I would like to dedicate my thesis to my parents and my family. Without their constant support and encouragement, I would not have made it to where I am today. Secondly, I would like to dedicate this body of work to Malay Ghosh, my advisor. It has been an honor to work with someone who has dedicated his entire life to research, teaching and enriching the lives of students. It is my greatest wish to be able to mentor students and contribute to statistical research in the way he has. Finally, I dedicate my thesis to all of the faculty, students, staff, and friends that made this dissertation possible. I have been humbled by this truly amazing journey.
ACKNOWLEDGMENTS

I would like to thank my advisor, Malay Ghosh, for being a constant daily reminder of what kind of researcher and teacher I want to be when I graduate from Florida. I have been incredibly grateful for the opportunity to learn under him and see my development in research grow from his endless knowledge not just in small area estimation but in all areas of statistics. Finally, I greatly admire his work ethic, enthusiasm for learning, and humility that has in return shaped who I am not only as a researcher but also a person who wants to serve my community. I will never forget what Professor Ghosh has done for me the past five years at the University of Florida.

I would also like to thank Hani Doss for being one of the best professors I have had in my graduate career and for pushing me to excel and stick with my determination to get my PhD. Professor Doss’ honesty and opinion on decisions in my career have been an invaluable asset to me. I am extremely grateful to him for all the time, energy, and encouragement he has given me over the years at the University of Florida. Furthermore, I would like to Mike Daniels for his insight that has lead to my work improving.

Finally, I would like to thank my parents for constantly being loving, supportive and encouraging of my goals and dreams. I would not be the person I am today without them and I would not be as far along in my academic career without their constant encouragement of science, mathematics, and statistics since I was very young.

The research was partially supported by NSF Grant SES 1026165 and the United States Census Bureau Dissertation Fellowship Program. The views expressed here are those of the authors and do not reflect those of the United States Census Bureau.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>4</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>7</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>8</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>9</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>11</td>
</tr>
<tr>
<td>2 LITERATURE REVIEW</td>
<td>13</td>
</tr>
<tr>
<td>2.1 Small Area Estimation</td>
<td>13</td>
</tr>
<tr>
<td>2.2 Area-Level Models</td>
<td>13</td>
</tr>
<tr>
<td>2.3 Unit-Level Models</td>
<td>16</td>
</tr>
<tr>
<td>2.4 BLUP and EBLUP</td>
<td>17</td>
</tr>
<tr>
<td>2.5 MSE of the BLUP and EBLUP</td>
<td>19</td>
</tr>
<tr>
<td>2.6 Benchmarking</td>
<td>21</td>
</tr>
<tr>
<td>2.7 Constrained Bayes Estimation</td>
<td>26</td>
</tr>
<tr>
<td>2.8 Proposed Research</td>
<td>27</td>
</tr>
<tr>
<td>3 ONE-STAGE BENCHMARKING</td>
<td>29</td>
</tr>
<tr>
<td>3.1 Development of the Estimators</td>
<td>29</td>
</tr>
<tr>
<td>3.2 Relationship with Some Existing Estimators</td>
<td>33</td>
</tr>
<tr>
<td>3.3 Benchmarking with Both Mean and Variability Constraints</td>
<td>35</td>
</tr>
<tr>
<td>3.4 An Illustrative Example</td>
<td>38</td>
</tr>
<tr>
<td>4 TWO-STAGE BENCHMARKING</td>
<td>43</td>
</tr>
<tr>
<td>4.1 Two-Stage Benchmarking Results</td>
<td>43</td>
</tr>
<tr>
<td>4.2 An Example</td>
<td>49</td>
</tr>
<tr>
<td>5 ON ESTIMATION OF MSES OF BENCHMARKED EB ESTIMATORS</td>
<td>55</td>
</tr>
<tr>
<td>5.1 Benchmarked EB Estimators</td>
<td>56</td>
</tr>
<tr>
<td>5.2 Second-Order Approximation to MSE</td>
<td>57</td>
</tr>
<tr>
<td>5.3 Estimator of MSE Approximation</td>
<td>66</td>
</tr>
<tr>
<td>5.4 Parametric Bootstrap Estimator of the Benchmarked EB Estimator</td>
<td>68</td>
</tr>
<tr>
<td>5.5 An Application</td>
<td>69</td>
</tr>
<tr>
<td>6 SUMMARY REMARKS AND FUTURE WORK</td>
<td>77</td>
</tr>
<tr>
<td>APPENDIX: LEMMAS FOR CHAPTER 5</td>
<td>79</td>
</tr>
</tbody>
</table>
REFERENCES ......................................................... 83
BIOGRAPHICAL SKETCH ............................................ 86
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-1</td>
<td>Benchmarking Statistics for ASEC CPS</td>
<td>40</td>
</tr>
<tr>
<td>4-1</td>
<td>Table of estimates using Theorem 1</td>
<td>52</td>
</tr>
<tr>
<td>5-1</td>
<td>Table of estimates for 2000</td>
<td>73</td>
</tr>
<tr>
<td>5-2</td>
<td>Table of estimates for 1997</td>
<td>75</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>3-1</td>
<td>Change in Estimators due to Benchmarking for 1998</td>
<td>42</td>
</tr>
<tr>
<td>3-2</td>
<td>Change in Estimators due to Benchmarking for 2001</td>
<td>42</td>
</tr>
</tbody>
</table>
BA YES AND EMPIRICAL BA YES BENCHMARKING FOR SMALL AREA ESTIMATION

By

Rebecca C. Steorts

May 2012

Chair: Malay Ghosh
Major: Department of Statistics

Small area estimation has become increasingly popular due to growing demand for such statistics. In order to produce estimates of adequate precision for these small areas, it is often necessary to borrow strength from other related areas. The resulting model-based estimates may not aggregate to the more reliable direct estimates at the higher level, which may be politically problematic. Adjusting model-based estimates to correct this problem is known as benchmarking.

In this dissertation, we propose a general class of benchmarked Bayes estimators that can be expressed in the form of a Bayesian adjustment applicable to any estimator, linear or nonlinear. We also derive a second set of estimators under an additional constraint that benchmarks the weighted variability. We illustrate this work using U.S. Census Bureau data. Furthermore, we extend one-stage benchmarking to a two-stage procedure and illustrate this using data from the National Health Interview Survey. Finally, we obtain the benchmarked empirical Bayes (EB) estimator. Our goal is to see how much mean squared error (MSE) is lost due to due benchmarking. Furthermore, we find an asymptotically unbiased estimator of this MSE and compare it to the second-order approximation of the MSE of the EB estimator or equivalently of the MSE of the empirical best linear unbiased predictor (EBLUP), which was derived by Prasad and Rao (1990). Moreover, using methods similar to those of Butar and Lahiri (2003), we compute a parametric bootstrap estimator of the MSE of the benchmarked EB estimator under
the Fay-Herriot model and compare it to the MSE of the benchmarked EB estimator found by a second-order approximation. Finally, we illustrate our methods using SAIPE data from the U.S. Census Bureau.
CHAPTER 1
INTRODUCTION

Small area estimation has become increasingly popular over the past decade due to a growing need for more reliable small area statistics. Estimation based on direct estimators usually has large standard errors and coefficients of variation. In order to produce estimates for small areas, it is necessary to borrow strength from related areas to form indirect estimators that increase the effective sample size, and hence, increase the precision.

It is widely known that small area estimation needs explicit (or at least implicit) use of models that provide a link to related small areas through supplementary data such as census counts or other administrative data. Model-based estimates often differ widely from the direct estimates, especially for areas with small sample sizes. When aggregated, the model-based estimates are often very different from the overall estimates for the larger geographical areas based on the direct estimates, where the latter is believed to be quite reliable. Furthermore, an overall agreement with the direct estimates at an aggregate level may be sometimes politically necessary to convince legislators of the utility of small area estimators.

We address this problem by benchmarking, which requires the aggregated model-based estimates to match the corresponding direct estimate at a higher level. Model-based estimates do not benchmark to reliable direct survey estimates for large areas. Because of this behavior and to protect against possible model failure and overshrinkage, we benchmark the model-based estimates in such a way that they add up to the direct estimates for the larger area. We propose constrained Bayes procedures which meet the objective of benchmarking the weighted mean as well as the weighted variability. We extend this idea to two-stage benchmarking using a single model. Finally, we obtain the benchmarked empirical Bayes estimator under specific modeling assumptions and some mild regularity conditions. A subsequent goal is to see how much
mean squared error (MSE) is lost due to benchmarking. Furthermore, we find an asymptotically unbiased estimator of this MSE and compare it to the second-order approximation of the MSE of the EB estimator or equivalently of the MSE of the empirical best linear unbiased predictor (EBLUP), which was derived by Prasad and Rao (1990). Moreover, using methods similar to those of Butar and Lahiri (2003), we compute a parametric bootstrap estimator of the MSE of the benchmarked EB estimator under the Fay-Herriot model and compare it to the MSE of the benchmarked EB estimator found by a second-order approximation. Finally, we illustrate our methods using SAIPE data from the U.S. Census Bureau.

Before proceeding with our proposed methods, we first discuss the literature that greatly motivates our work.
CHAPTER 2
LITERATURE REVIEW

Before presenting our proposed methods on Bayes and Empirical Bayes benchmarking, we first introduce small area estimation and present the basic area- and unit-level models. We then provide some background on the best linear unbiased predictor (BLUP) and the empirical (BLUP) as well as the MSE of each of these quantities of interest. Finally, we explain the idea of benchmarking and provide a review of many of the previously proposed benchmarked estimators in the literature.

2.1 Small Area Estimation

Sample surveys are generally constructed so that they yield reliable direct estimates for large areas or domains. However, in small areas, direct estimators tend to lead to large standard errors and coefficients of variation due to small sample sizes in the domains. Thus, in finding estimates for small areas, it is necessary to borrow strength from neighboring or related areas, which increases the effective sample size and hence the precision of the estimates. It is well known now that the estimates are based on either explicit (or implicit) models that link the small area through data (such as census counts or administrative records).

2.2 Area-Level Models

You and Rao (2002) and Rao (2003) provided the best explanation for area-level models. Small area models are classified into either area-level or unit-level models. We first cover area-level models. A basic area-level model assumes that some function $\theta_i = g(\mu_i)$ of the small area mean $\mu_i$ is related to the area-specific auxiliary data $x_i = (x_{i1}, ..., x_{ip})^T$ through a linear model with area-specific random effect $u_i$,

$$\theta_i = g(\mu_i) = x_i^T \beta + u_i, \quad i = 1, \ldots, m. \quad (2.2.1)$$
Note that $\beta$ is a $p \times 1$ vector of known regression parameters, $u_i \sim iid (0, \sigma^2_u)$, and $m$ is the number of small areas. Normality of the random effects $u_i$ is often taken as a common assumption. Throughout our work, we take $g(\cdot)$ as the identity function.

In the basic area-level model, we assume the direct estimate $\hat{\theta}_i$ of the small area mean $\mu_i$ is available whenever the area sample size $n_i > 1$. The direct estimate $\hat{\theta}_i$ is usually design-unbiased meaning $E(\hat{\theta}_i) = \theta_i$. It is usual to assume

$$\hat{\theta}_i = \theta_i + e_i, \quad i = 1, \ldots, m \tag{2.2.2}$$

where $e_i$ denote the sampling errors associated with the transformed direct estimator $\hat{\theta}_i$. It is usual to assume that the $e_i$ are independent normal random variables with $E(e_i|\theta_i) = 0$ and $\text{Var}(e_i|\theta_i) = D_i$. When we combine the two models, they produce a linear mixed effects model. That is, we obtain the linear mixed model of Fay and Herriot (1979)

$$\hat{\theta}_i = x_i^T \beta + u_i + e_i, \quad i = 1, \ldots, m. \tag{2.2.3}$$

This model involves design-based random variables $e_i$ as well as model-based random variables $u_i$. Methods in empirical Bayes (EB) or hierarchical Bayes (HB) can employ the Fay-Herriot model to obtain improved model-based small area estimates of $\mu_i$ and $\theta_i$ (Ghosh and Rao (1994); and Rao (1999)).

We finally note that $E(e_i|\theta_i) = 0$ may not be a valid assumption if the sample size $n_i$ is too small and $\theta_i$ is a nonlinear function of $\mu_i$. We also mention that the sampling variance $D_i$ in the Fay-Herriot model is usually assumed to be known. This is needed in order to avoid any identifiability problems.
First, assuming normality and known $\sigma_u^2$, the basic area-level model can be written as

\[
\hat{\theta}_i|\theta_i \overset{ind}{\sim} N(\theta_i, D_i) \\
\theta_i|\beta, \sigma_u^2 \overset{ind}{\sim} N(x_i^T \beta, \sigma_u^2), \quad i = 1, \ldots, m \\
\pi(\beta) \propto 1.
\] (2.2.4)

If we assume that $\sigma_u^2$ is unknown we simply replace the last line in model (2.2.4) with

\[
\pi(\beta, \sigma_u^2) \propto \pi(\beta)\pi(\sigma_u^2). 
\] (2.2.5)

Furthermore, when $\sigma_u^2$ is known and $\pi(\beta) \propto 1$, the hierarchical (HB) and best linear unbiased predictor (BLUP) methods (under normality) lead to the same point estimates and measures of variability. This will be explained more in Section 2.4.

We now consider empirical Bayes (EB) estimation for area-level models. Assuming normality, the two-stage HB model is

\[
\hat{\theta}_i|\theta_i \overset{ind}{\sim} N(\theta_i, D_i), \quad i = 1, \ldots, m \\
\theta_i \overset{ind}{\sim} N(x_i^T \beta, \sigma_u^2).
\] (2.2.6)

The optimal estimate of $\theta_i$ is given by the Bayes estimate

\[
\hat{\theta}_i^B = E[\theta_i|\hat{\theta}_i, \beta, \sigma_u^2] = B_i\hat{\theta}_i + (1 - B_i)x_i^T \beta,
\]

where $B_i = \sigma_u^2(\sigma_u^2 + D_i)^{-1}$. This follows from

\[
\theta_i|\hat{\theta}_i, \beta, \sigma_u^2 \overset{ind}{\sim} N(\hat{\theta}_i^B, B_i D_i).
\]

The estimate $\hat{\theta}_i^B$ is the Bayes estimate under squared error loss and is optimal in the sense that its MSE is smaller than the MSE of any other estimator of $\theta_i$, linear or nonlinear. The Bayes estimate depends on $\beta$ and $\sigma_u^2$, which must be estimated from the
marginal distribution
\[ \hat{\theta}_i \overset{\text{ind}}{\sim} N(x_i^T \beta, \sigma_u^2 + D_i) \]

using maximum likelihood (MLE), restricted maximum likelihood (REML), or method of moments (MOM). We denote these estimators by \( \hat{\sigma}_u^2 \) and \( \hat{\beta} \). We obtain the EB estimator of \( \theta_i \) from \( \hat{\theta}_i^B \) by substituting \( \hat{\beta} \) for \( \beta \) and \( \hat{\sigma}_u^2 \) for \( \sigma_u^2 \). Then
\[
\hat{\theta}^{EB}_i = \hat{\theta}^B_i(\hat{\beta}, \hat{\sigma}_u^2) = \hat{\beta}_i \hat{\theta}_i + (1 - \hat{\beta}_i)x_i^T \hat{\beta}.
\]

We note that the EB estimator is identical to the empirical BLUP (EBLUP) estimator \( \hat{\theta}_i^H \), which will be explained in detail in Section 2.4.

### 2.3 Unit-Level Models

Next, we turn to the unit-level model as presented by Battese et al. (1988), Datta and Ghosh (1991), and Rao (2003). We assume unit-specific auxiliary data
\[
x_{ij} = (x_{i1}, \ldots, x_{ip})^T
\]
is available for each population element \( j \) in each small area \( i \). We assume \( \hat{\theta}_{ij} \) is related to \( x_{ij} \) through the one-fold nested error linear regression model
\[
\hat{\theta}_{ij} = x_{ij}^T \beta + u_i + e_{ij}; \ j = 1, \ldots, N_i; \ i = 1, \ldots, m.
\]

The area-specific effects \( u_i \overset{\text{ind}}{\sim} (0, \sigma_u^2) \). We assume \( e_{ij} = k_{ij} \tilde{e}_{ij} \), where \( k_{ij} \) are known constants and \( \tilde{e}_{ij} \overset{\text{ind}}{\sim} (0, \sigma_e^2) \) and are independent of the \( u_i \). Furthermore, normality of the \( u_i \) and \( e_{ij} \) is often assumed. We assume a sample \( s_i \) of size \( n_i \) is taken from the \( N_i \) units in each of the \( i \) areas and the sample units are assumed to follow the population model.

We now apply a HB approach to the basic unit-level model assuming \( \sigma_u^2 \) and \( \sigma_e^2 \) are both known. Then
\[
\hat{\theta}_{ij}|\beta, u_i, \sigma_e^2 \overset{\text{ind}}{\sim} N(x_{ij}^T \beta + u_i, k_{ij} \sigma_e^2), \ j = 1, \ldots, n_i; \ i = 1, \ldots, m
\]
\[
u_i|\sigma_u^2 \overset{\text{ind}}{\sim} N(0, \sigma_u^2)
\]
\[\pi(\beta) \propto 1.\]
When $\sigma_u^2$ and $\sigma_e^2$ are both unknown, we replace $\pi(\beta) \propto 1$ with $\pi(\beta, \sigma_u^2, \sigma_e^2) \propto \pi(\sigma_u^2)\pi(\sigma_e^2)$ in the HB model (2.3.2).

2.4 BLUP and EBLUP

We first present the basic theory for a general linear mixed model. We suppose that the sample data $\theta$ (an $m \times 1$ vector of sample observations) obeys the general linear mixed model

$$\theta = X\beta + Zu + e. \quad (2.4.1)$$

Here $X$ and $Z$ are known $m \times p$ and $m \times h$ matrices and $\beta$ is a $p \times 1$ vector of regression parameters. Furthermore, $u, e$ are independently distributed with means 0 and covariances $G, R$ depending on variance parameters $\delta = (\delta_1, \ldots, \delta_q)^T$. We assume $\delta$ belongs to a specified subset of Euclidean $q$-space such that $V(\hat{\beta}) = V = V(\delta) = R + ZGZ^T$ is nonsingular for all $\delta$ in the subset.

We are interested in estimating a linear combination

$$\mu = l^T\beta + m^Tu$$

where $l, m$ are vectors of constants. A linear estimator of $\mu$ is $\hat{\mu} = a^T\hat{\theta} + b$ for known $a, b$. The estimator $\hat{\mu}$ is model-unbiased for $\mu$ meaning that $E_M[\hat{\mu}] = E_M[\mu]$, where $E_M$ represents expectation with respect to model (2.4.1). The MSE of $\hat{\mu}$ is $\text{MSE}[\hat{\mu}] = E[(\hat{\mu} - \mu)^2]$. We are interested in finding the best linear unbiased predictor (BLUP) estimator which minimizes the MSE in the class of linear unbiased estimators $\hat{\mu}$.

For known $\delta$, the BLUP estimator of $\mu$ is given by

$$\hat{\mu}^{\text{BLUP}} = t(\delta, \hat{\theta}) = l^T\tilde{\beta} + m^T\tilde{u} = l^T\tilde{\beta} + m^TGZ^TV^{-1}(\hat{\theta} - X\tilde{\beta}), \quad (2.4.2)$$

where

$$\tilde{\beta} = (X^TV^{-1}X)^{-1}X^TV^{-1}\hat{\theta} \quad (2.4.3)$$
is the BLUP of $\beta$. Moreover,

$$\hat{u} = GZ^T V^{-1}(\hat{\theta} - X\tilde{\beta})$$

(2.4.4)

is the BLUP of $u$. This theory was proposed by Henderson (1950) and appeared in Robinson (1991), Rao (2003), and others.

The BLUP estimator $\hat{\mu}_i^H = t(\hat{\delta}, \hat{\theta})$ given in (2.4.2) depends on the variance parameters $\delta$, which are typically unknown in applications. If we replace $\delta$ by $\hat{\delta}$, we obtain a two-stage estimator $\hat{\mu}_i^H = t(\hat{\delta}, \hat{\theta})$, which we call the empirical BLUP estimator (EBLUP).

We now present the BLUP and EBLUP estimators for the Fay-Herriot model given in (2.2.4). We define $\tilde{\theta}_i$ and $\hat{\theta}_i$ as the BLUP and EBLUP estimators of $\theta_i$ respectively. We assume that $D_i$ and $\sigma_u^2$ are known. Model (2.2.3) is a special case of (2.4.1). Using the general form of the model, we find that $Z = I$, $G = \sigma_u^2 I$, $R = \text{Diag}(D_1, \ldots, D_m)$, $V = \text{Diag}(\sigma_u^2 + D_1, \ldots, \sigma_u^2 + D_m)$. Also, $\mu_i = \theta_i = x_i^T \beta + u_i$ so that $l^T = X$ and $m^T = I$. Making the appropriate substitutions into (2.4.2), we find that

$$\tilde{\theta}_i = B_i \hat{\theta}_i + (1 - B_i) x_i^T \tilde{\beta},$$

(2.4.5)

where $B_i = \sigma_u^2 (\sigma_u^2 + D_i)^{-1}$ and

$$\tilde{\beta} = \left[ \sum_i x_i x_i^T (\sigma_u^2 + D_i)^{-1} \right]^{-1} \left[ \sum_i x_i \hat{\theta}_i (\sigma_u^2 + D_i)^{-1} \right].$$

(2.4.6)

It is clear from (2.4.5) that the BLUP estimator $\tilde{\theta}_i^H$ can be expressed as a weighted average of the direct estimator $\hat{\theta}_i$ and the regression-synthetic estimator $x_i^T \beta$, where the weight $B_i$ ($0 \leq B_i \leq 1$) measures the uncertainty in modeling the $\theta_i$, i.e., the model variance $\sigma_u^2$ relative to the total variance $\sigma_u^2 + D_i$.

The BLUP estimator depends on $\sigma_u^2$, which is unknown in practical applications. Replacing $\sigma_u^2$ by an estimate $\hat{\sigma}_u^2$ results in an empirical BLUP (EBLUP) estimator $\tilde{\theta}_i^H$,
where
\[
\hat{\theta}_i^H = B_i \hat{\theta} + (1 - B_i) x_i^T \hat{\beta}.
\] (2.4.7)

Also, \(\hat{B}_i\) and \(\hat{\beta}\) are the estimated values of \(B_i\) and \(\tilde{\beta}\) when \(\sigma_u^2\) is replaced by \(\hat{\sigma}_u^2\).

Finally, when \(\sigma_u^2\) is known model (2.2.4) is assumed, straightforward calculations show that
\[
\theta_i | \hat{\theta}, \beta, \sigma_u^2 \sim N(B_i \hat{\theta} + (1 - B_i) x_i^T \beta, B_i D_i).
\]

It can be shown that
\[
\beta | \hat{\theta} \sim N(\tilde{\beta}, (X^T V^{-1} X)^{-1}).
\]

Using these two facts and the formulas for iterated expectation and variance, we can show that
\[
\theta_i | \hat{\theta} \sim N(\tilde{\theta}_i^H, \text{MSE}(\tilde{\theta}_i^H)),
\] (2.4.8)

where \(\tilde{\theta}_i^H = B_i \hat{\theta} + (1 - B_i) x_i^T \tilde{\beta}\) and \(\text{MSE}(\tilde{\theta}_i^H) = B_i D_i + (1 - B_i)^2 x_i^T (X^T V^{-1} X)^{-1} x_i\).

This implies that when \(\sigma_u^2\) is known and when taking a flat prior for \(\beta\), the HB and BLUP approaches under normality lead to the same point estimates and measures of variability. Similarly, when \(\sigma_u^2\) is known and when taking a flat prior for \(\beta\), the EB and EBLUP approaches under normality lead to the same point estimates and measures of variability.

### 2.5 MSE of the BLUP and EBLUP

Rao (2003) presented the MSE of the BLUP \(\tilde{\theta}_i^H\) assuming the basic area-level model (2.2.3). A more general version of the MSE can be found in Section 6.3 of Rao (2003) for the general mixed linear model. However, for the area-level model, the MSE of \(\tilde{\theta}_i^H\) is given by
\[
\text{MSE}(\tilde{\theta}_i^H) = E(\tilde{\theta}_i^H - \theta_i)^2 = g_{1i}(\sigma_u^2) + g_{2i}(\sigma_u^2).
\] (2.5.1)
where \( g_1(\sigma^2_u) = B_i D_i \) and

\[
g_2(\sigma^2_u) = (1 - B_i)^2 x_i^T (X^T X)^{-1} x_i.
\]

The first term \( g_1(\sigma^2_u) \) is of order \( O(1) \), and the second term \( g_2(\sigma^2_u) \) is of order \( O(m^{-1}) \) for large \( m \) assuming the following regularity conditions:

(i) \( D_i \) and \( \sigma^2_u \) are uniformly bounded

(ii) \( \sup_{1 \leq i \leq m} x_i^T (X^T X)^{-1} x_i = O(m^{-1}) \).

Recall the BLUP estimator \( \tilde{\theta}_i^H \) depends on variance component \( \sigma^2_u \) which is unknown in practical applications. We can replace \( \sigma^2_u \) with an estimator \( \hat{\sigma}^2_u \), hence obtaining an EBLUP estimator \( \hat{\theta}_i^H \) as defined in (2.4.7).

We now turn to the work of Prasad and Rao (1990) and focus on model (2.2.4). We note that the normality assumption is not needed to derive the two-stage estimator \( \hat{\theta}_i^H \). We have already mentioned that \( \tilde{\theta}_i^H = t_i(\sigma^2_u, \hat{\theta}) \) depends on variance component \( \sigma^2_u \), which is generally unknown. We usually estimate \( t_i(\sigma^2_u, \hat{\theta}) \) by replacing \( \sigma^2_u \) with an asymptotically consistent estimator. Prasad and Rao (1990) presented an unbiased quadratic estimator of \( \sigma^2_u \) in the model (2.2.4)

\[
\tilde{\sigma}^2_u = (m - p)^{-1} \left[ \sum_{i=1}^m \hat{u}_i - \sum_{i=1}^m D_i (1 - x_i^T (X^T X)^{-1} x_i) \right],
\]

(2.5.2)

where \( \hat{u}_i = \hat{\theta}_i - x_i^T \hat{\beta} \) and we now take \( \hat{\beta} = (X^T X)^{-1} X^T \hat{\theta} \). The two-stage estimator \( t_i(\hat{\sigma}^2_u, \hat{\theta}_i) \) is an empirical Bayes estimator of \( \theta_i \) under normality (Fay and Herriot (1979)). We add that it is possible for \( \hat{\sigma}^2_u \) to take positive or negative values. Hence, we instead define a new estimator \( \hat{\sigma}^2_u = \max\{\tilde{\sigma}^2_u, 0\} \). This ensures that the two-stage estimator \( \hat{\theta}_i^H \) will have finite expectation. It may also be noted that \( P(\tilde{\sigma}^2_u \leq 0) \) tends to zero as \( m \to \infty \).

Prasad and Rao (1990) obtained estimators of the MSE approximation under normality. Their Appendix (Theorem A.2 and Theorem A.3) shows the expectation of the...
MSE estimator is correct to $O(m^{-1})$. They show

\[
\text{MSE}[t_i(\hat{\sigma}^2_u, \hat{\theta}_i)] = g_1(\sigma^2_u) + g_2(\sigma^2_u) + g_3(\sigma^2_u) + o(m^{-1})
\] (2.5.3)

where

\[
g_1(\sigma^2_u) = \sigma^2_u D_i(\sigma^2_u + D_i)^{-1}, \quad g_2(\sigma^2_u) = D_i^2(\sigma^2_u + D_i)^{-2} x_i^T (X^T V^{-2} X)^{-1} x_i,
\]

and

\[
g_3(\sigma^2_u) = D_i^2(\sigma^2_u + D_i)^{-3} V(\hat{\sigma}^2_u), \quad V(\hat{\sigma}^2_u) = 2m^{-1}[\sigma^4_u + 2m^{-1} \sigma^2_u \sum_i D_i + m^{-1} \sum_i D_i^2] + o(m^{-1}).
\]

This is estimated by

\[
\widehat{\text{MSE}}[t_i(\hat{\sigma}^2_u, \hat{\theta}_i)] = g_1(\hat{\sigma}^2_u) + g_2(\hat{\sigma}^2_u) + 2g_3(\hat{\sigma}^2_u)
\] (2.5.4)

since

\[
E[g_1(\hat{\sigma}^2_u)] = g_1(\sigma^2_u) - g_3(\sigma^2_u) + o(m^{-1})
\]

\[
E[g_2(\hat{\sigma}^2_u)] = g_2(\sigma^2_u) + o(m^{-1})
\]

\[
E[g_3(\hat{\sigma}^2_u)] = g_3(\sigma^2_u) + o(m^{-1}).
\]

### 2.6 Benchmarking

As we have mentioned already, small area estimation needs explicit or implicit use of models. Such model-based estimates may differ widely from direct estimates, especially in areas where the sample size is particularly low. Problems occur when we aggregate model-based estimates since they often do not agree with the overall direct estimate for a larger geographical area. Furthermore, an overall agreement with the direct estimators at some higher level may sometimes be politically necessary to convince legislators of the utility of small area estimates. This problem can be more severe in the event of model failure as often there is no check for validity of the assumed model.

One way to avoid the problem just described is by benchmarking. This entails adjusting the model-based estimates such that the aggregate estimates for the small areas match that for the larger geographical area. A simple illustration of this is to modify
the model-based state-level estimates so that the aggregate matches the national estimates. The most popular method utilized is the raking or ratio adjustment method, which amounts to multiplying the model-based estimates by a constant data-dependent factor so the weighted total agrees with the direct estimate.

For example, we mention the Small Area Income and Poverty Estimates Program (SAIPE) estimates of the U.S. Census Bureau based on the American Community Survey (ACS) data. Here, estimates are controlled so that the overall weighted estimates agree with the corresponding state estimates. Even though these estimates are model-based, they are quite close to the direct estimates.

We now discuss some of the existing benchmarking literature (mostly frequentist) for small area estimation. Battese et al. (1988) considered prediction of areas under corn and soybeans for 12 counties in Iowa based on the 1978 June Enumerative Survey and LANDSAT satellite data from the United States Department of Agriculture (USDA). Predicting crop areas for small domains such as counties had not been attempted at this time due to a lack of available data from farm surveys for these domains (Battese et al. (1988)). A survey regression predictor was found for the 12 counties’ mean crop areas which had a small variance. However, Battese et al. (1988) pointed out that it is desirable to modify the individual county predictors such that the weighted sum equals the unbiased survey regression predictor for the total area.

You et al. (2004) considered the area-level model of Fay and Herriot (1979) in (2.2.3). They suggested benchmarking the model-based HB estimates in such a way that they will add up to the direct estimates for the large areas. We refer to the HB estimator of \( \theta_i \) by \( \hat{\theta}_i^B = E(\theta_i|\hat{\theta}_i) \). Denote a generic small area predictor by \( \hat{\theta}_i^P \). A small area predictor satisfies the benchmarked property if

\[
\sum_{i=1}^{m} w_i \hat{\theta}_i^P = \sum_{i=1}^{m} w_i \hat{\theta}_i \quad (2.6.1)
\]

where \( w_i \) are normalized sampling weights such that \( \sum_{i=1}^{m} w_i \hat{\theta}_i \) is design consistent.
For instance, the ratio or raked estimator (R) can be obtained as

$$\hat{\theta}_i^R = \hat{\theta}_i^B \frac{\sum_{j=1}^{m} \hat{\theta}_j}{\sum_{j=1}^{m} \hat{\theta}_j^B}. \quad (2.6.2)$$

In many situations survey estimators of the larger geographical area are thought to have adequate precision. However, it may be desirable to use the design consistent estimator of the larger geographical area and require that the weighted sum of the small area estimators equal the design consistent estimator (Wang et al. (2008)). To explain design consistency, we first must define design-unbiased estimators.

Suppose we have a population total $\theta$ and we observe the $\hat{\theta}$ values associated with a sample $s$. In a design based approach, an estimator $\hat{\theta}^E$ of $\theta$ is said to be design-unbiased if $E_p(\hat{\theta}^E) = \sum_s p(s)\hat{\theta}_s = \theta$. The design variance of $\hat{\theta}^E$ is denoted by $V_p(\hat{\theta}^E) = \left(\hat{\theta}^E - E_p(\hat{\theta}^E)\right)^2$. We say that $\hat{\theta}^E$ is design consistent if it is design-unbiased (or its design bias tends to zero as the sample size increases) and if $V_p(\hat{\theta}^E)$ tends to zero as the sample size increases (Rao (2003)).

Wang et al. (2008) considered model (2.2.3). Assuming the variance components $\sigma_u^2$ and $D_i$ are known, the best linear unbiased predictor (estimator) (BLUP) of $\beta$ is

$$\bar{\beta} = (X^TV^{-1}X)^{-1}X^TV^{-1}\hat{\theta}$$

$$= \left[\sum_i (\sigma_u^2 + D_i)^{-1}x_ix_i^T\right]^{-1}\left[\sum_i (\sigma_u^2 + D_i)^{-1}x_i\hat{\theta}_i\right]$$

where $X^T = (x_1, \ldots, x_m)$, $\hat{\theta}^T = (\hat{\theta}_1, \ldots, \hat{\theta}_m)$, and $V = \text{Var}(\hat{\theta}) = \text{Diag}(\sigma_u^2 + D_1, \ldots, \sigma_u^2 + D_m)$. Recall the BLUP of $\theta_i$ is

$$\hat{\theta}_i^H = B_i\hat{\theta}_i + (1 - B_i)x_i^T\bar{\beta}, \quad (2.6.3)$$

where

$$B_i = \sigma_u^2(\sigma_u^2 + D_i)^{-1}. \quad (2.6.4)$$
We now review some benchmarking procedures in the literature. To do so, let \( \tilde{\theta}^H \) denote the BLUP predictor of \( \theta \) defined in (2.6.3), where \( \tilde{\theta}^H = X\hat{\beta} + \hat{u} \).

Note that \( \hat{\beta} \) and \( \hat{u} \) are any solutions to

\[
\begin{bmatrix}
X^T\Sigma_d^{-1}X & X^T\Sigma_d^{-1} \\
\Sigma_d^{-1}X & \Sigma_d^{-1} + \Sigma_u^{-1}
\end{bmatrix}
\begin{bmatrix}
\beta \\
u
\end{bmatrix}
= \begin{bmatrix}
X^T\Sigma_d^{-1}\hat{\theta} \\
\Sigma_d^{-1}\hat{\theta}
\end{bmatrix}
\]

(2.6.5)

where \( \Sigma_u = \sigma_u^2 I_m \) and \( \Sigma_d = \text{Diag}(D_1, \ldots, D_m) \). Moreover, finding a solution to the mixed model equation in (2.6.5) is equivalent to finding a solution to

\[
\min_{\beta, u} \left\{ [\hat{\theta} - X\beta - u]^T\Sigma_d^{-1}[\hat{\theta} - X\beta - u] + u^T\Sigma_u^{-1}u \right\}.
\]

(2.6.6)

Pfeffermann and Barnard (1991) proposed a modified predictor \( \hat{\theta}^{PB} = X\hat{\beta}^{PB} + \hat{u}^{PB} \), where \( \hat{\beta}^{PB} \) and \( \hat{u}^{PB} \) are the solutions to the minimization problem in (2.6.6) subject to the constraint

\[
\sum_i w_i (x_i^T \hat{\beta}^{PB} + \hat{u}^{PB}) = \sum_i w_i \hat{\theta}_i.
\]

(2.6.7)

This leads to the estimator

\[
\hat{\theta}_i^{PB} = \hat{\theta}_i^H + [\text{Var}(\hat{\theta})]^{-1}\text{Cov}(\hat{\theta}_i^H, \tilde{\theta}) \left[ \sum_{j=1}^m w_j \hat{\theta}_j - \tilde{\theta} \right]
\]

(2.6.8)

where \( \tilde{\theta} = \sum_{i=1}^m w_i \hat{\theta}_i^H \), \( \text{Cov}(\hat{\theta}_i^H, \tilde{\theta}) = w_i B_i D_i + \sum_{j=1}^m w_j (1 - B_i) (1 - B_j) x_i^T V(\hat{\beta}) x_j \), and \( \text{Var}(\tilde{\theta}) = \sum_{i=1}^m w_i \text{Cov}(\hat{\theta}_i^H, \tilde{\theta}) \).

Furthermore, Isaki et al. (2000) imposed the restriction by a procedure that basically constructs the best predictors of \( n - 1 \) quantities that are uncorrelated with \( \sum_i w_i \hat{\theta}_i \). After matrix computations, the Isaki-Tsay-Fuller (ITF) estimator can be written as

\[
\hat{\theta}_i^{ITF} = \hat{\theta}_i^H + \left[ \sum_{j=1}^m w_j^2 \text{Var}(\hat{\theta}_j) \right]^{-1} w_i \text{Var}(\hat{\theta}_j) \left( \sum_{j=1}^m w_j \hat{\theta}_j - \sum_{j=1}^m w_j \hat{\theta}_j^H \right)
\]

(2.6.9)

where \( \text{Var}(\hat{\theta}_i) \) estimates \( \sigma_u^2 + D_i \).
The Pfeffermann-Barnard (PB) (2.6.8) and ITF (2.6.9) estimators can be written in the form
\[
\hat{\theta}_i^a = \hat{\theta}_i + a_i \left( \sum_{j=1}^{m} w_j \hat{\theta}_j - \sum_{j=1}^{m} w_j \hat{\theta}_j^P \right) \tag{2.6.10}
\]
where \( \sum_i w_i a_i = 1 \). Wang et al. (2008) described the restriction in (2.6.1) to be just an adjustment issue. That is, in order to force \( \hat{\theta}_i^a \) to satisfy (2.6.1), \( \sum_{j=1}^{m} w_j \hat{\theta}_j - \sum_{j=1}^{m} w_j \hat{\theta}_j^P \) must be allocated to the small area estimator \( \hat{\theta}_j^P \) using \( a_i \).

You and Rao (2002) proposed an estimator of \( \beta \) such that the resulting estimators or predictors satisfy constraint (2.6.1) under a unit-level model. You and Rao called such estimators self-calibrated. Wang, Fuller, and Qu (2008) applied the methods of You and Rao (2002) to the area-level model to find
\[
\hat{\theta}_{YR}^i = \hat{B}_i \hat{\theta}_i + (1 - \hat{B}_i) x_i^T \hat{\beta}_{YR} \tag{2.6.11}
\]
where \( \hat{\beta}_{YR} = \left[ \sum_i w_i (1 - \hat{B}_i) x_i x_i^T \right]^{-1} \sum_i w_i (1 - \hat{B}_i) x_i \hat{\theta}_i \). An estimator \( \hat{\theta}_j^P \) is said to be self-calibrated if \( \sum_{j=1}^{m} w_j \hat{\theta}_j - \sum_{j=1}^{m} w_j \hat{\theta}_j^P = 0 \). For example, the You and Rao estimator (YR) (2.6.11) is self-calibrated. This implies that the YR estimator is of the form (2.6.10), where \( \hat{\theta}_i^a = \hat{\theta}_i \).

Wang et al. (2008) unified the previous work done for a weighted quadratic loss and found the BLUP for \( \theta \) that satisfies constraint (2.6.1). They defined the loss function
\[
Q(\hat{\theta}_i^a) = \sum_{i=1}^{m} \phi_i E(\hat{\theta}_i^a - \theta_i)^2 \tag{2.6.12}
\]
where \( \phi_i > 0 \) are a known set of weights. Using the random effects model defined in (2.2.3), define \( \hat{\theta}_i^a \) to be the unique predictor among all linear unbiased predictors satisfying (2.6.1) that minimizes (2.6.12). Then
\[
\hat{\theta}_i^a = \hat{\theta}_i^U + a_i \left( \sum_{j=1}^{m} w_j \hat{\theta}_j - \sum_{j=1}^{m} w_j \hat{\theta}_j^P \right) \tag{2.6.13}
\]
where \( \tilde{a}_i = (\sum_j \phi_j^{-1} w_j^2)^{-1} \phi_j^{-1} w_i \). It is noted that when the variance components are unknown, the variance components are replaced with reasonable estimators to obtain the EBLUP \( \hat{\theta}_j^H \). Then the modified estimator or predictor of Wang, Fuller, and Qu becomes

\[
\hat{\theta}_i^a = \hat{\theta}_i^H + \tilde{a}_i \left( \sum_{j=1}^m w_j \hat{\theta}_j - \sum_{j=1}^m w_j \hat{\theta}_j^H \right).
\]  

(2.6.14)

Recall the loss function (2.6.12), where the choice of the weights \( \phi_i \) depends on the problem at hand. For instance, more weight may be given to more important areas. It is often the case that \( \phi_i \) is a function of the variance components, or we can choose \( \phi_i \) in order to obtain desirable properties. For example, choosing \( \phi_i = [\text{Var}(\hat{\theta}_i)]^{-1} \) yields the ITF estimators. On the other hand, choosing \( \phi_i = w_i [\text{Cov}(\tilde{\theta}_i, \hat{\theta}_i)]^{-1} \), where \( \tilde{\theta} = \sum_j w_j \tilde{\theta}_j^H \), leads to the Pfeffermann and Barnard (1991) estimator in (2.6.8). Moreover, \( \phi_i = [\text{Var}(\hat{\theta}_i^H)]^{-1} \) leads to the estimator by Battese et al. (1988).

### 2.7 Constrained Bayes Estimation

The work of Louis (1984) and Ghosh (1992) motivates our work on one-stage benchmarking. Under any quadratic loss the Bayes estimates are usually the posterior means of the parameters of interest, however, the point of analysis with many parameters is usually to obtain another objective. This objective may be to have the parameters be above or below a cutoff point or to have the first two moments of the sample match the true moments. Ghosh (1992) showed

\[
E \left[ \sum_{i=1}^m (\theta_i - \bar{\theta})^2 \bigg| x \right] > \sum_{i=1}^m (\hat{\theta}_i^B(x) - \hat{\theta}_i^B(x))^2.
\]  

(2.7.1)

where \( \hat{\theta}_i^B(x) = E(\theta|x) \). This means that the Bayes estimates are too close together compared to the true parameter values. Louis (1984) commented that this behavior was due to overshrinking of the observed data towards the prior means.

These papers illustrated that the empirical histogram of the posterior means of a set of parameters of interest is underdispersed as compared to the posterior histogram of

26
the same set of parameters. Hence, adjustment of Bayes estimators is needed in order to meet the twin objective of accuracy and closeness of the histogram of the estimates with the posterior estimate of the parameter histogram. In other words, Louis (1984) and Ghosh (1992) wanted to find estimators \( e_i \) of \( \theta_i \) which satisfy

\[
E\left[ \sum_{i=1}^{m} \theta_i | x \right] = \sum_{i=1}^{m} e_i \quad \text{and} \quad E\left[ \sum_{i=1}^{m} (\theta_i - \bar{\theta})^2 | x \right] = \sum_{i=1}^{m} (e_i - \bar{e})^2.
\]  

(2.7.2)  

(2.7.3)

Ghosh (1992) provided a general class of constrained Bayes estimators in a general context that are obtained by matching the first two moments of the histogram of the estimates to the posterior expectations of the first two moments of the parameters. The quadratic loss is minimized subject to those conditions.

Thus, taking the motivating work of Louis (1984) and Ghosh (1992), we propose a general loss function without any distributional assumptions and benchmark the weighted mean as well as the weighted mean and weighted variability. We will soon see that many of the benchmarked estimators already mentioned are special cases of this general theory we present for an area-level model.

2.8 Proposed Research

Our work extends that of Wang et al. (2008) to a more general setting. Wang et al. (2008) derived a best linear unbiased estimator (BLUP) subject to the standard benchmarking constraint while minimizing a weighted squared error loss. They restricted their attention to a simple random effects model, only considered linear estimators of small area means, and did not consider multivariate settings. In contrast, in our work (Datta et al. (2011)), we develop more general benchmarked Bayes estimators, where the form of the Bayes estimate can be linear or nonlinear. We also derive a second set of estimators under a second constraint that benchmarks the weighted variability. We can extend our results to multivariate settings and use a more general loss
function. Both Wang et al. (2008) and Datta et al. (2011) are able to show that many previously proposed estimators are special cases of their respective estimator. Then in Ghosh and Steorts (2011), we extend the work above to a two-stage benchmarking procedure under one single model. Finally, since model-based estimates borrow strength, they usually show an improvement over the direct estimates in terms of mean squared error (MSE). In Steorts and Ghosh (2012), we determine how much of this advantage is lost due to benchmarking under specific modeling assumptions. This question was posed through simulation studies using an augmented model by Wang et al. (2008), however they did not derive any probabilistic results. They showed that the MSE of the benchmarked EB estimator was slightly larger than the MSE of the EB estimator for their simulation studies. In our work, we derive a second-order approximation of the MSE of the benchmarked EB estimator and then find an asymptotically unbiased estimate of this MSE. Finally, we derive a parametric bootstrap estimator of the MSE of the benchmarked EB estimator and analyze our methodology using data from the Small Area and Income and Poverty Estimation data from the U.S. Census Bureau.
CHAPTER 3
ONE-STAGE BENCHMARKING

As we have already mentioned, model-based estimates can differ widely from direct estimates, especially for areas with very low sample sizes. While model-based estimates are useful, one potential difficulty with such estimates is that when aggregated, the overall estimate for a larger geographical area may be quite different from the corresponding direct estimate. One way to avoid this problem is by benchmarking, which amounts to modifying these model-based estimates so that we get the same aggregate estimate for the larger geographical area. Currently the most popular approach is the so-called “raking” or ratio adjustment method, which involves multiplying all the small area estimates by a constant factor so that the weighted total agrees with the direct estimate. The raking approach is ad hoc, although, we give it a constrained Bayes interpretation.

Our objective is to develop a general class of Bayes estimators which achieves the necessary benchmarking. For definiteness, we will concentrate only on area-level models. As we will see later, many of the currently proposed benchmarked estimators including the raked ones belong to the proposed class of Bayes estimators. In particular, some of the estimators proposed in Pfeffermann and Barnard (1991), Isaki et al. (2000), Wang et al. (2008) and You et al. (2004) are members of this class.

3.1 Development of the Estimators

Let \( \hat{\theta}_1, \ldots, \hat{\theta}_m \) denote the direct estimators of the \( m \) small area means \( \theta_1, \ldots, \theta_m \). We write \( \hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_m)^T \) and \( \theta = (\theta_1, \ldots, \theta_m)^T \). Initially, we seek the benchmarked Bayes estimator \( \hat{\theta}^{BM1} = (\hat{\theta}^{BM1}_1, \ldots, \hat{\theta}^{BM1}_m)^T \) of \( \theta \) such that \( \sum_{i=1}^m w_i \hat{\theta}^{BM1}_i = t \), where either \( t \) is prespecified from some other source or \( t = \sum_{i=1}^m w_i \hat{\theta}_i \). The \( w_i \) are given weights attached to the direct estimator \( \hat{\theta}_i \), and without any loss of generality, \( \sum_{i=1}^m w_i = 1 \). These weights may depend on \( \hat{\theta} \) (which is most often not the case), but do not depend on \( \theta \). For example, we may take \( w_i = N_i / \sum_{j=1}^m N_j \), where the \( N_i \) are the population sizes for the \( m \) small areas.
A Bayesian approach to this end is to minimize the posterior expectation of the weighted squared error loss \( \sum_{i=1}^{m} \phi_i E[(\theta_i - e_i)^2|\hat{\theta}] \) with respect to the \( e_i \) satisfying \( \bar{e}_w = \sum_{i=1}^{m} w_ie_i = t \). The \( \phi_i \) may be the same as the \( w_i \), but that need not always be the case. Also, like \( w_i \), \( \phi_i \) may depend on \( \hat{\theta} \), but not on \( \theta \). Wang et al. (2008) considered the same loss but restricted themselves to a simple random effects model and consider only linear estimators of small area means. Indeed, in that special case, the Bayesian estimator proposed reduces to that of the former. It should be noted though that the Bayesian adjustment proposed is applicable to any Bayes estimator, linear or nonlinear.

The \( \phi_i \) can be regarded as weights for a multiple-objective decision process. That is, each specific weight is relevant only to the decision-maker for the corresponding small area, who may not be concerned with the weights related to decision-makers in other small areas. Combining losses in such situations in a linear fashion is discussed for example in Berger (1985).

We now prove a theorem which provides a solution to our problem. A few notations are needed before stating the theorem. Let \( \hat{\theta}_i^B \) denote the posterior mean of \( \theta_i \), \( i = 1, \ldots, m \), under a certain prior. The vector of posterior means and the corresponding weighted average are denoted respectively by \( \hat{\theta}_B^i = (\hat{\theta}_B^1, \ldots, \hat{\theta}_B^m)^T \) and \( \bar{\hat{\theta}}_B^w = \sum_{i=1}^{m} w_i \hat{\theta}_i^B \). Also, let \( r = (r_1, \ldots, r_m)^T \), where \( r_i = w_i/\phi_i \), \( i = 1, \ldots, m \), and \( s = \sum_{i=1}^{m} w_i^2/\phi_i \). Then we have the following theorem.

**Theorem 1.** The minimizer \( \hat{\theta}^{BM1} \) of \( \sum_{i=1}^{m} \phi_i E[(e_i - \theta_i)^2|\hat{\theta}] \) subject to \( \bar{e}_w = t \) is given by

\[
\hat{\theta}^{BM1} = \hat{\theta}^B + s^{-1}(t - \bar{\hat{\theta}}_B^w)r.
\]

**Proof.** First rewrite \( \sum_{i=1}^{m} \phi_i E[(e_i - \theta_i)^2|\hat{\theta}] = \sum_{i=1}^{m} \phi_i V(\theta_i|\hat{\theta}) + \sum_{i=1}^{m} \phi_i (e_i - \hat{\theta}_i^B)^2 \). Now the problem reduces to minimization of \( \sum_{i=1}^{m} \phi_i (e_i - \hat{\theta}_i^B)^2 \) subject to \( \bar{e}_w = t \). A Lagrangian multiplier approach provides the solution. But then we need to show in addition that the
solution provides a minimizer and not a maximizer. Alternately, we can use the identity
\[ \sum_{i=1}^{m} \phi_i(e_i - \hat{\theta}_i^B)^2 = \sum_{i=1}^{m} \phi_i \{ e_i - \hat{\theta}_i^B - s^{-1}(t - \bar{\theta}_w^B) r_i \}^2 + s^{-1}(t - \bar{\theta}_w^B)^2. \] (3.1.2)

The solution is now immediate from (3.1.2).

Remark 1: The constrained Bayes benchmarked estimators $\hat{\theta}_1^{BM}$ as given in (1) can also be viewed as limiting Bayes estimators under the loss
\[ L(\theta, e) = \sum_{i=1}^{m} \phi_i(\theta_i - e_i)^2 + \lambda(t - \bar{\theta}_w)^2, \] (3.1.3)
where $e = (e_1, \ldots, e_m)^T$, and $\lambda > 0$ is the penalty parameter. Like the $\phi_i$, the penalty parameter $\lambda$ can differ for different policy makers. The Bayes estimator of $\theta$ under the above loss (after some algebra) is given by

\[ \hat{\theta}_\lambda^B = \hat{\theta}^B + (s + \lambda^{-1})^{-1}(t - \bar{\theta}_w^B)r. \]

Clearly, when $\lambda \to \infty$, i.e., when we invoke the extreme penalty for not having the exact equality $\bar{\theta}_w = t$, we get the estimator given in (3.1.1). Otherwise, $\lambda$ serves as a trade-off between $t$ and $\bar{\theta}_w$ since

\[ w^T \hat{\theta}_\lambda^B = \frac{s\lambda}{s\lambda + 1} t + \frac{1}{s\lambda + 1} \bar{\theta}_w. \]

Remark 2: The balanced loss of Zellner (1986), Zellner (1988), and Zellner (1994) is not quite the same as the one in Remark 1 and is given by

\[ L(\theta, e) = \sum_{i=1}^{m} \phi_i(\theta_i - e_i)^2 + \lambda \sum_{i=1}^{m} (\hat{\theta}_i - e_i)^2. \]

This leads to the Bayes estimator $\hat{\theta}^B + \lambda(I + \phi)^{-1}(\hat{\theta} - \hat{\theta}^B)$, where $I$ is the identity matrix and $\phi = \text{Diag}(\phi_1, \ldots, \phi_m)$, which is a compromise between the Bayes estimator $\hat{\theta}^B$ and the direct estimator $\hat{\theta}$ of $\theta$ and converges to the direct estimator as $\lambda \to \infty$ and to the Bayes estimator when $\lambda \to 0$. 
The posterior risk of \( \hat{\theta}^{BM1} \) under the given loss in Theorem 1 simplifies to \( \sum_{i=1}^{m} \phi_i [V(\theta_i) + s^{-1}(t - \bar{\theta}_w)^2 r_i^2] \). Hence the excess posterior risk due to adjustment of the Bayes estimator if the assumed prior were “true” is given by \( \sum_{i} \phi_i s^{-1}(t - \bar{\theta}_w)^2 r_i^2 \). When \( \phi_i = w_i \) and \( \sum w_i = 1 \), the expression further simplifies to \( \frac{1}{m}(t - \bar{\theta}_w)^2 \). However, with a prior different from the assumed one, it is possible to have a lower posterior risk of the adjusted Bayes estimator than the Bayes estimator.

To see this in a very simple setting, consider the case where \( \hat{\theta}_i | \theta_i \sim N(\theta_i, 1) \) and \( \theta_i \sim N(0, \sigma_u^2) \), \( i = 1, \ldots, m \). Then the Bayes estimator of \( \theta \) is \( \hat{\theta}^B = (1 - B)\hat{\theta} \), where \( B = (1 + \sigma_u^2)^{-1} \). Also, if \( \phi_i = w_i \) and \( \sum w_i = 1 \), then \( r_i = 1 \) for all \( i = 1, \ldots, m \), and \( s = 1 \). Further, if \( t = \bar{\theta}_w \), as often is the case with internal benchmarking, then \( \hat{\theta}^{BM1} = (1 - B)\hat{\theta} + B\bar{\theta}_w \mathbf{1}_m \), where \( \mathbf{1}_m \) denotes an \( m \)-component vector with each element equal to one. If instead we have the prior \( \theta_i \sim N(0, \sigma_v^2) \), and \( B_0 = (1 + \sigma_v^2)^{-1} \), then after some simplification, the posterior risk of \( \hat{\theta}^B \) is \( (1 - B_0) + (B - B_0)^2 \sum_{i=1}^{m} w_i (\hat{\theta}_i - \bar{\theta}_w)^2 + (B - B_0)^2 \bar{\theta}_w^2 \), while that of \( \hat{\theta}^{BM1} \) is \( (1 - B_0) + (B - B_0)^2 \sum_{i=1}^{m} w_i (\hat{\theta}_i - \bar{\theta}_w)^2 + B_0^2 \bar{\theta}_w^2 \). Now \( \hat{\theta}^{BM1} \) has smaller posterior risk than that of \( \hat{\theta}^B \) if and only if \( |B - B_0|/B_0 > 1 \) which is quite possible if \( B_0 \) is very small compared to \( B \), i.e., if \( \sigma_v^2 \) is much larger than \( \sigma_u^2 \).

We now provide a generalization of Theorem 1 where we consider multiple constraints instead of one single constraint. As an example, for the SAIPE county-level analysis, one may need to control the county estimates in each state so that their weighted total agrees with the corresponding state estimates. We now consider a more general quadratic loss given by

\[
L(\theta, e) = (e - \theta)^T \Omega(e - \theta),
\]

where \( \Omega \) is a positive definite matrix. The following theorem provides a Bayesian solution for the minimization of \( E[L(\theta, e)|\hat{\theta}] \) subject to the constraint \( W^T e = t \), where \( t \) is a \( q \)-component vector and \( W \) is an \( m \times q \) matrix of rank \( q < m \).

**Theorem 2.** The constrained Bayesian solution under the loss 3.1.4 is given by
\[ \hat{\theta}^{MBM} = \hat{\theta}^B + \Omega^{-1} W (W^T \Omega^{-1} W)^{-1} (t - \bar{\hat{\theta}}^B), \]

where \( \bar{\hat{\theta}}^B_w = W^T \hat{\theta}^B \).

**Proof.** First write
\[ E[(e - \theta)^T \Omega (e - \theta)|\hat{\theta}] = E[(\theta - \hat{\theta}^B)^T \Omega (\theta - \hat{\theta}^B)|\hat{\theta}] + (e - \hat{\theta}^B)^T \Omega (e - \hat{\theta}^B). \]
Hence, the problem reduces to minimization of \( (e - \hat{\theta}^B)^T \Omega (e - \hat{\theta}^B) \) with respect to \( e \) subject to \( W^T e = t \). The result follows from the identity
\[
(e - \hat{\theta}^B)^T \Omega (e - \hat{\theta}^B) = [(e - \hat{\theta}^B - \Omega^{-1} W (W^T \Omega^{-1} W)^{-1} (t - \bar{\hat{\theta}}^B))^T
\times
\Omega[(e - \hat{\theta}^B - \Omega^{-1} W (W^T \Omega^{-1} W)^{-1} (t - \bar{\hat{\theta}}^B)]
\times
(t - \bar{\hat{\theta}}^B)^T (W^T \Omega^{-1} W)^{-1} (t - \bar{\hat{\theta}}^B). \]

\[ \square \]

The choice of the weight matrix \( \Omega \) usually depends on the experimenter depending on how much penalty she/he is willing to assign for a misspecified estimator. In the special case of a diagonal \( \Omega \), Wang et al. (2008) have argued in favor of \( \omega_i = [\text{Var}(\hat{\theta}_i)]^{-1} \).

### 3.2 Relationship with Some Existing Estimators

We now show how some of the existing benchmarked estimators follow as special cases of earlier proposed Bayes estimators. Indeed, our proposed class of Bayes estimators includes some of the raked benchmarked estimators as well as some of the other benchmarked estimators proposed by several authors.

**Example 1:** It is easy to see why the raked Bayes estimators, considered for example in You and Rao (2004), belong to the general class of estimators proposed in Theorem 1. If we choose (possibly quite artificially) \( \phi_i = w_i / \hat{\theta}_i^B \), \( i = 1, \ldots, m \), with \( \hat{\theta}_i^B > 0 \) for all \( i = 1, \ldots, m \), then \( r = \hat{\theta}^B \) and \( s = \bar{\hat{\theta}}^B_w \). Consequently, the constrained Bayes estimator proposed in Theorem 1 simplifies to \( t / \bar{\hat{\theta}}^B_w \hat{\theta}^B \), which is the raked Bayes estimator. In particular, we can take \( t = \bar{\hat{\theta}}_w \). We may also note that this choice of the \( \phi_i \)'s is different from the one in Wang et al. (2008) who considered \( \phi_i = w_i / \hat{\theta}_i. \)
Example 2: The next example considers the usual random effects model as considered in Fay and Herriot (1979) or Pfeffermann and Nathan (1981). Under this model, \( \tilde{\theta}_i | \theta_i \overset{\text{ind}}{\sim} N(\theta_i, D_i) \) and \( \theta_i \overset{\text{ind}}{\sim} N(x_i^T \sigma_u^2) \), with the \( D_i > 0 \) known. For the HB approach, we then use the the prior \( \pi(\beta, \sigma_u^2) = 1 \) although other priors are also possible as long as the posteriors are proper. The HB estimators \( E(\theta | \tilde{\theta}) \) cannot be obtained analytically, but it is possible to find them numerically either through Markov chain Monte Carlo (MCMC) or through numerical integration. Denoting the HB estimators by \( \hat{\beta}_i^B \), we can obtain the benchmarked Bayes estimators \( \hat{\theta}_i^{BM1} \) by applying Theorem 1. With an empirical Bayes approach, we estimate \( \tilde{\sigma}_i^2 = \sigma_u^2 + D_i \), \( i = 1, \ldots, m \), from the marginal independent \( N(x_i^T \beta, V_i) \) distribution of the \( \hat{\theta}_i \). Isaki et al. (2000) suggested that \( \varphi_i = \tilde{V}_i^{-1} \), \( i = 1, \ldots, m \).

Example 3: Wang et al. (2008) considered a slightly varied form of the Fay-Herriot random effects model, with the only change that the marginal variance of the \( \theta_i \) are now \( z_i^2 \sigma_u^2 \), where the \( z_i \) are known. They do not assume normality, but they restricted their attention to the class of linear estimators of \( \theta \) and benchmark the best linear unbiased predictor (BLUP) of \( \theta \) when \( \sigma_u^2 \) is known. For this example, the benchmarked estimators given in (3.2.2) of Wang et al. (2008) can be derived from Theorem 1. First for known \( \sigma_u^2 \), consider the uniform prior for \( \beta \). Write \( B_i = D_i / (D_i + z_i^2 \sigma_u^2) \), \( B = \text{Diag}(B_1, \ldots, B_m) \), \( \Sigma = \text{Diag}(D_1 + z_1^2 \sigma_u^2, \ldots, D_m + z_m^2 \sigma_u^2) \), \( x^T = (x_1, \ldots, x_m) \), and \( \tilde{\beta} = (x^T \Sigma^{-1} x)^{-1} x^T \Sigma^{-1} \), assuming \( x \) to be a matrix with full column rank. Then the Bayes estimator of \( \theta \) is \( \tilde{\theta}^B = B \hat{\theta} + (I - B)x^T \tilde{\beta} \), which is the same as the BLUP of \( \theta \) as well. Now identify the \( r_i \) in this paper with the \( a_i \) of Wang et al. (2008) to get (19) in their paper.

As shown in Wang et al. (2008), the Pfeffermann and Barnard (1991) estimator belongs to their (and accordingly our) general class of estimators where we choose \( \phi_i = w_i / \text{Cov}(\hat{\theta}_i^B, \tilde{\beta}_w^B) \), where the covariance is calculated over the joint distribution of \( \tilde{\theta} \) and \( \theta \), treating \( \beta \) as an unknown but fixed parameter. Then \( r \) contains the elements of \( \text{Cov}(\hat{\theta}_i^B, \tilde{\beta}^B) \) as its components, while \( s = V(\tilde{\beta}_w^B) \).
Instead of the constrained Bayes estimators as given in (3.1.1), it is possible to obtain constrained empirical Bayes (EB) estimators as well when we estimate the prior parameters from the marginal distribution of \( \hat{\theta} \) (after integrating out \( \theta \)). The resulting EB estimators are given by

\[
\hat{\theta}^{EBM1} = \hat{\theta}^{EB} + s^{-1}(t - \bar{\theta}_w^E)r.
\]  

(3.2.1)

where \( \hat{\theta}^{EB} = (\hat{\theta}_1^{EB}, \ldots, \hat{\theta}_m^{EB})^T \) is an EB estimator of \( \theta \) and \( \bar{\theta}_w^E = \sum_{i=1}^m w_i \hat{\theta}_i^{EB} \).

Remark 3: In the model as considered in Example 3, for unknown \( \sigma_u^2 \), we get estimators of \( \beta \) and \( \sigma_u^2 \) simultaneously from the marginals \( \hat{\theta}_i \sim N(x_i^T \beta, D_i + z_i^2 \sigma_u^2) \) (Fay and Herriot (1979); Prasad and Rao (1990); Datta and Lahiri (2000); Datta et al. (2005)). Denoting the estimator of \( \sigma_u^2 \) by \( \hat{\sigma}_u^2 \), we estimate \( \Sigma \) by \( \hat{\Sigma} = \text{Diag}(D_1 + z_1^2 \hat{\sigma}_u^2, \ldots, D_m + z_m^2 \hat{\sigma}_u^2) \), \( \beta \) by \( \hat{\beta} = (X^T \hat{\Sigma}^{-1} X)^{-1} X^T \hat{\Sigma}^{-1} \hat{\theta} \) and \( B \) by \( \hat{B} = D \hat{\Sigma}^{-1} \), where \( D = \text{Diag}(D_1, \ldots, D_m) \). Denoting the resulting EB estimator of \( \theta \) by \( \hat{\theta}_{EB} \), we get

\[
\hat{\theta}_{EB} = (I_m - \hat{B}) \hat{\theta} + \hat{B} X \hat{\beta}.
\]  

(3.2.2)

The benchmarked EB estimator is now obtained from (3.2.1).

Remark 4: The benchmarked EB estimator as given in (5) includes the one given in Isaki et al. (2000), where we take \( \phi_i \) as the reciprocal of the \( i \)th diagonal element of \( \hat{\Sigma} \) for all \( i = 1, \ldots, m \). Another option is to take \( \phi_i \) as the reciprocal of an estimator of \( V(\hat{\theta}_i^B) \), with the variance computed once again under the joint distribution of \( \hat{\theta} \) and \( \theta \), treating \( \beta \) as an unknown but fixed parameter.

### 3.3 Benchmarking with Both Mean and Variability Constraints

There are situations that demand benchmarking not only the first moment of the Bayes estimators but their variability as well. We will address this issue in the special case when \( \phi_i = cw_i \) for some \( c > 0 \), \( i = 1, \ldots, m \). In this case, \( \hat{\theta}_i^{BM1} \) given in (3.1.1) simplifies to \( \hat{\theta}_i^B + (t - \bar{\theta}_w^B) \) for all \( i = 1, \ldots, m \). This itself is not a very desirable estimator since then \( \sum_{i=1}^m w_i((\hat{\theta}_i^{BM1} - t)^2 = \sum_{i=1}^m w_i(\hat{\theta}_i^B - \bar{\theta}_w^B)^2 \). Ghosh (1992) showed that \( \sum_{i=1}^m w_i(\hat{\theta}_i^B - \bar{\theta}_w^B)^2 < \sum_{i=1}^m w_i E[(\theta_i - \bar{\theta}_w)^2|\theta] \). In other words, the weighted ensemble
variability of the estimators $\hat{\theta}_i^{BM1}$ is an underestimate of the posterior expectation of the corresponding weighted ensemble variability of the population parameters. To address this issue, or from other considerations, we will consider estimators $\hat{\theta}_i^{BM2}$, $i = 1, \ldots, m$, which satisfy two constraints, namely, (i) $\sum_{i=1}^{m} w_i \hat{\theta}_i^{BM2} = t$ and (ii) $\sum_{i=1}^{m} w_i (\hat{\theta}_i^{BM2} - t)^2 = H$, where $H$ is a preassigned number taken from some other source, for example from census data, or is taken as $\sum_{i=1}^{m} w_i E[(\theta_i - \bar{\theta}_w)^2 | \hat{\theta}]$ more in the spirit of Louis (1984) and Ghosh (1992). Subject to these two constraints, we minimize $\sum_{i=1}^{m} w_i E[(\theta_i - e_i)^2 | \hat{\theta}]$. The following theorem provides the resulting estimator.

**Theorem 3.** Subject to (i) and (ii), the benchmarked Bayes estimators of $\theta_i$, $i = 1, \ldots, m$, are given by

$$\hat{\theta}_i^{BM2} = t + a_{CB}(\hat{\theta}_i^B - \bar{\theta}_w^B),$$

(3.3.1)

where $a_{CB}^2 = H / \sum_{i=1}^{m} w_i (\hat{\theta}_i^B - \bar{\theta}_w^B)^2$. Note that $a_{CB} \geq 1$ when $H = \sum_{i=1}^{m} w_i E[(\theta_i - \bar{\theta}_w)^2 | \hat{\theta}]$.

**Proof.** As in Theorem 1, the problem reduces to minimization of $\sum_{i=1}^{m} w_i (e_i - \hat{\theta}_i^B)^2$. We will write

$$\sum_{i=1}^{m} w_i (e_i - \hat{\theta}_i^B)^2 = \sum_{i=1}^{m} w_i [(e_i - \bar{e}_w) - (\hat{\theta}_i^B - \bar{\theta}_w^B)]^2 + (\bar{e}_w - \bar{\theta}_w^B)^2.$$  

(3.3.2)

Now define two discrete random variables $Z_1$ and $Z_2$ such that

$$P(Z_1 = e_i - \bar{e}_w, Z_2 = \hat{\theta}_i^B - \bar{\theta}_w^B) = w_i,$$

$i = 1, \ldots, m$. Hence,

$$\sum_{i=1}^{m} w_i [(e_i - \bar{e}_w) - (\hat{\theta}_i^B - \bar{\theta}_w^B)]^2 = V(Z_1) + V(Z_2) - 2Cov(Z_1, Z_2)$$

which is minimized when the correlation between $Z_1$ and $Z_2$ equals 1, i.e., when

$$e_i - \bar{e}_w = a(\hat{\theta}_i^B - \bar{\theta}_w^B) + b,$$

(3.3.3)

$i = 1, \ldots, m$, with $a > 0$. Multiplying both sides of (3.3.3) by $w_i$ and summing over $i = 1, \ldots, m$, we get $b = 0$. Next, squaring both sides of (3.3.3), then multiplying both
sides by \( w_i \) and summing over \( i = 1, \ldots, m \), we get 
\[
H = a^2 \sum_{i=1}^{m} w_i (\hat{\theta}_i^B - \bar{\theta}_w^B)^2 
\]
due to condition (ii). Finally, by condition (i), the result follows from (3.3.3).

Remark 5: As in the case of Theorem 1, it is possible to work with arbitrary \( \phi_i \) rather than \( \phi_i = w_i \) for all \( i = 1, \ldots, m \). But then we do not get a closed form minimizer, although it can be shown that such a minimizer exists. We can also provide an algorithm for finding this minimizer numerically.

The multiparameter extension of the above result proceeds as follows. Suppose now \( \hat{\theta}_1, \ldots, \hat{\theta}_m \) are the \( q \)-component direct estimators of the small area means \( \theta_1, \ldots, \theta_m \).

We may generalize the constraints (i) and (ii) as
\[
(iM) \quad \bar{e}_w = \sum_{i=1}^{m} w_i e_i = t \quad \text{for some specified } t \text{ and}

(iiM) \quad \sum_{i=1}^{m} w_i (e_i - \bar{e}_w)(e_i - \bar{e}_w)^T = H,
\]
where \( H \) is a positive definite (possibly data-dependent matrix) and is often taken as
\[
\sum_{i=1}^{m} w_i E[(\theta_i - \bar{\theta}_w)(\theta_i - \bar{\theta}_w)^T | \hat{\theta}_w].
\]
The second condition is equivalent to
\[
c^T \left\{ \sum_{i=1}^{m} w_i (e_i - \bar{e}_w)(e_i - \bar{e}_w)^T \right\} c = c^T H c
\]
for every \( c = (c_1, \ldots, c_q)^T \neq 0 \), which simplifies to
\[
\sum_{i=1}^{m} w_i \{c^T (e_i - \bar{e}_w)\}^2 = c^T H c.
\]
An argument similar to before now leads to 
\[
c^T \hat{\theta}_i^{BM2} = c^T \bar{\theta}_w + a_{CB} c^T (\hat{\theta}_i^B - \bar{\theta}_w^B)
\]
for every \( c \neq 0 \), where \( \hat{\theta}_i^B \) is the posterior mean of \( \theta_i \), \( \bar{\theta}_w^B = \sum_{i=1}^{m} w_i \hat{\theta}_i^B \), and
\[
a_{CB}^2 = c^T H c / \sum_{i=1}^{m} w_i \{c^T (\hat{\theta}_i^B - \bar{\theta}_w^B)\}^2.
\]
The coordinatewise benchmarked Bayes estimators are now obtained by letting \( c = (1, 0, \ldots, 0)^T, \ldots, (0, 0, \ldots, 1)^T \) in succession.

The proposed approach can be extended also to a two-stage benchmarking somewhat similar to what is considered by Pfeffermann and Tiller (2006). To cite an example, consider the SAIPE scenario where one wants to estimate the number of poor school children in different counties within a state, as well as those numbers within the different school districts in all these counties. Let \( \hat{\theta}_i \) denote the Current Population
Survey (CPS) estimate of \( \theta_i \), the true number of poor school children for the \( i \)th county and \( \hat{\theta}_i^B \) the corresponding Bayes estimate, namely the posterior mean. Subject to the constraints 
\[
\bar{e}_w = \sum_{i=1}^m w_i \hat{\theta}_i = \bar{\theta}_w, \quad \text{and} \quad \sum_{i=1}^m w_i (e_i - \bar{e}_w)^2 = H,
\]
the benchmarked Bayes estimate for \( \theta_i \) in the \( i \)th county is \( \hat{\theta}_i^{BM2} \) as given in (3.3.1).

Next, suppose that \( \hat{\xi}_{ij} \) is the CPS estimator of \( \xi_{ij} \), the true number of poor school children for the \( j \)th school district in the \( i \)th county, and \( \eta_{ij} \) is the weight attached to the direct CPS estimator of \( \xi_{ij} \), \( j = 1, \ldots, n_i \). We seek estimators \( e_{ij} \) of \( \xi_{ij} \) such that (i) 
\[
\bar{e}_{in} = \sum_{j=1}^{n_i} \eta_{ij} e_{ij} = \bar{\xi}_{in,\text{BENCH}}, \quad \text{the benchmarked estimator of} \quad \bar{\xi}_{in} = \sum_{j=1}^{n_i} \eta_{ij} \xi_{ij}, \quad \text{and} \quad \text{(ii) } \sum_{j=1}^{n_i} \eta_{ij} (e_{ij} - \bar{e}_{in})^2 = H^*_i \text{ for some preassigned } H^*_i, \text{ where again } H^*_i \text{ can be taken as } \sum_{j=1}^{n_i} \eta_{ij} E[(\xi_{ij} - \bar{\xi}_{in})^2|\bar{x}_{ii}], \text{ } \bar{x}_{ii} \text{ being the vector with elements } \xi_{ij}. \text{ A benchmarked estimator similar to (3.3.1) can now be found for the } \xi_{ij} \text{ as well.}

### 3.4 An Illustrative Example

The motivation behind this example is primarily to illustrate how the proposed Bayesian approach can be used for real-life data. The Small Area Income and Poverty Estimates (SAIPE) program at the U.S. Bureau of the Census produces model-based estimates of the number of poor school-aged children (5–17 years old) at the national, state, county, and school district levels. The school district estimates are benchmarked to the state estimates by the Department of Education to allocate funds under the No Child Left Behind Act of 2001. In the SAIPE program, the model-based state estimates are benchmarked to the national school-aged poverty rate using the ratio adjustment method. The number of poor school-aged children has been collected from the Annual Social and Economic Supplement (ASEC) of the CPS from 1995 to 2004, while ACS estimates have been used beginning in 2005. Additionally, the model-based county estimates are benchmarked to the model-based state estimates in a hierarchical fashion, once again using ratio adjustments. In this section, we will consider three sets of risk function weights \( \phi_i \) that will be used to benchmark the estimated state poverty rates based on Theorem 1. We will also benchmark using the results from Theorem 3.
In the SAIPE program, the state model for poverty rates in school-aged children follows the basic Fay-Herriot framework (see e.g. Bell (1999)),

\[ \hat{\theta}_i = \theta_i + e_i \]  
\[ \theta_i = x_i^T \beta + u_i \]

where \( \theta_i \) is the true state-level poverty rate, \( \hat{\theta}_i \) is the direct survey estimate (from CPS ASEC), \( e_i \) is the sampling error term with assumed known variance \( D_i \), \( x_i \) are the predictors, \( \beta \) is the vector of regression coefficients, and \( u_i \) is the model error with constant variance \( \sigma_u^2 \). The explanatory variables in the model are IRS income tax–based pseudo-estimate of the child poverty rate, IRS non-filer rate, food stamp rate, and the residual term from the regression of the 1990 Census estimated child poverty rate. The posterior means and variances of \((\beta, \theta_i, \sigma_u^2)\) are estimated using a rejection sampling–type algorithm proposed by Everson and Morris (2000). Multivariate normal distributions are required for the first two levels of the hierarchy. See their paper for further details.

The state estimates were benchmarked to the CPS direct estimate of the national school-aged child poverty rate until 2004. The weights \( w_i \) to calibrate the state’s poverty rates to the national poverty rate are proportional to the population estimates of the number of school-aged children in each state. We utilize three different sets of risk function weights \( \phi_i \) to benchmark the estimated state poverty rates based on Theorem 1. The first set of weights will be the weights used in the benchmarking, i.e., \( \phi_i = w_i \). The second set of weights creates the ratio-adjusted benchmarked estimators \( \phi_i = w_i / \hat{\theta}_i^B \) (Example 1). The third set of weights uses the results from Pfeffermann and Barnard (1991) where \( \phi_i = w_i / \text{Cov}(\hat{\theta}_i^B, \hat{\theta}_w^B) \) (Example 3). Let this set of benchmarked estimates be denoted as \( \hat{\theta}^{(1)}, \hat{\theta}^{(r)} \) and \( \hat{\theta}^{(PB)} \), respectively. Finally, we benchmark the state poverty estimates using the results from Theorem 3 and denote the estimator as \( \hat{\theta}^{(3)} \).
Table 3-1. Benchmarking Statistics for ASEC CPS

<table>
<thead>
<tr>
<th>year</th>
<th>$t$ (est)</th>
<th>$\bar{\hat{\theta}}_B$ (est)</th>
<th>$a_{CB}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>18.2</td>
<td>17.9</td>
<td>1.04</td>
</tr>
<tr>
<td>1997</td>
<td>17.9</td>
<td>17.8</td>
<td>1.04</td>
</tr>
<tr>
<td>1998</td>
<td>17.0</td>
<td>16.8</td>
<td>1.08</td>
</tr>
<tr>
<td>1999</td>
<td>15.2</td>
<td>14.9</td>
<td>1.07</td>
</tr>
<tr>
<td>2000</td>
<td>15.6</td>
<td>15.4</td>
<td>1.05</td>
</tr>
<tr>
<td>2001</td>
<td>15.3</td>
<td>15.1</td>
<td>1.05</td>
</tr>
</tbody>
</table>

For benchmarking, as given by Theorems 1 and 3, the key summary quantities are $t = \sum_i w_i \hat{\theta}_i$, $\bar{\hat{\theta}}_B = \sum_i w_i \hat{\theta}_B^B$, and $a_{CB}$. As noted earlier in Theorem 3, the choice of $H$ as suggested in this theorem implies that $a_{CB} \geq 1$. Six years of historical data from the CPS and the SAIPE program are analyzed and benchmarked using the four criteria mentioned above. Table 3-1 gives the key quantities for these six years. The hierarchical Bayes estimates underestimate the benchmarked poverty rate for all years given. Even if the estimate $\bar{\hat{\theta}}_B^B$ is close to the benchmarked value $t$, there is still a strong desire to have exact agreement between the quantities when producing official statistics.

Figure 3-1 shows the differences of the various benchmarked estimates from the hierarchical Bayes estimate $\hat{\theta}^B$ made for the year 1998 when the overall poverty level had to be raised to agree with the national direct estimate. We also show this particular year since it illustrates when $a_{CB}$ is the largest at 1.08, corresponding to a very steep slope for $\hat{\theta}^{(3)}$ and illustrating that when the Bayes estimate is large we find a large linear increase, whereas when the Bayes estimate is quite small, we actually find that the estimates are adjusted downwards. We contrast this with $\hat{\theta}^{(r)}$, where we find that a large value of the Bayes estimate results in a small linear increase, whereas a small value of the Bayes estimate results in a small linear adjustment. The behavior of $\hat{\theta}^{(1)}$ is fairly obvious in that it the Bayes estimate is being adjusted by the same amount for each small area. Finally, observing $\hat{\theta}^{(PB)}$, we find that it is somewhat similar to $\hat{\theta}^{(r)}$, however, it is not a linear estimator.
We compare this plot to Figure 3-2. This figure shows the differences for year 2001, where the overall poverty level had to be raised to obtain agreement. Similar comments can be made here as were made for year 1998. The main difference however is regarding \( \hat{\theta}^{(3)} \). From Figure 3-2, we can see that the slope is not as steep, which is due to the fact that \( a_{CB} \) is 1.04. However, we still notice that when the Bayes estimates are small we can get adjustments that are negative (as was found also in year 2001).

In addition, if the differences for each of the benchmarked estimators \( \hat{\theta}^{(1)}, \hat{\theta}^{(r)}, \) and \( \hat{\theta}^{(2)} \) from the HB estimator are plotted versus the the value of the Bayes estimator, the differences for each benchmarked estimator will fall along a straight line. The lines for all three estimators will pass through the point \((\bar{\hat{\theta}}^B_w, t - \bar{\hat{\theta}}^B_w)\). In fact, these benchmarked estimators can be written in the form:

\[
\hat{\theta}^{BM}_i = t + \alpha (\hat{\theta}_i^B - \bar{\hat{\theta}}^B_w)
\]

where \( \alpha = 1 \) for \( \hat{\theta}^{(1)} \), \( \alpha = t/\bar{\hat{\theta}}^B_w \) for \( \hat{\theta}^{(r)} \), and \( \alpha = a_{CB} \) for \( \hat{\theta}^{(3)} \). The slopes of the lines in Figures 1 and 2 for differences in the benchmarked estimates from \( \hat{\theta}_i^B \) are \( \alpha - 1 \). The slopes for \( \hat{\theta}_i^{(1)} \) and \( \hat{\theta}_i^{(r)} \) depend on whether the benchmarked total \( t \) is larger or smaller than the model-based estimate \( \bar{\hat{\theta}}^B_w \). However, since \( a_{CB} \geq 1 \), the slope for the difference \( \hat{\theta}_i^{(3)} - \bar{\hat{\theta}}^B_w \) will always be non-negative. The Pfeffermann-Barnard benchmarked estimator does not follow this form. However, it does show a trend in a similar direction as the benchmarked estimator \( \hat{\theta}_i^{(r)} \) based on ratio adjustment.
Figure 3-1. Change in Estimators due to Benchmarking for 1998

Figure 3-2. Change in Estimators due to Benchmarking for 2001
CHAPTER 4
TWO-STAGE BENCHMARKING

We extend the work of Datta et al. (2011) and propose two-stage benchmarking based on a nested error regression model. For example, in the Small Area Income and Poverty Estimation (SAIPE) project of the U.S. Census Bureau, one is required to benchmark the state estimates so that their aggregate matches the national estimate. Also, the county estimates within each state are benchmarked so that the aggregated benchmarked county estimates match the corresponding benchmarked state estimates.

The current approach to two-stage benchmarking uses two separate models, for example, one for the states and the other for the counties, to achieve the necessary benchmarking. In contrast, we consider one single model to obtain the Bayes estimates and then adjust the Bayes estimates to achieve benchmarking at both levels. The need for a single model is often felt in many small area problems. For instance, the National Agricultural Statistical Service (NASS) of the United States Department of Agriculture (USDA) requires simultaneous benchmarking of region-level agricultural cash rent estimates to the state estimates, and the county-level estimates to the corresponding region estimates.

4.1 Two-Stage Benchmarking Results

Let \( \theta_{ij} \) denote the true parameter of interest for the \( j \)th unit in the \( i \)th area, and let \( \hat{\theta}_{ij}^B \) denote its Bayes estimator under a certain prior \( (j = 1, ..., n_i; \ i = 1, ..., m) \). For a given set of normalized weights \( w_{ij}, (\sum_{j=1}^{n_i} w_{ij} = 1 \text{ for all } i) \), let \( \bar{\theta}_{iw} = \sum_j w_{ij} \theta_{ij} \) denote the true weighted mean for the \( i \)th area. For example, \( \theta_{ij} \) may be the true proportion of poor school children in the \( j \)th county in the \( i \)th state, and \( w_{ij} \) may denote the true proportion of school children for the \( j \)th county in the \( i \)th state.

Also, let \( \eta_i (i = 1, ..., m) \) denote the normalized weights for the \( i \)th area, where \( \sum_i \eta_i = 1 \). We want to find estimates \( \hat{\theta}_{ij} \) for \( \theta_{ij} \) and \( e_i \) for \( \bar{\theta}_{iw} \) such that \( \sum_i \eta_i e_i = p \) and \( \sum_j w_{ij} \hat{\theta}_{ij} = e_i \) for all \( i \). For example, \( p \) might be the national proportion of poor school
children. Throughout we will use the notation \( \hat{\theta} = (\hat{\theta}_{11}, \ldots, \hat{\theta}_{1n_1}, \ldots, \hat{\theta}_{m1}, \ldots, \hat{\theta}_{mn_m})^T \) and
\( e = (e_1, \ldots, e_m)^T \). Also, \( \hat{\theta}_B \) denotes the Bayes estimate of \( \theta_{ij} \) under a certain prior
\((i = 1, \ldots, m; j = 1, \ldots, n_i)\). Our objective is to minimize the weighted squared error loss
\( L(\theta, \hat{\theta}, e) \) with respect to \( \hat{\theta}_B \) and \( e_i \), where

\[
L(\theta, \hat{\theta}, e) = \sum_i \sum_j \xi_{ij}(\hat{\theta}_{ij} - \theta_{ij})^2 + \sum_i \zeta_{i}(e_i - \bar{\theta}_{iw})^2
\]

subject to the restrictions (i) \( \sum_i \eta_i e_i = p \) and (ii) \( \sum_j w_{ij} \hat{\theta}_{ij} = e_i \) for all \( i \). Note that the
weights \( \xi_{ij} \) and \( \zeta_i \) need not be the same as \( w_{ij} \) and \( \eta_i \) respectively. Let \( s_i = \sum_j w_{ij}^2 \xi_{ij}^{-1} \) for
all \( i \). We now prove the following theorem.

**Theorem 4.** The minimizer of \( \mathbb{E}[L(\theta, \hat{\theta}, e)|\text{data}] \) with respect to \( \hat{\theta} \) and \( e \) subject to the
restrictions (i) and (ii) is given by

(a) \( \hat{\theta}_y = \hat{\theta}_y^B + \frac{(p - \bar{\theta}_w)\eta_i(1 + \zeta_i s_i)^{-1}w_{ij}s_{ij}^{-1}}{\sum_{k=1}^m \eta_k^2 s_k(1 + \zeta_k s_k)^{-1}} \) for all \( i, j \).

(b) \( e_i = \bar{\theta}_{iw}^B + \frac{(p - \bar{\theta}_w)\eta_i s_i(1 + \zeta_i s_i)^{-1}}{\sum_{k=1}^m \eta_k^2 s_k(1 + \zeta_k s_k)^{-1}} \) for all \( i \).

**Proof.** First, we write

\[
\mathbb{E}[L(\theta, \hat{\theta}, e)|\text{data}] = \mathbb{E}\left[ \sum_i \sum_j \xi_{ij}(\hat{\theta}_{ij} - \theta_{ij})^2 + \sum_i \zeta_{i}(\bar{\theta}_{iw} - \bar{\theta}_{iw}^B + \bar{\theta}_{iw}^B - e_i)^2|\text{data} \right]
\]

\[
= \sum_i \sum_j \xi_{ij} \mathbb{V}(\hat{\theta}_{ij}|\text{data}) + \sum_i \sum_j \xi_{ij}(\hat{\theta}_{ij} - \hat{\theta}_{ij}^B)^2
\]

\[
+ \sum_i \zeta_{i} \mathbb{V}(\bar{\theta}_{iw}|\text{data}) + \sum_i \zeta_{i}(\bar{\theta}_{iw}^B - e_i)^2.
\]

In view of (4.1.2), the problem reduces to minimization of

\[
g = \sum_i \sum_j \xi_{ij}(\hat{\theta}_{ij} - \hat{\theta}_{ij}^B)^2 + \sum_i \zeta_{i}(e_i - \bar{\theta}_{iw}^B)^2 - 2 \sum_i \lambda_{1i}(\sum_j w_{ij} \hat{\theta}_{ij} - e_i) - 2 \lambda_{2}(\sum_i \eta_i e_i - p)
\]

(4.1.3)
with respect to \( \hat{\theta}_{ij} \) and \( e_i \) (for all \( i, j \)) where the \( \lambda_1 \) and \( \lambda_2 \) are the Lagrangian multipliers.

From (4.1.3), we find

\[
\frac{\partial g}{\partial \hat{\theta}_{ij}} = 2\xi_{ij}(\hat{\theta}_{ij} - \hat{\theta}^B_{ij}) - 2\lambda_{1i}w_{ij} \tag{4.1.4}
\]

\[
\frac{\partial g}{\partial e_i} = 2\zeta_i(e_i - \bar{\hat{\theta}}^B_{iw}) + 2\lambda_{1i} - 2\lambda_{2i} \tag{4.1.5}
\]

Solving \( \frac{\partial g}{\partial \theta_{ij}} = 0 \) implies \( \hat{\theta}_{ij} = \hat{\theta}^B_{ij} + \lambda_{1i}w_{ij} \xi_{ij}^{-1} \). Now invoking \( e_i = \sum_j w_{ij} \hat{\theta}_{ij} \), we find that \( e_i = \bar{\hat{\theta}}^B_{iw} + \lambda_{1i}s_i \). That is, \( \lambda_{1i} = (e_i - \bar{\hat{\theta}}^B_{iw})s_i^{-1} \).

Next, we solve \( \frac{\partial g}{\partial e_i} = 0 \), which implies

\[
\zeta_i(e_i - \bar{\hat{\theta}}^B_{iw}) + \lambda_{1i} - \lambda_{2i} = 0 \iff \zeta_i(e_i - \bar{\hat{\theta}}^B_{iw}) + s^{-1}_i(e_i - \bar{\hat{\theta}}^B_{iw}) = \lambda_{2i} \iff
\]

\[
(e_i - \bar{\hat{\theta}}^B_{iw})(1 + s_i\zeta_i) = \lambda_{2i}s_i \iff e_i = \bar{\hat{\theta}}^B_{iw} + \lambda_{2i}s_i(1 + s_i\zeta_i)^{-1}. \tag{4.1.6}
\]

Recall that \( \sum_i \eta_i e_i = p \). Apply this to (4.1.6). Then

\[
p = \bar{\hat{\theta}}^B_{iw} + \lambda_2 \sum_i \eta_i^2 s_i(1 + s_i\zeta_i)^{-1}, \tag{4.1.7}
\]

which implies \( \lambda_2 = \frac{p - \bar{\hat{\theta}}^B_{iw}}{\sum_k \eta_k^2 s_k(1 + s_k\zeta_k)^{-1}} \). Then by (4.1.6) and (4.1.7), we get part (b) of the theorem.

Next by \( \hat{\theta}_{ij} = \hat{\theta}^B_{ij} + \lambda_{1i}w_{ij}\xi_{ij}^{-1} \) and \( \lambda_{1i} = (e_i - \bar{\hat{\theta}}^B_{iw})s_i^{-1} \), we get

\[
\hat{\theta}_{ij} = \hat{\theta}^B_{ij} + (e_i - \bar{\hat{\theta}}^B_{iw})w_{ij}s_i^{-1}\xi_{ij}^{-1}. \tag{4.1.8}
\]

Now apply (b) to get (a) from (8).

\[ \square \]

Remark 6: The result simplifies somewhat when \( \xi_i = w_{ij} \) for all \( i, j \). Then \( s_i = 1 \) for all \( i \) so that

\[
\hat{\theta}_{ij} = \hat{\theta}^B_{ij} + \frac{(p - \bar{\hat{\theta}}^B_{iw})\eta_i(1 + \zeta_i)^{-1}}{\sum_{k=1}^m \eta_k^2(1 + \zeta_k)^{-1}}
\]

\[
e_i = \bar{\hat{\theta}}^B_{iw} + \frac{(p - \bar{\hat{\theta}}^B_{iw})\eta_i(1 + \zeta_i)^{-1}}{\sum_{k=1}^m \eta_k^2(1 + \zeta_k)^{-1}}.
\]
Further simplification is possible when $\zeta_i = \eta_i$ ($i = 1, \ldots, m$).

Often, in addition to controlling the averages as in Theorem 1, one wants to control variability of the estimates as well. The following theorem provides a partial answer to this problem. We take $\xi_{ij} = w_{ij}$.

**Theorem 5.** The minimizing solution of $E[\sum_i \sum_j \xi_{ij}(\hat{\theta}_{ij} - \theta_{ij})^2 + \sum_i \zeta_i(e_i - \bar{\theta}_{iw})^2 | \text{data}]$ subject to (i) $\sum_j w_{ij} \hat{\theta}_{ij} = e_i$, (ii) $\sum_j w_{ij}(\hat{\theta}_{ij} - e_i)^2 = h_i$, and (iii) $\sum_i \eta_i e_i = p$ is given by

(a) $\hat{\theta}_{ij} = \hat{\theta}_{ij}^B + \frac{(p - \bar{\theta}_{iw}^B)\eta_i(1 + \zeta_i)^{-1}}{\sum_k \eta_k^2(1 + \zeta_k)^{-1}} + (h_i d_i^{-1})^{\frac{1}{2}} (\hat{\theta}_{ij}^B - \bar{\theta}_{iw}^B)$

(b) $e_i = \bar{\theta}_{iw}^B + \frac{(p - \bar{\theta}_{iw}^B)\eta_i(1 + \zeta_i)^{-1}}{\sum_k \eta_k^2(1 + \zeta_k)^{-1}}$.

where $d_i = \sum_j w_{ij}(\hat{\theta}_{ij}^B - \bar{\theta}_{iw}^B)^2$.

**Proof.** Similar to Theorem 1, the problem reduces to minimization of

$$g = \sum_i \sum_j w_{ij}(\hat{\theta}_{ij} - \hat{\theta}_{ij}^B)^2 + \sum_i \zeta_i(e_i - \bar{\theta}_{iw}^B)^2 - 2 \sum_i \lambda_1 \left( \sum_j w_{ij} \hat{\theta}_{ij} - e_i \right)$$

$$- \sum_i \lambda_2 \left( \sum_j w_{ij}(\hat{\theta}_{ij} - e_i)^2 - h_i \right) - 2 \lambda_3 \sum_i \eta_i e_i - p \quad (4.1.9)$$

with respect to $\hat{\theta}_{ij}$ and $e_i$ subject to (i)–(iii), where again $\lambda_1$, $\lambda_2$, and $\lambda_3$, are the Lagrangian multipliers. We begin with

$$0 = \frac{\partial g}{\partial \hat{\theta}_{ij}} = 2w_{ij}(\hat{\theta}_{ij} - \hat{\theta}_{ij}^B) - 2 \lambda_1 w_{ij} - 2 \lambda_3 w_{ij}(\hat{\theta}_{ij} - e_i) \quad (4.1.10)$$

$$0 = \frac{\partial g}{\partial e_i} = 2 \zeta_i(e_i - \bar{\theta}_{iw}^B) + 2 \lambda_1 - 2 \lambda_2 \eta_i. \quad (4.1.11)$$

By (4.1.10), we find

$$\hat{\theta}_{ij} = \hat{\theta}_{ij}^B + \lambda_1 + \lambda_3(\hat{\theta}_{ij} - e_i). \quad (4.1.12)$$

This implies

$$e_i = \bar{\theta}_{iw}^B + \lambda_1. \quad (4.1.13)$$
Combining (4.1.12) and (4.1.13), \( \hat{\theta}_j - e_i = \hat{\theta}_j^B - \hat{\theta}_w^B + \lambda_3(\hat{\theta}_j - e_i) \). Now we apply constraint (ii) to find \( (1 - \lambda_3)^2 h_i = \sum_j W_{ij}(\hat{\theta}_j^B - \hat{\theta}_w^B)^2 = d_i \), which implies \( \lambda_3 = 1 - (d_i h_i^{-1})^{\frac{1}{2}} \).

This in return implies that

\[
\hat{\theta}_j - e_i = \hat{\theta}_j^B - \hat{\theta}_w^B + \left[ 1 - (d_i h_i^{-1})^{\frac{1}{2}} \right] (\hat{\theta}_j - e_i)
\]

\[ \iff \hat{\theta}_j - e_i = (h_i d_i^{-1})^{\frac{1}{2}} (\hat{\theta}_j^B - \hat{\theta}_w^B). \] (4.1.14)

From (4.1.11), we know that \( e_i = \hat{\theta}_w^B - \lambda_1 \zeta_i^{-1} + \lambda_2 \eta_i \zeta_i^{-1} \). Now by (4.1.13),

\[ e_i = \hat{\theta}_w^B - (e_i - \hat{\theta}_w^B) \zeta_i^{-1} + \lambda_2 \eta_i \zeta_i^{-1}. \] This leads to \( \lambda_2 \eta_i = (e_i - \hat{\theta}_w^B)(1 + \zeta_i) \), which implies \( p = \hat{\theta}_w^B + \lambda_2 \sum_k \eta_k^2 (1 + \zeta_k)^{-1} \).

Then

\[ e_i = \hat{\theta}_w^B + \lambda_2 \eta_i (1 + \zeta_i)^{-1} \]

\[ \implies e_i = \hat{\theta}_w^B + \frac{(p - \hat{\theta}_w^B)}{\sum_k \eta_k^2 (1 + \zeta_k)^{-1}} \eta_i (1 + \zeta_i)^{-1}. \] (4.1.15)

We combine (4.1.14) and (4.1.15) to get the the above theorem.

We now extend Theorem 1 to a multiparameter setting. Consider the following notation. Let \( \theta = (\theta_{11}, \ldots, \theta_{1n_1}, \ldots, \theta_{mn_1}, \ldots, \theta_{mn_m})^T \), \( \hat{\theta} = (\hat{\theta}_{11}, \ldots, \hat{\theta}_{1n_1}, \ldots, \hat{\theta}_{mn_1}, \ldots, \hat{\theta}_{mn_m})^T \), and \( e = (e_1, \ldots, e_m)^T \). Define \( \hat{\theta}_w = \sum_j W_{ij} \theta_{ij} \) and \( \hat{\theta}_w = \sum_i \Gamma_i \hat{\theta}_w \), where \( W_{ij} \) and \( \Gamma_i \) are known positive definite matrices (for all \( i \) and \( j \)). We denote the Bayes estimators by \( \hat{\theta}_w^B = (\hat{\theta}_{11}^B, \ldots, \hat{\theta}_{1n_1}^B, \ldots, \hat{\theta}_{mn_1}^B, \ldots, \hat{\theta}_{mn_m}^B)^T \). We also define \( \hat{\theta}_w^B = \sum_j W_{ij} \hat{\theta}_{ij}^B \) and \( \hat{\theta}_w^B = \sum_i \Gamma_i \hat{\theta}_w^B \).

**Theorem 6.** Consider the loss function \( L(\theta, \hat{\theta}, e) = \sum_i \sum_j (\hat{\theta}_{ij} - \theta_{ij})^T \Lambda_{ij} (\hat{\theta}_{ij} - \theta_{ij}) + \sum_i (e_i - \hat{\theta}_w)^T \Omega_i (e_i - \hat{\theta}_w) \), where \( \Lambda_{ij} \) and \( \Omega_i \) are positive definite matrices. Then the minimizing solution of \( E[L(\theta, \hat{\theta}, e)|\text{data}] \) subject to (i) \( \sum_j W_{ij} \hat{\theta}_{ij} = e_i \), and (ii) \( \sum_i \Gamma_i e_i = p \) is given by

(a) \( \hat{\theta}_{ij} = \hat{\theta}_{ij}^B + \Lambda_{ij}^{-1} W_{ij} s_i^{-1} (\Omega_i + s_i^{-1})^{-1} \Gamma_i R^{-1} (p - \hat{\theta}_w^B) \)

(b) \( e_i = \hat{\theta}_w^B + (\Omega_i + s_i^{-1})^{-1} \Gamma_i R^{-1} (p - \hat{\theta}_w^B) \),

where \( s_i = \sum_j W_{ij} \Lambda_{ij}^{-1} W_{ij} \) and \( R = \sum_i \Gamma_i (\Omega_i + s_i^{-1})^{-1} \Gamma_i \).

Note that \( \Lambda_{ij} \) need not be the same as \( W_{ij} \). Similarly, \( \Omega_i \) need not be the same as \( \Gamma_i \).

Also, \( p \) is given.
Proof. By standard results, the problem reduces to minimization of
\[ g = \sum_i \sum_j (\hat{\theta}_{ij} - \hat{\theta}_{ij}^B)^T \Lambda_{ij} (\hat{\theta}_{ij} - \hat{\theta}_{ij}^B) + \sum_i (\epsilon_i - \tilde{\epsilon}_{iw}^B)^T \Omega_i (\epsilon_i - \tilde{\epsilon}_{iw}^B) \]

\[-2 \sum_i \lambda_{ii}^1 (\sum_j W_{ij} \hat{\theta}_{ij} - \epsilon_i) - 2 \lambda_2^2 (\sum_j \Gamma_i \epsilon_i - p),\]

where \( \lambda_{ii}^1 \) and \( \lambda_2^2 \) are the Lagrange multipliers. We solve
\[ 0 = \frac{\partial g}{\partial \hat{\theta}_{ij}} = 2\Lambda_{ij}(\hat{\theta}_{ij} - \hat{\theta}_{ij}^B) - 2W_{ij} \lambda_{ii}^1 \quad \forall \ i, j \quad (4.1.16) \]
\[ 0 = \frac{\partial g}{\partial \epsilon_i} = 2\Omega_i (\epsilon_i - \tilde{\epsilon}_{iw}^B) + 2 \lambda_{ii}^1 - 2 \Gamma_i \lambda_2^2 \quad \forall \ i. \quad (4.1.17) \]

From (4.1.16),
\[ \hat{\theta}_{ij} = \hat{\theta}_{ij}^B + \Lambda_{ij}^{-1} W_{ij} \lambda_{ii}^1 \implies \epsilon_i = \tilde{\epsilon}_{iw}^B + s_i \lambda_{ii}^1 \iff \lambda_{ii}^1 = s_i^{-1}(\epsilon_i - \tilde{\epsilon}_{iw}^B) \quad \forall \ i. \quad (4.1.18) \]

From (4.1.17),
\[ \epsilon_i = \tilde{\epsilon}_{iw}^B + \Omega_i^{-1} \Gamma_i \lambda_2^2 - \Omega_i^{-1} \lambda_{ii}^1 \iff \]
\[ \epsilon_i = \tilde{\epsilon}_{iw}^B + \Omega_i^{-1} \Gamma_i \lambda_2^2 - \Omega_i^{-1} s_i^{-1}(\epsilon_i - \tilde{\epsilon}_{iw}^B) \iff \]
\[ \Omega_i^{-1} \Gamma_i \lambda_2^2 = (I + \Omega_i^{-1} s_i^{-1})(\epsilon_i - \tilde{\epsilon}_{iw}^B) \implies \]
\[ \epsilon_i = \tilde{\epsilon}_{iw}^B + (\Omega_i + s_i^{-1})^{-1} \Gamma_i \lambda_2^2 \quad (4.1.19) \]

Applying constraint (ii), we find \( p = \tilde{\theta}_{i}^B + \sum_i \Gamma_i (\Omega_i + s_i^{-1})^{-1} \Gamma_i \lambda_2^2 = \tilde{\theta}_{w}^B + R \lambda_2 \), where
\( R := \sum_i \Gamma_i (\Omega_i + s_i^{-1})^{-1} \Gamma_i \). This implies that \( \lambda_2 = R^{-1}(p - \tilde{\epsilon}_{iw}^B) \). From (4.1.19),
\[ \epsilon_i = \tilde{\epsilon}_{iw}^B + (\Omega_i + s_i^{-1})^{-1} \Gamma_i R^{-1}(p - \tilde{\epsilon}_{iw}^B). \quad (4.1.20) \]

Combining (4.1.18) and (4.1.20),
\[ \hat{\theta}_{ij} = \hat{\theta}_{ij}^B + \Lambda_{ij}^{-1} W_{ij} \lambda_{ii}^1 \iff \]
\[ \hat{\theta}_{ij} = \hat{\theta}_{ij}^B + \Lambda_{ij}^{-1} W_{ij} s_i^{-1}(\epsilon_i - \tilde{\epsilon}_{iw}^B) \implies \]
\[ \hat{\theta}_{ij} = \hat{\theta}_{ij}^B + \Lambda_{ij}^{-1} W_{ij} s_i^{-1}(\Omega_i + s_i^{-1})^{-1} \Gamma_i R^{-1}(p - \tilde{\epsilon}_{iw}^B). \quad (4.1.21) \]
The result follows from (4.1.20) and (4.1.21).

4.2 An Example

This section considers small area/domain estimation of the proportion of persons without health insurance for several domains of the Asian subpopulation. Our goal is to benchmark the aggregated probabilities that a person does not have health insurance to the corresponding domain values. We also benchmark the domain estimates to match the overall population estimates.

The small domains were constructed on the basis of age, sex, race, and the region where each person lives. The National Health Interview Survey (NHIS) data provides the individual-level binary response data as well as the individual-level covariates (Ghosh, Kim, Sinha, Maiti, Katzoff and Parsons (2009)). We have information on the main response variable of interest, whether or not a person has health insurance. More information on this data is described in Ghosh, Kim, Sinha, Maiti, Katzoff and Parsons (2009). Moreover, when targeting specific subpopulations crossclassified by demographic characteristics, direct estimates are usually accompanied with large standard errors and coefficients of variation. Hence, a procedure such as the one proposed in Section 2 is appropriate.

The Asian group is made up of the following four groups: Chinese, Filipino, Asian Indian, and others such as Koreans, Vietnamese, Japanese, Hawaiian, Samoan, Guamanian, etc. Individuals from these subpopulations are assigned to specific domains depending on their age, gender, and the region they come from. There are three age groups (0–17, 18–64, and 65+). Furthermore, there are two genders, four races, and four regions that depend on the size of the metropolitan statistical area (< 499,999; 500,000–999,999; 1,000,000–2,499,999; > 2,500,000). Hence, there are \(4 \times 2 \times 3 \times 4 = 96\) total domains.

In Ghosh, Kim, Sinha, Maiti, Katzoff and Parsons (2009) a stepwise regression procedure is used for this data set to reach a final model with an intercept term and
three covariates: family size, education level, and total family income. Since we are using this data for mainly illustrative purposes, we use the same covariates. The model structure is as follows:

\[ y_{ij} \mid \theta_{ij} \sim \text{Bin}(1, \theta_{ij}) \forall i, j \]

\[ \text{logit}(\theta_{ij}) = x_{ij}^T \beta + u_i \forall i, j \]

\[ u_i \sim \text{iid } N(0, \sigma_u^2) \forall i \]

\[ \beta \sim \text{Uniform}(\mathbb{R}^p) \]

\[ \sigma_u^2 \sim \text{Gamma}(c, d). \]

Here \( y_{ij} \) is the response of the \( j \)th unit in the \( i \)th small domain (i.e., whether or not a person has health insurance) and \( \theta_{ij} \) is the probability that person \( j \) in unit \( i \) does not have health insurance.

In our analysis of the data, we first find the Bayes estimates and associated standard errors (of the small domains) using Markov chain Monte Carlo (MCMC). Let \( \hat{\theta}_{ij}^{(m)} \) denote the sampled value of \( \theta_{ij} \) from the MCMC output generated from the \( m \)th draw, where there are \( M \) total draws. The Monte Carlo estimate of \( E(\theta_{ij} \mid y) = M^{-1} \sum_{m=1}^{M} \hat{\theta}_{ij}^{(m)}. \) The Monte Carlo estimate of \( \text{Cov}(\theta_{ij}, \theta_{ij}^T | y) = M^{-1} \sum_{m=1}^{M} (\hat{\theta}_{ij}^{(m)} \hat{\theta}_{ij}^{(m)T}) - (M^{-1} \sum_{m=1}^{M} \hat{\theta}_{ij}^{(m)}) (M^{-1} \sum_{m=1}^{M} \hat{\theta}_{ij}^{(m)T}). \) Then \( E(\hat{\theta}_{iw} | y) = \sum_{j=1}^{n_i} w_{ij} E(\theta_{ij} | y) \) and \( \text{Var}(\hat{\theta}_{iw} | y) = \sum_{j=1}^{n_i} \sum_{j'=1}^{n_i} w_{ij} w_{ij'} \text{Cov}(\theta_{ij}, \theta_{ij'}). \) We will then find the two-stage benchmarked Bayes estimates using Theorem 1 and considering the weights below. Furthermore, to measure the variability associated with the two-stage benchmarked Bayes estimates at the area levels, we use the posterior mean squared error (PMSE), where

\[ \text{PMSE} \left( \hat{\theta}_{ij}^{\text{BENCH}} \right) = E \left[ (\hat{\theta}_{ij}^{\text{BENCH}} - \theta_i)^2 | \hat{\theta} \right]. \]  \hspace{1cm} (4.2.1)

Then the PMSE of \( e_i \) is given by

\[ \text{PMSE}(e_i) = (e_i - \bar{\theta}_{iw})^2 + \text{Var}(\bar{\theta}_{iw} | \hat{\theta}). \]  \hspace{1cm} (4.2.2)
We consider in our analysis the weights $w_{ij} = \frac{w_i^*}{\sum_j w_{ij}^*}$, where $w_i^*$ is the final person weight in the NHIS dataset, and $\eta_i = \frac{\sum_j w_{ij}^*}{\sum_i \sum_j w_{ij}^*}$. This leads to $p = \frac{\sum_i \sum_j w_{ij}^* y_{ij}}{\sum_i \sum_j w_{ij}^*} = \sum_i \sum_j (w_{ij} y_{ij}) \eta_i$.

We also take $\xi_{ij} = V^{-1}(\theta_{ij}|\text{data})$ and $\zeta_i = V^{-1}(\sum_j w_{ij} \theta_{ij}|\text{data})$. Simplification of the area-level estimate can be found by inserting these quantities into Theorem 1.

Table 4-1 contains the direct, hierarchical Bayes, and benchmarked Bayes estimates and associated posterior RMSEs (PRMSE) for each of the small domains. It also contains the percent increase in the PRMSE in the benchmarked Bayes estimates compared to the HB estimates. The direction of adjustment depends on the sign of $p - \bar{\eta}_w^B$, and the amount of adjustment depends on the relative magnitudes of $\eta_i s_i (1 + \zeta_i s_i)^{-1}$.

For the given dataset, the adjustments are always positive since $p > \bar{\theta}_w^B$. Also, with the present choice of weights, the percent increase in PRMSE over the HB estimators is quite small, which somewhat justifies the choice of the given HB model. Moreover, the amount of adjustment is typically more for domains with smaller sample sizes as compared to those with larger samples as we would like to see.

Consider, for example, domain 12 with a sample size of 11 and $se(\text{HB})$ equal to 0.056. Since the constraint weight $\zeta_i$ is inversely proportional to the posterior variance, we expect to see a larger adjustment in this domain, which is precisely what occurs. We also note that this domain also shows the largest percent increase in RPMSE, which is 1.616. Similar expected behavior occurs for domain 79 with sample size 122, which ends up having the smallest overall adjustment.
Table 4-1. Table of estimates using Theorem 1

<table>
<thead>
<tr>
<th>Domain</th>
<th>$n_i$</th>
<th>Direct</th>
<th>HB</th>
<th>se(HB)</th>
<th>$e_i$</th>
<th>PRMSE</th>
<th>% inc*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>0.126</td>
<td>0.135</td>
<td>0.04889</td>
<td>0.141</td>
<td>0.04927</td>
<td>0.784</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>24</td>
<td>0.063</td>
<td>0.088</td>
<td>0.03198</td>
<td>0.091</td>
<td>0.03213</td>
<td>0.467</td>
</tr>
<tr>
<td>4</td>
<td>28</td>
<td>0.146</td>
<td>0.150</td>
<td>0.04471</td>
<td>0.158</td>
<td>0.04538</td>
<td>1.514</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>0.138</td>
<td>0.125</td>
<td>0.04138</td>
<td>0.132</td>
<td>0.04193</td>
<td>1.320</td>
</tr>
<tr>
<td>6</td>
<td>17</td>
<td>0.112</td>
<td>0.130</td>
<td>0.04290</td>
<td>0.135</td>
<td>0.04318</td>
<td>0.662</td>
</tr>
<tr>
<td>7</td>
<td>78</td>
<td>0.097</td>
<td>0.108</td>
<td>0.02746</td>
<td>0.110</td>
<td>0.02757</td>
<td>0.396</td>
</tr>
<tr>
<td>8</td>
<td>66</td>
<td>0.274</td>
<td>0.214</td>
<td>0.04228</td>
<td>0.219</td>
<td>0.04265</td>
<td>0.880</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>0.173</td>
<td>0.127</td>
<td>0.04820</td>
<td>0.132</td>
<td>0.04846</td>
<td>0.542</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>0</td>
<td>0.143</td>
<td>0.05422</td>
<td>0.150</td>
<td>0.05471</td>
<td>0.908</td>
</tr>
<tr>
<td>11</td>
<td>7</td>
<td>0</td>
<td>0.140</td>
<td>0.05159</td>
<td>0.147</td>
<td>0.05208</td>
<td>0.941</td>
</tr>
<tr>
<td>12</td>
<td>11</td>
<td>0.335</td>
<td>0.164</td>
<td>0.05630</td>
<td>0.175</td>
<td>0.05721</td>
<td>1.616</td>
</tr>
<tr>
<td>13</td>
<td>7</td>
<td>0</td>
<td>0.133</td>
<td>0.04613</td>
<td>0.126</td>
<td>0.04657</td>
<td>0.969</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>0</td>
<td>0.097</td>
<td>0.04217</td>
<td>0.100</td>
<td>0.04229</td>
<td>0.288</td>
</tr>
<tr>
<td>15</td>
<td>27</td>
<td>0</td>
<td>0.075</td>
<td>0.02837</td>
<td>0.077</td>
<td>0.02846</td>
<td>0.315</td>
</tr>
<tr>
<td>16</td>
<td>29</td>
<td>0.113</td>
<td>0.129</td>
<td>0.03963</td>
<td>0.134</td>
<td>0.03992</td>
<td>0.723</td>
</tr>
<tr>
<td>17</td>
<td>27</td>
<td>0.12</td>
<td>0.125</td>
<td>0.03925</td>
<td>0.130</td>
<td>0.03955</td>
<td>0.757</td>
</tr>
<tr>
<td>18</td>
<td>14</td>
<td>0</td>
<td>0.093</td>
<td>0.03571</td>
<td>0.097</td>
<td>0.03589</td>
<td>0.492</td>
</tr>
<tr>
<td>19</td>
<td>77</td>
<td>0.131</td>
<td>0.125</td>
<td>0.02931</td>
<td>0.128</td>
<td>0.02943</td>
<td>0.426</td>
</tr>
<tr>
<td>20</td>
<td>75</td>
<td>0.223</td>
<td>0.188</td>
<td>0.03762</td>
<td>0.192</td>
<td>0.03788</td>
<td>0.697</td>
</tr>
<tr>
<td>21</td>
<td>3</td>
<td>0</td>
<td>0.100</td>
<td>0.04185</td>
<td>0.104</td>
<td>0.04204</td>
<td>0.449</td>
</tr>
<tr>
<td>22</td>
<td>6</td>
<td>0</td>
<td>0.115</td>
<td>0.04532</td>
<td>0.119</td>
<td>0.04554</td>
<td>0.489</td>
</tr>
<tr>
<td>23</td>
<td>8</td>
<td>0</td>
<td>0.153</td>
<td>0.05529</td>
<td>0.160</td>
<td>0.05579</td>
<td>0.902</td>
</tr>
<tr>
<td>24</td>
<td>9</td>
<td>0</td>
<td>0.119</td>
<td>0.04514</td>
<td>0.124</td>
<td>0.04542</td>
<td>0.625</td>
</tr>
<tr>
<td>25</td>
<td>10</td>
<td>0</td>
<td>0.094</td>
<td>0.03777</td>
<td>0.098</td>
<td>0.03799</td>
<td>0.576</td>
</tr>
<tr>
<td>26</td>
<td>6</td>
<td>0</td>
<td>0.095</td>
<td>0.04043</td>
<td>0.098</td>
<td>0.04056</td>
<td>0.329</td>
</tr>
<tr>
<td>27</td>
<td>32</td>
<td>0.098</td>
<td>0.114</td>
<td>0.03617</td>
<td>0.118</td>
<td>0.03636</td>
<td>0.547</td>
</tr>
<tr>
<td>28</td>
<td>23</td>
<td>0</td>
<td>0.083</td>
<td>0.03176</td>
<td>0.086</td>
<td>0.03193</td>
<td>0.531</td>
</tr>
<tr>
<td>29</td>
<td>25</td>
<td>0.187</td>
<td>0.122</td>
<td>0.04021</td>
<td>0.127</td>
<td>0.04050</td>
<td>0.711</td>
</tr>
<tr>
<td>30</td>
<td>23</td>
<td>0.226</td>
<td>0.145</td>
<td>0.04482</td>
<td>0.150</td>
<td>0.04509</td>
<td>0.602</td>
</tr>
<tr>
<td>31</td>
<td>71</td>
<td>0.118</td>
<td>0.118</td>
<td>0.02996</td>
<td>0.121</td>
<td>0.03004</td>
<td>0.282</td>
</tr>
<tr>
<td>32</td>
<td>50</td>
<td>0.109</td>
<td>0.109</td>
<td>0.03134</td>
<td>0.112</td>
<td>0.03152</td>
<td>0.567</td>
</tr>
<tr>
<td>33</td>
<td>2</td>
<td>0</td>
<td>0.115</td>
<td>0.04834</td>
<td>0.119</td>
<td>0.04856</td>
<td>0.439</td>
</tr>
<tr>
<td>34</td>
<td>2</td>
<td>0</td>
<td>0.119</td>
<td>0.04902</td>
<td>0.123</td>
<td>0.04911</td>
<td>0.189</td>
</tr>
<tr>
<td>35</td>
<td>8</td>
<td>0.108</td>
<td>0.117</td>
<td>0.04439</td>
<td>0.122</td>
<td>0.04471</td>
<td>0.720</td>
</tr>
<tr>
<td>36</td>
<td>7</td>
<td>0</td>
<td>0.125</td>
<td>0.04829</td>
<td>0.130</td>
<td>0.04851</td>
<td>0.456</td>
</tr>
<tr>
<td>37</td>
<td>9</td>
<td>0.062</td>
<td>0.098</td>
<td>0.03949</td>
<td>0.103</td>
<td>0.03987</td>
<td>0.966</td>
</tr>
<tr>
<td>38</td>
<td>17</td>
<td>0</td>
<td>0.073</td>
<td>0.02929</td>
<td>0.076</td>
<td>0.02937</td>
<td>0.279</td>
</tr>
<tr>
<td>39</td>
<td>24</td>
<td>0.117</td>
<td>0.136</td>
<td>0.04386</td>
<td>0.142</td>
<td>0.04423</td>
<td>0.836</td>
</tr>
<tr>
<td>40</td>
<td>20</td>
<td>0</td>
<td>0.096</td>
<td>0.03532</td>
<td>0.100</td>
<td>0.03560</td>
<td>0.816</td>
</tr>
<tr>
<td>41</td>
<td>50</td>
<td>0.163</td>
<td>0.156</td>
<td>0.03848</td>
<td>0.160</td>
<td>0.03875</td>
<td>0.703</td>
</tr>
</tbody>
</table>

* in PRMSE
<table>
<thead>
<tr>
<th>Domain</th>
<th>( n_i )</th>
<th>Direct</th>
<th>HB</th>
<th>se(HB)</th>
<th>( c_i )</th>
<th>PRMSE</th>
<th>% inc*</th>
</tr>
</thead>
<tbody>
<tr>
<td>42</td>
<td>38</td>
<td>0.141</td>
<td>0.122</td>
<td>0.03592</td>
<td>0.126</td>
<td>0.03612</td>
<td>0.558</td>
</tr>
<tr>
<td>43</td>
<td>76</td>
<td>0.104</td>
<td>0.114</td>
<td>0.02822</td>
<td>0.116</td>
<td>0.02829</td>
<td>0.251</td>
</tr>
<tr>
<td>44</td>
<td>73</td>
<td>0.142</td>
<td>0.121</td>
<td>0.02961</td>
<td>0.124</td>
<td>0.02975</td>
<td>0.468</td>
</tr>
<tr>
<td>45</td>
<td>2</td>
<td>0</td>
<td>0.129</td>
<td>0.05255</td>
<td>0.135</td>
<td>0.05284</td>
<td>0.546</td>
</tr>
<tr>
<td>46</td>
<td>3</td>
<td>0</td>
<td>0.096</td>
<td>0.04157</td>
<td>0.099</td>
<td>0.04171</td>
<td>0.345</td>
</tr>
<tr>
<td>47</td>
<td>10</td>
<td>0</td>
<td>0.098</td>
<td>0.03858</td>
<td>0.101</td>
<td>0.03872</td>
<td>0.371</td>
</tr>
<tr>
<td>48</td>
<td>7</td>
<td>0</td>
<td>0.125</td>
<td>0.04711</td>
<td>0.130</td>
<td>0.04739</td>
<td>0.596</td>
</tr>
<tr>
<td>49</td>
<td>10</td>
<td>0.087</td>
<td>0.116</td>
<td>0.04245</td>
<td>0.121</td>
<td>0.04275</td>
<td>0.692</td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>0</td>
<td>0.126</td>
<td>0.05070</td>
<td>0.131</td>
<td>0.05096</td>
<td>0.510</td>
</tr>
<tr>
<td>51</td>
<td>23</td>
<td>0.038</td>
<td>0.089</td>
<td>0.03253</td>
<td>0.092</td>
<td>0.03264</td>
<td>0.355</td>
</tr>
<tr>
<td>52</td>
<td>21</td>
<td>0.243</td>
<td>0.143</td>
<td>0.04584</td>
<td>0.149</td>
<td>0.04625</td>
<td>0.907</td>
</tr>
<tr>
<td>53</td>
<td>31</td>
<td>0.114</td>
<td>0.125</td>
<td>0.03789</td>
<td>0.129</td>
<td>0.03814</td>
<td>0.671</td>
</tr>
<tr>
<td>54</td>
<td>18</td>
<td>0.202</td>
<td>0.158</td>
<td>0.04941</td>
<td>0.164</td>
<td>0.04971</td>
<td>0.601</td>
</tr>
<tr>
<td>55</td>
<td>74</td>
<td>0.094</td>
<td>0.094</td>
<td>0.02554</td>
<td>0.096</td>
<td>0.02562</td>
<td>0.312</td>
</tr>
<tr>
<td>56</td>
<td>83</td>
<td>0.204</td>
<td>0.186</td>
<td>0.03699</td>
<td>0.191</td>
<td>0.03729</td>
<td>0.814</td>
</tr>
<tr>
<td>57</td>
<td>2</td>
<td>0</td>
<td>0.140</td>
<td>0.05667</td>
<td>0.145</td>
<td>0.05688</td>
<td>0.363</td>
</tr>
<tr>
<td>58</td>
<td>1</td>
<td>0</td>
<td>0.093</td>
<td>0.04170</td>
<td>0.096</td>
<td>0.04177</td>
<td>0.163</td>
</tr>
<tr>
<td>59</td>
<td>2</td>
<td>0</td>
<td>0.094</td>
<td>0.04115</td>
<td>0.097</td>
<td>0.04125</td>
<td>0.232</td>
</tr>
<tr>
<td>60</td>
<td>8</td>
<td>0.112</td>
<td>0.132</td>
<td>0.04798</td>
<td>0.138</td>
<td>0.04831</td>
<td>0.689</td>
</tr>
<tr>
<td>61</td>
<td>16</td>
<td>0.202</td>
<td>0.156</td>
<td>0.05206</td>
<td>0.163</td>
<td>0.05260</td>
<td>1.040</td>
</tr>
<tr>
<td>62</td>
<td>3</td>
<td>0.301</td>
<td>0.158</td>
<td>0.06362</td>
<td>0.165</td>
<td>0.06395</td>
<td>0.516</td>
</tr>
<tr>
<td>63</td>
<td>33</td>
<td>0.055</td>
<td>0.093</td>
<td>0.03118</td>
<td>0.096</td>
<td>0.03131</td>
<td>0.421</td>
</tr>
<tr>
<td>64</td>
<td>28</td>
<td>0.105</td>
<td>0.118</td>
<td>0.03743</td>
<td>0.122</td>
<td>0.03764</td>
<td>0.560</td>
</tr>
<tr>
<td>65</td>
<td>33</td>
<td>0.126</td>
<td>0.133</td>
<td>0.03899</td>
<td>0.138</td>
<td>0.03928</td>
<td>0.731</td>
</tr>
<tr>
<td>66</td>
<td>13</td>
<td>0.393</td>
<td>0.207</td>
<td>0.06393</td>
<td>0.216</td>
<td>0.06460</td>
<td>1.038</td>
</tr>
<tr>
<td>67</td>
<td>70</td>
<td>0.079</td>
<td>0.096</td>
<td>0.02635</td>
<td>0.098</td>
<td>0.02642</td>
<td>0.283</td>
</tr>
<tr>
<td>68</td>
<td>75</td>
<td>0.179</td>
<td>0.159</td>
<td>0.03425</td>
<td>0.163</td>
<td>0.03444</td>
<td>0.537</td>
</tr>
<tr>
<td>69</td>
<td>1</td>
<td>1</td>
<td>0.146</td>
<td>0.06075</td>
<td>0.149</td>
<td>0.06081</td>
<td>0.107</td>
</tr>
<tr>
<td>70</td>
<td>2</td>
<td>0.361</td>
<td>0.157</td>
<td>0.05946</td>
<td>0.162</td>
<td>0.05968</td>
<td>0.372</td>
</tr>
<tr>
<td>71</td>
<td>4</td>
<td>0</td>
<td>0.104</td>
<td>0.04291</td>
<td>0.108</td>
<td>0.04307</td>
<td>0.384</td>
</tr>
<tr>
<td>72</td>
<td>2</td>
<td>0</td>
<td>0.226</td>
<td>0.08082</td>
<td>0.235</td>
<td>0.08134</td>
<td>0.643</td>
</tr>
<tr>
<td>73</td>
<td>45</td>
<td>0.271</td>
<td>0.208</td>
<td>0.04578</td>
<td>0.215</td>
<td>0.04632</td>
<td>1.189</td>
</tr>
<tr>
<td>74</td>
<td>10</td>
<td>0</td>
<td>0.096</td>
<td>0.03854</td>
<td>0.100</td>
<td>0.03873</td>
<td>0.485</td>
</tr>
<tr>
<td>75</td>
<td>83</td>
<td>0.149</td>
<td>0.134</td>
<td>0.03002</td>
<td>0.137</td>
<td>0.03017</td>
<td>0.489</td>
</tr>
<tr>
<td>76</td>
<td>59</td>
<td>0.113</td>
<td>0.126</td>
<td>0.03247</td>
<td>0.129</td>
<td>0.03263</td>
<td>0.505</td>
</tr>
<tr>
<td>77</td>
<td>68</td>
<td>0.338</td>
<td>0.270</td>
<td>0.04638</td>
<td>0.276</td>
<td>0.04685</td>
<td>1.024</td>
</tr>
<tr>
<td>78</td>
<td>39</td>
<td>0.098</td>
<td>0.107</td>
<td>0.03331</td>
<td>0.111</td>
<td>0.03346</td>
<td>0.452</td>
</tr>
<tr>
<td>79</td>
<td>122</td>
<td>0.11</td>
<td>0.106</td>
<td>0.02348</td>
<td>0.108</td>
<td>0.02355</td>
<td>0.285</td>
</tr>
<tr>
<td>80</td>
<td>125</td>
<td>0.308</td>
<td>0.293</td>
<td>0.03805</td>
<td>0.298</td>
<td>0.0383</td>
<td>0.651</td>
</tr>
<tr>
<td>81</td>
<td>7</td>
<td>0</td>
<td>0.122</td>
<td>0.04681</td>
<td>0.126</td>
<td>0.04704</td>
<td>0.484</td>
</tr>
<tr>
<td>82</td>
<td>12</td>
<td>0</td>
<td>0.097</td>
<td>0.03752</td>
<td>0.101</td>
<td>0.03771</td>
<td>0.498</td>
</tr>
<tr>
<td>83</td>
<td>13</td>
<td>0.049</td>
<td>0.122</td>
<td>0.04428</td>
<td>0.128</td>
<td>0.04458</td>
<td>0.678</td>
</tr>
</tbody>
</table>

* in PRMSE
<table>
<thead>
<tr>
<th>Domain</th>
<th>( n_i )</th>
<th>Direct</th>
<th>HB</th>
<th>se(HB)</th>
<th>( c_i )</th>
<th>PRMSE</th>
<th>% inc*</th>
</tr>
</thead>
<tbody>
<tr>
<td>84</td>
<td>4</td>
<td>0</td>
<td>0.126</td>
<td>0.04915</td>
<td>0.131</td>
<td>0.04940</td>
<td>0.506</td>
</tr>
<tr>
<td>85</td>
<td>32</td>
<td>0.189</td>
<td>0.172</td>
<td>0.04641</td>
<td>0.178</td>
<td>0.04674</td>
<td>0.699</td>
</tr>
<tr>
<td>86</td>
<td>10</td>
<td>0.135</td>
<td>0.132</td>
<td>0.04844</td>
<td>0.137</td>
<td>0.04878</td>
<td>0.700</td>
</tr>
<tr>
<td>87</td>
<td>52</td>
<td>0.192</td>
<td>0.157</td>
<td>0.03765</td>
<td>0.162</td>
<td>0.03800</td>
<td>0.925</td>
</tr>
<tr>
<td>88</td>
<td>65</td>
<td>0.153</td>
<td>0.140</td>
<td>0.03346</td>
<td>0.143</td>
<td>0.03366</td>
<td>0.584</td>
</tr>
<tr>
<td>89</td>
<td>71</td>
<td>0.285</td>
<td>0.226</td>
<td>0.04132</td>
<td>0.231</td>
<td>0.04164</td>
<td>0.763</td>
</tr>
<tr>
<td>90</td>
<td>57</td>
<td>0.086</td>
<td>0.103</td>
<td>0.02965</td>
<td>0.106</td>
<td>0.02979</td>
<td>0.446</td>
</tr>
<tr>
<td>91</td>
<td>153</td>
<td>0.149</td>
<td>0.135</td>
<td>0.02387</td>
<td>0.136</td>
<td>0.02394</td>
<td>0.297</td>
</tr>
<tr>
<td>92</td>
<td>138</td>
<td>0.308</td>
<td>0.277</td>
<td>0.03542</td>
<td>0.281</td>
<td>0.03563</td>
<td>0.593</td>
</tr>
<tr>
<td>93</td>
<td>10</td>
<td>0</td>
<td>0.121</td>
<td>0.04513</td>
<td>0.126</td>
<td>0.04547</td>
<td>0.761</td>
</tr>
<tr>
<td>94</td>
<td>16</td>
<td>0.067</td>
<td>0.107</td>
<td>0.03885</td>
<td>0.110</td>
<td>0.03901</td>
<td>0.404</td>
</tr>
<tr>
<td>95</td>
<td>18</td>
<td>0.108</td>
<td>0.144</td>
<td>0.04668</td>
<td>0.150</td>
<td>0.04707</td>
<td>0.816</td>
</tr>
<tr>
<td>96</td>
<td>14</td>
<td>0.111</td>
<td>0.138</td>
<td>0.04731</td>
<td>0.143</td>
<td>0.04759</td>
<td>0.595</td>
</tr>
</tbody>
</table>

* in PRMSE
CHAPTER 5
ON ESTIMATION OF MSES OF BENCHMARKED EB ESTIMATORS

Small area estimation has become increasingly popular recently due to a growing demand for such statistics. It is well known that direct small-area estimators usually have large standard errors and coefficients of variation. In order to produce estimates for these small areas, it is necessary to borrow strength from other related areas. Accordingly, model-based estimates often differ widely from the direct estimates, especially for areas with small sample sizes. One problem that arises in practice is that the model-based estimates do not aggregate to the more reliable direct survey estimates. Agreement with the direct estimates is often a political necessity to convince legislators regarding the utility of small area estimates. The process of adjusting model-based estimates to correct this problem is known as benchmarking. Another key benefit of benchmarking is protection against model misspecification as pointed out by You, Rao and Dick (2004) and Datta, Ghosh, Steorts and Maples (2011).

In recent years, the literature on benchmarking has grown in small area estimation. Among others, Pfeffermann and Barnard (1991); You and Rao (2003); You, Rao and Dick (2004); and Pfeffermann and Tiller (2006) have made an impact on the continuing development of this field. Specifically, Wang, Fuller and Qu (2008) provided a frequentist method wherein an augmented model was used to construct a best linear unbiased predictor (BLUP) that automatically satisfies the benchmarking constraint. In addition, Datta, Ghosh, Steorts and Maples (2011) developed very general benchmarked Bayes estimators, which covered most of the earlier estimators that have been motivated from either a frequentist or Bayesian perspective. In particular, they found benchmarked Bayes estimators under the celebrated Fay and Herriot (1979) model.

Due to the fact that they borrow strength, model-based estimates typically show a substantial improvement over direct estimates in terms of mean squared error (MSE). It
is of particular interest to determine how much of this advantage is lost by constraining the estimates through benchmarking. The aforementioned work of Wang, Fuller and Qu (2008) examined this question through simulation studies but did not derive any probabilistic results. They showed that the MSE of the benchmarked EB estimator was slightly larger than the MSE of the EB estimator for their simulation studies. In Section 5.2 we derive a second-order approximation of the MSE of the benchmarked Bayes EB estimator to show that the increase due to benchmarking is $O(m^{-1})$, where $m$ is the number of small areas.

In this paper, we are concerned with the basic area-level model of Fay and Herriot (1979). We obtain benchmarked empirical Bayes estimators, and these estimators are proposed in Section 5.1. In Section 5.2, we derive a second-order asymptotic expansion of the MSE of the benchmarked EB estimator. In Section 5.3, we then find an estimator of this MSE and compare it to the second-order approximation of the MSE of the EB estimator or equivalently the MSE of the EBLUP, which was derived by Prasad and Rao (1990). Finally, in Section 5.4, using methods similar to those of Butar and Lahiri (2003), we compute a parametric bootstrap estimate of the mean squared error of the benchmarked EB estimate under the Fay-Herriot (1979) model and compare it to our estimates from Section 5.1. Section 5.5 contains an application based on Small Area Income and Poverty Estimation Data (SAIPE) from the U.S. Census Bureau.

### 5.1 Benchmarked EB Estimators

Consider the area-level random effects model

$$
\hat{\theta}_i = \theta_i + e_i, \quad \theta_i = x_i^T \beta + u_i; \quad i = 1, \ldots, m;
$$

(5.1.1)

where $e_i$ and $u_i$ are mutually independent with $e_i \sim_{iid} N(0, D_i)$ and $u_i \sim_{iid} N(0, \sigma^2_u)$. This model was first considered in the context of estimating income for small areas (population less than 1000) by Fay and Herriot (1979). In model 5.1.1, $D_i$ are known as
are the $p \times 1$ design vectors $x_i$. However, the vector of regression coefficients $\beta_{p \times 1}$ is unknown.

When the variance component $\sigma^2_u$ is known and $\beta$ has a uniform prior on $\mathbb{R}^p$, then the Bayes estimator of $\theta_i$ is given by $\hat{\theta}_i^B = (1 - B_i)\hat{\theta}_i + B_i x_i^T \tilde{\beta}$ where $B_i = D_i(\sigma^2_u + D_i)^{-1}$, $\tilde{\beta} \equiv \tilde{\beta}(\sigma^2_u) = (X^T V^{-1}X)^{-1}X^T V^{-1}\tilde{\theta}$, and $V = \text{Diag}(\sigma^2_u + D_1, \ldots, \sigma^2_u + D_m)$. Suppose now we want to match the weighted average of some estimates $e_i$ to the weighted average of the direct estimates, which we denote by $t$. We will assume for our calculations that $t = \bar{\hat{\theta}}_w$. We denote the normalized weights by $w_i$, so that $\sum w_i = 1$. Under the loss $L(\theta, e) = \sum w_i(\theta - e)^2$ and subject to $\sum w_i e_i = \sum w_i \hat{\theta}_i$, the benchmarked Bayes estimator as derived in Datta, Ghosh, Steorts and Maples (2011) is given by

$$\hat{\theta}_i^{BM1} = \hat{\theta}_i^B + (\bar{\hat{\theta}}_w - \bar{\hat{\theta}}_w^B); \quad i = 1, \ldots, m. \quad (5.1.2)$$

In more realistic settings, $\sigma^2_u$ is unknown. Define $P_X = X(X^T X)^{-1}X^T$, $h_i = x_i^T (X^T X)^{-1}x_i$, $\hat{u}_i = \hat{\theta}_i - x_i^T \hat{\beta}$, and $\tilde{\beta} = (X^T X)^{-1}X^T \hat{\theta}$. In this paper, we consider the simple moment estimator given by $\hat{\sigma}^2_u = \max\{0, \hat{\sigma}^2_u\}$ where $\hat{\sigma}^2_u = (m - p)^{-1} \left[ \sum_{i=1}^m \hat{u}_i^2 - \sum_{i=1}^m D_i(1 - h_i) \right]$, which is given in Prasad and Rao (1990). Then the empirical benchmarked Bayes estimator of $\theta_i$ is given by

$$\hat{\theta}_i^{EBM1} = \hat{\theta}_i^{EB} + (\bar{\hat{\theta}}_w - \bar{\hat{\theta}}_w^{EB}), \quad (5.1.3)$$

where $\hat{\theta}_i^{EB} = (1 - B_i)\hat{\theta}_i + B_i x_i^T \tilde{\beta}(\hat{\sigma}^2_u); \quad \tilde{\beta} = D_i(\hat{\sigma}^2_u + D_i)^{-1}; \quad i = 1, \ldots, m$. The objective of the next two sections will be to obtain the MSE of the benchmarked empirical Bayes estimator correct up to $O(m^{-1})$ and also to find an estimator of the MSE correct to the same order.

### 5.2 Second-Order Approximation to MSE

Wang et al. (2008) construct a simulation study to compare the MSE of the benchmarked EB estimator to the MSE of the EB estimator. In this section, we derive a second order expansion for the MSE of the benchmarked Bayes estimator under the same
regularity conditions and assuming the standard benchmarking constraint. That is, assuming the model proposed in Section 5.1, we obtain a second-order approximation to the MSE of the empirical benchmarked Bayes estimator. Define $h_{ij}^V = x_i^T (X^TV^{-1}X)^{-1}x_j$ and assume that $\sigma_u^2 > 0$. The following regularity conditions are necessary for establishing Theorem 7:

(i) $0 < D_L \leq \inf_{1 \leq i \leq m} D_i \leq \sup_{1 \leq i \leq m} D_i \leq D_U < \infty$;

(ii) $\max_{1 \leq i \leq m} h_{ii} = O(m^{-1})$; and

(iii) $\max_{1 \leq i \leq m} w_i = O(m^{-1})$.

Condition (iii) requires a kind of homogeneity of the small areas, in particular, that there do not exist a few large areas which dominate the remaining small areas in terms of the $w_i$. Conditions (i) and (ii) are similar to those of Prasad and Rao (1990) and are very often assumed in the small area estimation literature. The resulting approximation is given in Theorem 7.

**Theorem 7.** Assume regularity conditions (i)–(iii) hold. Then $E[(\hat{\theta}^EBM1_i - \theta_i)^2] = g_1(\sigma_u^2) + g_2(\sigma_u^2) + g_3(\sigma_u^2) + g_4(\sigma_u^2) + o(m^{-1})$, where

$$
\begin{align*}
    g_1(\sigma_u^2) &= \sum_{i=1}^{m} \sigma_u^2 B_i
    \\
    g_2(\sigma_u^2) &= \sum_{i=1}^{m} h_{ii}^V
    \\
    g_3(\sigma_u^2) &= \sum_{i=1}^{m} D_i^{-1} \text{Var}(\tilde{\sigma}_u^2)
    \\
    g_4(\sigma_u^2) &= \sum_{i=1}^{m} w_i^2 B_i^2 V_i - \sum_{i=1}^{m} \sum_{j=1}^{m} w_i w_j B_i B_j h_{ij}^V,
\end{align*}
$$

and where $\text{Var}(\tilde{\sigma}_u^2) = 2(m - p)^{-2} \sum_{k=1}^{m} (\sigma_u^2 + D_k)^2 + o(m^{-1})$. 

58
Proof. Observe

\[ E[(\hat{\theta}_i^{EBM1} - \theta_i)^2] = E[(\hat{\theta}_i^B - \theta_i)^2] + E[(\hat{\theta}_i^{EBM1} - \hat{\theta}_i^B)^2] \]
\[ = E[(\hat{\theta}_i^B - \theta_i)^2] + E[(\hat{\theta}_i^B - \hat{\theta}_i^B + \bar{t}^{EB}_w - \bar{t}^{EB}_w - t)^2] \]
\[ = E[(\hat{\theta}_i^B - \theta_i)^2] + E[(\hat{\theta}_i^{EB} - \hat{\theta}_i^B)^2] + E[(\bar{t}^{EB}_w - t)^2] \]
\[ - 2E[(\hat{\theta}_i^{EB} - \hat{\theta}_i^B)(\bar{t}^{EB}_w - \bar{t}^B_w)] - 2E[(\hat{\theta}_i^{EB} - \hat{\theta}_i^B)(\bar{t}^B_w - t)] \]
\[ + 2E[(\bar{t}^{EB}_w - \bar{t}^B_w)(\bar{t}^B_w - t)]. \quad (5.2.1) \]

Next, we observe that \( E[(\hat{\theta}_i^B - \theta_i)^2] + E[(\hat{\theta}_i^{EB} - \hat{\theta}_i^B)^2] = g_{1i}(\sigma_u^2) + g_{2i}(\sigma_u^2) + g_{3i}(\sigma_u^2) + o(m^{-1}) \)
by Prasad and Rao (1990), where

\[ g_{1i}(\sigma_u^2) = B_i \sigma_u^2 \]
\[ g_{2i}(\sigma_u^2) = B_i^2 h_{ii}^\gamma \]
\[ g_{3i}(\sigma_u^2) = B_i^3 D_i^{-1}\text{Var}(\hat{\theta}_i^B). \]

It may be noted that while \( g_{1i}(\sigma_u^2) = O(1) \), both \( g_{2i}(\sigma_u^2) \) and \( g_{3i}(\sigma_u^2) \) are of order \( O(m^{-1}) \)
as shown in Prasad and Rao (1990). We will show that \( E[(\bar{t}^{EB}_w - t)^2] = g_4(\sigma_u^2) = O(m^{-1}) \),
whereas the remaining four terms of expression (5.2.1) are of order \( o(m^{-1}) \).

First we show that \( E[(\bar{t}^{EB}_w - t)^2] = g_4(\sigma_u^2) \). We observe \( \bar{t}^{EB}_w - t = -\sum_{i=1}^{m} w_i B_i (\hat{\theta}_i - x_i^T \bar{\beta}) \)
and consider

\[ E[(\bar{t}^{EB}_w - t)^2] = E \left[ \left( \sum_{i=1}^{m} w_i B_i (\hat{\theta}_i - x_i^T \bar{\beta}) \right)^2 \right] \]
\[ = \sum_{i=1}^{m} w_i^2 B_i^2 E[(\hat{\theta}_i - x_i^T \bar{\beta})^2] + \sum w_i w_j B_i B_j E[(\hat{\theta}_i - x_i^T \bar{\beta})(\hat{\theta}_j - x_j^T \bar{\beta})] \]
\[ = \sum_{i=1}^{m} w_i^2 B_i^2 (V_i - h_{ii}^\gamma) + \sum_{i \neq j} w_i w_j B_i B_j (-h_{ij}^\gamma) \]
\[ = \sum_{i=1}^{m} w_i^2 B_i^2 V_i - \sum_{i=1}^{m} \sum_{j=1}^{m} w_i w_j B_i B_j h_{ij}^\gamma. \quad (5.2.2) \]
We may note that the expression on the right hand side of (5.2.2) is $O(m^{-1})$ since $\max_{1 \leq i \leq m} h_{ii} = O(m^{-1})$, which implies that $\max_{1 \leq i \leq j \leq m} h_{ij}^V = O(m^{-1})$.

Next, we return to (5.2.1) and show that $E[(\hat{\theta}_{w}^{EB} - \hat{\theta}_{w}^{B})^2] = o(m^{-1})$. Consider

$$E[(\hat{\theta}_{w}^{EB} - \hat{\theta}_{w}^{B})^2] = \sum_{i} w_i^2 E\left[(\hat{\theta}_{i}^{EB} - \hat{\theta}_{i}^{B})^2\right] + 2 \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} w_i w_j E\left[(\hat{\theta}_{i}^{EB} - \hat{\theta}_{i}^{B})(\hat{\theta}_{j}^{EB} - \hat{\theta}_{j}^{B})\right]$$

$$= 2 \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} w_i w_j E\left[(\hat{\theta}_{i}^{EB} - \hat{\theta}_{i}^{B})(\hat{\theta}_{j}^{EB} - \hat{\theta}_{j}^{B})\right] + o(m^{-1})$$

(5.2.3)

since $\sum_{i} w_i^2 E[(\hat{\theta}_{i}^{EB} - \hat{\theta}_{i}^{B})^2] = o(m^{-1})$. The latter holds because $E[(\hat{\theta}_{i}^{EB} - \hat{\theta}_{i}^{B})^2] = g_2(\sigma_u^2) + g_3(\sigma_u^2) = O(m^{-1})$, $\max_{1 \leq i \leq m} w_i = O(m^{-1})$, and $\sum_{i} w_i = 1$. Thus, it suffices to show $E\left[(\hat{\theta}_{i}^{EB} - \hat{\theta}_{i}^{B})(\hat{\theta}_{j}^{EB} - \hat{\theta}_{j}^{B})\right] = o(m^{-1})$ for all $i \neq j$, and we do so by expanding $\hat{\theta}_{i}^{EB}$ about $\hat{\theta}_{i}^{B}$. For simplicity of notation, denote $\frac{\partial \hat{\theta}_{i}^{B}}{\partial \sigma_u^2} = \frac{\partial \hat{\theta}_{i}^{B}(\sigma_u^2)}{\partial \sigma_u^2}$ and $\frac{\partial^2 \hat{\theta}_{i}^{B}}{\partial (\sigma_u^2)^2} = \frac{\partial^2 \hat{\theta}_{i}^{B}(\sigma_u^2)}{\partial (\sigma_u^2)^2}$.

This results in

$$\hat{\theta}_{i}^{EB} - \hat{\theta}_{i}^{B} = \frac{\partial \hat{\theta}_{i}^{B}}{\partial \sigma_u^2} (\hat{\sigma}_u^2 - \sigma_u^2) + \frac{1}{2} \frac{\partial^2 \hat{\theta}_{i}^{B}}{\partial (\sigma_u^2)^2} (\hat{\sigma}_u^2 - \sigma_u^2)^2$$

for some $\sigma_u^2$ between $\sigma_u^2$ and $\hat{\sigma}_u^2$. The expansion of $\hat{\theta}_{j}^{EB}$ about $\hat{\theta}_{j}^{B}$ is similar.

We now consider $E[(\hat{\theta}_{i}^{EB} - \hat{\theta}_{i}^{B})(\hat{\theta}_{j}^{EB} - \hat{\theta}_{j}^{B})]$ for $i \neq j$. Notice that

$$E[(\hat{\theta}_{i}^{EB} - \hat{\theta}_{i}^{B})(\hat{\theta}_{j}^{EB} - \hat{\theta}_{j}^{B})] = E\left[\frac{\partial \hat{\theta}_{i}^{B}}{\partial \sigma_u^2} \frac{\partial \hat{\theta}_{j}^{B}}{\partial \sigma_u^2} (\hat{\sigma}_u^2 - \sigma_u^2)\right] + \frac{1}{2} E\left[\frac{\partial \hat{\theta}_{i}^{B}}{\partial \sigma_u^2} \frac{\partial^2 \hat{\theta}_{j}^{B}}{\partial (\sigma_u^2)^2} (\hat{\sigma}_u^2 - \sigma_u^2)\right]$$

$$+ \frac{1}{2} E\left[\frac{\partial^2 \hat{\theta}_{i}^{B}}{\partial (\sigma_u^2)^2} \frac{\partial \hat{\theta}_{j}^{B}}{\partial \sigma_u^2} (\hat{\sigma}_u^2 - \sigma_u^2)\right] + \frac{1}{4} E\left[\frac{\partial^2 \hat{\theta}_{i}^{B}}{\partial (\sigma_u^2)^2} \frac{\partial^2 \hat{\theta}_{j}^{B}}{\partial (\sigma_u^2)^2} (\hat{\sigma}_u^2 - \sigma_u^2)^2\right]$$

$$:= R_0 + R_1 + R_2 + R_3.$$

In $R_1$, consider that

$$E\left[\frac{\partial \hat{\theta}_{i}^{B}}{\partial \sigma_u^2} \frac{\partial^2 \hat{\theta}_{j}^{B}}{\partial (\sigma_u^2)^2} (\hat{\sigma}_u^2 - \sigma_u^2)\right] = E\left[\frac{\partial \hat{\theta}_{i}^{B}}{\partial \sigma_u^2} \frac{\partial \hat{\theta}_{j}^{B}}{\partial \sigma_u^2} (\hat{\sigma}_u^2 - \sigma_u^2)^3 I(\hat{\sigma}_u^2 > 0)\right]$$

$$- E\left[\frac{\partial \hat{\theta}_{i}^{B}}{\partial \sigma_u^2} \frac{\partial \hat{\theta}_{j}^{B}}{\partial \sigma_u^2} (\sigma_u^2)^3 I(\hat{\sigma}_u^2 \leq 0)\right].$$

(5.2.4)
Observe that
\[
E \left[ \frac{\partial \hat{\theta}_i^B}{\partial \sigma_u^2} \frac{\partial^2 \hat{\theta}_j^B}{\partial^2 \sigma_u^2} (\sigma_u^2 - \sigma_u^2)^3 \right] \leq \sigma_u^6 E^{1/2} \left[ \left( \frac{\partial \hat{\theta}_i^B}{\partial \sigma_u^2} \right)^4 \right] E^{1/2} \left[ \sup_{\sigma_u^2 \geq 0} \left( \frac{\partial^2 \hat{\theta}_j^B}{\partial (\sigma_u^2)^2} \right)^4 \right] P^{1/2} (\sigma_u^2 \leq 0) \\
\leq \sigma_u^6 E^{1/2} \left[ \left( \frac{\partial \hat{\theta}_i^B}{\partial \sigma_u^2} \right)^4 \right] E^{1/2} \left[ \sup_{\sigma_u^2 \geq 0} \left( \frac{\partial^2 \hat{\theta}_j^B}{\partial (\sigma_u^2)^2} \right)^4 \right] P^{1/2} (\sigma_u^2 \leq 0) \\
= o(m^{-r})
\]

for all \( r > 0 \) by Lemma 2 (ii) and 3, which we have proved in Appendix A, and since \( P(\sigma_u^2 \leq 0) = O(m^{-r}) \) \( \forall r > 0 \) as proved in Lemma A.6 of Prasad and Rao (1990). We next consider
\[
E \left[ \frac{\partial \hat{\theta}_i^B}{\partial \sigma_u^2} \frac{\partial^2 \hat{\theta}_j^B}{\partial^2 \sigma_u^2} (\sigma_u^2 - \sigma_u^2)^3 \right] = E \left[ \frac{\partial \hat{\theta}_i^B}{\partial \sigma_u^2} \frac{\partial^2 \hat{\theta}_j^B}{\partial^2 \sigma_u^2} (\sigma_u^2 - \sigma_u^2)^3 \right] \\
- E \left[ \frac{\partial \hat{\theta}_i^B}{\partial \sigma_u^2} \frac{\partial^2 \hat{\theta}_j^B}{\partial^2 \sigma_u^2} (\sigma_u^2 - \sigma_u^2)^3 \right] = O(m^{-r}) \text{ since } P(\sigma_u^2 \leq 0) = O(m^{-r}) \text{ } \forall r > 0.
\]
Continuing along, we next observe that
\[
E \left[ \frac{\partial \hat{\theta}_i^B}{\partial \sigma_u^2} \frac{\partial^2 \hat{\theta}_j^B}{\partial^2 \sigma_u^2} (\sigma_u^2 - \sigma_u^2)^3 \right] \leq E^{1/2} \left[ \left( \frac{\partial \hat{\theta}_i^B}{\partial \sigma_u^2} \right)^4 \right] E^{1/2} \left[ \sup_{\sigma_u^2 \geq 0} \left( \frac{\partial^2 \hat{\theta}_j^B}{\partial (\sigma_u^2)^2} \right)^4 \right] E^{1/2} \left[ (\sigma_u^2 - \sigma_u^2)^6 \right] \\
\leq E^{1/2} \left[ \left( \frac{\partial \hat{\theta}_i^B}{\partial \sigma_u^2} \right)^4 \right] E^{1/2} \left[ \sup_{\sigma_u^2 \geq 0} \left( \frac{\partial^2 \hat{\theta}_j^B}{\partial (\sigma_u^2)^2} \right)^4 \right] E^{1/2} \left[ (\sigma_u^2 - \sigma_u^2)^6 \right] = O(m^{-3/2})
\]
since \( E[(\sigma_u^2 - \sigma_u^2)^2] = O(m^{-r}) \) for any \( r \geq 1 \) by Lemma A.5 in Prasad and Rao (1990).

This proves that \( R_1 = o(m^{-1}) \) since \( \max_{1 \leq i \leq m} w_i = O(m^{-1}) \). By symmetry, \( R_2 \) is also \( o(m^{-1}) \). Finally, we show that \( R_3 \) is \( o(m^{-1}) \). Using a similar calculation involving \( R_1 \), we can show that
\[
E \left[ \frac{\partial^2 \hat{\theta}_i^B}{\partial \sigma_u^2} \frac{\partial^2 \hat{\theta}_j^B}{\partial^2 \sigma_u^2} (\sigma_u^2 - \sigma_u^2)^4 \right] = E \left[ \frac{\partial^2 \hat{\theta}_i^B}{\partial \sigma_u^2} \frac{\partial^2 \hat{\theta}_j^B}{\partial^2 \sigma_u^2} (\sigma_u^2 - \sigma_u^2)^4 \right] + o(m^{-r}).
\]
Observe now that
\[
E \left[ \frac{\partial^2 \hat{B}_{i*}}{(\partial \sigma_u^2)^2} \frac{\partial^2 \hat{B}_{j*}}{(\partial \sigma_u^2)^2} (\hat{\sigma}_u^2 - \sigma_u^2)^4 \right] \leq E^{\frac{1}{2}} \left[ \left\{ \frac{\partial^2 \hat{B}_{i*}}{(\partial \sigma_u^2)^2} \right\}^4 \right] \cdot E^{\frac{1}{2}} \left[ \left\{ \frac{\partial^2 \hat{B}_{j*}}{(\partial \sigma_u^2)^2} \right\}^4 \right] \cdot E^{\frac{1}{2}} \left[ (\hat{\sigma}_u^2 - \sigma_u^2)^8 \right]
\]
\[
\leq E^{\frac{1}{2}} \sup_{\hat{\sigma}_u^2 \geq 0} \left\{ \frac{\partial^2 \hat{B}_{i*}}{(\partial \sigma_u^2)^2} \right\}^4 \cdot E^{\frac{1}{2}} \sup_{\hat{\sigma}_u^2 \geq 0} \left\{ \frac{\partial^2 \hat{B}_{j*}}{(\partial \sigma_u^2)^2} \right\}^4 \cdot E^{\frac{1}{2}} \left[ (\hat{\sigma}_u^2 - \sigma_u^2)^8 \right]
\]
\[
= O(m^{-2}).
\]
Plugging this back into the expression in (5.2.6), we find that
\[
E \left[ \frac{\partial^2 \hat{B}_{i*}}{(\partial \sigma_u^2)^2} \frac{\partial^2 \hat{B}_{j*}}{(\partial \sigma_u^2)^2} (\hat{\sigma}_u^2 - \sigma_u^2)^4 \right] = o(m^{-1}).
\]
Hence, $R_3$ is $o(m^{-1})$. Finally, by calculations similar to those used for expression (5.2.4), we find that
\[
R_0 = E \left[ \frac{\partial^2 \hat{B}_{i*}}{\partial \sigma_u^2} \frac{\partial^2 \hat{B}_{j*}}{\partial \sigma_u^2} (\hat{\sigma}_u^2 - \sigma_u^2)^2 \right] = E \left[ \frac{\partial^2 \hat{B}_{i*}}{\partial \sigma_u^2} \frac{\partial^2 \hat{B}_{j*}}{\partial \sigma_u^2} (\hat{\sigma}_u^2 - \sigma_u^2)^2 \right] + o(m^{-1}).
\]
Define $\Sigma = V - X(X^T V^{-1}X)^{-1}X^T = (I - P_X^V)V$, where $P_X = X(X^T V^{-1}X)^{-1}X^T$. Also, define $P_X^V = X(X^T V^{-1}X)^{-1}X^T V^{-1}$ and let $e_i$ represent the $i$th unit vector. We can show
\[
\frac{\partial^2 \hat{B}}{\partial \sigma_u^2} = B_ie_i^T \Sigma V^{-2} \hat{u},
\]
where $\hat{u} = \hat{\theta} - X\hat{\beta}$. Define $A_{ij} = B_i B_j V^{-2} \Sigma e_i e_j^T \Sigma V^{-2}$ and consider
\[
E \left[ \frac{\partial^2 \hat{B}_{i*}}{\partial \sigma_u^2} \frac{\partial^2 \hat{B}_{j*}}{\partial \sigma_u^2} (\hat{\sigma}_u^2 - \sigma_u^2)^2 \right] = E[\hat{u}^T A_{ij} \hat{u} (\hat{\sigma}_u^2 - \sigma_u^2)^2] = \text{Cov}(\hat{u}^T A_{ij} \hat{u}, (\hat{\sigma}_u^2 - \sigma_u^2)^2) + E[\hat{u}^T A_{ij} \hat{u}] E[(\hat{\sigma}_u^2 - \sigma_u^2)^2].
Using Lemma 4 and the relation \((I - P_X)\Sigma = (I - P_X)V\),

\[
\text{Cov}(\tilde{u}^T A_j \tilde{u}, (\tilde{\sigma}_u^2 - \sigma_u^2)^2) = (m - p)^{-2} \text{Cov}(\tilde{u}^T A_j \tilde{u}, [\tilde{u}^T (I - P_X)\tilde{u} - \text{tr}((I - P_X)V)]^2) \\
= (m - p)^{-2} \text{Cov}(\tilde{u}^T A_j \tilde{u}, [\tilde{u}^T (I - P_X)\tilde{u}]^2) \\
- 2(m - p)^{-2} \text{Cov}(\tilde{u}^T A_j \tilde{u}, \tilde{u}^T (I - P_X)\tilde{u}) \text{tr}((I - P_X)V) \\
= (m - p)^{-2} \left\{ 4 \text{tr}\{A_j V (I - P_X)V\} \text{tr}((I - P_X)V) \\
+ 8 \text{tr}\{A_j V (I - P_X)V (I - P_X)V\} \\
- 4 \text{tr}\{A_j V (I - P_X)V\} \text{tr}((I - P_X)V) \right\} \\
= 8(m - p)^{-2} \text{tr}\{A_j V (I - P_X)V (I - P_X)V\}. \\
= 8(m - p)^{-2} B_i B_j e_j^T \Sigma V^{-1} (I - P_X)V (I - P_X)V^{-1} \Sigma e_i,
\]

where \(\text{tr}\) denotes the trace. Observe that \((I - P_X)V^{-1}\Sigma = I - (P_X^V)^T\) and \((I - P_X^V)V(I - (P_X^V)^T) = \Sigma\). Then

\[
\text{Cov}(\tilde{u}^T A_j \tilde{u}, (\tilde{\sigma}_u^2 - \sigma_u^2)^2) = 8(m - p)^{-2} B_i B_j e_j^T \Sigma V^{-1} (I - P_X)V (I - P_X)V^{-1} \Sigma e_i \]

\[
= 8(m - p)^{-2} B_i B_j e_j^T (I - P_X^V)V (I - (P_X^V)^T) e_i \]

\[
= 8(m - p)^{-2} B_i B_j e_j^T \Sigma e_i \]

\[
= 8(m - p)^{-2} B_i B_j e_j^T V e_i + O(m^{-3}) = O(m^{-3})
\]

since the first term is zero because \(i \neq j\) and \(V\) is diagonal. We now calculate

\[
E[\tilde{u}^T A_j \tilde{u}] = \text{tr}\{B_i B_j V^{-2} \Sigma e_i e_j^T \Sigma V^{-2} \Sigma\} = B_i B_j e_j^T \Sigma V^{-2} V^{-2} \Sigma e_i.
\]

Observe that \(\Sigma V^{-2} \Sigma = I - (P_X^V)^T - P_X^V + P_X^V (P_X^V)^T\). Then after some computations, we find that \(E[\tilde{u}^T A_j \tilde{u}] = B_i B_j e_j^T V^{-1} e_i + O(m^{-1}) = O(m^{-1})\) since \(i \neq j\). By Lemma 5,

\[
E[(\tilde{\sigma}_u^2 - \sigma_u^2)^2] = 2(m - p)^{-2} \sum_{k=1}^m (\sigma_u^2 + D_k)^2 + O(m^{-2}).
\]

Then

\[
E[\tilde{u}^T A_j \tilde{u}] E[(\tilde{\sigma}_u^2 - \sigma_u^2)^2] = o(m^{-1})
\]
since \( i \neq j \). This implies that \( R_0 = o(m^{-1}) \), which in turn implies that

\[
E[(\hat{\theta}_i^{EB} - \hat{\theta}_i^B)(\hat{\theta}_j^{EB} - \hat{\theta}_j^B)] = o(m^{-1}) \text{ for } i \neq j,
\]

(5.2.8)

since \( R_0, R_1, R_2, \) and \( R_3 \) are all \( o(m^{-1}) \). Finally, this and (5.2.3) establishes that \( E[(\hat{\theta}_w^{EB} - \tilde{\theta}_w^B)^2] = o(m^{-1}) \).

We now return to (5.2.1) and show that \( E[(\hat{\theta}_i^{EB} - \hat{\theta}_i^B)(\bar{\theta}_w^{EB} - \bar{\theta}_w^B)] = o(m^{-1}) \). By the Cauchy-Schwarz inequality, we find that

\[
E[(\hat{\theta}_i^{EB} - \hat{\theta}_i^B)(\bar{\theta}_w^{EB} - \bar{\theta}_w^B)] \leq E^{\frac{1}{2}}[(\hat{\theta}_i^{EB} - \hat{\theta}_i^B)^2] E^{\frac{1}{2}}[(\bar{\theta}_w^{EB} - \bar{\theta}_w^B)^2] = o(m^{-1})
\]

since the first term is \( O(m^{-1/2}) \) and the second term is \( o(m^{-1/2}) \).

For the next term of (5.2.1), we are interested in showing that \( E[(\hat{\theta}_i^{EB} - \hat{\theta}_i^B)(\bar{\theta}_w^B - t)] = o(m^{-1}) \). First, by Taylor expansion, we find that

\[
\hat{\theta}_i^{EB} - \hat{\theta}_i^B = \frac{\partial \hat{\theta}_i^{EB}}{\partial \sigma_u^2}(\sigma_u^2 - \hat{\sigma}_u^2) + \frac{1}{2} \frac{\partial^2 \hat{\theta}_i^{EB}}{\partial (\sigma_u^2)^2}(\sigma_u^2 - \hat{\sigma}_u^2)^2
\]

for some \( \sigma_u^2 \) between \( \sigma_u^2 \) and \( \hat{\sigma}_u^2 \). Observe that \( \bar{\theta}_w^B - t = -\sum_i w_i B_i (\hat{\theta}_i - x_i^T \tilde{\beta}) \). Then

\[
E[(\hat{\theta}_i^{EB} - \hat{\theta}_i^B)(\bar{\theta}_w^B - t)] = -\sum_j w_j B_j E \left[ \frac{\partial \hat{\theta}_j^B}{\partial \sigma_u^2}(\sigma_u^2 - \hat{\sigma}_u^2)(\hat{\theta}_j - x_j^T \tilde{\beta}) \right]
\]

\[
- \frac{1}{2} \sum_j w_j B_j E \left[ \frac{\partial^2 \hat{\theta}_j^B}{\partial (\sigma_u^2)^2}(\sigma_u^2 - \hat{\sigma}_u^2)^2(\hat{\theta}_j - x_j^T \tilde{\beta}) \right] := R_4 + R_5.
\]
Observe

\[
E \left[ \frac{\partial \hat{\theta}_i^B}{\partial \sigma_u^2} (\hat{\sigma}_u^2 - \sigma_u^2)(\hat{\theta}_j - \mathbf{x}_j^T \tilde{\mathbf{B}}) \right] = -\sigma_u^2 E \left[ \frac{\partial \hat{\theta}_i^B}{\partial \sigma_u^2} (\hat{\theta}_j - \mathbf{x}_j^T \tilde{\mathbf{B}}) I(\hat{\sigma}_u^2 \leq 0) \right] + o(m^{-r}) \]

(5.2.9)

+ \frac{E \left[ \frac{\partial \hat{\theta}_i^B}{\partial \sigma_u^2} (\hat{\sigma}_u^2 - \sigma_u^2)(\hat{\theta}_j - \mathbf{x}_j^T \tilde{\mathbf{B}}) I(\hat{\sigma}_u^2 > 0) \right]}{E \left[ \frac{\partial \hat{\theta}_i^B}{\partial \sigma_u^2} (\hat{\sigma}_u^2 - \sigma_u^2)(\hat{\theta}_j - \mathbf{x}_j^T \tilde{\mathbf{B}}) \right]}

= E \left[ \frac{\partial \hat{\theta}_i^B}{\partial \sigma_u^2} (\hat{\sigma}_u^2 - \sigma_u^2)(\hat{\theta}_j - \mathbf{x}_j^T \tilde{\mathbf{B}}) \right] - E \left[ \frac{\partial \hat{\theta}_i^B}{\partial \sigma_u^2} (\hat{\sigma}_u^2 - \sigma_u^2)(\hat{\theta}_j - \mathbf{x}_j^T \tilde{\mathbf{B}}) I(\hat{\sigma}_u^2 \leq 0) \right] + o(m^{-r})

since we may observe that \( E \left[ \frac{\partial \hat{\theta}_i^B}{\partial \sigma_u^2} (\hat{\sigma}_u^2)(\hat{\theta}_j - \mathbf{x}_j^T \tilde{\mathbf{B}}) I(\hat{\sigma}_u^2 \leq 0) \right] = o(m^{-r}) \) and

\[
E \left[ \frac{\partial \hat{\theta}_i^B}{\partial \sigma_u^2} (\hat{\sigma}_u^2)(\hat{\theta}_j - \mathbf{x}_j^T \tilde{\mathbf{B}}) I(\hat{\sigma}_u^2 \leq 0) \right] = o(m^{-r}).
\]

Now, observe that \( \frac{\partial \hat{\theta}_i^B}{\partial \sigma_u^2} = B_i \mathbf{e}_i^T \Sigma V^{-2} \tilde{\mathbf{u}} \) and define \( D_{ij} = B_i V^{-2} \Sigma \mathbf{e}_i \mathbf{e}_j^T \). Then by calculations similar to those in expression (5.2.7), we find

\[
E \left[ \frac{\partial \hat{\theta}_i^B}{\partial \sigma_u^2} (\hat{\sigma}_u^2)(\hat{\theta}_j - \mathbf{x}_j^T \tilde{\mathbf{B}}) \right] = \text{Cov}(\tilde{\mathbf{u}}^T \mathbf{D}_{ij} \tilde{\mathbf{u}}, \hat{\sigma}_u^2 - \sigma_u^2)
\]

\[
= (m-p)^{-1} \text{Cov}(\tilde{\mathbf{u}}^T \mathbf{D}_{ij} \tilde{\mathbf{u}}, \tilde{\mathbf{u}}^T (I-P_X) \tilde{\mathbf{u}} - \text{tr}((I-P_X) V))
\]

\[
= 2(m-p)^{-1} \text{tr}\{D_{ij} V(I-P_X) V\}
\]

\[
= 2(m-p)^{-1} \text{tr}\{B_i V^{-2} \Sigma \mathbf{e}_i \mathbf{e}_j^T V(I-P_X) V\}
\]

\[
= 2(m-p)^{-1} B_i \mathbf{e}_j^T V(I-P_X) V^{-1} \Sigma \mathbf{e}_i
\]

\[
= 2(m-p)^{-1} B_i \mathbf{e}_j^T V(I-P_X)^T \Sigma \mathbf{e}_i
\]

\[
= 2(m-p)^{-1} B_i \mathbf{e}_j^T V \mathbf{e}_i - h_{ij} V \]

\[
= 2(m-p)^{-1} B_i \mathbf{e}_j^T V \mathbf{e}_i + o(m^{-1}).
\]
Using the expression derived above, we find that
\[ \sum_j w_j B_j E \left[ \frac{\partial \hat{\theta}_i^B}{\partial \sigma_u^2} (\hat{\sigma}_u^2 - \sigma_u^2) (\hat{\theta}_j - x_j^T \hat{\beta}) \right] = 2(m - p)^{-1} B_i^2 w_i (\sigma_u^2 + D_i) + o(m^{-1}) = o(m^{-1}). \]

Hence, \( R_4 \) is \( o(m^{-1}) \). We now show that \( R_5 = o(m^{-1}) \). By calculations similar to those in expression (5.2.9),
\[ \sum_j w_j B_j E \left[ \frac{\partial^2 \hat{\theta}_i^B}{\partial (\sigma_u^2)^2} (\hat{\sigma}_u^2 - \sigma_u^2)^2 (\hat{\theta}_j - x_j^T \hat{\beta}) \right] = \sum_j w_j B_j E \left[ \frac{\partial^2 \hat{\theta}_i^B}{\partial (\sigma_u^2)^2} (\hat{\sigma}_u^2 - \sigma_u^2)^2 (\hat{\theta}_j - x_j^T \hat{\beta}) \right] + o(m^{-r}). \]

Recall that \( E \left[ \left\{ \sum_j w_j B_j (\hat{\theta}_j - x_j^T \hat{\beta}) \right\}^2 \right] = O(m^{-1}) \) by (5.2.2). Now consider
\[ \sum_j w_j B_j E \left[ \frac{\partial^2 \hat{\theta}_i^B}{\partial (\sigma_u^2)^2} (\hat{\sigma}_u^2 - \sigma_u^2)^2 (\hat{\theta}_j - x_j^T \hat{\beta}) \right] \leq E^\frac{1}{2} \left[ \left\{ \frac{\partial^2 \hat{\theta}_i^B}{\partial (\sigma_u^2)^2} \right\}^4 \right] E^\frac{1}{2} \left[ \left\{ \sum_j w_j B_j (\hat{\theta}_j - x_j^T \hat{\beta}) \right\}^2 \right] \leq E^\frac{1}{2} \left[ \left\{ \sup_{\sigma_u^2 \geq 0} \frac{\partial^2 \hat{\theta}_i^B}{\partial (\sigma_u^2)^2} \right\}^4 \right] E^\frac{1}{2} \left[ \left\{ \sum_j w_j B_j (\hat{\theta}_j - x_j^T \hat{\beta}) \right\}^2 \right] = O(m^{-3/2}). \]

by Lemma 2(ii), by Theorem A.5 of Prasad and Rao (1990), and by expression (5.2.2). Thus, \( R_5 \) is \( o(m^{-1}) \), and \( E[(\hat{\theta}_i^{EB} - \hat{\theta}_j^{EB})(\hat{\theta}_i^{EB} - t)] = o(m^{-1}) \).

For the last term in (5.2.1), we use the the Cauchy-Schwartz inequality to show
\[ E[(\hat{\theta}_i^{EB} - \hat{\theta}_j^{EB})(\hat{\theta}_i^{EB} - t)] \leq E^\frac{1}{2}[(\hat{\theta}_i^{EB} - \hat{\theta}_j^{EB})^2] E^\frac{1}{2}[(\hat{\theta}_i^{EB} - t)^2] = o(m^{-1}). \]

This concludes the proof of the theorem. \( \square \)

### 5.3 Estimator of MSE Approximation

We now obtain an estimator of the MSE approximation for the Fay-Herriot model (assuming normality). Theorem 8 shows that the expectation of the MSE estimator is correct to \( O(m^{-1}) \).
Lemma 1: Suppose that

$$\sup_{t \in T} |h'(t)| = O(m^{-1}) \quad (5.3.1)$$

for some interval $T \subseteq \mathbb{R}$. If $\hat{\sigma}_u^2, \sigma_u^2 \in T$ w.p. 1, then $E[h(\hat{\sigma}_u^2)] = h(\sigma_u^2) + o(m^{-1})$.

Proof. Consider the expansion $h(\hat{\sigma}_u^2) = h(\sigma_u^2) + h'(\sigma_u^2)(\hat{\sigma}_u^2 - \sigma_u^2) + o(\hat{\sigma}_u^2 - \sigma_u^2)^2$ for some $\sigma_u^2$ between $\hat{\sigma}_u^2$ and $\sigma_u^2$. Then $\sigma_u^2 \in T$ a.s., and $h'(\sigma_u^2) \leq \sup_{t \in T} |h'(t)|$ a.s. as well. This implies $E[h'(\sigma_u^2)(\hat{\sigma}_u^2 - \sigma_u^2)] \leq \sup_{t \in T} |h'(t)|E[|\hat{\sigma}_u^2 - \sigma_u^2|] = O(m^{-3/2})$ by equation (5.3.1) and since $E[\hat{\sigma}_u^2 - \sigma_u^2] \leq E\left[\frac{1}{2}(\hat{\sigma}_u^2 - \sigma_u^2)^2\right]$. Hence, if assumption (5.3.1) holds, then $E[h(\hat{\sigma}_u^2)] = h(\sigma_u^2) + o(m^{-1})$. \hfill \square

Theorem 8. $E[g_{1i}(\hat{\sigma}_u^2) + g_{2i}(\hat{\sigma}_u^2) + 2g_{3i}(\hat{\sigma}_u^2) + g_{4}(\hat{\sigma}_u^2)] = g_{1i}(\sigma_u^2) + g_{2i}(\sigma_u^2) + g_{3i}(\sigma_u^2) + g_{4}(\sigma_u^2) + o(m^{-1})$, where $g_{1i}(\sigma_u^2), g_{2i}(\sigma_u^2), g_{3i}(\sigma_u^2)$, and $g_{4}(\sigma_u^2)$ are defined in Theorem 1.

Proof. By Theorem A.3 in Prasad and Rao (1990), $E[g_{1i}(\hat{\sigma}_u^2) + g_{2i}(\hat{\sigma}_u^2) + 2g_{3i}(\hat{\sigma}_u^2)] = g_{1i}(\sigma_u^2) + g_{2i}(\sigma_u^2) + g_{3i}(\sigma_u^2) + o(m^{-1})$. In addition, we consider $E[g_{4}(\hat{\sigma}_u^2)]$, where $g_{4}(\sigma_u^2) = \sum_{i=1}^{m} w_i^2 B_i^2 V_i - \sum_{i=1}^{m} \sum_{j=1}^{m} w_i w_j B_i B_j h_y^V =: g_{41}(\sigma_u^2) + g_{42}(\sigma_u^2)$. We first show that the derivatives of $g_{41}(\sigma_u^2)$ and $g_{42}(\sigma_u^2)$ satisfy assumption (5.3.1). Let $T = [0, \infty)$. Consider

$$\sup_{\sigma_u^2 \geq 0} \left| \frac{\partial g_{41}(\sigma_u^2)}{\partial \sigma_u^2} \right| = \sup_{\sigma_u^2 \geq 0} \sum_{i=1}^{m} w_i^2 B_i^2 = O(m^{-1}).$$

It can be shown that $\frac{\partial B_i B_j}{\partial \sigma_u^2} = -B_i B_j D_j^{-1} - B_i B_j D_i^{-1}$ and $(X^T V^{-1} X)^{-1} \leq (X^T V^{-2} X)^{-1} D_i^{-1}$. Observe that

$$\left| \frac{\partial g_{42}(\sigma_u^2)}{\partial \sigma_u^2} \right| \leq \sum_{i=1}^{m} \sum_{j=1}^{m} w_i w_j \left[ |B_i D_i^{-1} h_y^V| + |B_j D_j^{-1} h_y^V| \right] + B_i B_j x^T (X^T V^{-1} X)^{-1} X^T V^{-2} X (X^T V^{-1} X)^{-1} x_i$$

$$\leq 3m^2 \left( \max_{1 \leq i \leq m} w_i \right)^2 D_i^{-1} B_i (\sigma_u^2 + D_u)(\max_{1 \leq i \leq m} h_i)$$

$$\leq 3m^2 \left( \max_{1 \leq i \leq m} w_i \right)^2 D_i^{-1} D_u (\sigma_u^2 + D_u)^{-1} (\sigma_u^2 + D_u)(\max_{1 \leq i \leq m} h_i)$$

$$= 3m^2 \left( \max_{1 \leq i \leq m} w_i \right)^2 D_i^{-1} D_u (1 + D_u D_i^{-1})(\max_{1 \leq i \leq m} h_i) = O(m^{-1}).$$
This implies that \( \sup_{\sigma^2 \geq 0} \left| \frac{\partial g_{42}(\sigma^2)}{\partial \sigma^2} \right| = O(m^{-1}) \). Since the derivatives of \( g_{41}(\sigma^2) \) and \( g_{42}(\sigma^2) \) satisfy assumption (5.3.1), we know that \( E[g_4(\hat{\sigma}^2)] = g_4(\sigma^2) + o(m^{-1}) \).

\[ \square \]

5.4 Parametric Bootstrap Estimator of the Benchmarked EB Estimator

In this section, we extend the methods of Butar and Lahiri (2003) to find a parametric bootstrap estimator of the MSE of the benchmarked empirical Bayes estimator. Under the proposed model, the expectation of the proposed measure of uncertainty of the benchmarked empirical Bayes estimator is correct up to order \( O(m^{-1}) \).

To introduce the parametric bootstrap method, consider the following model:

\[
\hat{\theta}_i^* | u_i^* \overset{\text{ind}}{\sim} N(\mathbf{x}_i^T \tilde{\beta} + u_i^* , D_i) \]
\[
u_i^* \overset{\text{ind}}{\sim} N(0, \hat{\sigma}^2_u).
\]

(5.4.1)

As explained in Butar and Lahiri (2003), we use the parametric bootstrap twice. We first use it to estimate \( g_{41}(\hat{\sigma}^2) \), \( g_{42}(\hat{\sigma}^2) \), and \( g_4(\hat{\sigma}^2) \) by correcting the bias of \( g_{41}(\hat{\sigma}^2) \), \( g_{42}(\hat{\sigma}^2) \), and \( g_4(\hat{\sigma}^2) \). We then use it again to estimate \( E[(\hat{\theta}_i^{EB} - \hat{\theta}_i^{B})^2] = g_{3i}(\sigma^2_u) + o(m^{-1}) \).

Butar and Lahiri (2003) derived a parametric bootstrap estimator for the MSE of the EB estimator under the Fay and Herriot (1979) model. Using Theorem A.1 of their paper, they show that the bootstrap estimator \( V_i^{BOOT} \) is

\[
V_i^{BOOT} = 2[g_{41}(\hat{\sigma}^2) + g_{42}(\hat{\sigma}^2) - E_*(g_{41}(\hat{\sigma}^2) + g_{42}(\hat{\sigma}^2)) + E_*(\hat{\theta}_i^{EB*} - \hat{\theta}_i^{EB})^2],
\]

(5.4.2)

where \( E_* \) denotes the expectation computed with respect to the model given in (5.4.1), and \( \hat{\theta}_i^{EB*} = (1 - B_i(\hat{\sigma}_u^{*2}))\hat{\theta}_i + B_i(\hat{\sigma}_u^{*2})x_i^T \hat{\beta} \). Following their work, we propose a parametric bootstrap estimator of the MSE of benchmarked EB estimator which is a simple extension of (5.4.2).

We propose to estimate \( g_{41}(\sigma^2_u) + g_{42}(\sigma^2_u) + g_4(\sigma^2_u) \) by

\[
2[g_{41}(\hat{\sigma}^2_u) + g_{42}(\hat{\sigma}^2_u) + g_4(\hat{\sigma}^2_u) - E_*(g_{41}(\hat{\sigma}^2_u) + g_{42}(\hat{\sigma}^2_u) + g_4(\hat{\sigma}^2_u))]
\]
and then estimate $E[(\hat{\theta}_i^{EB} - \hat{\theta}_i^B)]^2$ by $E_*[(\hat{\theta}_i^{EB*} - \hat{\theta}_i^{EB})^2]$. Thus, our proposed estimator of $\text{MSE}[\hat{\theta}_i^{EBM1}]$ is given by

$$V_{i}^{B-BOOT} = 2[g_1(\hat{\sigma}_u^2) + g_2(\hat{\sigma}_u^2) + g_4(\hat{\sigma}_u^2)] - E_*[g_1(\hat{\sigma}_u^2) + g_2(\hat{\sigma}_u^2) + g_4(\hat{\sigma}_u^2)]$$

$$+ E_*[(\hat{\theta}_i^{EB*} - \hat{\theta}_i^{EB})^2].$$

We now show that the expectation of $V_{i}^{B-BOOT}$ is correct up to $O(m^{-1})$.

**Theorem 9.** $E[V_{i}^{B-BOOT}] = \text{MSE}[\hat{\theta}_i^{EBM1}] + o(m^{-1})$.

**Proof.** First, by Theorem A.1 in Butar and Lahiri (2003), we note that

$$E_*[g_1(\hat{\sigma}_u^2)] = g_1(\hat{\sigma}_u^2) - g_3(\hat{\sigma}_u^2) + o_p(m^{-1}),$$

$$E_*[g_2(\hat{\sigma}_u^2)] = g_2(\hat{\sigma}_u^2) + o_p(m^{-1}), \text{ and}$$

$$E_*[(\hat{\theta}_i^{EB*} - \hat{\theta}_i^{EB})^2] = g_5(\hat{\sigma}_u^2) + o_p(m^{-1}),$$

where $g_5(\hat{\sigma}_u^2) = [B_i(\hat{\sigma}_u^2)]^4D_i^{-2}(\hat{\theta}_i - X^T\tilde{\tilde{\beta}}(\hat{\sigma}_u^2))^2$. Also, $E_*[g_4(\hat{\sigma}_u^2)] = g_4(\hat{\sigma}_u^2) + o_p(m^{-1})$, which follows along the lines of the proof of Theorem A.2(b) of Datta and Lahiri (2000). Applying these results and Theorem 2 of this paper, we find

$$V_{i}^{B-BOOT} = g_1(\hat{\sigma}_u^2) + g_2(\hat{\sigma}_u^2) + g_3(\hat{\sigma}_u^2) + g_4(\hat{\sigma}_u^2) + g_5(\hat{\sigma}_u^2) + o_p(m^{-1}).$$

This implies that

$$E[V_{i}^{B-BOOT}] = g_1(\sigma_u^2) + g_2(\sigma_u^2) + g_3(\sigma_u^2) + g_4(\sigma_u^2) + o(m^{-1})$$

since $E[g_5(\hat{\sigma}_u^2)] = g_3(\sigma_u^2) + o(m^{-1})$ by Butar and Lahiri (2003).

### 5.5 An Application

In this section, we consider an example where we compute estimates of the MSE of the EB estimator and the benchmarked EB estimator. We also compute an estimate of the MSE of the EB estimator using a parametric bootstrap procedure for the Fay-Herriot model as described in Section 5.
We consider data from 1997 and 2000 from the Small Area Income and Poverty Estimates (SAIPE) program at the U.S. Census Bureau, which produces model-based estimates of the number of poor school-aged children (5–17 years old) at the national, state, county, and district levels. The school district estimates are benchmarked to the state estimates by the Department of Education to allocate funds under the No Child Left Behind Act of 2001. In the SAIPE program, the model-based state estimates are benchmarked to the national school-aged poverty rate using the benchmarked estimator in (5.1.3). The number of poor school-aged children has been collected from the Annual Social and Economic Supplement (ASEC) of the CPS from 1995 to 2004, while ACS estimates have been used beginning in 2005. Additionally, the model-based county estimates are benchmarked to the model-based state estimates using the benchmarked estimator in (5.1.3).

In the SAIPE program, the state model for poverty rates in school-aged children follows the basic Fay-Herriot (1979) framework where
\[ \hat{\theta}_i = \theta_i + e_i \quad \text{and} \quad \theta_i = x_i^T \beta + u_i \]
where \( \theta_i \) is the true state level poverty rate, \( \hat{\theta}_i \) is the direct survey estimate (from CPS ASEC), \( e_i \) is the sampling error term with assumed known variance \( D_i > 0 \), \( x_i \) are the predictors, \( \beta \) is the unknown vector of regression coefficients, and \( u_i \) is the model error with unknown variance \( \sigma_u^2 \). The explanatory variables in the model are IRS income tax–based pseudo-estimate of the child poverty rate, IRS non-filer rate, food stamp rate, and the residual term from the regression of the 1990 Census estimated child poverty rate. We estimate \( \beta \) using the weighted least squares estimate
\[ \tilde{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} \hat{\theta}, \]
and we estimate \( \sigma_u^2 \) using the modified moment estimator \( \hat{\sigma}_u^2 \) from Section 2.

As shown in Table 5-1, the estimated MSE of the EB estimator, \( \text{mse}(\hat{\theta}_{EB}^i) \), compared to the estimated MSE of the benchmarked EB estimator, \( \text{mse}(\hat{\theta}_{EBM}^i) \), differs by the constant \( g_4(\sigma_u^2) \), which is 0.015. This constant is effectively the increase in MSE that we suffer from benchmarking, and we see that in this case this quantity is small.
(compared to the values of the MSEs). Generally speaking this quantity is expected to be small since $g_4(\sigma^2_u) = O(m^{-1})$.

It should be noted that in the proofs of our paper above and in Prasad and Rao (1990), we take advantage of the fact that

$$P(\tilde{\sigma}^2_u \leq 0) = O(m^{-r}) \forall r > 0.$$  

Practically speaking, this fact should correspond to the aforementioned probability being very small. However, if this is not the case for some particular dataset with fixed $m$, then our theoretical derivations of the MSE of the benchmarked EB estimator and the derivations in Prasad and Rao (1990) under the Fay-Herriot model may be unreliable. We now illustrate this concept with SAIPE data from the U.S. Census Bureau.

In Table 5-1 and Table 5-2, we define $\text{mse}^B$ and $\text{mse}^{BB}$ as the bootstrap estimates of the MSE of the EB estimator and the benchmarked EB estimator respectively. We consider two years for illustrative purposes from the SAIPE dataset (years 1997 and 2000). Table 5-1 illustrates when $\tilde{\sigma}^2_u$ is not too close to zero, being 3.08. When we perform the bootstrapping, we resample $\tilde{\sigma}^2_u$ 10,000 times in order to calculate $\text{mse}^B$ and $\text{mse}^{BB}$, and the proportion of resampled values of $\tilde{\sigma}^2_u$ that are negative is 0.034. Since this proportion is small, we can see that the bootstrap estimate of the MSE of the benchmarked EB estimator and the EB estimator are somewhat close to each other. This is best understood by considering the concept behind our bootstrapping approach. Consider the behavior of $g_{1i}(\sigma^2_u)$, the only term that is $O(1)$. Ordinarily, $g_{1i}(\tilde{\sigma}^2_u)$ underestimates $g_{1i}(\sigma^2_u)$, and $E_s[g_{1i}(\tilde{\sigma}^2_u)]$ underestimates $g_{1i}(\tilde{\sigma}^2_u)$. The basic idea is that we use the amount by which $E_s[g_{1i}(\tilde{\sigma}^2_u)]$ underestimates $g_{1i}(\tilde{\sigma}^2_u)$ as an approximation of the amount by which $g_{1i}(\tilde{\sigma}^2_u)$ underestimates $g_{1i}(\sigma^2_u)$.

We run into a problem with the 1997 data, where $g_{1i}(\tilde{\sigma}^2_u)$ is 0, since now $E_s[g_{1i}(\tilde{\sigma}^2_u)]$ overestimates $g_{1i}(\tilde{\sigma}^2_u)$. Recall that

$$V_i^{\text{B-BOOT}} = g_{1i}(\tilde{\sigma}^2_u) + \{g_{1i}(\tilde{\sigma}^2_u) - E_s[g_{1i}(\tilde{\sigma}^2_u)]\} + O(m^{-1}).$$

71
Since $g_{1i}(\hat{\sigma}_{ui}^2)$ is 0 and is the dominating term of $\mathcal{V}_i^{\text{B-BOOT}}$, many of the estimated MSEs of the benchmarked bootstrapped estimator (mse$^{\text{BB}}$) will be negative. Also, observe this same behavior holds true for the bootstrapped estimator proposed by Butar and Lahiri (2003), which we denote by mse$^\text{B}$. 
Table 5-1. Table of estimates for 2000

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\theta}_i$</th>
<th>$\hat{\theta}_{iEB}$</th>
<th>$\hat{\theta}_{iEBM1}$</th>
<th>mse($\hat{\theta}_i$)</th>
<th>mse($\hat{\theta}_{iEB}$)</th>
<th>mse($\hat{\theta}_{iEBM1}$)</th>
<th>mse$^B$</th>
<th>mse$^{BB}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19.38</td>
<td>19.05</td>
<td>19.23</td>
<td>10.78</td>
<td>3.16</td>
<td>3.17</td>
<td>3.01</td>
<td>3.03</td>
</tr>
<tr>
<td>2</td>
<td>9.36</td>
<td>9.90</td>
<td>10.08</td>
<td>2.16</td>
<td>1.59</td>
<td>1.61</td>
<td>1.56</td>
<td>1.58</td>
</tr>
<tr>
<td>3</td>
<td>21.21</td>
<td>19.51</td>
<td>19.69</td>
<td>4.05</td>
<td>2.45</td>
<td>2.47</td>
<td>2.50</td>
<td>2.51</td>
</tr>
<tr>
<td>4</td>
<td>22.94</td>
<td>21.12</td>
<td>21.30</td>
<td>16.07</td>
<td>3.52</td>
<td>3.53</td>
<td>3.42</td>
<td>3.43</td>
</tr>
<tr>
<td>5</td>
<td>18.79</td>
<td>18.48</td>
<td>18.66</td>
<td>1.29</td>
<td>1.12</td>
<td>1.13</td>
<td>1.07</td>
<td>1.09</td>
</tr>
<tr>
<td>6</td>
<td>11.09</td>
<td>11.65</td>
<td>11.83</td>
<td>4.14</td>
<td>2.22</td>
<td>2.24</td>
<td>2.10</td>
<td>2.12</td>
</tr>
<tr>
<td>7</td>
<td>8.60</td>
<td>8.99</td>
<td>9.17</td>
<td>2.78</td>
<td>1.92</td>
<td>1.93</td>
<td>1.80</td>
<td>1.82</td>
</tr>
<tr>
<td>8</td>
<td>13.46</td>
<td>13.09</td>
<td>13.27</td>
<td>5.09</td>
<td>2.35</td>
<td>2.37</td>
<td>2.19</td>
<td>2.21</td>
</tr>
<tr>
<td>9</td>
<td>24.90</td>
<td>22.53</td>
<td>22.71</td>
<td>13.75</td>
<td>5.96</td>
<td>5.97</td>
<td>5.95</td>
<td>5.96</td>
</tr>
<tr>
<td>10</td>
<td>16.58</td>
<td>17.25</td>
<td>17.43</td>
<td>1.68</td>
<td>1.36</td>
<td>1.38</td>
<td>1.45</td>
<td>1.47</td>
</tr>
<tr>
<td>11</td>
<td>16.04</td>
<td>16.97</td>
<td>17.15</td>
<td>5.70</td>
<td>2.46</td>
<td>2.48</td>
<td>2.36</td>
<td>2.37</td>
</tr>
<tr>
<td>12</td>
<td>13.01</td>
<td>12.63</td>
<td>12.81</td>
<td>6.26</td>
<td>3.05</td>
<td>3.07</td>
<td>2.88</td>
<td>2.89</td>
</tr>
<tr>
<td>13</td>
<td>19.00</td>
<td>16.59</td>
<td>16.77</td>
<td>3.44</td>
<td>2.09</td>
<td>2.11</td>
<td>2.50</td>
<td>2.51</td>
</tr>
<tr>
<td>14</td>
<td>16.54</td>
<td>14.56</td>
<td>14.74</td>
<td>3.69</td>
<td>2.05</td>
<td>2.07</td>
<td>2.24</td>
<td>2.26</td>
</tr>
<tr>
<td>15</td>
<td>10.47</td>
<td>11.25</td>
<td>11.42</td>
<td>2.73</td>
<td>1.76</td>
<td>1.77</td>
<td>1.75</td>
<td>1.76</td>
</tr>
<tr>
<td>16</td>
<td>8.76</td>
<td>9.43</td>
<td>9.61</td>
<td>1.88</td>
<td>1.40</td>
<td>1.42</td>
<td>1.46</td>
<td>1.47</td>
</tr>
<tr>
<td>17</td>
<td>12.70</td>
<td>12.38</td>
<td>12.56</td>
<td>2.49</td>
<td>1.66</td>
<td>1.67</td>
<td>1.56</td>
<td>1.57</td>
</tr>
<tr>
<td>18</td>
<td>14.40</td>
<td>16.28</td>
<td>16.46</td>
<td>9.35</td>
<td>3.10</td>
<td>3.11</td>
<td>3.05</td>
<td>3.07</td>
</tr>
<tr>
<td>19</td>
<td>23.65</td>
<td>21.27</td>
<td>21.45</td>
<td>9.61</td>
<td>3.34</td>
<td>3.35</td>
<td>3.26</td>
<td>3.27</td>
</tr>
<tr>
<td>20</td>
<td>7.41</td>
<td>9.87</td>
<td>10.05</td>
<td>8.55</td>
<td>3.16</td>
<td>3.17</td>
<td>3.20</td>
<td>3.22</td>
</tr>
<tr>
<td>21</td>
<td>7.02</td>
<td>8.45</td>
<td>8.63</td>
<td>1.75</td>
<td>1.35</td>
<td>1.36</td>
<td>1.94</td>
<td>1.96</td>
</tr>
<tr>
<td>22</td>
<td>14.01</td>
<td>11.98</td>
<td>12.16</td>
<td>3.31</td>
<td>2.11</td>
<td>2.12</td>
<td>2.43</td>
<td>2.44</td>
</tr>
<tr>
<td>23</td>
<td>12.96</td>
<td>13.04</td>
<td>13.22</td>
<td>2.43</td>
<td>1.64</td>
<td>1.65</td>
<td>1.53</td>
<td>1.55</td>
</tr>
<tr>
<td>24</td>
<td>9.77</td>
<td>9.49</td>
<td>9.67</td>
<td>1.27</td>
<td>1.07</td>
<td>1.08</td>
<td>1.04</td>
<td>1.06</td>
</tr>
<tr>
<td>25</td>
<td>14.16</td>
<td>20.15</td>
<td>20.33</td>
<td>12.69</td>
<td>3.64</td>
<td>3.66</td>
<td>4.13</td>
<td>4.15</td>
</tr>
<tr>
<td>26</td>
<td>9.41</td>
<td>12.34</td>
<td>12.52</td>
<td>6.63</td>
<td>2.63</td>
<td>2.64</td>
<td>2.85</td>
<td>2.87</td>
</tr>
<tr>
<td>27</td>
<td>19.60</td>
<td>17.54</td>
<td>17.72</td>
<td>10.08</td>
<td>3.16</td>
<td>3.18</td>
<td>3.07</td>
<td>3.09</td>
</tr>
<tr>
<td>28</td>
<td>10.73</td>
<td>10.86</td>
<td>11.04</td>
<td>2.08</td>
<td>1.52</td>
<td>1.53</td>
<td>1.43</td>
<td>1.44</td>
</tr>
<tr>
<td>29</td>
<td>11.88</td>
<td>13.13</td>
<td>13.31</td>
<td>3.50</td>
<td>2.23</td>
<td>2.24</td>
<td>2.26</td>
<td>2.28</td>
</tr>
<tr>
<td>30</td>
<td>7.01</td>
<td>6.94</td>
<td>7.12</td>
<td>1.08</td>
<td>0.95</td>
<td>0.97</td>
<td>0.90</td>
<td>0.91</td>
</tr>
<tr>
<td>31</td>
<td>10.68</td>
<td>10.77</td>
<td>10.95</td>
<td>1.72</td>
<td>1.33</td>
<td>1.35</td>
<td>1.26</td>
<td>1.27</td>
</tr>
<tr>
<td>32</td>
<td>23.59</td>
<td>25.24</td>
<td>25.42</td>
<td>12.21</td>
<td>3.91</td>
<td>3.92</td>
<td>3.86</td>
<td>3.88</td>
</tr>
<tr>
<td>33</td>
<td>19.23</td>
<td>18.49</td>
<td>18.66</td>
<td>2.31</td>
<td>1.72</td>
<td>1.73</td>
<td>1.68</td>
<td>1.69</td>
</tr>
<tr>
<td>34</td>
<td>17.99</td>
<td>16.54</td>
<td>16.72</td>
<td>10.52</td>
<td>2.83</td>
<td>2.85</td>
<td>2.71</td>
<td>2.73</td>
</tr>
<tr>
<td>35</td>
<td>12.40</td>
<td>11.88</td>
<td>12.06</td>
<td>7.64</td>
<td>2.90</td>
<td>2.91</td>
<td>2.74</td>
<td>2.76</td>
</tr>
<tr>
<td>36</td>
<td>13.87</td>
<td>13.08</td>
<td>13.26</td>
<td>2.97</td>
<td>1.86</td>
<td>1.88</td>
<td>1.79</td>
<td>1.81</td>
</tr>
<tr>
<td>37</td>
<td>21.73</td>
<td>19.74</td>
<td>19.92</td>
<td>14.30</td>
<td>3.19</td>
<td>3.20</td>
<td>3.09</td>
<td>3.10</td>
</tr>
<tr>
<td>38</td>
<td>18.40</td>
<td>14.71</td>
<td>14.89</td>
<td>8.61</td>
<td>3.46</td>
<td>3.47</td>
<td>3.57</td>
<td>3.59</td>
</tr>
<tr>
<td>39</td>
<td>10.53</td>
<td>10.98</td>
<td>11.16</td>
<td>1.21</td>
<td>1.02</td>
<td>1.03</td>
<td>1.10</td>
<td>1.11</td>
</tr>
<tr>
<td>40</td>
<td>11.10</td>
<td>12.06</td>
<td>12.24</td>
<td>6.33</td>
<td>2.72</td>
<td>2.73</td>
<td>2.60</td>
<td>2.61</td>
</tr>
<tr>
<td>i</td>
<td>$\hat{\theta}_i$</td>
<td>$\hat{\theta}^{EB}_i$</td>
<td>$\hat{\theta}^{EBM1}_i$</td>
<td>mse($\hat{\theta}_i$)</td>
<td>mse($\hat{\theta}^{EB}_i$)</td>
<td>mse($\hat{\theta}^{EBM1}_i$)</td>
<td>$\text{mse}^B$</td>
<td>$\text{mse}^{BB}$</td>
</tr>
<tr>
<td>----</td>
<td>-----------------</td>
<td>-----------------------</td>
<td>----------------------</td>
<td>------------------</td>
<td>---------------------</td>
<td>---------------------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>41</td>
<td>14.43</td>
<td>16.39</td>
<td>16.57</td>
<td>4.53</td>
<td>2.37</td>
<td>2.38</td>
<td>2.51</td>
<td>2.53</td>
</tr>
<tr>
<td>42</td>
<td>13.14</td>
<td>13.44</td>
<td>13.61</td>
<td>4.18</td>
<td>2.40</td>
<td>2.41</td>
<td>2.24</td>
<td>2.26</td>
</tr>
<tr>
<td>43</td>
<td>21.64</td>
<td>17.34</td>
<td>17.52</td>
<td>8.15</td>
<td>2.93</td>
<td>2.95</td>
<td>3.19</td>
<td>3.21</td>
</tr>
<tr>
<td>44</td>
<td>20.02</td>
<td>19.84</td>
<td>20.02</td>
<td>1.62</td>
<td>1.34</td>
<td>1.35</td>
<td>1.24</td>
<td>1.26</td>
</tr>
<tr>
<td>45</td>
<td>10.02</td>
<td>10.57</td>
<td>10.75</td>
<td>3.42</td>
<td>2.12</td>
<td>2.13</td>
<td>2.01</td>
<td>2.03</td>
</tr>
<tr>
<td>46</td>
<td>12.80</td>
<td>11.31</td>
<td>11.49</td>
<td>2.62</td>
<td>1.87</td>
<td>1.88</td>
<td>2.03</td>
<td>2.04</td>
</tr>
<tr>
<td>47</td>
<td>7.26</td>
<td>10.02</td>
<td>10.20</td>
<td>3.35</td>
<td>1.94</td>
<td>1.96</td>
<td>2.72</td>
<td>2.73</td>
</tr>
<tr>
<td>48</td>
<td>10.82</td>
<td>11.10</td>
<td>11.28</td>
<td>2.40</td>
<td>1.66</td>
<td>1.67</td>
<td>1.57</td>
<td>1.59</td>
</tr>
<tr>
<td>49</td>
<td>15.22</td>
<td>17.14</td>
<td>17.32</td>
<td>6.41</td>
<td>3.19</td>
<td>3.21</td>
<td>3.23</td>
<td>3.24</td>
</tr>
<tr>
<td>50</td>
<td>14.98</td>
<td>12.28</td>
<td>12.46</td>
<td>4.80</td>
<td>2.45</td>
<td>2.47</td>
<td>2.75</td>
<td>2.76</td>
</tr>
<tr>
<td>51</td>
<td>13.22</td>
<td>13.02</td>
<td>13.20</td>
<td>3.00</td>
<td>1.96</td>
<td>1.98</td>
<td>1.82</td>
<td>1.84</td>
</tr>
</tbody>
</table>
Table 5-2. Table of estimates for 1997

<table>
<thead>
<tr>
<th></th>
<th>( \hat{\theta}_i )</th>
<th>( \hat{\theta}^{EB}_i )</th>
<th>( \hat{\theta}^{EBM}_i )</th>
<th>\text{mse}(( \hat{\theta}_i ))</th>
<th>\text{mse}(( \hat{\theta}^{EB}_i ))</th>
<th>\text{mse}(( \hat{\theta}^{EBM}_i ))</th>
<th>\text{mse}^B</th>
<th>\text{mse}^{BB}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25.16</td>
<td>21.38</td>
<td>21.56</td>
<td>15.72</td>
<td>1.38</td>
<td>1.41</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>2</td>
<td>10.99</td>
<td>14.94</td>
<td>15.11</td>
<td>10.44</td>
<td>2.12</td>
<td>2.14</td>
<td>0.66</td>
<td>0.68</td>
</tr>
<tr>
<td>3</td>
<td>23.35</td>
<td>20.89</td>
<td>21.06</td>
<td>11.84</td>
<td>1.68</td>
<td>1.70</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>4</td>
<td>23.32</td>
<td>22.18</td>
<td>22.35</td>
<td>13.85</td>
<td>1.90</td>
<td>1.92</td>
<td>0.37</td>
<td>0.38</td>
</tr>
<tr>
<td>5</td>
<td>23.55</td>
<td>22.71</td>
<td>22.88</td>
<td>2.39</td>
<td>5.92</td>
<td>5.94</td>
<td>1.12</td>
<td>1.13</td>
</tr>
<tr>
<td>6</td>
<td>9.14</td>
<td>13.12</td>
<td>13.29</td>
<td>6.38</td>
<td>2.19</td>
<td>2.22</td>
<td>0.36</td>
<td>0.38</td>
</tr>
<tr>
<td>7</td>
<td>10.34</td>
<td>13.39</td>
<td>13.56</td>
<td>9.85</td>
<td>2.08</td>
<td>2.10</td>
<td>0.39</td>
<td>0.41</td>
</tr>
<tr>
<td>8</td>
<td>15.54</td>
<td>13.06</td>
<td>13.23</td>
<td>17.56</td>
<td>0.91</td>
<td>0.94</td>
<td>-0.47</td>
<td>-0.45</td>
</tr>
<tr>
<td>9</td>
<td>35.85</td>
<td>32.43</td>
<td>32.60</td>
<td>4.92</td>
<td>4.95</td>
<td>3.49</td>
<td>3.50</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>18.34</td>
<td>19.59</td>
<td>19.76</td>
<td>3.70</td>
<td>3.71</td>
<td>3.74</td>
<td>0.40</td>
<td>0.41</td>
</tr>
<tr>
<td>11</td>
<td>23.52</td>
<td>20.53</td>
<td>20.70</td>
<td>12.93</td>
<td>1.16</td>
<td>1.19</td>
<td>-0.38</td>
<td>-0.37</td>
</tr>
<tr>
<td>12</td>
<td>18.98</td>
<td>13.72</td>
<td>13.89</td>
<td>20.87</td>
<td>2.45</td>
<td>2.48</td>
<td>1.24</td>
<td>1.26</td>
</tr>
<tr>
<td>13</td>
<td>17.56</td>
<td>13.64</td>
<td>13.82</td>
<td>12.38</td>
<td>1.70</td>
<td>1.73</td>
<td>0.23</td>
<td>0.25</td>
</tr>
<tr>
<td>14</td>
<td>14.57</td>
<td>15.72</td>
<td>15.89</td>
<td>3.56</td>
<td>3.45</td>
<td>3.47</td>
<td>-0.06</td>
<td>-0.05</td>
</tr>
<tr>
<td>15</td>
<td>11.07</td>
<td>12.53</td>
<td>12.70</td>
<td>7.58</td>
<td>1.84</td>
<td>1.86</td>
<td>-0.23</td>
<td>-0.22</td>
</tr>
<tr>
<td>16</td>
<td>11.09</td>
<td>11.21</td>
<td>11.38</td>
<td>8.49</td>
<td>1.74</td>
<td>1.76</td>
<td>-0.24</td>
<td>-0.22</td>
</tr>
<tr>
<td>17</td>
<td>11.01</td>
<td>13.48</td>
<td>13.65</td>
<td>9.34</td>
<td>1.61</td>
<td>1.63</td>
<td>-0.15</td>
<td>-0.14</td>
</tr>
<tr>
<td>18</td>
<td>23.12</td>
<td>20.78</td>
<td>20.95</td>
<td>13.98</td>
<td>1.37</td>
<td>1.40</td>
<td>-0.12</td>
<td>-0.11</td>
</tr>
<tr>
<td>19</td>
<td>21.08</td>
<td>24.15</td>
<td>24.32</td>
<td>15.19</td>
<td>1.80</td>
<td>1.82</td>
<td>0.40</td>
<td>0.42</td>
</tr>
<tr>
<td>20</td>
<td>13.18</td>
<td>12.44</td>
<td>12.61</td>
<td>13.63</td>
<td>2.09</td>
<td>2.11</td>
<td>0.56</td>
<td>0.57</td>
</tr>
<tr>
<td>21</td>
<td>9.90</td>
<td>13.16</td>
<td>13.33</td>
<td>9.28</td>
<td>1.65</td>
<td>1.67</td>
<td>-0.03</td>
<td>-0.01</td>
</tr>
<tr>
<td>22</td>
<td>19.66</td>
<td>14.38</td>
<td>14.56</td>
<td>7.66</td>
<td>2.46</td>
<td>2.48</td>
<td>1.02</td>
<td>1.04</td>
</tr>
<tr>
<td>23</td>
<td>13.78</td>
<td>16.86</td>
<td>17.03</td>
<td>4.04</td>
<td>3.11</td>
<td>3.13</td>
<td>0.38</td>
<td>0.39</td>
</tr>
<tr>
<td>24</td>
<td>14.34</td>
<td>10.11</td>
<td>10.28</td>
<td>9.91</td>
<td>1.64</td>
<td>1.67</td>
<td>0.16</td>
<td>0.17</td>
</tr>
<tr>
<td>25</td>
<td>20.58</td>
<td>22.30</td>
<td>22.47</td>
<td>15.07</td>
<td>2.42</td>
<td>2.45</td>
<td>0.97</td>
<td>0.99</td>
</tr>
<tr>
<td>26</td>
<td>18.90</td>
<td>15.11</td>
<td>15.28</td>
<td>15.24</td>
<td>1.00</td>
<td>1.03</td>
<td>-0.37</td>
<td>-0.35</td>
</tr>
<tr>
<td>27</td>
<td>17.00</td>
<td>18.60</td>
<td>18.77</td>
<td>12.95</td>
<td>1.37</td>
<td>1.40</td>
<td>-0.21</td>
<td>-0.19</td>
</tr>
<tr>
<td>28</td>
<td>9.72</td>
<td>9.62</td>
<td>9.79</td>
<td>7.18</td>
<td>2.24</td>
<td>2.26</td>
<td>0.09</td>
<td>0.10</td>
</tr>
<tr>
<td>29</td>
<td>14.06</td>
<td>12.94</td>
<td>13.12</td>
<td>10.23</td>
<td>1.71</td>
<td>1.74</td>
<td>-0.06</td>
<td>-0.04</td>
</tr>
<tr>
<td>30</td>
<td>10.94</td>
<td>6.72</td>
<td>6.89</td>
<td>11.35</td>
<td>1.88</td>
<td>1.91</td>
<td>0.50</td>
<td>0.52</td>
</tr>
<tr>
<td>31</td>
<td>14.66</td>
<td>13.28</td>
<td>13.45</td>
<td>5.52</td>
<td>2.48</td>
<td>2.51</td>
<td>-0.03</td>
<td>-0.01</td>
</tr>
<tr>
<td>32</td>
<td>29.69</td>
<td>24.44</td>
<td>24.61</td>
<td>13.18</td>
<td>2.62</td>
<td>2.65</td>
<td>1.38</td>
<td>1.40</td>
</tr>
<tr>
<td>33</td>
<td>23.76</td>
<td>22.85</td>
<td>23.02</td>
<td>3.10</td>
<td>4.76</td>
<td>4.79</td>
<td>0.94</td>
<td>0.95</td>
</tr>
<tr>
<td>34</td>
<td>13.90</td>
<td>16.58</td>
<td>16.75</td>
<td>5.70</td>
<td>2.29</td>
<td>2.31</td>
<td>-0.01</td>
<td>-0.01</td>
</tr>
<tr>
<td>35</td>
<td>18.19</td>
<td>13.64</td>
<td>13.81</td>
<td>11.92</td>
<td>1.81</td>
<td>1.84</td>
<td>0.48</td>
<td>0.50</td>
</tr>
<tr>
<td>36</td>
<td>13.91</td>
<td>13.64</td>
<td>13.81</td>
<td>3.95</td>
<td>3.07</td>
<td>3.10</td>
<td>-0.25</td>
<td>-0.23</td>
</tr>
<tr>
<td>37</td>
<td>16.09</td>
<td>21.50</td>
<td>21.68</td>
<td>11.14</td>
<td>1.52</td>
<td>1.54</td>
<td>0.24</td>
<td>0.26</td>
</tr>
<tr>
<td>38</td>
<td>12.60</td>
<td>13.43</td>
<td>13.60</td>
<td>10.35</td>
<td>2.53</td>
<td>2.56</td>
<td>0.83</td>
<td>0.84</td>
</tr>
<tr>
<td>39</td>
<td>14.61</td>
<td>13.92</td>
<td>14.09</td>
<td>3.73</td>
<td>3.40</td>
<td>3.42</td>
<td>-0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>40</td>
<td>20.37</td>
<td>14.60</td>
<td>14.77</td>
<td>18.53</td>
<td>1.04</td>
<td>1.07</td>
<td>-0.15</td>
<td>-0.14</td>
</tr>
<tr>
<td>(i)</td>
<td>(\hat{\theta}_i)</td>
<td>(\hat{\theta}_{i,EB}^{EB})</td>
<td>(\hat{\theta}_{i,EBM}^{i})</td>
<td>(\text{mse}(\hat{\theta}_i))</td>
<td>(\text{mse}(\hat{\theta}_{i,EB}^{EB}))</td>
<td>(\text{mse}(\hat{\theta}_{i,EBM}^{i}))</td>
<td>(\text{mse}^B)</td>
<td>(\text{mse}^{BB})</td>
</tr>
<tr>
<td>------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>41</td>
<td>18.74</td>
<td>21.21</td>
<td>21.38</td>
<td>14.57</td>
<td>1.49</td>
<td>1.52</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>42</td>
<td>12.87</td>
<td>15.77</td>
<td>15.94</td>
<td>12.94</td>
<td>1.98</td>
<td>2.01</td>
<td>0.46</td>
<td>0.47</td>
</tr>
<tr>
<td>43</td>
<td>16.09</td>
<td>16.10</td>
<td>16.27</td>
<td>11.94</td>
<td>1.92</td>
<td>1.95</td>
<td>0.28</td>
<td>0.30</td>
</tr>
<tr>
<td>44</td>
<td>21.95</td>
<td>21.38</td>
<td>21.55</td>
<td>3.38</td>
<td>4.05</td>
<td>4.07</td>
<td>0.38</td>
<td>0.40</td>
</tr>
<tr>
<td>45</td>
<td>11.27</td>
<td>9.76</td>
<td>9.93</td>
<td>9.45</td>
<td>2.28</td>
<td>2.31</td>
<td>0.50</td>
<td>0.51</td>
</tr>
<tr>
<td>46</td>
<td>11.15</td>
<td>10.10</td>
<td>10.27</td>
<td>11.95</td>
<td>2.45</td>
<td>2.48</td>
<td>0.86</td>
<td>0.88</td>
</tr>
<tr>
<td>47</td>
<td>16.40</td>
<td>14.96</td>
<td>15.13</td>
<td>11.51</td>
<td>1.20</td>
<td>1.22</td>
<td>-0.49</td>
<td>-0.47</td>
</tr>
<tr>
<td>48</td>
<td>12.26</td>
<td>13.17</td>
<td>13.34</td>
<td>9.33</td>
<td>1.85</td>
<td>1.87</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>49</td>
<td>18.76</td>
<td>22.25</td>
<td>22.42</td>
<td>13.73</td>
<td>3.81</td>
<td>3.83</td>
<td>2.46</td>
<td>2.48</td>
</tr>
<tr>
<td>50</td>
<td>7.60</td>
<td>11.87</td>
<td>12.04</td>
<td>6.41</td>
<td>2.74</td>
<td>2.76</td>
<td>0.97</td>
<td>0.98</td>
</tr>
<tr>
<td>51</td>
<td>11.74</td>
<td>11.70</td>
<td>11.87</td>
<td>8.86</td>
<td>2.08</td>
<td>2.10</td>
<td>0.17</td>
<td>0.19</td>
</tr>
</tbody>
</table>
CHAPTER 6
SUMMARY REMARKS AND FUTURE WORK

Small area estimation first grew substantially during World War II but is flourishing again due to its necessity in government agencies worldwide. We briefly describe the importance of small area estimation in Chapter 1 and how benchmarking specifically has played an important role in research. Furthermore, in Chapter 2, we review the important models used in small area estimation in this body of work, and most relevantly those of Fay and Herriot (1979), Louis (1984), Prasad and Rao (1990), Ghosh (1992), Louis (1984), You and Rao (2002), You and Rao (2003), and Wang, Fuller and Qu (2008).

In Chapter 3 we develop a general class of benchmarked Bayes estimators under a general loss function where we can benchmark the weighted mean or both the weighted mean and weighted variability. Moreover, our results do not assume any distribution assumptions, and the form of our estimator can be linear or nonlinear. We can also extend our results to multiparameter settings using a general loss function. Finally, many of the previously proposed estimators from the literature result as special cases of the benchmarked Bayes estimator. We illustrate our methods using SAIPE data from the U.S. Census Bureau.

Next, in Chapter 4, we extend the above work to a two-stage benchmarking procedure under one model. For example, we are able to benchmark the weighted means and find closed-form solutions, and we are able to benchmark both the weighted means and weighted unit-level variability and find closed-form solutions. We can extend these to multiparameter settings. However, going beyond our present work indicates that a solution exists but not in closed form. We study the behavior of our estimates by looking at the proportion of people who do not have health insurance for an Asian subpopulation as studied by the NHIS in 2000.
Finally, in Chapter 5, we find the benchmarked EB estimator under the Fay-Herriot model (assuming normality) and assuming the standard benchmarking constraint. Under some mild regularity conditions, we determine how much mean squared error is lost due to benchmarking. Specifically, using a second-order expansion, we show that the error lost due to benchmarking is $O(m^{-1})$, where $m$ is the number of small areas. We then find an asymptotically unbiased estimator of this MSE. In addition, using methods similar to those in Butar and Lahiri (2003), we derive a parametric bootstrap estimator of the MSE of the benchmarked EB estimator. We investigate the properties of our results using SAIPE data and illustrate interesting behavior that occurs when the estimated value of $\sigma^2_u$ is 0 as occurs in 1997.

In terms of future work, we believe that our work in benchmarking just scratches the surface. Benchmarking in spatial settings has not been explored, and such work would find immediate application whenever response variables show correlation due to geographic proximity or similarity. We also believe there is much more to be done in two-stage benchmarking including the derivation of closed-form solutions or numerical algorithms if closed-form solutions cannot be obtained. Furthermore, this idea could extend to multi-stage procedures. For example, we might wish to benchmark county-level estimates to district-level estimates, district-level estimates to state-level estimates, and then state-level estimates to the national-level estimate. Finally, we should attempt to find closed-form expressions for the MSE of the benchmarked EB estimator under more complicated constraints and models, as well as under different variance component estimators, in order to generalize our results on the amount of error lost due to benchmarking.
APPENDIX: LEMMAS FOR CHAPTER 5

Lemma 2: Let $r > 0$ be arbitrary. Then

(i) $E \left[ \frac{\partial \hat{\theta}_i^B}{\partial \sigma_u^2} \right]^{2r} = O(1)$, and

(ii) $E \left[ \sup_{\sigma_u^2 \geq 0} \frac{\partial^2 \hat{\theta}_i^B}{\partial (\sigma_u^2)^2} \right]^{2r} = O(1)$.

Proof of Lemma 2(i). Recall $\tilde{u} = \hat{\theta} - X\tilde{\beta}$ and define $u = \hat{\theta} - X\beta$. Recall $\hat{\theta}_i^B = (1 - B_i)\hat{\theta}_i + B_ix_i^T\tilde{\beta}$. Since $\frac{\partial \beta}{\partial \sigma_u^2} = -(X^TV^{-1}X)^{-1}X^TV^{-2}\hat{\theta}$, we can easily show that

$\frac{\partial \hat{\theta}_i^B}{\partial \sigma_u^2} = [B_i^2D_i^{-1}e_i^T - B_ix_i^T(X^TV^{-1}X)^{-1}X^TV^{-2}]\tilde{u}$. This implies that

$$\left| \frac{\partial \hat{\theta}_i^B}{\partial \sigma_u^2} \right| \leq |D_i^{-1}e_i^T\tilde{u}| + |x_i^T(X^TV^{-1}X)^{-1}X^TV^{-2}\tilde{u}|.$$

Then

$$\left| \frac{\partial \hat{\theta}_i^B}{\partial \sigma_u^2} \right|^{2r} \leq 2^{2r-1}D_i^{-2r}e_i^T\tilde{u}^{2r} + 2^{2r-1}|x_i^T(X^TV^{-1}X)^{-1}X^TV^{-2}\tilde{u}|^{2r}$$

$$\leq 2^{2r-1}D_i^{-2r}|\hat{\theta}_i - x_i^T\tilde{\beta}|^{2r} + 2^{2r-1}\left[ x_i^T(X^TV^{-1}X)^{-1}x_i\tilde{u}^T\tilde{v}^{-3}\tilde{u} \right]^{2r}$$

$$\leq 2^{2r-1}D_i^{-2r}|\hat{\theta}_i - x_i^T\tilde{\beta}|^{2r} + 2^{2r-1}\left[ (\max_{1 \leq i \leq m} h_i)(\sigma_u^2 + D_u)(\sigma_u^2 + D_L)^{-1}D_i^{-1}u^TV^{-1}u \right]^{2r}$$

$$\leq 2^{2r-1}D_i^{-2r}|\hat{\theta}_i - x_i^T\tilde{\beta}|^{2r} + 2^{2r-1}\left[ (1 + D_uD_L^{-1})D_i^{-1}(\max_{1 \leq i \leq m} h_i)u^TV^{-1}u \right]^{2r} = O_p(1).$$

since $\hat{\theta}_i - x_i^T\tilde{\beta} \sim N(0, V_i)$, $\max_{1 \leq i \leq m} h_i = O(m^{-1})$, and $u^TV^{-1}u \sim \chi_m^2$. This implies that

$E \left[ \frac{\partial \hat{\theta}_i^B}{\partial \sigma_u^2} \right]^{2r} = O(1)$.

Proof of Lemma 2(ii). Recall $P_X^V = X(X^TV^{-1}X)^{-1}X^TV^{-1}$, $\tilde{u} = \hat{\theta} - X\tilde{\beta}$, and $u = \hat{\theta} - X\beta$. Knowing $\frac{\partial \beta}{\partial \sigma_u^2} = -(X^TV^{-1}X)^{-1}X^TV^{-2}\hat{\theta}$, we find that

$$\frac{\partial^2 \hat{\theta}_i^B}{\partial (\sigma_u^2)^2} = -2B_i^3D_i^{-2}e_i^T\tilde{u} + 2B_i^2D_i^{-1}e_i^TP_X^VP_X^V\tilde{u}$$

$$+ 2B_i^2e_i^TP_X^VP_X^V\tilde{u} - 2B_i^2e_i^TP_X^VP_X^V\tilde{u}.$$
Then we find
\[
\left| \frac{\partial^2 \hat{B}}{\partial (\sigma_u^2)^2} \right| \leq |2B_i^3 D_i^{-2} e_i^T \tilde{u}| + |2B_i^2 D_i^{-1} e_i^T P^X V^{-1} \tilde{u}| \\
+ |2B_i e_i^T P^X V^{-2} \tilde{u}| + |2B_i e_i^T P^X V^{-1} P^X V^{-1} \tilde{u}|.
\] (A.0.1)

Define \( \Omega = (X^T V^{-1} X)^{-1}(X^T V^{-2} X)(X^T V^{-1} X)^{-1}(X^T V^{-2} X)(X^T V^{-1} X)^{-1} \). Using our
expression in (A.0.1), we find that
\[
\left| \frac{\partial^2 \hat{B}}{\partial (\sigma_u^2)^2} \right|^{2r} \leq 4^{3r-1} \left[ B_i^3 D_i^{-2} e_i^T \tilde{u}^{2r} + B_i^4 r D_i^{-2r} [x_i^T (X^T V^{-1} X)^{-1} x_i \tilde{u}^T V^{-3} \tilde{u}]^r \\
+ B_i^2 [x_i^T (X^T V^{-1} X)^{-1} x_i \tilde{u}^T V^{-5} \tilde{u}]^r + B_i [x_i^T \Omega x_i \tilde{u}^T V^{-3} \tilde{u}]^r \right] \\
\leq 4^{3r-1} \left[ B_i^r D_i^{-4r} |\hat{\beta} - x_i^T \beta|^{2r} + D_i^{-3r} [(1 + D_i D_i^{-1}) (\max_{1 \leq i \leq m} h_i) \tilde{u}^T V^{-1} \tilde{u}]^r \\
+ D_i^{-3r} [(1 + D_i D_i^{-1}) (\max_{1 \leq i \leq m} h_i) \tilde{u}^T V^{-1} \tilde{u}]^r \right] \\
= 4^{3r-1} \left[ B_i^r D_i^{-4r} |\hat{\beta} - x_i^T \beta|^{2r} + 3D_i^{-3r} [(1 + D_i D_i^{-1}) (\max_{1 \leq i \leq m} h_i) \tilde{u}^T V^{-1} \tilde{u}]^r \right].
\] (A.0.2)

From equation (A.0.2) and since \( \tilde{u}^T V^{-1} \tilde{u} \leq u^T V^{-1} u \), it follows that
\[
\sup_{\sigma_u^2 \geq 0} \left| \frac{\partial^2 \hat{B}}{\partial (\sigma_u^2)^2} \right| \leq \sup_{\sigma_u^2 \geq 0} \left[ 4^{3r-1} \left( B_i^r D_i^{-4r} |\hat{\beta} - x_i^T \beta|^{2r} + 3 \left[ D_i^{-3r} (1 + D_i D_i^{-1}) (\max_{1 \leq i \leq m} h_i) u^T V^{-1} u \right]^r \right) \right] \\
\leq 4^{3r-1} D_i^{-4r} \sup_{\sigma_u^2 \geq 0} B_i^r |\hat{\beta} - x_i^T \beta|^{2r} + 3D_i^{-r} (1 + D_i D_i^{-1}) (\max_{1 \leq i \leq m} h_i) \sup_{\sigma_u^2 \geq 0} (u^T V^{-1} u)^r \\
\leq 4^{3r-1} D_i^{-4r} \sup_{\sigma_u^2 \geq 0} \left( \frac{|\hat{\beta} - x_i^T \beta|}{\sigma_u^2 + D_i} \right)^{2r} + 3(D_i + D_i) (\max_{1 \leq i \leq m} h_i) \sup_{\sigma_u^2 \geq 0} (u^T V^{-1} u)^r
\]
where for all $\sigma_u^2 \geq 0$, \[ \frac{\hat{\theta}_i - x_i^T \beta}{(\sigma_u^2 + D_i)^{1/2}} \stackrel{d}{=} |Z|^{2r} \text{ and } Z \sim N(0, 1). \] Also, for all $\sigma_u^2 \geq 0$, \[ (u^T V^{-1} u) \stackrel{d}{=} W_m, \text{ where } W_m \sim \chi_m^2. \] This implies that \[
abla
\begin{align*}
E \left[ \sup_{\sigma_u^2 \geq 0} \left| \frac{\partial^2 \hat{\theta}}{\partial (\sigma_u^2)^2} \right|^{2r} \right] &\leq 4^{3r-1} D_L^{-4r} \left[ D'_u E[|Z|^{2r}] + 3(D_L + D_U)'( \max_{1 \leq i \leq m} h_i )' E[W_m'] \right] = O(1).
\end{align*}
\]

Recall that $u = \hat{\theta} - X\beta \sim N(0, V)$. We have the following collection of results:

**Lemma 3:** Let $r > 0$ and assume $\max_{1 \leq i \leq m} x_i^T \beta = O(1)$. Then

\[ ||\hat{\theta} - X\beta||^{2r} = O_p(m') \text{ and } E \left[ ||\hat{\theta} - X\beta||^{2r} \right] = O(m').\]

**Proof of Lemma 3.** Recall $\bar{u} = \hat{\theta} - X\beta$. Then $\Sigma^{-1/2} \bar{u} \sim N(0, I)$. Let $W = \bar{u}^T \Sigma^{-1} \bar{u} \sim \chi_m^2$ and observe $\bar{u}^T \Sigma^{-1} \bar{u} \geq \bar{u}^T V^{-1} \bar{u} \geq \bar{u}^T (\sigma_u^2 + D_U)^{-1} \bar{u} = (\sigma_u^2 + D_U)^{-1} ||\bar{u}||^2$. This implies that $||\bar{u}||^{2r} \leq (\sigma_u^2 + D_U)' W' = O_p(m')$. Also,

\[ E ||\bar{u}||^{2r} \leq E \left[ (\sigma_u^2 + D_U)' W' \right] = O(m'). \]

**Lemma 4:** Let $z \sim N_p(0, \Sigma)$ with matrices $A_{p \times p}$ and $B_{p \times p}$, where $B$ is symmetric. Then

(i) $\text{Cov}(z^T Az, z^T Bz) = 2\text{tr}(A \Sigma B \Sigma)$

(ii) $\text{Cov}(z^T Az, (z^T Bz)^2) = 4\text{tr}(A \Sigma B \Sigma)\text{tr}(B^2) + 8\text{tr}(A \Sigma B \Sigma B^2).$

**Proof of (i).** See Searle (1971, pg. 51)

**Proof of (ii).** First, let $\Sigma = I_\rho$. By the Spectral Decomposition Theorem, define $D := PBP^T$, where $P$ is orthogonal and $D$ is diagonal with eigenvalues $\lambda_i$. Define $C := PAP^T$.

We know that $z^T Bz = z^T P^T D P z$ and $z^T A z = z^T P^T C P z$. Also, since $z \sim N_p(0, I_\rho)$ and $z \stackrel{d}{=} P z$, $\text{Cov}(z^T Cz, (z^T Dz)^2) = \text{Cov}(z^T Az, (z^T Bz)^2)$. Then by the above and algebra, we can show

\[ E \left[ (z^T Dz)^2 \right] = 2\text{tr}(B^2) + \text{tr}(B)^2 \]
and $E \left[ z^T C z (z^T D z)^2 \right] = 8 \text{tr}(AB^2) + 2 \text{tr}(A) \text{tr}(B^2) + 4 \text{tr}(AB) \text{tr}(B) + \text{tr}(A) \text{tr}(B)^2$. Hence,

$$
\text{Cov} \left( z^T A z, (z^T B z)^2 \right) = 8 \text{tr}(AB^2) + 4 \text{tr}(AB) \text{tr}(B). \tag{A.0.3}
$$

Now we assume general $\Sigma$ and let $w = \Sigma^{-1/2} z \sim N_p(0, I_p)$. By (A.0.3), we observe that

$$
\text{Cov}(z^T A z, (z^T B z)^2) = \text{Cov}(w^T \Sigma^{1/2} A \Sigma^{1/2} w, (w^T \Sigma^{1/2} B \Sigma^{1/2} w)^2)
= 4 \text{tr}(A \Sigma B \Sigma) \text{tr}(B \Sigma) + 8 \text{tr}(A \Sigma B \Sigma B \Sigma).
$$

**Lemma 5:** $E[(\hat{\sigma}_u^2 - \sigma_u^2)^2] = 2(m - p)^{-2} \sum_{i=1}^m (\sigma_u^2 + D_i)^2 + O(m^{-2})$.

**Proof.** Observe $m - p = \text{tr}\{I - P_X\}$ and define $d = \sum_i D_i (1 - h_i) = \text{tr}\{(I - P_X)D\}$, where $D = \text{Diag}\{D_i\}$. Also, recall $\tilde{u} = \tilde{\theta} - X \tilde{\beta}$. Then

$$
E[(\hat{\sigma}_u^2 - \sigma_u^2)^2] = (m - p)^{-2} E \left[ \left\{ \tilde{u}^T (I - P_X) \tilde{u} - \sigma_u^2 (m - p) - d \right\}^2 \right]
= (m - p)^{-2} E \left[ \left\{ \tilde{u}^T (I - P_X) \tilde{u} - \text{tr}\{(I - P_X) V\} \right\}^2 \right]
= (m - p)^{-2} E \left[ \left\{ \tilde{u}^T (I - P_X) \tilde{u} \right\}^2 \right] - 2 \text{tr}\{(I - P_X) V\} E[\tilde{u}^T (I - P_X) \tilde{u}]
+ \text{tr}\{(I - P_X) V\}^2 \right] = 2(m - p)^{-2} \text{tr}\{(I - P_X) V(I - P_X) V\}.
$$

Using matrix manipulations, it is easy to show that

$$
E[(\hat{\sigma}_u^2 - \sigma_u^2)^2] = 2(m - p)^{-2} \sum_{i=1}^m (\sigma_u^2 + D_i) \left[ (\sigma_u^2 + D_i) 
+ x_i^T (X^T X)^{-1} X^T VX (X^T X)^{-1} x_i^T - 2(\sigma_u^2 + D_i)^2 h_i^V \right]
= 2(m - p)^{-2} \sum_{i=1}^m (\sigma_u^2 + D_i)^2 + O(m^{-2}).
$$

82
REFERENCES


BIOGRAPHICAL SKETCH

Rebecca Carter Steorts graduated from Davidson College in 2005 with a B.S. in Mathematics. She then completed an M.S. in Mathematical Sciences at Clemson University in 2007. Finally, she received her Ph.D. from the University of Florida in the Department of Statistics in 2012 under the direction of Malay Ghosh. Her current areas of interest include small area estimation, survey methodology, Bayesian methodology, and decision theory. During her time at the University of Florida she received the United States Census Bureau Dissertation Fellowship, and she was also awarded the UF Innovation through Institutional Integration (I-Cubed) Program (funded by the NSF) Teaching Award for development of a new course to the department curriculum in spring 2011. She plans to join the Department of Statistics at Carnegie Mellon University after graduation as a Visiting Assistant Professor.