0.628	0.623
	0.5	0.792	0.792	0.787	0.781
	0.429	0.897	0.897	0.899	0.894
2.333	0.7	0.723	0.723	0.726	0.724
	0.609	0.947	0.947	0.952	0.948
	0.538	1.120	1.120	1.130	1.120

5 THE EFFECTS OF NONNORMAL ERRORS

5.1 Introduction

The results of Chapter 2 through Chapter 4 are all based on normal sampling errors. The formulas for the MSE of the EBLUP estimator are slightly different if the sampling errors are not normally distributed because $Var(\widehat{\sigma}^2_{ei} - \sigma^2_{ei})$ is different. Lahiri and Rao (1995) discussed the robustness of the MSE estimator with non-normally distributed b_i. We will study the robustness of the MSE estimator we derived in Chapter 3 with respect to non-normal sampling errors in this chapter.

Another impact of non-normal sampling errors is that the EBLUP is biased if the sampling errors are not symmetrically distributed. For the random effects model, we consider the estimator defined in (3.43)

$$\widehat{y}_i = \widehat{\gamma}_i Y_i + (1 - \widehat{\gamma}_i)\boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}}_O, \tag{5.1}$$

where $\widehat{\gamma}_i = (\widehat{\sigma}^2_{b,m} + \widehat{\sigma}^2_{ei})^{-1}\widehat{\sigma}^2_{b,m}$, which is defined in (3.44). The only difference between $\widehat{y}_i$ and the EBLUP $\widehat{y}_i^H$ is the estimator of $\boldsymbol{\beta}$ used. The $\widehat{y}_i$ uses $\widehat{\boldsymbol{\beta}}_O = (\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{Y}$ as the estimator of $\boldsymbol{\beta}$. The EBLUP $\widehat{y}_i^H$ uses $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\widehat{\boldsymbol{V}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\widehat{\boldsymbol{V}}^{-1}\boldsymbol{Y}$ as the estimator of $\boldsymbol{\beta}$. If the $\boldsymbol{W}_m$ is chosen to be close to $\boldsymbol{V}^{-1}$, $\boldsymbol{W}_m$ should be reasonably close to $\widehat{\boldsymbol{V}}^{-1}$. For simplicity, we will discuss the bias of $\widehat{y}_i$, which has been discussed extensively in Chapter 3 and Chapter 4, instead of the bias of the EBLUP $\widehat{y}_i^H$. Note that

$$\widehat{y}_i = \gamma_i Y_i + (1 - \gamma_i)\boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}}_O + (\widehat{\gamma}_i - \gamma_i)(Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}) - (\widehat{\gamma}_i - \gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta}) \tag{5.2}$$

Since

$$E\left\{\gamma_i Y_i + (1-\gamma_i)\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}}_O\right\} = \boldsymbol{x}_i^T\boldsymbol{\beta},$$

the possible bias of $\widehat{y}_i$ is

$$E\left\{(\widehat{\gamma}_i - \gamma_i)(Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}) - (\widehat{\gamma}_i - \gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})\right\}. \tag{5.3}$$

We will assess this bias for non-symmetric sampling errors $e_i, i = 1,\ldots,m$ in this chapter.

5.2 Approximation to bias for non-normal sampling errors

We use Taylor expansion to get an approximation to the bias of $\widehat{y}_i$ defined in (5.3). We assume that the conditions of Theorem 3.3 hold except the assumptions about the distributions of b_i and e_i. We assume that $b_i, i = 1,\ldots,m$ are independent and identically distributed (iid) symmetric random variables with $E(b_i) = 0, V(b_i) = \sigma_b^2$, and finite eighth moment. We assume the e_i's are independent and identically distributed sampling errors with $E(e_i) = 0$, $V(e_i) = \sigma_{ei}^2$, and finite eighth moment. We also assume that all b_i are independent of all e_i. By Taylor expansion, we have

$$\begin{aligned}
E\left\{(\widehat{\gamma}_i - \gamma_i)(Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})\right\} &= E\{(\sigma_b^2+\sigma_{ei}^2)^{-2}[\sigma_{ei}^2(\widehat{\sigma}_{b,m}^2 - \sigma_b^2) - \sigma_b^2(\widehat{\sigma}_{ei}^2 - \sigma_{ei}^2)] \\
&\quad +(\sigma_b^2+\sigma_{ei}^2)^{-3}[-\sigma_{ei}^2(\widehat{\sigma}_{b,m}^2 - \sigma_b^2)^2 + \sigma_b^2(\widehat{\sigma}_{ei}^2 - \sigma_{ei}^2)^2 \\
&\quad +(\sigma_b^2-\sigma_{ei}^2)(\widehat{\sigma}_{b,m}^2 - \sigma_b^2)(\widehat{\sigma}_{ei}^2 - \sigma_{ei}^2)](b_i+e_i)\} + O(r_m) \\
&= E\{(\sigma_b^2+\sigma_{ei}^2)^{-2}[\sigma_{ei}^2(\widehat{\sigma}_{b,m}^2 - \sigma_b^2) - \sigma_b^2(\widehat{\sigma}_{ei}^2 - \sigma_{ei}^2)] \\
&\quad +(\sigma_b^2+\sigma_{ei}^2)^{-3}[-\sigma_{ei}^2(\widehat{\sigma}_{b,m}^2 - \sigma_b^2)^2 + \sigma_b^2(\widehat{\sigma}_{ei}^2 - \sigma_{ei}^2)^2 \\
&\quad +(\sigma_b^2-\sigma_{ei}^2)(\widehat{\sigma}_{b,m}^2 - \sigma_b^2)(\widehat{\sigma}_{ei}^2 - \sigma_{ei}^2)]e_i\} + O(r_m).
\end{aligned} \tag{5.4}$$

In the sequence of samples defined in Chapter 3, $\widehat{\sigma}_{ei}^2 = O_p(1)$ and $\widehat{\sigma}_{ei}^2 - \sigma_{ei}^2 = O_p(d^{-0.5})$. It is desirable that $E\{(\widehat{\sigma}_{ei}^2 - \sigma_{ei}^2)e_i\}$ is of order $O(d^{-1})$. However, this is not necessarily true for any $\widehat{\sigma}_{ei}^2$ and e_i. We will consider the following scenario. For each small area i, we observe $Y_{ij}, j = 1,\ldots,d_i+1$. The Y_{ij} have common mean y_i and independent sampling error ϵ_{ij}, where $\epsilon_{ij} = (d_i + 1)^{\frac{1}{3}}U$ and U has mean zero and variance $\sigma^2_{U_i}$. We assume that U has finite fifth moment. Let $\kappa_{3i} = E(\epsilon^3_{ij}), j = 1, \ldots, d_i + 1$. We have $\kappa_{3i} = E\{(d_i + 1)U^3\} = O(d)$. Let $\kappa_{5i} = E(\epsilon^5_{ij}), j = 1, \ldots, d_i + 1$. We have $\kappa_{5i} = E\left\{(d_i + 1)^{\frac{5}{3}}U^5\right\} = O(d^{\frac{5}{3}}) = O(d^2)$. In other words, $Y_{ij} = y_i + \epsilon_{ij}$. Let $Y_i = \overline{Y}_{i.} = (d_i + 1)^{-1}\sum_{j=1}^{d_i+1} Y_{ij}$ and $e_i = \bar{\epsilon}_{i.} = (d_i + 1)^{-1}\sum_{j=1}^{d_i+1} \epsilon_{ij}$, which has mean zero and variance $\sigma^2_{ei} = (d_i + 1)^{-\frac{1}{3}}\sigma^2_{U_i}$. The random effects model is

$$Y_i = y_i + e_i, \tag{5.5}$$

$$y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + b_i. \tag{5.6}$$

The unbiased estimator of σ^2_{ei} is

$$\widehat{\sigma}^2_{ei} = (d_i + 1)^{-1}d_i^{-1}\sum_{j=1}^{d_i+1}(Y_{ij} - Y_i)^2 = (d_i + 1)^{-1}d_i^{-1}\sum_{j=1}^{d_i+1}(\epsilon_{ij} - e_i)^2.$$

We have

$$\begin{aligned}
E\left\{(\widehat{\sigma}^2_{ei} - \sigma^2_{ei})e_i\right\} &= E\left\{\widehat{\sigma}^2_{ei}e_i\right\} \\
&= (d_i + 1)^{-1}d_i^{-1}E\left\{(\sum_{j=1}^{d_i+1} \epsilon^2_{ij})e_i - (d_i + 1)e_i^3\right\} \\
&= (d_i + 1)^{-1}d_i^{-1}E\left\{(d_i + 1)^{-1}\sum_{j=1}^{d_i+1} \epsilon^3_{ij} - (d_i + 1)^{-2}\sum_{j=1}^{d_i+1} \epsilon^3_{ij}\right\} \\
&= (d_i + 1)^{-2}\kappa_{3i}.
\end{aligned} \tag{5.7}$$

and

$$E\left\{(\widehat{\sigma}^2_{b,m} - \sigma^2_b)e_i\right\} = m^{-1}E\left\{(\widehat{\sigma}^2_{ei} - \sigma^2_{ei})e_i\right\} = m^{-1}(d_i + 1)^{-2}\kappa_{3i}. \tag{5.8}$$

Since $|\kappa_{3i}| = O(d)$, we have

$$E\left\{(\widehat{\sigma}^2_{ei} - \sigma^2_{ei})e_i\right\} = O(d^{-1}) \tag{5.9}$$

and

$$E\left\{(\widehat{\sigma}^2_{b,m} - \sigma^2_b)e_i\right\} = O(m^{-1}d^{-1}). \tag{5.10}$$

Note that

$$E\left\{(\widehat{\sigma}_{ei}^2-\sigma_{ei}^2)^2\epsilon_i\right\} = E\left\{(\widehat{\sigma}_{ei}^2)^2\epsilon_i - 2\sigma_{ei}^2\widehat{\sigma}_{ei}^2\epsilon_i + \sigma_{ei}^4\epsilon_i\right\}$$
$$= (d_i+1)^{-2}d_i^{-2}E\left\{\left(\sum_{j=1}^{d_i+1} e_{ij}^2\right)^2 \epsilon_i - 2(d_i+1)\left(\sum_{j=1}^{d_i+1} e_{ij}^2\right) e_i^3 + (d_i+1)^2 e_i^5\right\}$$
$$-2\sigma_{ei}^2(d_i+1)^{-2}\kappa_{3i}. \tag{5.11}$$

Let $\epsilon_{ij} = (d_i+1)^{-\frac{1}{3}}e_{ij}$. Since ϵ_{ij} have finite fifth moment. we have

$$E\left\{\left((d_i+1)^{-1}\sum_{j=1}^{d_i+1}\epsilon_{ij}\right)^5\right\} = O(d^{-3})$$

by Theorem 3.1. Therefore,

$$\begin{aligned}(d_i+1)^{-2}d_i^{-2}E\left\{(d_i+1)^2e_i^5\right\} &= d_i^{-2}(d_i+1)^{\frac{5}{3}}E\left\{\left((d_i+1)^{-1}\sum_{j=1}^{d_i+1}\epsilon_{ij}\right)^5\right\} \\ &= d_i^{-2}(d_i+1)^{\frac{5}{3}}O(d^{-3}) = O(d^{-3}). \end{aligned} \tag{5.12}$$

Note that

$$\begin{aligned}(d_i+1)\left(\sum_{j=1}^{d_i+1}e_{ij}^2\right)e_i^3 &= (d_i+1)^{-1}\left(\sum_{j=1}^{d_i+1}e_{ij}^2\right)\left(\sum_{j=1}^{d_i+1}\sum_{k=1}^{d_i+1}e_{ij}e_{ik}\right)e_i \\ &= (d_i+1)^{-1}\left[\left(\sum_{j=1}^{d_i+1}e_{ij}^2\right)^2 e_i + \left(\sum_{j=1}^{d_i+1}e_{ij}^2\right)\left(\sum_{j=1}^{d_i+1}\sum_{k\neq j}e_{ij}e_{ik}\right)e_i\right].\end{aligned}$$

and

$$\begin{aligned}E\left\{\left(\sum_{j=1}^{d_i+1}e_{ij}^2\right)^2 e_i\right\} &= (d_i+1)^{-1}E\left\{2\left(\sum_{j=1}^{d_i+1}\sum_{k\neq j}e_{ij}^2e_{ik}^3\right)+\sum_{j=1}^{d_i+1}e_{ij}^5\right\} \\ &= 2d_i(d_i+1)\sigma_{ei}^2\kappa_{3i}+\kappa_{5i}. \end{aligned} \tag{5.13}$$

and

$$\begin{aligned}E\left\{\left(\sum_{j=1}^{d_i+1}e_{ij}^2\right)\left(\sum_{j=1}^{d_i+1}\sum_{k\neq j}e_{ij}e_{ik}\right)e_i\right\} &= (d_i+1)^{-1}E\left\{2\left(\sum_{j=1}^{d_i+1}e_{ij}^2\right)\left(\sum_{j=1}^{d_i+1}\sum_{k\neq j}e_{ij}e_{ik}^2\right)\right\} \\ &= 2(d_i+1)^{-1}E\left\{\sum_{j=1}^{d_i+1}\sum_{k\neq j}e_{ij}^3e_{ik}^2\right\} \\ &= 2d_i(d_i+1)\sigma_{ei}^2\kappa_{3i}. \end{aligned} \tag{5.14}$$

Therefore,

$$\begin{aligned}
&(d_i+1)^{-2}d_i^{-2}E\left\{\left(\sum_{j=1}^{d_i+1}\epsilon_{ij}^2\right)^2\epsilon_i-2(d_i+1)\left(\sum_{j=1}^{d_i+1}\epsilon_{ij}^2\right)\epsilon_i^3\right\}\\
&=(d_i+1)^{-2}d_i^{-2}E\left\{(1-2(d_i+1)^{-1})\left(\sum_{j=1}^{d_i+1}\epsilon_{ij}^2\right)^2\epsilon_i\right.\\
&\qquad\left.-2(d_i+1)^{-1}\left(\sum_{j=1}^{d_i+1}\epsilon_{ij}^2\right)\left(\sum_{j=1}^{d_i+1}\sum_{k\neq j}\epsilon_{ij}\epsilon_{ik}\right)\epsilon_i\right\}\\
&=2d_i^{-1}(d_i+1)^{-2}(d_i-3)\sigma_{ei}^2\kappa_{3i}+d_i^{-2}(d_i+1)^{-3}(d_i-1)\kappa_{5i}. \qquad (5.15)
\end{aligned}$$

By (5.11), (5.12), and (5.15), we have

$$E\left\{(\widehat{\sigma}_{ei}^2-\sigma_{ei}^2)^2\epsilon_i\right\}=-6\sigma_{ei}^2d_i^{-1}(d_i+1)^{-2}\kappa_{3i}+d_i^{-2}(d_i+1)^{-3}(d_i-1)\kappa_{5i}+O(d^{-3}). \quad (5.16)$$

Since $\kappa_{3i}=O(d)$ and $\kappa_{5i}=O(d^2)$, we have

$$E\left\{(\widehat{\sigma}_{ei}^2-\sigma_{ei}^2)^2e_i\right\}=O(d^{-2}). \quad (5.17)$$

We also have

$$E\left\{(\widehat{\sigma}_{b,m}^2-\sigma_b^2)(\widehat{\sigma}_{ei}^2-\sigma_{ei}^2)e_i\right\}=m^{-1}E\left\{(\widehat{\sigma}_{ei}^2-\sigma_{ei}^2)^2e_i\right\}=O(m^{-1}d^{-2}), \quad (5.18)$$

and

$$E\left\{(\widehat{\sigma}_{b,m}^2-\sigma_b^2)^2e_i\right\}=m^{-2}E\left\{(\widehat{\sigma}_{ei}^2-\sigma_{ei}^2)^2e_i\right\}=O(m^{-2}d^{-2}). \quad (5.19)$$

By (5.4), (5.7), (5.8), (5.16), (5.18), and (5.19), we have

$$\begin{aligned}
E\left\{(\widehat{\gamma}_i-\gamma_i)(Y_i-\boldsymbol{x}_i^T\boldsymbol{\beta})\right\} &= -(\sigma_b^2+\sigma_{ei}^2)^{-2}\sigma_b^2E\left\{(\widehat{\sigma}_{ei}^2-\sigma_{ei}^2)\epsilon_i\right\}+O(r_m)\\
&= -(d_i+1)^{-2}(\sigma_b^2+\sigma_{ei}^2)^{-2}\sigma_b^2\kappa_{3i}+O(r_m), \qquad (5.20)
\end{aligned}$$

where $r_m=O(max(m^{-1}d^{-1},d^{-2}))$.

By (3.55), we have

$$\boldsymbol{x}_j^T(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}\boldsymbol{x}_kw_{kk,m}\le m^{-1}C_x^2(\sum_{i=1}^p\sum_{j=1}^p a_{2ij})w_{kk,m}=m^{-1}C_4$$

for some constant C_4. Therefore.

$$\begin{aligned}
&\left|E\left\{(\widehat{\gamma}_i - \gamma_i)\boldsymbol{x}_j^T(\widehat{\boldsymbol{\beta}}_O - \boldsymbol{\beta})\right\}\right| \\
&= \left|E[(\widehat{\gamma}_i - \gamma_i)\sum_{k=1}^{m}\boldsymbol{x}_j^T(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}\boldsymbol{x}_k w_{kk.m}(b_k + e_k)]\right| \\
&\le m^{-1}C_4 E\left|\{(\widehat{\gamma}_i - \gamma_i)(b_i + e_i)\}\right| + m^{-1}C_4\sum_{k\ne i} E\left|\{(\widehat{\gamma}_i - \gamma_i)(b_k + e_k)\}\right| \\
&= O(m^{-1}d^{-1}) + m^{-1}(m-1)O(m^{-1}d^{-1}) = O(m^{-1}d^{-1}). \qquad (5.21)
\end{aligned}$$

Hence. an order $O(r_m)$ approximation to the bias in (5.3) is

$$-(d_i + 1)^{-2}(\sigma_b^2 + \sigma_{ei}^2)^{-2}\sigma_b^2\kappa_{3i}. \qquad (5.22)$$

5.3 Simulation study

We use the settings of the simulation study in Chapter 3. The only difference is that e_{ij} is not normally distributed. In this section. $e_{ij} \sim (d+1)^{0.5}6^{-0.5}\sigma_{ei}(\chi_3^2 - 3)$. which has mean zero and variance $(d+1)\sigma_{ei}^2$. We set one third of the σ_{ei}^2 equal to 1.0. one third of the σ_{ei}^2 equal to 1.5. and one third of the σ_{ei}^2 equal to 2.0 and generate samples for $\sigma_b^2 = 0.1, 0.25, \frac{3}{7}, \frac{2}{3}, 1, 1.5, \frac{7}{3}$. For each of the parameter settings. we generated 1000 samples for (d, m)= (5, 36), (9, 99), and (14, 225).

The samples are used for two purposes. The first purpose is to assess the bias of $\widehat{y}_i$ under non-normal sampling errors. The theoretical approximation to the bias defined in (5.22) is

$$B = -4[6(d+1)]^{-0.5}(\sigma_b^2 + \sigma_{ei}^2)^{-2}\,\sigma_b^2\,\sigma_{ei}^3. \qquad (5.23)$$

An estimator of the theoretical approximation is

$$\widehat{B} = -4[6(d+1)]^{-0.5}(\widehat{\sigma}_b^2 + \widehat{\sigma}_{ei}^2)^{-2}\,\widehat{\sigma}_b^2\,\widehat{\sigma}_{ei}^3. \qquad (5.24)$$

We have three distinct σ_{ei}^2. We divide the data into three groups according to the values of σ_{ei}^2. For each sample, we compute the mean of the simulated biases of $\widehat{y}_i$ for each of the three groups. the theoretical bias defined in (5.23) for each of the three groups. and the estimated biases defined in (5.24) for each of the three groups. We then compute the mean of the 1000 samples for these values. We also compute the variance of the 1000 samples of the simulated bias of $\widehat{y}_i$. We want to determine if the simulated "bias" is really a bias and is not due to the random variation. For $(d, m, \sigma_b^2, \sigma_{ei}^2) = (5, 36, 0.1, 1)$. the mean of the simulated bias of $\widehat{y}_i$ is -0.0647 and the corresponding standard error is 0.007. The t-statistic is 9.18. Therefore. the simulated bias of $\widehat{y}_i$ is significantly different from zero. Similarly. the t-statistic is 10.66 for $(d, m, \sigma_b^2, \sigma_{ei}^2) = (5, 36, 0.25, 1)$. We computed a test for all the $(d, m, \sigma_b^2, \sigma_{ei}^2)$ combinations. The conclusion is that the bias of $\widehat{y}_i$ is significantly different from zero.

We discussed in Chapter 4 that the ratio adjusted estimator described in Chapter 2 is in the same family of restricted estimators as $\widehat{y}_i^M$ defined in (4.41). These restricted estimators make an adjustment such that the (weighted) sum or mean of the adjusted small areas estimators is equal to the direct survey estimator of the total or mean for a larger area. Another feature of restricted estimators such as $\widehat{y}_i^M$ is that the restriction removes most of the bias in the small area estimator $\widehat{y}_i$. We made this claim in Chapter 2. For each sample, we compute the mean of the simulated biases of $\widehat{y}_i^M$. the restricted estimator of y_i, for each of the three groups. We then compute the mean and variance of the 1000 simulated biases of $\widehat{y}_i^M$. The t-statistic for bias of $\widehat{y}_i^M$ is 0.915 when $(d, m, \sigma_b^2, \sigma_{ei}^2) = (5, 36, 0.1, 1)$ and is 1.34 $(d, m, \sigma_b^2, \sigma_{ei}^2) = (5, 36, 0.25, 1)$. There are a few cases where the t-statistic for the bias of $\widehat{y}_i^M$ is significantly different from zero. For example, the t-statistic is 3.38 for $(d, m, \sigma_b^2, \sigma_{ei}^2) = (9, 99, 2.33, 2)$. So we can not claim that $\widehat{y}_i^M$ is an unbiased estimator of y_i. However, the bias of $\widehat{y}_i^M$ is much smaller than that of $\widehat{y}_i$.

The means of the simulated biases of $\widehat{y}_i$, theoretical bias of $\widehat{y}_i$, the estimated biases of $\widehat{y}_i$, and the simulated biases of $\widehat{y}_i^M$are shown in Figure 5.1 through Figure 5.3. Each Figure corresponds to a (d, m) combination and consists of three plots. The first plot is for $\sigma_{\epsilon i}^2 = 1$: the second plot is for $\sigma_{\epsilon i}^2 = 1.5$: and the third plot is for $\sigma_{\epsilon i}^2 = 2$. In each plot, the mean values are plotted against the values of σ_b^2.

Note that the bias approximation defined in (5.22) does not depend on m explicitly. From these figures, we can see that the theoretical bias defined in (5.22) is a reasonable approximation to the simulated bias even for $d = 5$. For $d = 14$, the simulated biases are very close to the theoretical biases and the estimated biases. The following observations are reasonable if we keep (5.22) in mind. The bias decreases when d increases. For a given $\sigma_{\epsilon i}^2$ and (d, m), the bias increases first then decreases as σ_b^2 increases. The claim that the restriction removes most of the bias in the small area estimator $\widehat{y}_i$ is also supported by these figures. We can see that the simulated biases of the restricted estimator $\widehat{y}_i^M$ is very close to zero from these figures.

The second purpose of the simulation study is to assess the impact of the bias on the $MSE(\widehat{y}_i - y_i)$ when sampling errors are not normally distributed. A naive approximation to the MSE for non-normal sampling errors is

$$MSE_N(\widehat{y}_i - y_i) = B^2 + MSE(\widehat{y}_i - y_i), \tag{5.25}$$

where B is defined in (5.23) and $MSE(\widehat{y}_i - y_i)$ is defined in (3.67). An estimator of $MSE_N(\widehat{y}_i - y_i)$ is

$$\widehat{MSE}_N(\widehat{y}_i - y_i) = \widehat{B}^2 + \widehat{MSE}_2(\widehat{y}_i - y_i), \tag{5.26}$$

where $\widehat{B}$ is defined in (5.24) and $\widehat{MSE}_2(\widehat{y}_i - y_i)$ is defined in (3.75).

For each sample, we compute the mean of the simulated MSE of $\widehat{y}_i$ for each group. We also compute the theoretical MSE defined in (5.25) and the estimated MSE defined in (5.26) for each of the three groups. We then compute the mean of the 1000 samples for these values.

First, we assess the impact of the bias on the $MSE(\widehat{y}_i - y_i)$. The square of the simulated bias is less than 4% of the simulated $MSE(\widehat{y}_i - y_i)$. Therefore, the contribution to $MSE(\widehat{y}_i - y_i)$ due to the bias of $\widehat{y}_i$ is small. If we use the restricted estimator $\widehat{y}_i^M$ to estimate y_i, which is an almost unbiased estimator of y_i, the $MSE(\widehat{y}_i^M - y_i)$ is slightly larger than $MSE(\widehat{y}_i - y_i)$. The $\widehat{y}_i$ is a better estimator of y_i in terms of MSE than $\widehat{y}_i^M$.

However, the bias becomes a concern if we are also interested in estimating $\bar{y}_. = \sum_{i=1}^m \omega_i y_i$. An example is the NRI. The direct survey estimator Y_i is not in the officially released NRI data. The Y_i is replaced by the small area estimator. If we use $\widehat{y}_i$ as the small area estimator for all areas, then NRI users will get the biased estimator of $\bar{y}_.$, $\widehat{\bar{y}}_. = \sum_{i=1}^m \omega_i \widehat{y}_i$, and the bias of $\widehat{\bar{y}}_.$ is not negligible compared to $MSE(\widehat{\bar{y}}_. - \bar{y}_.)$. To see this, we let $\omega_i = m^{-1}$ and compute the simulated bias, $\widehat{\bar{y}}_. - \bar{y}_.$, for the 1000 samples. We also compute the simulated variance of $\widehat{\bar{y}}_. - \bar{y}_.$ for the 1000 samples. The simulated $MSE(\widehat{\bar{y}}_. - \bar{y}_.)$ is defined to be the sum of the square of the simulated bias and the simulated variance of $\widehat{\bar{y}}_. - \bar{y}_.$. We also compute the mean of $\bar{Y}_. - \bar{y}_.$ and variance of $\bar{Y}_. - \bar{y}_.$, where $\bar{Y}_. = \sum_{i=1}^m \omega_i Y_i = m^{-1} \sum_{i=1}^m Y_i$. The results are reported in Table 5.1. From this table, we can see that the bias remains at about the same level as m increases. However, the variances of $\widehat{\bar{y}}_. - \bar{y}_.$ and $\bar{Y}_. - \bar{y}_.$ decrease as m increases. The square of the simulated bias using $\widehat{\bar{y}}_.$ to estimate $\bar{y}_.$ ranges from 20% to 200% of the variance of $\widehat{\bar{y}}_. - \bar{y}_.$. The percentage increases as m increases. The restricted estimator $\widehat{y}_i^M$ should be used to remove the overall bias.

To assess the robustness of the $MSE(\widehat{y}_i - y_i)$ when sampling errors are not normally distributed, we also computed the mean of the simulated MSE with normal errors for each of the three groups from Chapter 3. The ratios of the theoretical MSE, estimated MSE, and the simulated MSE with normal errors to the simulated MSE of non-normal errors are computed and the results are shown in Figure 5.4 through Figure 5.6. Each Figure corresponds to a (d, m) combination and consists of three plots. The first plot is for $\sigma_{ei}^2 = 1$; the second plot is for $\sigma_{ei}^2 = 1.5$; and the third plot is for $\sigma_{ei}^2 = 2$. In each plot, the ratios are plotted against the values of σ_b^2.

From these plots, we have the following observations. The naive theoretical MSE is slightly larger than the simulated MSE for normal errors, in general. The naive theoretical MSE underestimates the simulated MSE for non-normal errors when σ_b^2 is small and overestimates the simulated MSE for non-normal errors when σ_b^2 is large. The simulated MSE for normal errors underestimates the simulated MSE for non-normal errors when σ_b^2 is small. The maximum underestimation is less than 15% and the underestimation decreases as σ_b^2 increases. For large σ_b^2, e.g., $\sigma_b^2 = \frac{3}{2}$ and $\sigma_b^2 = \frac{7}{3}$, the simulated MSE for normal errors is very close to the simulated MSE for non-normal errors. Therefore, our approximation to $MSE(\widehat{y}_i - y_i)$ based upon normal errors is quite robust with respect to non-normal sampling errors. The estimated MSE for non-normal errors is slightly smaller than the simulated MSE for normal errors except for very small σ_b^2. The estimated MSE for non-normal errors also underestimates the simulated MSE for non-normal errors except for very small σ_b^2. The underestimation is usually between 2% and 10% and the underestimation decreases as σ_b^2 increases.

Table 5.1 Comparison of Monte Carlo Bias and MSE of $\hat{\bar{y}}_{.} - \bar{y}_{.}$ (1000 samples)

(d, m, σ_b^2)	Mean of $\hat{\bar{y}}_{.} - \bar{y}_{.}$	Variance of $\hat{\bar{y}}_{.} - \bar{y}_{.}$	MSE of $\hat{\bar{y}}_{.} - \bar{y}_{.}$	Mean of $\bar{Y}_{.} - \bar{y}_{.}$	Variance of $\bar{Y}_{.} - \bar{y}_{.}$
5,36,0.100	-0.0672	0.0454	0.0499	0.01094	0.0402
5,36,0.250	-0.0876	0.0485	0.0562	0.00551	0.0424
5,36,0.429	-0.1178	0.0480	0.0619	0.001694	0.0411
5,36,0.667	-0.1388	0.0453	0.0646	0.00259	0.0419
5,36,1.000	-0.1501	0.0422	0.0647	-0.00120	0.0414
5,36,1.500	-0.1588	0.0418	0.0670	-0.00660	0.0423
5,36,2.333	-0.1463	0.0380	0.0594	-0.00602	0.0398
9,99,0.100	-0.0568	0.0177	0.0209	-0.00120	0.0150
9,99,0.250	-0.0763	0.0175	0.0233	0.00175	0.0149
9,99,0.429	-0.1027	0.0172	0.0277	0.00211	0.0153
9,99,0.667	-0.1261	0.0158	0.0317	-0.00247	0.0148
9,99,1.000	-0.132	0.0137	0.0311	0.00104	0.0141
9,99,1.500	-0.1335	0.0148	0.0326	-0.00156	0.0143
9,99,2.333	-0.1290	0.0149	0.0315	-0.00458	0.0152
14,225,0.100	-0.0393	0.0084	0.0099	0.00043	0.0070
14,225,0.250	-0.0694	0.0077	0.0125	-0.00209	0.0070
14,225,0.429	-0.0915	0.0073	0.0157	-0.00317	0.0068
14,225,0.667	-0.0100	0.0069	0.0169	0.00303	0.0067
14,225,1.000	-0.1124	0.0065	0.0192	-0.00017	0.0066
14,225,1.500	-0.1159	0.0065	0.0200	-0.00107	0.0070
14,225,2.333	-0.1033	0.0061	0.0168	0.00422	0.0063

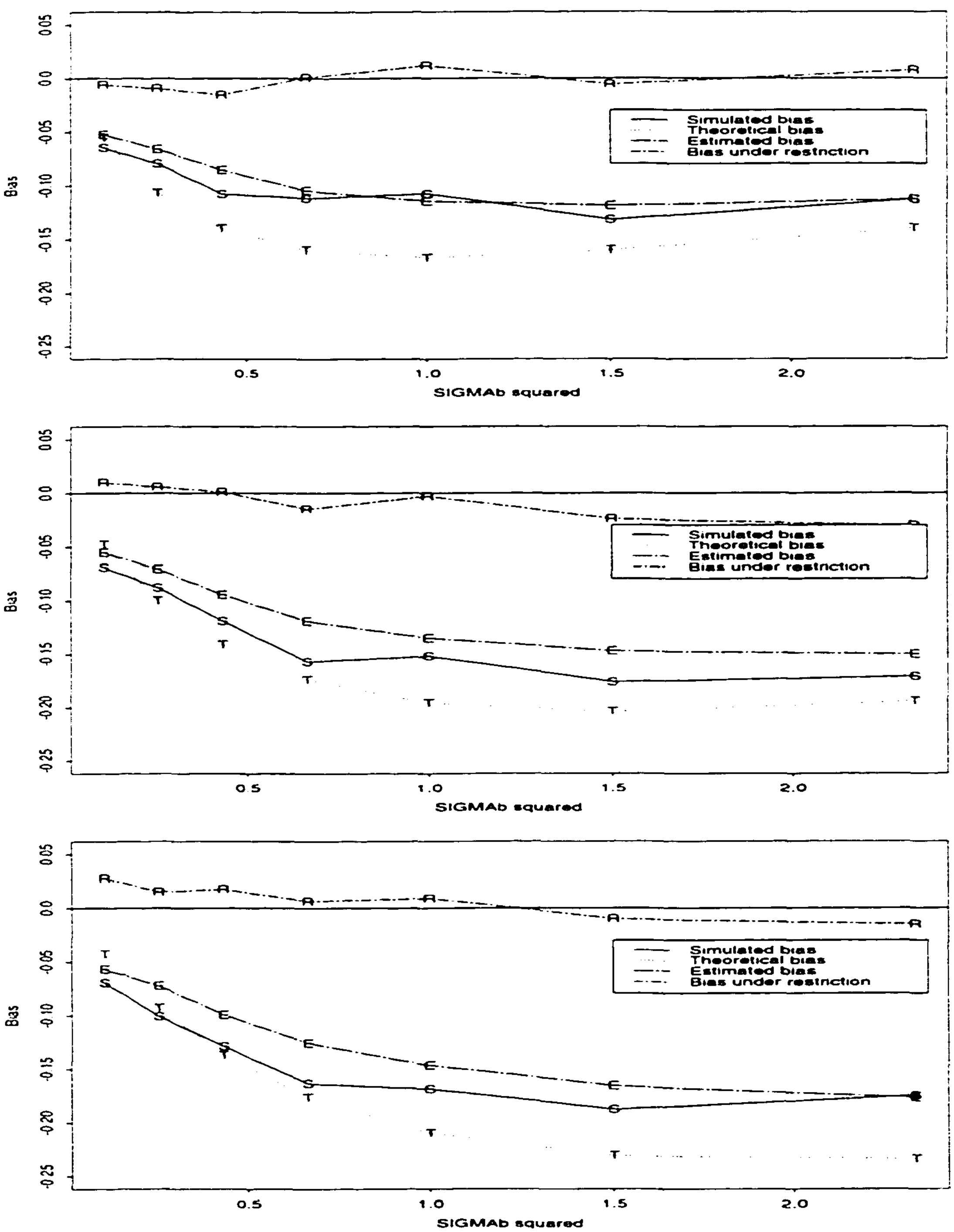

Figure 5.1 Monte Carlo bias of $\widehat{y}_i$ when $(m, d) = (36, 5)$ and e_i are centered χ^2_3 distribution with variance σ^2_{ei}. The plots are for $\sigma^2_{ei} = 1, 1.5$, and 2, respectively.

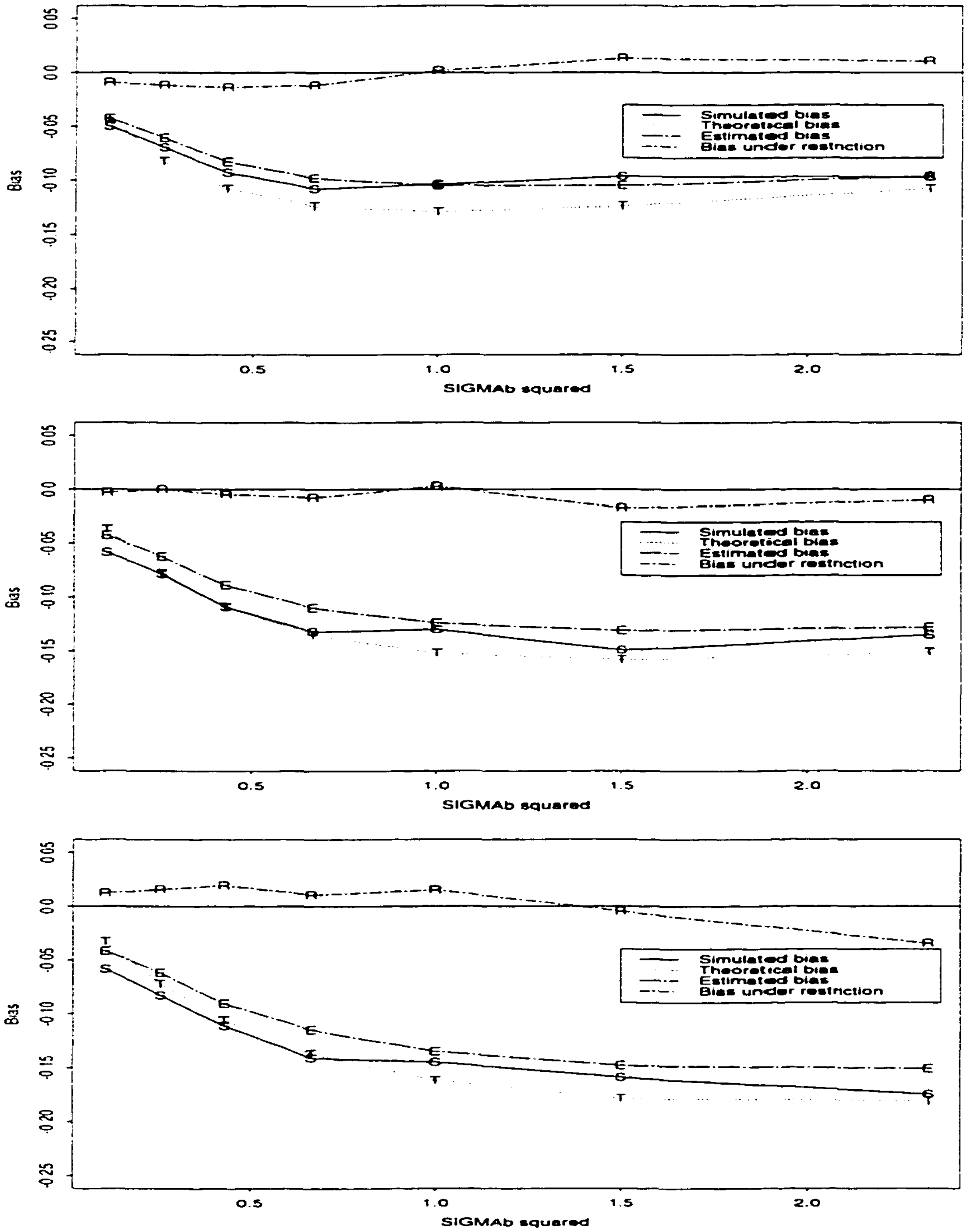

Figure 5.2 Monte Carlo bias of $\widehat{y}_i$ when $(m, d) = (99, 9)$ and e_i are centered χ^2_3 distribution with variance σ^2_{ei}. The plots are for $\sigma^2_{ei} = 1, 1.5$, and 2, respectively.

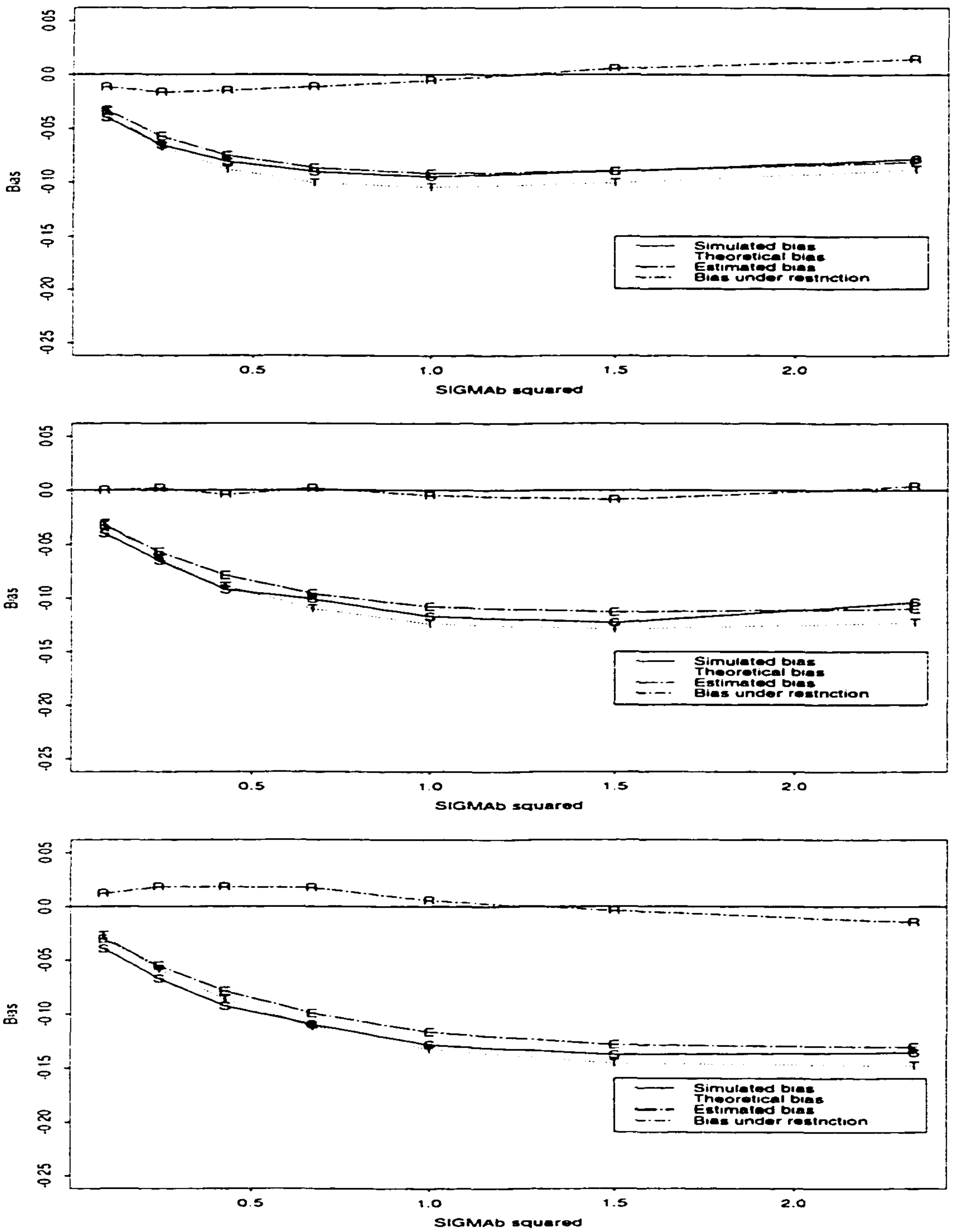

Figure 5.3 Monte Carlo bias of $\widehat{y}_i$ when $(m, d) = (225, 14)$ and ϵ_i are centered χ_3^2 distribution with variance σ_{ei}^2. The plots are for $\sigma_{ei}^2 = 1, 1.5$, and 2, respectively.

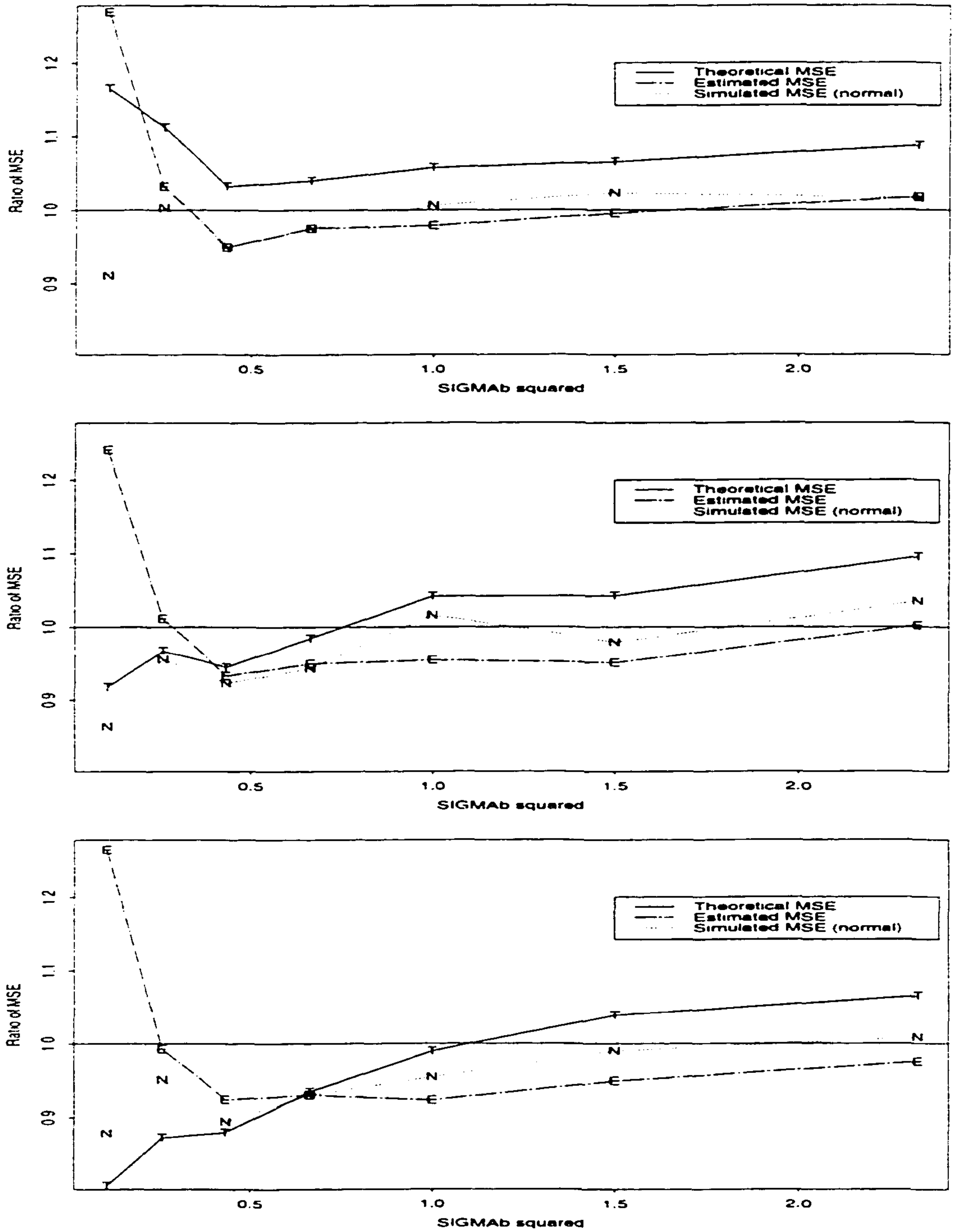

Figure 5.4 Monte Carlo MSE ratio when $(m, d) = (36, 5)$ and ϵ_i are centered χ^2_3 distribution with variance $\sigma^2_{\epsilon i}$. The plots are for $\sigma^2_{\epsilon i} = 1, 1.5$, and 2, respectively.

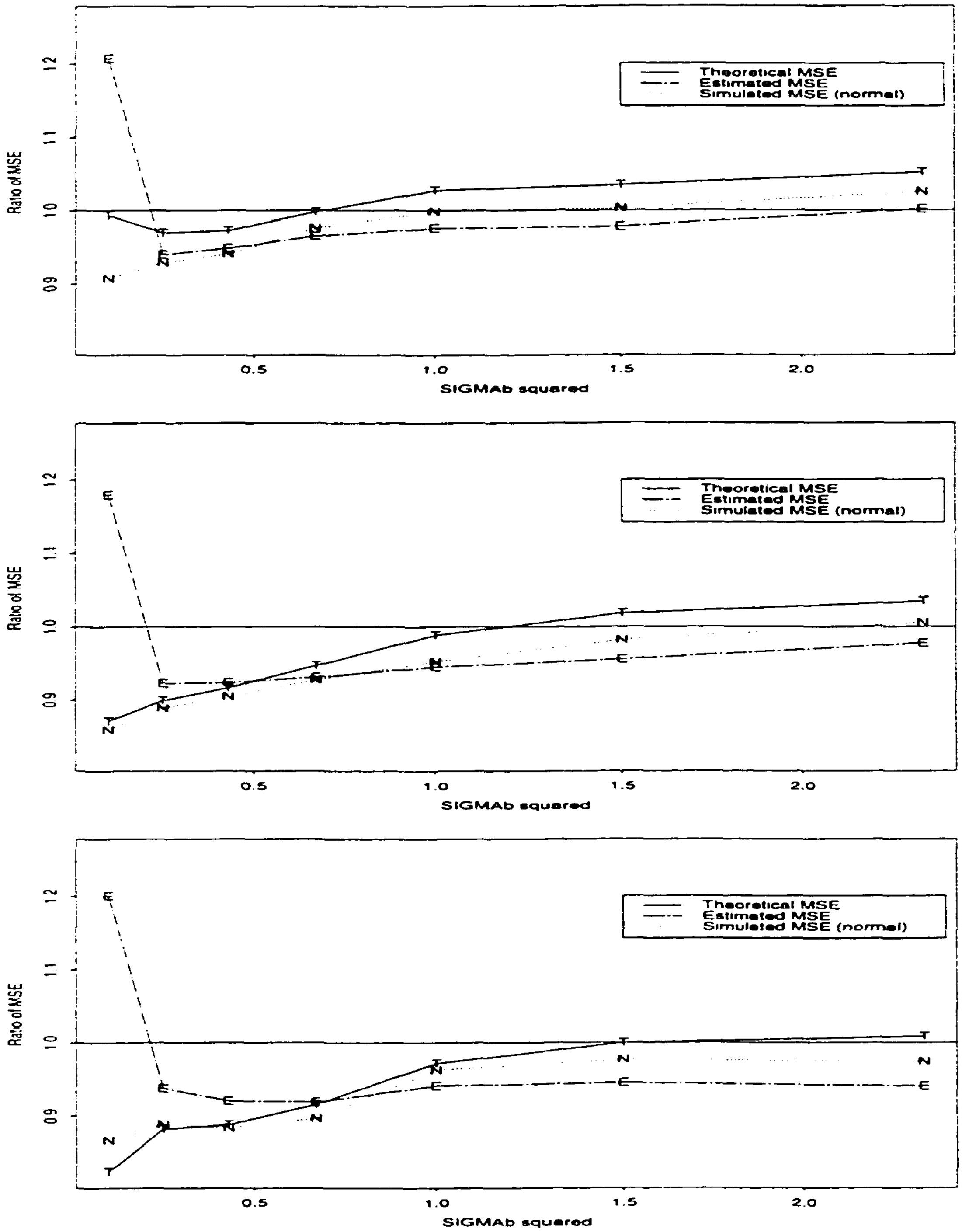

Figure 5.5 Monte Carlo MSE ratio when $(m, d) = (99, 9)$ and e_i are centered χ_3^2 distribution with variance σ_{ei}^2. The plots are for $\sigma_{ei}^2 = 1, 1.5$, and 2, respectively.

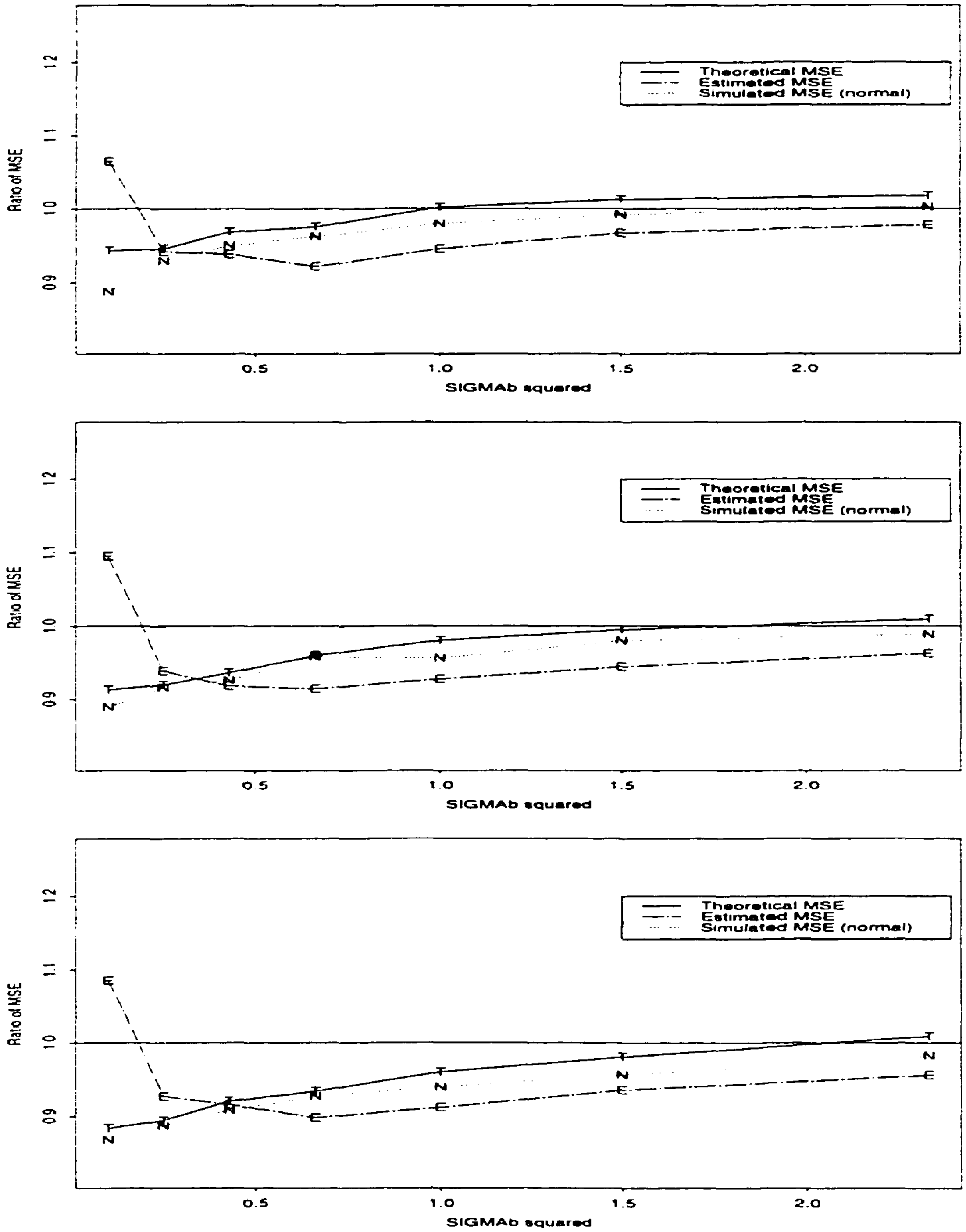

Figure 5.6 Monte Carlo MSE ratio when $(m, d) = (225, 14)$ and ϵ_i are centered χ_3^2 distribution with variance σ_{ei}^2. The plots are for $\sigma_{ei}^2 = 1, 1.5$, and 2, respectively.

6 SUMMARY

The random effects model for small area estimation is

$$y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + z_i b_i, \tag{6.1}$$

where y_i are unobservable small area means and

$$Y_i = y_i + e_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + z_i b_i + e_i, \qquad i = 1, \cdots, m, \tag{6.2}$$

where Y_i are direct survey estimators, $\boldsymbol{x}_i^T$ are known constants and the z_i are known positive constants. $\boldsymbol{\beta}$ is the vector of regression parameters, the b_i's are independent and identically distributed random variables with $E(b_i) = 0$ and $V(b_i) = \sigma_b^2$, and the e_i's are sampling errors with $E(e_i|y_i) = 0$ and $V(e_i|y_i) = \sigma_{ei}^2$. We are interested in estimating the small area means $y_i, i = 1, \ldots, m$. The empirical best linear unbiased predictor (EBLUP) is

$$\widehat{y}_i^H = \widehat{\gamma}_i Y_i + (1 - \widehat{\gamma}_i) \boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}}, \tag{6.3}$$

where

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T \widehat{\boldsymbol{V}} \boldsymbol{X})^{-1} \boldsymbol{X}^T \widehat{\boldsymbol{V}} \boldsymbol{Y}, \tag{6.4}$$

$$\widehat{\gamma}_i = (\widehat{\sigma}_b^2 + \widehat{\sigma}_{ei}^2)^{-1} \widehat{\sigma}_b^2. \tag{6.5}$$

$\widehat{\sigma}_b^2$ is the estimator of σ_b^2, and $\widehat{\sigma}_{ei}^2$ is the estimator of σ_{ei}^2.

In this dissertation, a practical application of EBLUP small area estimation in the National Resources Inventory, a large land survey of the nation's non-federal land area, is described. Several estimation issues raised by this application are discussed as motivation for the theoretical investigation of small area estimation.

For small area estimation, we not only need to estimate the small area statistics y_i but also need to derive the MSE of the small area estimators. Prasad and Rao (1990) proposed an approximation to $MSE(\widehat{y}_i^H - y_i)$ when the unknown variance components depend on a fixed number of parameters. Lahiri and Rao (1995) use simulation to show that the approximation proposed by Prasad and Rao is robust to departures from the assumption of normality of b_i under the random effects model (6.2) when the sampling variances σ_{ei}^2 are known.

In many practical applications, the σ_{ei}^2 are unknown and estimated. If the estimates $\widehat{\sigma}_{ei}^2$ depend on a fixed number of parameters, we can use Prasad and Rao's approximation

$$MSE(\widehat{y}_i^H - y_i) = \sigma_{ei}^2\gamma_i + (1-\gamma_i)^2\boldsymbol{x}_i^T V(\widehat{\boldsymbol{\beta}})\boldsymbol{x}_i + tr[(\boldsymbol{B}'\boldsymbol{V}\boldsymbol{B}E(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})']. \quad (6.6)$$

where

$$V(\widehat{\boldsymbol{\beta}}) \doteq (\boldsymbol{X}^T\widehat{\boldsymbol{V}}\boldsymbol{X})^{-1},$$

$$\boldsymbol{B} = [\partial\boldsymbol{\gamma}_i/\partial\theta_1, \ldots, \partial\boldsymbol{\gamma}_i/\partial\theta_p],$$

and $\boldsymbol{\gamma}_i$ is the vector with i-th element equal to $\widehat{\gamma}_i$ and other elements equal to zero.

However, it is difficult to find a good model for the estimator of sampling error variances $\widehat{\sigma}_{ei}^2$ in many applications. So we consider the EBLUP for y_i by using the individual direct variance estimator $\widehat{\sigma}_{ei}^2$ from each small area. The derivation procedure of the $MSE(\widehat{y}_i^H - y_i)$ in (6.6) outlined in Prasad and Rao (1990) is not appropriate for this problem because the dimension of the variance component parameter $\boldsymbol{\theta} = (\sigma_b^2, \sigma_{e1}^2, \ldots, \sigma_{em}^2)^T$ is $m+1$, which is not fixed. We studied the estimator

$$\widehat{y}_i = \widehat{\gamma}_i Y_i + (1-\widehat{\gamma}_i)\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}}_O, \quad (6.7)$$

where $\widehat{\boldsymbol{\beta}}_O = (\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{Y}$, $\boldsymbol{W}_m$ is a known fixed diagonal matrix, and $\widehat{\gamma}_i = (\widehat{\sigma}_b^2 + \widehat{\sigma}_{ei}^2)^{-1}\widehat{\sigma}_b^2$. By using a Taylor expansion, we have

$$\begin{aligned} MSE(\widehat{y}_i - y_i) &= \sigma_{ei}^2\gamma_i + (1-\gamma_i)^2\boldsymbol{x}_i^T V(\widehat{\boldsymbol{\beta}}_O)\boldsymbol{x}_i \\ &\quad + (\sigma_b^2 + \sigma_{ei}^2)V(\widehat{\gamma}_i) + O(r_m), \end{aligned} \quad (6.8)$$

where

$$V(\widehat{\boldsymbol{\beta}}_O) = (\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{V}\boldsymbol{W}_m\boldsymbol{X})(\boldsymbol{X}^T\boldsymbol{W}_m\boldsymbol{X})^{-1},$$

$$\boldsymbol{B} = [\partial\gamma_i/\partial\theta_1, \ldots, \partial\gamma_i/\partial\theta_p],$$

$$\begin{aligned} V(\widehat{\gamma}_i) &= (\sigma_b^2 + \sigma_{ei}^2)^{-4} 2\sigma_{ei}^4 \left\{ \sigma_b^4 d_i^{-1} + \frac{1}{2} V(\widehat{\sigma}_{b,m}^2) \right\} \\ &= (\sigma_b^2 + \sigma_{ei}^2)^{-4} \left\{ \sigma_b^4 V(\widehat{\sigma}_{ei}^2) + \sigma_{ei}^2 V(\widehat{\sigma}_{b,m}^2) \right\}, \end{aligned} \tag{6.9}$$

and $O(r_m) = \max(m^{-1.5}, m^{-1}d^{-1}, d^{-2})$. We also derived $E\{(\widehat{y}_i - y_i)(\widehat{y}_j - y_j)\}$.

Ignoring the fact that the dimension of $\boldsymbol{\theta}$ is not fixed and using Prasad and Rao's approximation (6.6) by letting $\boldsymbol{\theta} = (\sigma_b^2, \sigma_{e1}^2, \ldots, \sigma_{em}^2)^T$, we have

$$\begin{aligned} MSE(\widehat{y}_i^H - y_i) &= \sigma_{ei}^2\gamma_i + (1-\gamma_i)^2 \boldsymbol{x}_i^T V(\widehat{\boldsymbol{\beta}})\boldsymbol{x}_i + (\sigma_b^2 + \sigma_{ei}^2)^{-3} \left\{ \sigma_b^4 V(\widehat{\sigma}_{ei}^2) + \sigma_{ei}^2 V(\widehat{\sigma}_{b,m}^2) \right. \\ &\quad \left. - 2\sigma_b^2\sigma_{ei}^2 cov(\widehat{\sigma}_{b,m}^2, \widehat{\sigma}_{ei}^2) \right\}. \end{aligned} \tag{6.10}$$

Since $2\sigma_b^2\sigma_{ei}^2 cov(\widehat{\sigma}_{b,m}^2, \widehat{\sigma}_{ei}^2) = O(r_m)$, we omitted this term in our approximation. Therefore, our result is a generalization of Prasad and Rao's approximation when the σ_{ei}^2 are estimated individually.

Simulations are used to study the properties of the theoretical approximations. The simulations in other literature, such as that of Laird and Louis (1987), Prasad and Rao (1990), and Lahiri and Rao (1995), assume the σ_{ei}^2 are known constants. The only unknown variance parameter in their studies is σ_b^2. In our simulation, σ_{ei}^2 are unknown parameters. When $\gamma_i = (\sigma_b^2 + \sigma_{ei}^2)^{-1}\sigma_b^2$ is very small, there is a severe overestimation of $MSE(\widehat{y}_i^H - y_i)$ by simply replacing the parameters in the variance expression with their estimators. An estimator of $MSE(\widehat{y}_i - y_i)$ is developed that is superior to the related estimators in the literature when γ_i is small. Both our simulation and other simulations in the literature show that the bias of the variance estimator is smaller for larger γ_i.

In practical applications, it is often desirable that the weighted total of the small area estimators be equal to the the weighted total of the direct survey estimators. Therefore.

it is necessary to put a restriction on the weighted total of the small area estimators in many situations. A criterion that unifies the derivation of several restricted estimators is proposed. The estimator, denoted by $\widehat{y}_i^M$, that is the unique best linear unbiased estimator satisfying the criterion is derived. An approximation to $MSE(\widehat{y}_i^M - y_i)$ is presented. Simulation results indicate that the estimator of the mean square error yields a reasonable approximation to the true mean square error.

The EBLUP is a biased estimator of y_i if the sampling errors e_i are not symmetrically distributed. The approximate bias of $\widehat{y}_i$ for the non-symmetrically distributed e_i is derived. We used simulations to demonstrate that the restricted estimator $\widehat{y}_i^M$ reduces the bias for non-symmetric sampling errors. We also used simulation to assess the robustness of the theoretical and estimated $MSE(\widehat{y}_i - y_i)$ when sampling errors e_i are not normally distributed. For the particular choice of e_i that have a centered chi-square distribution with three degrees of freedom (i.e. the mean is zero), our simulation results indicate that the theoretical MSE underestimates the simulated MSE when σ_b^2 is small and overestimates the simulated MSE when σ_b^2 is large. The estimated MSE underestimates the simulated MSE except for very small σ_b^2. The underestimation is usually between 2% and 10% and the underestimation decreases as σ_b^2 increases. Therefore, our approximation for $MSE(\widehat{y}_i - y_i)$ based upon normal errors is robust with respect to non-normal sampling errors e_i.

BIBLIOGRAPHY

Arora, V., Lahiri, P., and Mukherjee, K. (1997). Empirical Bayes estimation of finite population means from complex survey. *Journal of American Statistical Association*, **92**, 1555-1562.

Battese, George E., Harter, Rachel M. and Fuller, Wayne A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of American Statistical Association*, **83**, 28-36.

Cressie, Noel (1992). REML estimation in empirical Bayes smoothing of census undercount. *Survey Methodology*, **18**, 75-94.

Datta, Gauri S. and Ghosh, Malay (1991). Bayesian prediction in linear models: applications to small area estimation. *Annals of Statistics*, **19**, 1748-1770.

DeSouza, Cynthia M. (1992). An approximate bivariate Bayesian method for analysing small frequencies. *Biometrics*, **48**, 1113-1130.

Dodd, Kevin W. (1999). Estimation of a distribution function from survey data. Unpublished Ph.D. dissertation, 53-55.

Ericksen, Eugene P., Kadane, Joseph B. and Tukey, John W. (1989). Adjusting the 1980 census of population and housing. *Journal of American Statistical Association*, **84**, 927-944.

Fay, R.E. and Herriot, R. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of American Statistical Association*, **74**, 269-277.

Isaki, C.T., Tsay J.H., and Fuller, W.A. (1999). Estimation of census adjustment factors. Unpublished manuscript.

Fuller, Wayne A. (1987). *Measurement Error Models*. New York: Wiley.

Fuller. Wayne A. (1996). *Introduction to Statistical Time Series (Second Edition).* New York: Wiley.

Fuller. Wayne A. (1997). Small area inference for binary variables in the National Health Interview Survey. *Journal of American Statistical Association.* **92**. 815-826.

Fuller. Wayne A.. Dodd. Kevin W. and Wang. Junyuan (2000). Estimation for the 1997 National Resources Inventory. *Techniqcal Report. Survey Section.*

Fuller. W. A. and Harter. R. M. (1987). The multivariate components of variance model for small area estimation. In *Small Area Statistics* (eds R. Platek. J. N. K. Rao. C. E. Särndal and M. P. Singh). 103-123. New York: Wiley.

Gelman. Andrew. Carlin. John B.. Stern. Hal S. and Rubin. Donald B. (1995). *Bayesian Data Analysis.* London: Chapman & Hall.

Ghosh. Malay and Lahiri. Parthasarathi (1987). Robust empirical Bayes estimation of means from stratified samples. *Journal of American Statistical Association*. **82**. 1153-1162.

Ghosh. M. and Rao. J. N. K. (1994). Small area estimation: an appraisal. *Statistical Science*. **9**. 55-93.

Ghosh. M. and Natarajan. K.. (1999). Small area estimation: a Bayesian perspective. In *Multivariate analysis. design of experiments and survey sampling.* (eds Ghosh. S.). 69-92. New York: Marcel Dekker.

Ghosh. Malay (1999). Some recent results on empirical Bayes methods for small area estimation. In *Small Area Estimation: Conference Proceedings of International Association of Survey Statisticians Satellite Conference*. 51-64. The Central Statistical Bureau of Latvia.

Henderson. C. R. (1950). Estimation of genetic parameters (abstract). *Annals of Mathematical Statistics*. **21**. 309-310.

Harville. D. A. (1976). Extension of the Gauss-Markov Theorem to include the estimation of random effects. *Annals of Statistics*. **4**. 384-395.

Harville. D. A. (1985). Decomposition of prediction error. *Journal of American Statistical Association*. **80**. 132-138.

Henderson, C. R., Kempthorne, O., Searle, S.R. and von Krosigk, C.M. (1959). The estimation of environmental and genetic trends from records subject to culling. *Biometrics*, **15**, 192.

Henderson, C. R. (1963). Selection index and expected genetic advance. In *Statistical Genetics and Plant Breeding*, (eds Hanson, W.D. and Robinson, H.F.), 141-163. National Academy of Sciences–National Research Council, Washington. Publication 982.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, **31**, 423-447.

Kackar, R. N. and Harville, D. A. (1984). Approximations for standard errors of estimators for fixed and random effects in mixed models. *Journal of American Statistical Association*, **79**, 853-862.

Lahiri, P. and Rao, J. N. K. (1995). Robust estimation of mean square error of small area estimators. *Journal of American Statistical Association*, **90**, 758-766.

Laird, N. M. and Louis, T. A. (1987). Empirical Bayes confidence intervals based on bootstrap samples. *Journal of American Statistical Association*, **82**, 739-750.

MacGibbon, Brenda and Tomberlin, Thomas J. (1989). Small area estimation of proportions via empirical Bayes techniques. *Survey Methodology*, **15**, 237-252.

Marker, David (1999). Orgnization of small area estimators using a generalized linear regression framework. *Journal of Official Statistics*, **15**, 1-24.

Marshall, Roger J. (1991). Mapping disease and mortality rates using empirical Bayes estimators. *Journal of Royal Statistical Society Series C*, **40**, 283-294.

Nusser, S. M. and Goebel, J. J. (1997). The National Resources Inventory: a long-term multi-resource monitoring programmme. *Environmental and Ecological Statistics*, **4**, 181-204.

Pfeffermann, Danny (1999). Small area estimation - big developments. In *Small Area Estimation: Conference Proceedings of International Association of Survey Statisticians Satellite Conference*, 51-64. The Central Statistical Bureau of Latvia.

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

Robinson, G. K. (1991). That BLUP is a good thing: the estimation of random effects. *Statistical Science*, **6**, 15-51.

Särndal, Carl-Eric and Hidiroglou, Michael A. (1989). Small domain estimation: a conditional analysis. *Journal of American Statistical Association*, **84**, 266-275.

Singh, A. C., Stukel, D. M. and Pfeffermann, D. (1998). Bayesian versus frequentist measures of error in small area estimation. *Journal of Royal Statistical Society Series B*, **60**, 377-396.

Snedecor, G. W. and Cochran, W. G. (1956). *Statistical Methods (Fifth Edition)*. Ames, Iowa: Iowa State Univeristy Press.

Wang, Junyuan, Fuller, Wayne A. and Opsomer, Jean (1999). Small area estimation in the National Resources Inventory. *American Statistical Association 1999 Proceedings of the Section on Survey Research Methods*, 650-654.

ACKNOWLEDGEMENTS

I take this opportunity to express my thanks to those who helped me with various aspects of this dissertation.

First and foremost, I thank Dr. Wayne A. Fuller for his guidance throughout this research, the writing of my Ph.D. dissertation, and my professional development. His insights in analyzing theoretical problems and real data have often inspired me. His words of encouragement renewed my hope of completing my graduate education. He also offered me many valuable professional opportunities over the past five years. I thank Dr. F. Jay Breidt for guidance in many areas of my research. I thank Dr. Hal Stern for conversations on the development and properties of the hierarchical Bayes model. I also thank my other committee members, Dr. Soumendra N. Lahiri and Dr. Ananda Weerasinghe, for their contributions to this work.

I thank my family for having faith in me and helping me through the hardest time of my life. They provided never-failing support and made life easier for me during the final months of my dissertation preparation. I express special appreciation to my dad, for teaching me to always pursue excellence; to my mother, for her love; and to my son, Franklin, for his smile.

To Yunfeng Li, Yongming Qu, Cong Chen, Zugeng Zheng and my other contemporaries: I express thanks for listening to my failures and my successes and for sharing theirs with me. We have shared much here in Ames. Thanks to Yongming Qu for helping me with C++ programming.

I would additionally like to thank Dr. Nusser for her guidance throughout the initial stages of my graduate career and giving me several professional development opportunities. I thank the NRI staff and survey graduate students for their cooperation. I also thank Dr. Krishna Athreya for his inspirational teaching style.