ABSTRACT

Title of dissertation: SMALL AREA ESTIMATION: AN EMPIRICAL BEST LINEAR UNBIASED PREDICTION APPROACH

Huilin Li
Doctor of Philosophy, 2007

Dissertation directed by: Professor Partha Lahiri
Joint Program of Survey Methodology

In a large scale survey, we are usually concerned with estimation of some characteristics of interest for a large area (e.g., a country). But we are frequently interested in estimating similar characteristics for a subpopulation using the same survey data. The direct survey estimator which utilizes data only from the small area of interest has been found to be highly unreliable due to small sample size. Model-based methods have been used in small area estimation in order to combine information available from the survey data and various administrative and census data.

We study the empirical best linear unbiased prediction (EBLUP) and its inferences under the general Fay-Herriot small area model. Considering that the currently used variance estimation methods could produce zero estimates, we propose the adjusted density method (ADM) following Morris' comments. This new method always produces positive estimates. Morris only suggested such adjustment to the restricted maximum likelihood. Asymptotic theory of ADM is unknown. We prove

the consistency for the ADM estimator. We also propose an alternate consistent ADM estimator by adjusting the maximum likelihood. By comparing these two ADM estimators both in theory and simulation, we find that the ADM estimator using maximum likelihood is better than the one using the restricted likelihood in terms of bias. We provide a concrete proof for the positiveness and consistency of both ADM estimators.

We also propose EBLUP estimator of θ_i where we use two ADM estimators of A. The associated second-order unbiased Taylor linearization MSE estimators are also proposed.

In addition, a new parametric bootstrap prediction interval method using ADM estimator is proposed. The positiveness of ADM estimators is emphasized in the construction of the prediction interval. We also show that the coverage probability of this new method is accurate up to $O(m^{-3/2})$.

Extensive Monte Carlo simulations are conducted. A data analysis for the SAIPE data set is also presented. The positiveness of ADM estimators plays a vital role here since for this data set the method-of-moments, REML, ML and FH methods could be all zero. We observe that ADM methods produce EBLUP's which generally put more weights to the direct survey estimates than the corresponding EBLUP's that use the other methods of variance component estimation.

SMALL AREA ESTIMATION:AN EMPIRICAL BEST LINEAR UNBIASED PREDICTION APPROACH

by

Huilin Li

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2007

Advisory Committee:
Dr. Partha Lahiri, Chair/Advisor
Dr. Francis Alt
Dr. Michael Cohen
Dr. Benjamin Kedem
Dr. Paul Smith

Dedication

To my beloved parents Xueqin Guo and Shaohua Li, husband Dongming Wei and daughter Ziqi Wei.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest appreciation to my advisor, Professor Partha Lahiri, for his support, patience and encouragement throughout my graduate studies. Partha is always a great source of knowledge and inspiration for me. From him, I not only learn how to do good research but also how to be a good statistician.

I thank Professor Paul Smith and Benjamin Kedem for their consistent help and advisory during my six year doctoral study in Department of Mathematics. I thank Dr. Michael Cohen for his knowledgeable guidance during my intern at National Academy of Science. I thank Professor Francis Alt for agreeing to serve on my dissertation committee. I also thank Professor Eric Slud, Abram Kagan, Grace Yang and Zongyou Li for their valuable teaching.

I would like to thank Santanu Pramanik for his constructive discussion and help on my thesis.

An invaluable friendship of Bo, Denise, Hyejin, Jianzhu, Shanshan, Weiran, Yan, Zhiwei, Ganesh, Guanhua, Linbao, Ning, Tinghui, and Yabing is much appreciated.

I would like to thank my parents for their dedication and love. Without them, I would never be able to pursue my Ph.D. studies.

Last but not the least, I want to thank my husband Dongming just for everything.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Small Area Estimation

There is a growing need to estimate some characteristics of interest for small geographic areas using a large scale sample survey that targets a much larger population. For example, U.S. Bureau of Labor Statistics publishes the unemployment rate not only for the entire country but also for all the fifty states and the District of Columbia using the Current Population Survey. Small area statistics are used by various government agencies for a variety of purposes, including formulating policies and programs, fund allocation and regional planning, etc.

Due to the small sample size, the direct survey estimator which utilizes data only from the small area of interest has been found to be highly unreliable. Model-based methods have been used in small area estimation in order to combine survey data with various administrative and census records. The use of mixed models is widespread in small area estimation because such models can explain different sources of errors. Broadly speaking, a mixed model has both fixed effects and the random effects components. The random area-specific effects can explain between area variation that is usually neglected in the corresponding regression model.

There are two main kinds of mixed models which are used in small area estimation. The first type is the unit level model that can be used when information at the unit level is available. A good example of the unit level model is the nested error regression model considered by Battess, Harter, and Fuller (1988). They used this model to combine farm survey data and satellite data in estimating area under corn and soybeans for 12 counties of north central Iowa. One good feature of the unit level model is that it can incorporate all sources of uncertainty, including the uncertainty incurred due to estimation of sampling variances. However, at the time of analysis, the data at the unit level are frequently not accessible for a variety of reasons. For example, in order to protect the confidentiality of respondents, survey agencies may not release important information for the respondents. To deal with this situation, the second type of small area models, known as area level models, become essential. The area level models relate the small area survey estimates to area-specific auxiliary variables. Fay and Herriot (1979) introduced a basic area level model to obtain model-based estimates for per-capita income for small places (with population less than $1,000$) in the United States. They used such a model to combine the U.S. Current Population Survey data with Internal Revenue Service data and housing data. The Fay-Herriot model has been extensively used in small area estimation and related problems. For example, in order to allocate more than $7 billion of funds annually for educationally disadvantaged students, the U.S. Department of Education uses the Fay-Herriot model to obtain small area income and poverty estimates (SAIPE) of poor school-age children for counties and school districts. We will focus on the general Fay-Herriot model throughout the dissertation.

1.2 The General Fay-Herriot Model

Let y_i be a direct survey estimator of the ith small area mean θ_i and $x_i = (x_{i1}, \cdots, x_{ip})$ be a $p \times 1$ vector of associated predictor variables. The general Fay-Herriot model may be written as the following two level model:

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

Level 2 (linking model): $\theta_i|A \overset{ind}{\sim} N(x_i'\beta, b_i A)$.

In the above model, level 1 is used to incorporate errors due to sampling. Level 2 is used to link the true small area means θ_i to a vector of p known auxiliary variables x_i, which are often obtained from various administrative and census records. In the above model, the sampling variances D_i are assumed to be known. In practice, they are estimated using the generalized variance function (GVF) method (see Wolter, 1985). In many applications, $b_i = 1$. In some applications, errors in the linking model may be heteroscedastic where the factor b_i can be used to introduce such heteroscedasticity. We assume that b_i is a known positive constant. In practice, b_i may be related to D_i or x_i. Note that in the above model, there are two types of parameters - the high dimensional parameters θ_i and the low dimensional parameters β and A, usually refereed to as hyperparameters. In small area estimation, θ_i is the main object of inference which involves estimation of the unknown hyperparameters.

Note that we can write the general Fay-Herriot model as the following linear mixed model

$$y_i = \theta_i + e_i = x_i'\beta + b_i v_i + e_i, i = 1, \cdots, m, \tag{1.1}$$

where area specific random effects $v_i \overset{ind}{\sim} N(0, A)$ and the sampling errors $e_i \overset{ind}{\sim} N(0, D_i)$. Define $X = (x_1, \cdots, x_m)'$, $Z = \text{diag}(b_1, \cdots, b_m)'$, $v = (v_1, \cdots, v_m)'$, $e = (e_1, \cdots, e_m)'$, and $Y = (y_1, \cdots, y_m)'$. The equation (1.1) can be rewritten as

$$Y = X\beta + Zv + e, \tag{1.2}$$

which is a special case of the general linear mixed model with block diagonal covariance structure. The variance-covariance matrix of Y is $\Sigma(A) = D + AR$, where $D = \text{diag}(D_1, \cdots, D_m)$ and $R = \text{diag}(b_1^2, \cdots, b_m^2)$.

1.3 Empirical Best Linear Unbiased Prediction

We are interested in infering about the small area means θ_i. Empirical best linear unbiased (EBLUP) methods are extensively discussed in the small area literature (see Rao, 2003). When A is known, the best linear unbiased prediction (BLUP) of θ_i, under the general Fay-Herriot model (1.1), is given by:

$$\hat{\theta}_i(Y; A) = (1 - B_i)y_i + B_i x_i' \tilde{\beta}(A), \tag{1.3}$$

where $B_i = D_i/(b_i^2 A + D_i), i = 1, \cdots, m$ and $\tilde{\beta}(A) = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y$.

The BLUP estimator $\tilde{\theta}_i(Y; A)$ depends on the unknown variance component A. We can estimate A from the the marginal distribution of Y. An EBLUP of θ_i is obtained by replacing A in BLUP by an estimator $\hat{A}$:

$$\hat{\theta}_i(Y; \hat{A}) = (1 - \hat{B}_i) y_i + \hat{B}_i x_i' \hat{\beta}, \tag{1.4}$$

where $\hat{B}_i = D_i/(b_i^2 \hat{A} + D_i), i = 1, \cdots, m$ and $\hat{\beta}$ is $\tilde{\beta}$ with A is replaced by $\hat{A}$.

Kacker and Harville (1981) showed that $\hat{\theta}_i$ is unbiased for θ_i, under the conditions: (i) $\mathrm{E}(\hat{\theta}_i)$ is finite; (ii) $\hat{A}$ is any even translation invariant estimator of A, that is, $\hat{A}(-Y) = \hat{A}(Y)$ and $\hat{A}(Y - Xb) = \hat{A}(Y)$ for all Y and b; (iii) the distributions of v and e are both symmetric around 0.

The EBLUP estimator $\hat{\theta}_i$ (1.4) is indeed a weighted average of a direct estimate y_i and a model estimate $x_i'\hat{\beta}$. The weight $\hat{B}_i$ depends on the estimate of the ratio between sampling variance D_i and model variance $b_i^2 A$.

1.4 Outline of the Dissertation

In Chapter 2 of this dissertation, we first review four currently used variance component estimation methods, including the method-of-moments which uses the weighted or unweighted residual sum of squares (Fay & Herriot, 1979; Prasad & Rao, 1990), maximum likelihood method and residual maximum likelihood method. The common shortcoming of all of the four methods is that they could produce zero estimates. Following Morris' comments in Jiang & Lahiri (2006), we propose a new adjusted density maximization method (ADM), which always produces positive estimates. Morris only suggested such adjustment to the restricted maximum likelihood. Asymptotic theory for ADM is unknown. In this chapter, we prove consistency for the ADM estimator. We also propose an alternate consistent ADM estimator by adjusting the maximum likelihood. By comparing these two ADM estimators both in theory and simulation, we find that the ADM estimator using maximum likelihood is better than the one using the restricted likelihood in terms of bias. We provide a concrete proof for the positiveness and consistency of both ADM estimators.

In Chapter 3, we propose an EBLUP estimator of θ_i where we use the two ADM estimators of A. Associated second-order unbiased Taylor linearization MSE estimators are also proposed.

In Chapter 4, a new parametric bootstrap prediction interval method using ADM estimators is proposed. The positiveness of ADM estimators is emphasized in the construction of the prediction interval. We also show that the coverage probability of this new method is accurate up to $O(m^{-3/2})$.

Finally, in Chapter 5, we present a data analysis for the SAIPE data set. The positiveness of the ADM estimators plays a vital role here since for this data set the method-of-moments, REML, ML and FH methods could be all zero. We observe that ADM methods produce EBLUP's which generally put more weight on the direct survey estimates than the corresponding EBLUP's that use the other methods of variance component estimation.

Chapter 2

Variance Component Estimation

In the small area estimation with the Fay-Herriot model, estimation of the variance component A comes as an intermediate step. Accurate estimation of A is necessary in order to obtain an efficient EBLUP for the small area means θ_i. A variety of estimators of A has been considered in the literature. The methods include the Prasad-Rao simple closed-form method-of-moments (Prasad and Rao, 1990), the Fay-Herriot iterated method-of-moments (Fay and Herriot, 1979), the maximum likelihood (ML) method and the residual maximum likelihood (REML) method. The estimators are all consistent for large m, under certain regularity conditions. However, all of these methods could produce zero estimate for A. When this happens, it creates problems in real data analysis since EBLUP reduces to the synthetic regression estimator that could lead to a serious overshrinkage problem.

In this Chapter, we first briefly review the four commonly used estimators of A. We then propose two new Adjusted Density Maximization (ADM) methods. The first method follows the original ideas of Morris (Morris 1988) and adjusts the residual maximum likelihood. The second method adjusts the maximum likelihood in an effort to reduce the bias of the ADM estimator based on the REML likelihood. One of the good features of the ADM estimators is that both produces strictly positive estimates of A which, in turn, ensures $0 < \hat{B}_i < 1$. Thus, the resulting

EBLUP of θ_i is never synthetic and is always a weighted combination of the direct estimator and the synthetic regression estimator. Also this strict positiveness of the ADM estimators enhances the good performance of parametric bootstrap prediction interval method which we will discuss in Chapter 4. The asymptotic theory of ADM estimators is also provided in this Chapter.

2.1 The Four Commonly Used Estimators of A

In this section, we briefly review four commonly used estimators of A: the Prasad-Rao simple method-of-moments estimator (Prasad & Rao, 1990), $\hat{A}^{\text{PR}}$; the Fay-Herriot method-of-moments estimator (Fay and Herriot, 1979), $\hat{A}^{\text{FH}}$; the maximum likelihood estimator (ML), $\hat{A}^{\text{ML}}$; and the residual maximum likelihood estimator (REML), $\hat{A}^{\text{RE}}$.

1. The Prasad-Rao Method-of-Moments Estimator

Prasad and Rao (1990) proposed a simple method-of-moments to estimate A. The method can be viewed as the well-known method of fitting constants, commonly refereed to as Henderson's method 3. The estimator is given by

$$\tilde{A}^{\text{PR}} = \frac{1}{m-p}[l'l - \sum D_i(1 - x_i'(X'X)^{-1}x_i)], \tag{2.1}$$

where $l = Y - X\hat{\beta}$ and $\hat{\beta} = (X'X)^{-1}X'Y$. Note that equation (2.1) could yield a negative estimate. In order to avoid the problem, Prasad and Rao proposed the following estimator of A: $\hat{A}^{\text{PR}} = \max(\tilde{A}, 0)$. Evidently, $\hat{A}^{\text{PR}}$ satisfies the conditions

$\hat{A}(-Y) = \hat{A}(Y)$ and $\hat{A}(Y - Xb) = \hat{A}(Y)$, so $\hat{A}^{\text{PR}}$ is an even translation invariant estimator of A. Under certain regularity conditions, the Prasad-Rao estimator $\hat{A}^{\text{PR}}$ is a consistent estimator of A for large m.

2. The Fay-Herriot Method-of-Moments Estimator

The Fay and Herriot (1979) estimator of A is based on the weighted least squares residual sum of squares. Using the best linear unbiased estimator of β,

$$\tilde{\beta}^{\text{W}} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y. \tag{2.2}$$

The estimator of $\hat{A}^{FH}$ is obtained by solving the following equation iteratively:

$$\sum_i \frac{(Y_i - x_i\tilde{\beta}^{\text{W}})^2}{b_i^2\tilde{A}^{\text{FH}} + D_i} = m - p. \tag{2.3}$$

The left side of equation (2.3) is the weighted residual sum of squares whose expectation, under the Fay-Herriot model, is identical to the right hand side. This is the main motivation for the Fay-Herriot estimator. Again the solution of (2.3) could be negative and so the following estimator is used in practice: $\hat{A}^{\text{FH}} = \max(\tilde{A}^{\text{FH}}, 0)$. It is easy to check that $\hat{A}^{\text{FH}}$ satisfies the conditions $\hat{A}(-Y) = \hat{A}(Y)$ and $\hat{A}(Y - Xb) = \hat{A}(Y)$, so $\hat{A}^{\text{FH}}$ is an even translation invariant estimator of A. Like the Prasad-Rao estimator, the Fay-Herriot estimator of A is consistent for large m, under certain regularity conditions. In the simulation experiment of Datta, Rao and Smith (2005), the Fay-Herriot estimator was found to be best among the four estimators of A in term of relative bias of the corresponding mean squared error estimator of EBLUP.

3. The REML Estimator

Patterson and Thompson (1971) proposed the restricted or residual maximum likelihood (REML) approach. The approach uses transformed data which do not include the inferences about the nuisance parameters β. The REML method produces unbiased variance component estimators no matter whether rank$(X) = p$ is fixed or goes to infinity. Jiang (1996) showed that the REML estimators are consistent and asymptotically normally distributed even when the normality assumptions in the mixed linear model do not hold.

Under the Fay-Herriot model (1.2), the restricted log likelihood has the form

$$l_{\mathrm{R}}(A) = c - (1/2)[\log(|F'\Sigma F|) + y'Py], \tag{2.4}$$

where c is a constant, $|F'\Sigma F|$ is the determinant of $F'\Sigma F$, F is any $m \times (m-p)$ matrix such that rank$(F) = m - p$ and $F'X = 0$, and

$$\begin{aligned} P &= F(F'\Sigma F)^{-1}F' && (2.5)\\ &= \Sigma^{-1} - \Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}. && (2.6) \end{aligned}$$

The first derivative of (2.4) is given as

$$\begin{aligned} \frac{\partial l_{\mathrm{R}}(A)}{\partial A} &= \frac{1}{2}[y'PRPy - \mathrm{tr}(PR)] && (2.7)\\ &= \frac{1}{2}[u'PRPu - \mathrm{tr}(PR)], && (2.8) \end{aligned}$$

where $u = y - X\beta$ and $R = \text{diag}(b_1^2, \cdots, b_m^2)$. The last equality is due to the fact $PX = 0$ which also shows that $\hat{A}^{\text{RM}}$ is translation invariant. The REML of A is obtained as $\hat{A}^{\text{RE}} = \max(\tilde{A}^{\text{RE}}, 0)$ where $\tilde{A}^{RE}$ is a solution to $(\partial/\partial A)l_{\text{R}}(A) = 0$.

4. ML Estimator

The maximum log-likelihood under the Fay-Herriot model (1.2) has the form

$$l_{\text{M}}(\beta, A) = c - (1/2)[\log(|\Sigma|) + (y - X\beta)'\Sigma^{-1}(y - X\beta)], \tag{2.9}$$

where c is a constant. By differentiating (4) with respect to A and β, we have

$$\begin{aligned} \frac{\partial l_{\text{M}}(\beta, A)}{\partial \beta} &= X'\Sigma^{-1}y - X'\Sigma^{-1}X\beta, & (2.10) \\ \frac{\partial l_{\text{M}}(\beta, A)}{\partial A} &= (1/2)[(y - X\beta)'\Sigma^{-1}R\Sigma^{-1}(y - X\beta) - \text{tr}(\Sigma^{-1}R)]. & (2.11) \end{aligned}$$

where $R = \text{diag}(b_1^2, \cdots, b_m^2)$. From (2.10), letting $(\partial/\partial\beta)l_{\text{M}}(\beta, A) = 0$, we obtain $\tilde{\beta}(A) = (X'\Sigma X)^{-1}X'\Sigma^{-1}y$. Replacing β by $\tilde{\beta}(A)$ in (2.9), we obtain the following profile log-likelihood after some algebra:

$$l_{\text{MP}}(A) = c - (1/2)[\log(|\Sigma|) + y'Py]. \tag{2.12}$$

The first derivative of (2.12) is given as

$$\begin{aligned} \frac{\partial l_{\text{MP}}(A)}{\partial A} &= \frac{1}{2}[y'PRPy - \text{tr}(\Sigma^{-1}R)] & (2.13) \\ &= \frac{1}{2}[u'PRPu - \text{tr}(\Sigma^{-1}R)], & (2.14) \end{aligned}$$

where $u = y - X\beta$ and $R = \text{diag}(b_1^2, \cdots, b_m^2)$. The last equality is due to the fact $PX = 0$, and it shows that $\hat{A}^{\text{ML}}$ is translation invariant. The maximum likelihood estimator of A is given by $\hat{A}^{\text{ML}} = \max(\tilde{A}^{\text{ML}}, 0)$, where $\tilde{A}^{\text{ML}}$ is a solution to $(\partial/\partial A) l_{\text{MP}}(A) = 0$. The estimator $\hat{A}^{\text{ML}}$ obtained from the profile maximum likelihood is the same as the one by solving maximum log-likelihood (2.9) (See Jiang, 2007).

Miller (1977) showed that the maximum likelihood estimators are consistent and asymptotically normal. Since the maximum likelihood variance estimators do not take account of uncertainty in estimates of fixed effects parameters β, they are generally biased. If $\text{rank}(X) = p$ is fixed, their biases diminish as m tends to ∞.

A well-known problem associated with all of the above four variance component estimation methods is that all could yield a zero estimate, especially when the number of small areas is small. The zero estimate of A yields $\hat{B}_i = 0$ and consequently the EBLUP estimator of θ_i reduces to the synthetic regression estimator (see Morris' discussion of Jiang and Lahiri, 2006). Moreover, the zero estimate of A creates problem in the parametric bootstrap prediction interval which we will discuss more in Chapter 4. In the next section, we will propose a new consistent and strictly positive estimator of A.

2.2 The Adjusted Density Method

The maximum likelihood estimator of A can be regarded as the posterior mode when the profile likelihood is treated as the posterior density of A and a normal approximation is made to the profile likelihood. For a large sample, the performance of the normal approximation is reasonably good. However, when the sample size is small, the posterior distribution of A is usually skewed. In addition, the range of A is $(0, \infty)$ instead of $(-\infty, \infty)$ suggested by the normal approximation. For small samples, the normal approximation performs poorly (see Morris' discussion of Jiang and Lahiri, 2006). Similar situation arises for the REML method. Here one can regard the restricted maximum likelihood density as the posterior density of A. The adjusted density method (ADM) we discuss in this section fits the posterior density of A by a gamma distribution instead of normal distribution.

The ADM estimator was first proposed, but not named, by Morris (1988). Morris suggested to use a suitable distribution in the Pearson family of distributions to approximate an univariate density, especially when the given density is skewed, bounded or semi-bounded, in which case the standard normal distribution does not work well. The Pearson family includes Gamma, Reciprocal Gamma, Beta, F, t, Normal distribution, and others. Given a unimodal density/likelihood $f(x)$, we can approximate the density by a distribution in the Pearson family using the range of the estimated parameter or the suspected direction of skewness. Multiplying $f(x)$ by $Q(x)$, the characterizing quadratic term for the selected Pearson distribution (see Table 1 of Morris, 1988), we can obtain the adjusted density/likelihood $Q(x)f(x)$. By computing the first two derivatives of the adjusted log-likelihood, we can get the first and second moment estimates of x.

Since the range of A is $(0, \infty)$, and the likelihood of A is almost always right-skewed (Morris' discussion in Jiang and Lahiri, 2006), Morris suggested to choose the

Gamma distribution as the approximating distribution in the Pearson family. Then the corresponding adjustment we need to do is to multiply the original likelihood by the Gamma quadratic function: A. Thus we maximize $A\mathrm{L}(A|Y)$ instead of $\mathrm{L}(A|Y)$ to obtain the new estimate of A, where $\mathrm{L}(A|Y)$ is the studied likelihood.

For general Fay-Herriot model (1.1), we have two options to estimate variance component A by applying the new adjusted density maximization method. We can start with the restricted likelihood and regard it as the posterior density of A, as suggested by Morris. Then the adjusted log-likelihood is given by

$$l_{\mathrm{AR}}(A) = c - (1/2)[\log(|F'\Sigma F|) + y'Py] + \log(A). \tag{2.15}$$

By solving $(\partial/\partial A)l_{\mathrm{AR}}(A) = 0$, we obtain the ADM estimator $\hat{A}^{\mathrm{AR}}$. We can also use the profile maximum likelihood, and regard it as the posterior density of A. Then the adjusted log-likelihood is given by

$$l_{\mathrm{AMP}}(A) = c - (1/2)[\log(|\Sigma|) + y'Py] + \log(A), \tag{2.16}$$

By solving $(\partial/\partial A)l_{\mathrm{AMP}}(A) = 0$, we obtain an alternate ADM estimator: $\hat{A}^{\mathrm{AM}}$.

Both $\hat{A}^{\mathrm{AR}}$ and $\hat{A}^{\mathrm{AM}}$ are even translation invariant estimators of A. Both ADM estimators satisfy the following two conditions: $\hat{A}(-Y) = \hat{A}(Y)$ and $\hat{A}(Y - Xb) = \hat{A}(Y)$ for all Y and b. This property follows from (2.14) and (2.8) and the fact that the difference between the first derivatives of ADM estimators and those of ML/REML estimators is the term $1/A$.

Both $\hat{A}^{\text{AR}}$ and $\hat{A}^{\text{AM}}$ are guaranteed to be strictly positive. This strict positiveness is very important in the practice. We shall come back to this issue later in Chapter 5 when we analyze the SAIPE data analysis.

The ADM estimators of A are not unbiased. The biases of these two ADM estimators are given by

$$\text{Bias}(\hat{A}^{\text{AR}}) = \text{E}[\hat{A}^{\text{AR}}] - A = \frac{2/A}{\text{tr}(\Sigma^{-2}R^2)} + o(m^{-1}), \tag{2.17}$$

$$\text{Bias}(\hat{A}^{\text{AM}}) = \text{E}[\hat{A}^{\text{AM}}] - A = \frac{\text{tr}(PR - \Sigma^{-1}R) + 2/A}{\text{tr}(\Sigma^{-2}R^2)} + o(m^{-1}). \tag{2.18}$$

Interestingly, under certain regularity conditions, we always have $\hat{A}^{\text{AR}} > \hat{A}^{\text{AM}}$ so that $\text{Bias}(\hat{A}^{\text{AR}}) > \text{Bias}(\hat{A}^{\text{AM}})$. The REML estimation of A has a bias of order $o(m^{-1})$. In contrast, the ML estimator of A suffers from a bias of order $O(m^{-1})$ and the order $O(m^{-1})$ term in the bias is negative; that is it underestimates. Thus, in terms of bias criterion, the REML method is superior to the ML method. The ADM methods correct the shortcomings of REML/ML, e.g. zero estimates and underestimation. However, the ADM REML corrects too much and produces larger biases than the corresponding ADM ML method. So in the sense of bias, ADM ML is better than ADM REML. In the EBLUP methodology, estimation of the shrinkage factor B_i is critical. Since $\hat{B}_i = D_i/(b_i^2\hat{A} + D_i)$ is a convex function of $\hat{A}$, the overestimations of ADM estimators of A make up the further underestimation of $\hat{B}_i$ caused by the convexity. This conclusion is also verified in the simulation.

Although ADM estimators are biased, the ADM REML estimator is asymptotically unbiased no matter what kind of conditions are assumed about $\text{rank}(X) = p$.

The ADM ML estimator, like the ML estimator, is asymptotically unbiased if $\mathrm{rank}(X) = p$ is fixed or bounded. Moreover, we have shown that ADM REML and ADM ML are both consistent estimators of A. These asymptotic properties support the usage of ADM estimators.

All detailed computations and technical proofs are deferred to the last section.

2.3 Monte Carlo Simulation Study

In this section, using a Monte Carlo simulation, we investigate the small sample performances of the various variance component estimators of A: $\hat{A}^{\mathrm{PR}}$, $\hat{A}^{\mathrm{FH}}$, $\hat{A}^{\mathrm{RE}}$, $\hat{A}^{\mathrm{ML}}$, $\hat{A}^{\mathrm{AR}}$ and $\hat{A}^{\mathrm{AM}}$, as well as the performances of the corresponding estimators of the shrinkage factor B_i.

Following Datta, Rao and Smith (2005), we generate $N = 10,000$ independent data sets $\{y_i, i = 1, \cdots, 15\}$ using a simplified Fay-Herriot model $y_i = \mu + v_i + e_i$, where $v_i \sim N(0, A = 1)$, and $e_i \sim N(0, D_i)$. Since all the estimators of A we considered here are translation invariant, we take $\mu = 0$, without loss generality. The 15 small areas are divided into 5 groups, each group containing 3 small areas with the same D_i values. In this simulation, we consider the following three D_i patterns:

1. Pattern 1: $D_i = (0.7, 0.6, 0.5, 0.4, 0.3)$;
2. Pattern 2: $D_i = (2.0, 0.6, 0.5, 0.4, 0.2)$;
3. Pattern 3: $D_i = (4.0, 0.6, 0.5, 0.4, 0.1)$.

The sampling variances D_i are more or less similar for pattern 1. The second and third patterns are more unbalanced than pattern 1, the D_i's are more dispersed in pattern 3 than in pattern 2.

We use the following criteria to compare and contrast different estimators of A and B_i. Let $\hat{A}^{(j)}$ be the estimator of true A in the jth simulation run.

- Bias: $\text{Bias}(\hat{A}) = \sum_{j=1}^{N} \hat{A}^{(j)}/N - A$;
- Relative Bias: $\text{RB}(\hat{A}) = \text{Bias}(\hat{A})/A$;
- Coefficient of Variation: $\text{CV}(\hat{A}) = \sqrt{\sum(\hat{A}^{(j)} - \bar{\hat{A}})^2/(N-1)}/\bar{\hat{A}}$, $\bar{\hat{A}} = \sum_{j=1}^{N} \hat{A}^{(j)}/N$;
- Absolute Relative Bias: $\text{ARB}(\hat{A}) = \sum_{j=1}^{N} |\hat{A}^{(j)} - A|/(NA)$;
- Mean Squared Error: $\text{MSE}(\hat{A}) = \sum_{j=1}^{N}(\hat{A}^{(j)} - A)^2/N$;

Since all the small areas are exchangeable in each group, we only report the group means for all the criteria.

From Tables 2.1–2.7, we conclude the following:

1. Table 2.1 reports the count of zero estimates for the six variance component estimation methods we considered in this dissertation. The ADM methods never yield zero estimates, because their positiveness is guaranteed in theory. All the other methods produce about $1\% - 2\%$ zero estimates for Pattern 1, and this count tends to increase as the variability of D becomes larger. For pattern 3, the Prasad-Rao method yields as high as 12% zero estimates. When the percentage of the zero estimates is so large, the validity of the inference from this variance estimate is in suspect.

2. Table 2.2 provides comparison of different estimators of A. Consistent with the theory, the ADM methods yield positive biases and the bias of the ADM ML is much smaller than that of ADM REML. The PR, FH and REML estimators of A produce almost zero biases in all the patterns, since all of them are approximately unbiased estimators. As the theory indicates, the ML method always underestimates. The above results are based on the "Bias" criteria. However, when we turn to "Absolute Relative Bias", although the ADM methods still have larger ARB than most of the non-ADM methods, the differences become smaller. As the variability of the D_i increases, the bias differences between the ADM estimators and the other competitors become smaller. The MSE of ADM estimators are a little bit larger than those of the others.

3. Tables 2.3-2.7 compare different estimators of the shrinkage factor B_i, an important factor in the EBLUP method. Regarding the bias, Tables 2.3-2.6 show that the ADM estimators, especially the ADM ML, outperform all other four estimators. All the non ADM methods have the problem of underestimation. In term of the MSE (Table 2.7), although all the methods have comparable MSE's, the ADM methods have the smaller ones than the other four methods.

2.4 Proofs

In this section, the letter r, with or without suffix, will be generic for constants.

2.4.1 Positiveness of the ADM Estimator

Lemma 2.1 *Let $f(x)$ be a continuous positive function and $\lim_{x\to\infty} xf(x) = 0$. Then for $g(x) = xf(x)$, there exists x_0 where $g(x_0) = \max_x g(x)$ and $x_0 > 0$.*

Proof: Since $f(x) > 0$, we have $g(x) \leq 0$ when $x \leq 0$ and $g(x) > 0$ when $x > 0$. In addition, $g(x) \to 0$ as $x \to \infty$. So there exists x_0 where $g(x_0) = \max_x g(x)$ and $x_0 > 0$.

Theorem 2.1 $\hat{A}^{\mathrm{AR}} > 0$ *and* $\hat{A}^{\mathrm{AM}} > 0$.

Proof: The positiveness of the ADM estimator can be easily derived from Lemma (2.1).

2.4.2 Asymptotic Properties of the ADM Estimators

In order to prove the asymptotic properties of the ADM REML and ADM ML estimators, we first obtain the asymptotic representations of $\hat{A}^{\mathrm{AR}} - A$ and $\hat{A}^{\mathrm{AM}} - A$ for the general Fay-Herriot model.

Theorem 2.2 *Under the general Fay-Herriot model (1.1), we have*

$$\hat{A}^{\mathrm{AR}} - A = \frac{y'PRPy - \mathrm{tr}(PR) + 2/A}{\mathrm{tr}(\Sigma^{-2}R^2)} + r_{\mathrm{R}}, \tag{2.19}$$

$$\hat{A}^{\mathrm{AM}} - A = \frac{y'PRPy - \mathrm{tr}(\Sigma^{-1}R) + 2/A}{\mathrm{tr}(\Sigma^{-2}R^2)} + r_{\mathrm{M}}, \tag{2.20}$$

where $\mathrm{E}(r_{\mathrm{R}}) = o(m^{-1})$ *and* $\mathrm{E}(r_{\mathrm{M}}) = o(m^{-1})$.

To prove Theorem 2.2, we need the following lemmas.

Lemma 2.2 *Let Q be a symmetric matrix, and $\xi \sim N(0, I)$. Then, for any $t \leq 2$, there is a constant c that only depends on t such that $\mathrm{E}|\xi' Q\xi - \mathrm{E}\xi' Q\xi|^t \leq c||Q||_2^t$. ($||Q||_2$ is defined as $[\mathrm{tr}(Q'Q)]^{1/2}$.)*

This is the Lemma 5.1 in Das, Jiang, and Rao (2004).

Corollary 2.1 *Under the General Fay-Herriot Model (1.2), for any positive integer t and k, $\mathrm{E}|y'(PR)^{k-1}Py - \mathrm{tr}((PR)^{k-1})|^t < \infty$.*

Proof: Let $V = [\sqrt{D}, \sqrt{A}Z]$. Then $u = Zv + e = V\xi$, where $\xi \sim N(0, I_{2m})$. Then $y'P^k y = u'(PR)^{k-1}Pu = \xi' V(PR)^{k-1}PV\xi$. By Lemma (2.2) and the fact that $\mathrm{E}(y'(PR)^{k-1}Py) = \mathrm{tr}((PR)^{k-1})$, we have for any $t \geq 2$, $\mathrm{E}|y'(PR)^{k-1}Py - \mathrm{tr}((PR)^{k-1})|^t < c||V(PR)^{k-1}PV||_2^t < \infty$.

When $t = 1$, $\mathrm{E}|y'(PR)^{k-1}Py - \mathrm{tr}((PR)^{k-1})| \leq 2\mathrm{tr}((PR)^{k-1}) < \infty$. ■

Lemma 2.3 *(Theorem 2.1 of Das, Jiang, and Rao, 2004) For any $\hat{\theta}$ which is obtained as a solution to a "score" equation of the form $(\partial/\partial\theta)l(\theta) = 0$, suppose that*

i) $l(\theta) = l(\theta, y)$ is three times continuously differentiable with respect to $\theta = (\theta_1, \cdots, \theta_s)'$, where $y = (y_1, \cdots, y_n)'$;

ii) $\theta_0 \in \Theta_0$, the interior of Θ;

iii) $-\infty < \limsup_{n\to\infty} \lambda_{\max}(G^{-1}HG^{-1}) < 0$, where $\lambda_{\max}$ means the largest eigenvalue, $H = \mathrm{E}[\partial^2 l(\theta)/\partial\theta^2|_{\theta_0}]$, and $G = \mathrm{diag}(g_1, \cdots, g_s)$ with $g_i > 0$, $1 \leq i \leq s$ such that $g_* = \min_{1\leq i\leq s} g_i \to \infty$ as $n \to \infty$;

iv) the t-th moments of the following are bounded $(t > 0)$*:*

$$\frac{1}{g_i}\left|\frac{\partial l(\theta)}{\partial \theta_i}\Big|_{\theta_0}\right|, \quad \frac{1}{\sqrt{g_i g_j}}\left|\frac{\partial^2 l(\theta)}{\partial \theta_i \partial \theta_j}\Big|_{\theta_0} - \mathrm{E}\left[\frac{\partial^2 l(\theta)}{\partial \theta_i \partial \theta_j}\Big|_{\theta_0}\right]\right|, \quad \frac{g_*}{g_i g_j g_k} M_{ijk}(y), \;\; 1 \le i, j, k \le s,$$

where $M_{ijk}(y) = \sup_{\theta \in S_\delta(\theta_0)} |\partial^3 l(\theta)/\partial\theta_i \partial\theta_j \partial\theta_k|$ *with* $S_\delta(\theta_0) = \{\theta : |\theta_i - \theta_{0i}| \le \delta g_*/g_i, 1 \le i \le s\}$ *for some* $\delta > 0$*.*

Then there exists $\hat{\theta}$ *such that for any* $0 < \rho < 1$*, there is a set* $\mathcal{B}$ *satisfying for large* n *and on* $\mathcal{B}$*,*

$$\hat{\theta} \in \Theta, \quad \frac{\partial l(\theta)}{\partial \theta}\Big|_{\hat{\theta}} = 0, \quad |G(\hat{\theta} - \theta_0)| < g_*^{1-\rho},$$

and

$$\hat{\theta} - \theta_0 = -H^{-1} a + r,$$

where $a = \partial l(\theta)/\partial\theta|_{\theta_0}$*, and* $|r| \le g_*^{-2\rho}\eta$ *with* $\mathrm{E}(\eta^t)$ *bounded; and* $\mathrm{P}(\mathcal{B}^c) \le c g_*^{-\tau t}$*, where* $\tau = (1/4) \wedge (1 - \rho)$ *and* c *is a constant.*

Remark: $l(\theta)$ could be any form of log-likelihood, including the restricted log-likelihood, the profile log-likelihood, the adjusted restricted likelihood or the adjusted profile log-likelihood.

Proof of Theorem (2.2): We prove this Theorem by verifying the four conditions of Lemma (2.3),

We first consider the ADM REML estimation. Note that $\hat{A}^{\mathrm{AR}}$ is the solution of $(\partial/\partial A) l_{\mathrm{AR}}(A) = 0$, where

$$l_{\mathrm{AR}}(A) = c - \frac{1}{2}\log|F'\Sigma F| - \frac{1}{2} y' P y + \log A. \qquad (2.21)$$

Then, by differentiating with respect to A three times, we have

$$\frac{\partial l_{\mathrm{AR}}}{\partial A} = \frac{1}{2}[y'PRPy - \mathrm{tr}(PR)] + \frac{1}{A}, \tag{2.22}$$

$$\frac{\partial^2 l_{\mathrm{AR}}}{\partial A^2} = -y'(PR)^2Py + \frac{1}{2}\mathrm{tr}((PR)^2) - \frac{1}{A^2}, \tag{2.23}$$

$$\frac{\partial^3 l_{\mathrm{AR}}}{\partial A^3} = 3y'(PR)^3Py - \mathrm{tr}((PR)^3) + \frac{2}{A^3}. \tag{2.24}$$

So, the first condition holds: l_{AR} is three times continuously differentiable.

Note that, by identities $PX = 0$ and $P\Sigma P = P$, we have $\mathrm{E}y'(PR)^kPy = \mathrm{tr}((PR)^k)$. Then, we have

$$\mathrm{E}(\partial^2 l_{\mathrm{AR}}/\partial A^2) = -(1/2)\mathrm{tr}((PR)^2) - 1/A^2. \tag{2.25}$$

Let $\max(b_i^2, D_i) \leq C, \;\; 1 \leq i \leq m$. It is easy to show that

$$\frac{m-p}{(A+1)^2} \leq \mathrm{tr}((PR)^2) \leq \frac{m-p}{A^2}. \tag{2.26}$$

If we take $G = \sqrt{m}$ (note that for the general Fay-Herriot model, there is only one g, so $G = g$), then the third condition holds: that is,

$$-\infty < \limsup_{m\to\infty} \lambda_{\max}(G^{-1}\mathrm{E}(\partial^2 l_{\mathrm{AR}}/\partial A^2)G^{-1}) < 0.$$

Finally, we need to verify the last condition—to show that the t^{th} moments of those three terms are bounded. For the first two terms, we have

$$E\Big(\frac{1}{\sqrt{m}}\Big|\frac{\partial l_{\rm AR}}{\partial A}\Big|\Big)^t = \frac{1}{(2\sqrt{m})^t}E\Big(\Big|y'PRPy - {\rm tr}(PR) + \frac{2}{A}\Big|\Big)^t, \quad \text{and}$$

$$E\Big(\frac{1}{\sqrt{m}}\Big|\frac{\partial^2 l_{\rm AR}}{\partial A^2} - E\Big(\frac{\partial^2 l_{\rm AR}}{\partial A^2}\Big)\Big|\Big)^t = \frac{1}{m^{t/2}}E\Big(\Big|y'(PR)^2Py - {\rm tr}((PR)^2)\Big|\Big)^t.$$

By Corollary 2.1, it is easy to show that those two terms are bounded for any $t > 0$.

Now, we examine the third term. When $\delta = A/2$, we have $A/2 \le \tilde{A} \le 3A/2$. With $\max(b_i^2, D_i) \le C,\ \ 1 \le i \le m$, $y'(\tilde{P}R)^3\tilde{P}y = u'(\tilde{P}R)^3\tilde{P}u \le c(2/(A)^4\sum_{i=1}^m u_i^2$, where $\tilde{P}$ is obtained when A is replaced by $\tilde{A}$ in P. Then we have

$$E[y'(\tilde{P}R)^3\tilde{P}y]^t \le c(\frac{2}{A})^{4t}E[\sum_{i=1}^m u_i^2]^t \le c(\frac{2}{A})^{4t}E[m\max_i(u_i^2)]^t < \infty. \qquad (2.27)$$

In addition, we have $(m-p)/(1+\tilde{A})^3 \le {\rm tr}((\tilde{P}R)^3) \le (m-p)/\tilde{A}^3$, then $E[{\rm tr}((\tilde{P}R)^3)]^g < \infty$. Thus, by (2.34),

$$E\Bigg(\frac{1}{m}\sup_{A/2\le\tilde{A}\le 3A/2}\Big|\frac{\partial^3 l_{\rm AR}}{\partial A^3}|_{\tilde{A}}\Big|\Bigg)^t = \frac{1}{m^t}E\Bigg(\sup_{A/2\le\tilde{A}\le 3A/2}\Big|3y'(\tilde{P}R)^3\tilde{P}y - {\rm tr}((\tilde{P}R)^3) + \frac{2}{\tilde{A}^3}\Big|\Bigg)^t$$

is bounded for any $t > 0$.

Since all the four conditions of Lemma 2.3 are satisfied, we have

$$\hat{A}^{\rm AR} - A = -E\Big[\frac{\partial^2 l_{\rm AR}}{\partial A^2}\Big]^{-1}\frac{\partial l_{\rm AR}}{\partial A} + r_{\rm R} \qquad (2.28)$$

$$= \frac{y'PRPy - {\rm tr}(PR) + 2/A}{{\rm tr}((PR)^2) + 2/A^2} + r_{\rm R} \qquad (2.29)$$

$$= \frac{y'PRPy - \mathrm{tr}(PR) + 2/A}{\mathrm{tr}(\Sigma^{-2}R^2)} + r_{\mathrm{R}}, \tag{2.30}$$

where $|r_{\mathrm{R}}| \leq m^{-2\rho}\eta$ with $\mathrm{E}(\eta^t)$ bounded. With $\rho = 1/2$, we have $\mathrm{E}(r_{\mathrm{R}}) = o(m^{-1})$.

Next we turn to the ADM ML estimation. Note that $\hat{A}^{\mathrm{AM}}$ is the solution of $(\partial/\partial A)l_{\mathrm{AMP}}(A) = 0$, where

$$l_{\mathrm{AMP}}(A) = c - \frac{1}{2}\log|\Sigma| - \frac{1}{2}y'Py + \log A. \tag{2.31}$$

Then, by differentiating with respect to A three times, we have

$$\frac{\partial l_{\mathrm{AMP}}}{\partial A} = \frac{1}{2}[y'P^2y - \mathrm{tr}(\Sigma^{-1})] + \frac{1}{A}, \tag{2.32}$$

$$\frac{\partial^2 l_{\mathrm{AMP}}}{\partial A^2} = -y'P^3y + \frac{1}{2}\mathrm{tr}(\Sigma^{-2}) - \frac{1}{A^2}, \tag{2.33}$$

$$\frac{\partial^3 l_{\mathrm{AMP}}}{\partial A^3} = 3y'P^4y - \mathrm{tr}(\Sigma^{-3}) + \frac{2}{A^3}. \tag{2.34}$$

So, the first condition holds — l_{AMP} is three times continuously differentiable.

As for the last two conditions, comparing the derivatives of l_{AMP} with l_{AR}, we can see that the differences are only $\mathrm{tr}(P^k) - \mathrm{tr}(\Sigma^{-k})$, for $k = 1, 2, 3$. These differences are obviously bounded and tend to zero when p is fixed. Thus, following the previous proof for l_{AR}, we can show that l_{AMP} also satisfies the conditions of Lemma 2.3. Then we have

$$\hat{A}^{\mathrm{AM}} - A = -\mathrm{E}\left[\frac{\partial^2 l_{\mathrm{AM}}}{\partial A^2}\right]^{-1}\frac{\partial l_{\mathrm{AM}}}{\partial A} + r_{\mathrm{M}} \tag{2.35}$$

$$= \frac{u'PRPu - \mathrm{tr}(\Sigma^{-1}R) + 2/A}{2\mathrm{tr}(PR)^2 - \mathrm{tr}(\Sigma^{-2}R^2) + 2/A^2} + r_{\mathrm{M}} \tag{2.36}$$

$$= \frac{u'PRPu - \mathrm{tr}(\Sigma^{-1}R) + 2/A}{\mathrm{tr}(\Sigma^{-2}R^2)} + r_{\mathrm{M}}, \qquad (2.37)$$

where $|r_{\mathrm{M}}| \leq m^{-2\rho}\eta$ with $\mathrm{E}(\eta^t)$ bounded. With $\rho = 1/2$, we have $\mathrm{E}(r_{\mathrm{M}}) = o(m^{-1})$.

■

Theorem 2.3 *We have*

$$\mathrm{Bias}(\hat{A}^{\mathrm{AR}}) = \mathrm{E}[\hat{A}^{\mathrm{AR}}] - A = \frac{2/A}{\mathrm{tr}(\Sigma^{-2}R^2)} + o(m^{-1}); \qquad (2.38)$$

$$\mathrm{Bias}(\hat{A}^{\mathrm{AM}}) = \mathrm{E}[\hat{A}^{\mathrm{AM}}] - A = \frac{\mathrm{tr}(PR - \Sigma^{-1}R) + 2/A}{\mathrm{tr}(\Sigma^{-2}R^2)} + o(m^{-1}). \qquad (2.39)$$

Proof: Take the expectation on (2.19) and (2.20), and use the fact that $\mathrm{E}y'(PR)^{k-1}Py = (PR)^{k-1}$. We obtain (2.38) and (2.39) immediately.

Remark: Define $W = \Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}$. Then $P = \Sigma^{-1} - W$. Note that W is positive definite, so that $\mathrm{tr}(PR) - \mathrm{tr}(\Sigma^{-1}R) = -\mathrm{tr}(WR) > 0$. So the sign of $\mathrm{Bias}(\hat{A}^{\mathrm{AM}})$ is undetermined. It could be positive, zero or negative, but it is always less than $\mathrm{Bias}(\hat{A}^{\mathrm{AR}})$.

Lemma 2.4 *Let $u \sim N(0, \Sigma)$. Then for a symmetric matrix A,*

$$\mathrm{E}[(u'Au)^2] = 2\mathrm{tr}[(A\Sigma)^2] + [\mathrm{tr}(A\Sigma)]^2.$$

The proof of this lemma is immediate from Lemma A.1 of Prasad and Rao (1990).

Theorem 2.4 *Both $\hat{A}^{\mathrm{AR}}$ and $\hat{A}^{\mathrm{AR}}$ are consistent estimators of A. In addition, we* *have*

$$\mathrm{E}(\hat{A}^{\mathrm{AR}} - A)^2 \doteq \mathrm{E}(\hat{A}^{\mathrm{AM}} - A)^2 \doteq \frac{2}{\mathrm{tr}(\Sigma^{-2}R^2)} + o(m^{-1}). \tag{2.40}$$

Remark: It is interesting to note that both ADM estimators have the same asymptotic variances as the REML and ML estimators.

Proof: Assume that $\mathrm{rank}(X) = p$ is fixed. Using (2.29) and (2.36), we have

(a)

$$\begin{aligned}
\mathrm{E}(\hat{A}^{\mathrm{AR}} - A)^2 &= \frac{1}{[\mathrm{tr}((PR)^2) + 2/A^2]^2}\mathrm{E}\Big(u'PRPu - \mathrm{tr}(PR) + \frac{2}{A}\Big)^2 + \mathrm{E}[r_{\mathrm{R}}^2] \\
&\quad + \frac{2}{[\mathrm{tr}((PR)^2) + 2/A^2]}\mathrm{E}[r_{\mathrm{R}} * (u'PRPu - \mathrm{tr}(PR) + \frac{2}{A})] \\
&= \frac{1}{[\mathrm{tr}((PR)^2) + 2/A^2]^2}\Big[\mathrm{E}(u'PRPu - \mathrm{tr}(PR))^2 + \frac{4}{A^2} \\
&\quad + \frac{4}{A}\mathrm{E}(u'PRPu - \mathrm{tr}(PR))\Big] + o(m^{-1}) \\
&= \frac{1}{[\mathrm{tr}((PR)^2) + 2/A^2]^2}\Big[\mathrm{E}(u'PRPu)^2 - [\mathrm{tr}(PR)]^2 + \frac{4}{A^2}\Big] + o(m^{-1}) \\
&= \frac{1}{[\mathrm{tr}((PR)^2) + 2/A^2]^2}\Big[2\mathrm{tr}(PRP\Sigma PRP\Sigma) + [\mathrm{tr}(PRP\Sigma)]^2 \\
&\quad - [\mathrm{tr}(PR)]^2 + \frac{4}{A^2}\Big] + o(m^{-1}) \\
&= \frac{1}{[\mathrm{tr}((PR)^2) + 2/A^2]^2}\Big[2\mathrm{tr}((PR)^2) + \frac{4}{A^2}\Big] + o(m^{-1}) \\
&= \frac{2}{\mathrm{tr}(P^2) + 2/A^2} + o(m^{-1}) \\
&= \frac{2}{\mathrm{tr}(\Sigma^{-2})} + o(m^{-1}).
\end{aligned}$$

About the order of the last two terms in the first line, since $|r_{\mathrm{R}}| \leq m^{-2\rho}\eta$ with $\mathrm{E}(\eta^t)$ bounded, obviously $\mathrm{E}[r_{\mathrm{R}}^2] = o(m^{-1})$. By the Cauchy inequality, the crossed term is also $o(m^{-1})$.

(b)

$$
\begin{aligned}
\mathrm{E}(\hat{A}^{\mathrm{AM}}-A)^2 &= \frac{1}{\triangle}\mathrm{E}\Big(u'PRPu-\mathrm{tr}(\Sigma^{-1}R)+\frac{2}{A}\Big)^2+o(m^{-1})\\
&= \frac{1}{\triangle}\mathrm{E}(u'PRPu-\mathrm{tr}(\Sigma^{-1}R))^2+\frac{4}{A^2}\\
&\quad +\frac{4}{A}\mathrm{E}(u'PRPu-\mathrm{tr}(\Sigma^{-1}R))+o(m^{-1})\\
&= \frac{1}{\triangle}\Big[2\mathrm{tr}((PR)^2)+[\mathrm{tr}(PR)]^2+[\mathrm{tr}(\Sigma^{-1}R)]^2\\
&\quad -2\mathrm{tr}(\Sigma^{-1}R)\mathrm{tr}(PR)+\frac{4}{A^2}+\frac{4}{A}(\mathrm{tr}(PR)-\mathrm{tr}(\Sigma^{-1}R))\Big]+o(m^{-1})\\
&= \frac{1}{\triangle}2\mathrm{tr}((PR)^2)+(\mathrm{tr}(PR)-\mathrm{tr}(\Sigma^{-1}R)+\frac{2}{A})^2+o(m^{-1})\\
&= \frac{2}{\mathrm{tr}(\Sigma^{-2}R^2)}+o(m^{-1}).
\end{aligned}
$$

where $\triangle=4(\mathrm{tr}(PR)^2-\mathrm{tr}(\Sigma^{-2}R^2)/2+1/A^2)^2$. The first line follows from arguments similar to these in part (a).

■

2.4.3 Comparison of $\hat{A}^{\mathrm{AR}}$ and $\hat{A}^{\mathrm{AM}}$

Theorem 2.5 *Assume the adjusted likelihood is unimodal. Then* $\hat{A}^{AR}>\hat{A}^{AM}>0$.

Proof: For this proof, to simplify the notation, we let $R=I$, which will not affect the validity of the proof for the general model. The positiveness part has already been proved in last subsection.

The ADM estimators $\hat{A}^{\text{AR}}$ and $\hat{A}^{\text{AM}}$ are the solutions of the following two equations respectively,

$$\frac{\partial l_{\text{AR}}}{\partial A} = \frac{1}{2}[u'P^2u - \text{tr}(P)] + \frac{1}{A} = 0, \tag{2.41}$$

$$\frac{\partial l_{\text{AMP}}}{\partial A} = \frac{1}{2}[u'P^2u - \text{tr}(\Sigma)] + \frac{1}{A} = 0. \tag{2.42}$$

Using (2.41), we get

$$A_1 = 2[\text{tr}(P(A_1)) - u'P(A_1)^2u]^{-1}.$$

Also, using (2.42), we get

$$A_2 = 2[\text{tr}(\Sigma(A_2)) - u'P(A_2)^2u]^{-1}.$$

Now, we define

$$f(x) = 2[\text{tr}(P(x)) - u'P(x)^2u]^{-1}; \tag{2.43}$$

$$g(x) = 2[\text{tr}(\Sigma(x)^{-1}) - u'P(x)^2u]^{-1} \tag{2.44}$$

Then, A_1 is just the abscissa of the interaction of the graphs of $y = f(x)$ and $y = x$. Also, A_2 is just the abscissa of the interaction of the graphs of $y = g(x)$ and $y = x$. Since $\text{tr}(P) < \text{tr}(\Sigma^{-1})$, we have $g(x) < f(x)$. Hence $g(A_1) < f(A_1) = A_1$.

Since the adjusted likelihoods are unimodal, there is only one possibility that when $x > A_1$, $x > f(x)$. i.e. when $x > A_1$, $f(x)$ is under the line $y = x$. Since we have already shown that $g(x) < f(x)$, the interaction point between $g(x)$ and $y = x$ should be on the left of A_1. Thus $A_1 < A_2$, i.e. $0 < \hat{A}^{AM} < \hat{A}^{AR}$. ■

Table 2.1: Percentages of Zero Estimates of A for Different Estimation Methods

	PR	FH	RE	ML	AR	AM
Pattern 1	0.98	1.57	0.84	2.48	0	0
Pattern 2	3.49	2.38	0.95	3.03	0	0
Pattern 3	12.15	4.11	0.99	3.96	0	0

Table 2.2: Comparison of Different Estimators of A

PATTERN 1						
	PR	FH	RE	ML	AR	AM
Bias	-0.001	0	-0.002	-0.102	0.363	0.224
RB %	-0.1	0	-0.2	-10.2	36.3	22.4
CV	0.006	0.006	0.006	0.006	0.005	0.005
ARB %	45.3	44.8	44.7	43.5	55.2	47.7
MSE	0.334	0.326	0.323	0.292	0.554	0.406
PATTERN 2						
	PR	FH	RE	ML	AR	AM
Bias	0.006	0.016	0.004	-0.104	0.404	0.250
RB %	0.6	1.6	0.4	-10.4	40.4	25.0
CV	0.007	0.006	0.006	0.006	0.005	0.005
ARB %	54.0	47.2	45.8	44.5	58.7	50.0
MSE	0.471	0.365	0.338	0.303	0.623	0.446
PATTERN 3						
Bias	0.029	0.027	-0.004	-0.118	0.399	0.235
RB %	2.9	2.7	-0.4	-11.8	39.9	23.5
CV	0.009	0.006	0.006	0.006	0.005	0.005
ARB %	72.5	50	45.4	44.6	58.9	50.0
MSE	0.896	0.430	0.333	0.304	0.641	0.451

Table 2.3: Biases of Different Estimators of B_i

PATTERN 1						
	PR	FH	RE	ML	AR	AM
G1	0.150	0.146	0.146	0.230	-0.116	-0.039
G2	0.159	0.154	0.154	0.238	-0.106	-0.032
G3	0.167	0.162	0.162	0.245	-0.093	-0.023
G4	0.172	0.167	0.167	0.249	-0.078	-0.013
G5	0.174	0.169	0.168	0.246	-0.061	-0.003
PATTERN 2						
	PR	FH	RE	ML	AR	AM
G1	0.094	0.065	0.069	0.141	-0.172	-0.091
G2	0.223	0.156	0.156	0.249	-0.121	-0.041
G3	0.236	0.165	0.164	0.257	-0.108	-0.031
G4	0.249	0.172	0.170	0.261	-0.091	-0.020
G5	0.257	0.170	0.163	0.246	-0.049	0.002
PATTERN 3						
	PR	FH	RE	ML	AR	AM
G1	0.059	0.024	0.032	0.087	-0.143	-0.077
G2	0.369	0.170	0.158	0.264	-0.115	-0.028
G3	0.398	0.181	0.165	0.273	-0.101	-0.018
G4	0.428	0.190	0.170	0.278	-0.085	-0.008
G5	0.480	0.180	0.130	0.226	-0.020	0.014

Table 2.4: Percentage Relative Biases of Different Estimators of B_i

PATTERN 1						
	PR	FH	RE	ML	AR	AM
G1	36.5	35.5	35.6	55.9	-28.2	-9.6
G2	42.4	41.2	41.2	63.6	-28.2	-8.4
G3	50.0	48.5	48.5	73.6	-28.0	-6.8
G4	60.4	58.6	58.4	87.1	-27.4	-4.6
G5	75.3	73.1	72.7	106.8	-26.2	-1.3
PATTERN 2						
	PR	FH	RE	ML	AR	AM
G1	14.1	9.7	10.3	21.1	-25.7	-13.6
G2	59.3	41.6	41.7	66.4	-32.4	-10.9
G3	70.9	49.4	49.2	77.0	-32.3	-9.3
G4	87.2	60.1	59.4	91.5	-31.9	-7.1
G5	154.4	101.9	97.7	147.8	-29.1	0.9
PATTERN 3						
	PR	FH	RE	ML	AR	AM
G1	7.4	3.1	4.0	10.8	-17.9	-9.6
G2	98.4	45.3	42.2	70.5	-30.6	-7.4
G3	119.3	54.2	49.6	81.8	-30.3	-5.5
G4	149.6	66.6	59.6	97.2	-29.7	-2.9
G5	527.7	198.0	142.6	249.0	-21.8	15.8

Table 2.5: Coefficients of Variation of Different Estimators of B_i

PATTERN 1						
	PR	FH	RE	ML	AR	AM
G1	0.011	0.010	0.010	0.010	0.009	0.009
G2	0.012	0.011	0.011	0.011	0.010	0.010
G3	0.013	0.013	0.012	0.013	0.011	0.010
G4	0.014	0.014	0.014	0.014	0.012	0.011
G5	0.017	0.017	0.016	0.017	0.013	0.013
PATTERN 2						
	PR	FH	RE	ML	AR	AM
G1	0.006	0.006	0.005	0.005	0.006	0.005
G2	0.014	0.012	0.012	0.012	0.010	0.010
G3	0.015	0.013	0.013	0.013	0.011	0.011
G4	0.017	0.015	0.014	0.015	0.012	0.012
G5	0.025	0.022	0.020	0.022	0.014	0.014
PATTERN 3						
	PR	FH	RE	ML	AR	AM
G1	0.005	0.004	0.003	0.003	0.004	0.003
G2	0.016	0.013	0.011	0.012	0.010	0.010
G3	0.018	0.014	0.012	0.013	0.011	0.011
G4	0.020	0.016	0.014	0.015	0.012	0.012
G5	0.037	0.037	0.029	0.034	0.017	0.018

Table 2.6: Percentage Absolute Relative Biases of Different Estimators of B_i

PATTERN 1						
	PR	FH	RE	ML	AR	AM
G1	91.1	89.7	89.5	97.5	74.1	71.9
G2	99.3	97.7	97.5	107.0	78.2	76.5
G3	109.5	107.7	107.4	119.1	83.0	81.8
G4	122.7	120.5	120.1	134.9	88.5	88.2
G5	140.8	138.2	137.5	157.0	95.1	95.9
PATTERN 2						
	PR	FH	RE	ML	AR	AM
G1	55.1	47.5	46.5	48.4	47.2	43.3
G2	123.7	101.8	99.6	110.6	80.7	78.7
G3	138.1	112.3	109.8	123.3	85.5	84.2
G4	157.4	126.0	122.9	140.1	91.1	90.7
G5	231.6	174.7	168.0	201.8	105.8	108.6
PATTERN 3						
	PR	FH	RE	ML	AR	AM
G1	42.1	29.3	27.1	27.9	30.4	27.2
G2	174.7	107.8	98.7	113.0	81.1	79.9
G3	198.5	119.4	108.6	126.3	85.9	85.5
G4	232.1	134.9	121.4	143.9	91.6	92.3
G5	621.5	277.2	214.9	303.9	117.9	126.2

Table 2.7: Mean Squared Error of Different Estimators of B_i

PATTERN 1						
	PR	FH	RE	ML	AR	AM
G1	0.813	0.811	0.810	0.828	0.772	0.771
G2	0.792	0.790	0.788	0.809	0.745	0.745
G3	0.757	0.755	0.753	0.776	0.704	0.706
G4	0.703	0.700	0.698	0.724	0.644	0.647
G5	0.620	0.617	0.614	0.643	0.557	0.560
PATTERN 2						
	PR	FH	RE	ML	AR	AM
G1	0.735	0.718	0.716	0.721	0.715	0.708
G2	0.843	0.797	0.792	0.817	0.747	0.748
G3	0.813	0.762	0.757	0.785	0.706	0.708
G4	0.764	0.708	0.702	0.734	0.646	0.648
G5	0.570	0.506	0.493	0.534	0.433	0.436
PATTERN 3						
	PR	FH	RE	ML	AR	AM
G1	0.534	0.508	0.504	0.505	0.511	0.505
G2	0.962	0.811	0.789	0.824	0.748	0.750
G3	0.947	0.779	0.753	0.793	0.707	0.710
G4	0.915	0.728	0.698	0.743	0.647	0.651
G5	0.618	0.365	0.303	0.373	0.255	0.257

Chapter 3

Uncertainty of EBLUP

We can measure the uncertainty of the EBLUP by its mean squared prediction error (MSPE), defined as $\text{MSPE}[\hat{\theta}_i(Y;\hat{A})] = \text{E}[\hat{\theta}_i(Y;\hat{A}) - \theta_i]^2$, where E is the expectation with respect to the joint distribution of Y and θ induced by the general Fay-Herriot model.

Note that

$$\begin{aligned} \text{MSPE}[\hat{\theta}_i(Y;\hat{A})] &= \text{E}[\hat{\theta}_i(Y;A) - \theta_i]^2 + \text{E}[\hat{\theta}_i(Y;\hat{A}) - \hat{\theta}_i(Y;A)]^2 \\ &\quad +2\text{E}[\hat{\theta}_i(Y;A) - \theta_i][\hat{\theta}_i(Y;\hat{A}) - \hat{\theta}_i(Y;A)], \end{aligned} \tag{3.1}$$

where $\hat{\theta}_i(Y;A)$ is the BLUP of θ_i defined in (1.3). Kackar and Harville (1984) showed that under the normality of the random effects v and e, the cross-product term of (3.1) vanishes, provided that the variance estimator $\hat{A}$ is a translation invariant and even function. We have already seen in Chapter 2 that the Prasad-Rao, Fay-Herriot, ML and REML and ADM estimators of A are all translation invariant and even functions. Therefore,

$$\begin{aligned} \text{MSPE}[\hat{\theta}_i(Y;\hat{A})] &= \text{E}[\hat{\theta}_i(Y;A) - \theta_i]^2 + \text{E}[\hat{\theta}_i(Y;\hat{A}) - \hat{\theta}_i(Y;A)]^2 \\ &= \text{MSPE}[\hat{\theta}_i(Y;A)] + \text{E}[\hat{\theta}_i(Y;\hat{A}) - \hat{\theta}_i(Y;A)]^2. \end{aligned} \tag{3.2}$$

The first term of (3.2) is the mean squared prediction error of the BLUP estimator, which is given by (see Rao, 2003; Henderson, 1975)

$$\text{MSPE}[\hat{\theta}_i(Y;A)] = g_{1i}(A) + g_{2i}(A), \tag{3.3}$$

where

$$g_{1i}(A) = \frac{b_i^2 A D_i}{b_i^2 A + D_i},$$

and

$$g_{2i}(A) = \frac{D_i^2}{(b_i^2 A + D_i)^2} x_i'(X'\Sigma^{-1}X)^{-1}x_i.$$

Note that $g_{1i}(A)$ and $g_{2i}(A)$ do not depend on the estimation method of A.

In order to derive the estimator of MSPE, the following regularity conditions are assumed throughout this chapter:

1. The elements of X are bounded such as $(X'\Sigma^{-1}X)^{-1} = O(m^{-1})$;

2. $0 < D_{\text{L}} \le D_i \le D_{\text{U}} < \infty$, and $0 < b_{\text{L}}^2 \le b_i^2 \le b_{\text{U}}^2 < \infty, \forall\ i = 1, \cdots, m$.

Under the above regularity conditions, it is easy to see that $g_{1i}(A)$ has order $O(1)$, and $g_{2i}(A)$ has order $O(m^{-1})$.

A naive MSPE estimator can be obtained when A in (3.3) is replaced by $\hat{A}$, i.e.

$$mspe_i^{\text{N}}(\hat{A}) = g_{1i}(\hat{A}) + g_{2i}(\hat{A}).$$

This naive MSPE estimator usually underestimates the true MSPE for two reasons. First, by neglecting the second term of (3.2), $mspe_i^{\text{N}}(\hat{A})$ does not include the variability caused by the estimation of the model parameters, which has order $O(m^{-1})$. Secondly, the naive MSPE estimator even underestimates the true MSPE of the BLUP, the order of underestimation being $O(m^{-1})$.

It is important to obtain an accurate estimator of MSPE to reflect the true variability associated with the EBLUP estimator. An estimator mspe_i is called a second-order unbiased estimator of MSPE_i if $\text{E}[\text{mspe}_i] = \text{MSPE}_i + o(m^{-1})$. Prasad and Rao (1990), Datta and Lahiri (2000), Das, Jiang, and Rao (2004), and Datta, Rao, and Smith (2005), etc., have studied the second-order unbiased (or nearly unbiased) MSPE estimators using various variance components estimators under various small area models. In this chapter, we first obtain a second-order approximation to the MSPE of EBLUP and then an estimator of MSPE correct to the same order when the ADM estimators of A are used under the general Fay-Herriot model (1.1).

In this chapter, the letter c, with or without suffix, will be generic for constants.

3.1 Mean Squared Prediction Error Approximation

The second term in (3.2) is the uncertainty due to the estimation of A. It has no closed-form expression and we will approximate it up to order $O(m^{-1})$, same as the order of $g_{2i}(A)$. Throughout this dissertation, we will use the notation r to denote a generic remainder term in a Taylor series expansion. By the Taylor series expansion of $\hat{\theta}_i$ around A, we obtain

$$\hat{\theta}_i(Y;\hat{A}) - \hat{\theta}_i(Y;A) = (\hat{A} - A)\frac{\partial \hat{\theta}_i(Y;A)}{\partial A} + r, \tag{3.4}$$

and

$$
\begin{aligned}
\frac{\partial \hat{\theta}_i(Y;A)}{\partial A} &= \frac{\partial}{\partial A}[B_i(y_i - x_i'\beta)] + \frac{\partial}{\partial A}[(1-B_i)x_i'(\tilde{\beta}-\beta)] \\
&= \frac{b_i^2 D_i}{(b_i^2 A + D_i)^2}(y_i - x_i'\beta) - \frac{b_i^2 D_i}{(b_i^2 A + D_i)^2} x_i'(X'\Sigma^{-1}X)^{-1}\Big[X'\Sigma^{-1}(Zv+e) \\
&\quad -(X'\Sigma^{-2}X)(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y - X'\Sigma^{-2}y\Big] + r \qquad (3.5) \\
&= \frac{b_i^2 D_i}{(b_i^2 A + D_i)^2}(y_i - x_i'\beta) + r. \qquad (3.6)
\end{aligned}
$$

The last equality is due to the fact that under the regularity condition 1, the second term of (3.5) has lower order than the first term. Then we have

$$
\hat{\theta}_i(Y;\hat{A}) - \hat{\theta}_i(Y;A) = (\hat{A}-A)\frac{b_i^2 D_i}{(b_i^2 A + D_i)^2}(y_i - x_i'\beta) + r. \qquad (3.7)
$$

Thus, we obtain the approximation of the second term in (3.2) as:

$$
\begin{aligned}
\mathrm{E}(\hat{\theta}_i(Y;\hat{A}) - \hat{\theta}_i(Y;A))^2 &= \frac{b_i^4 D_i^2}{(b_i^2 A + D_i)^4}\mathrm{E}[(\hat{A}-A)^2(y_i - x_i'\beta)^2] + o(m^{-1}), \qquad (3.8) \\
&= g_{3i}(A) + o(m^{-1}). \qquad (3.9)
\end{aligned}
$$

Remark: $\mathrm{E}(r) = o(m^{-1})$ requires that r is uniformly integrable. This can be proved following arguments similar to the ones given in Prasad and Rao (1990) and Lahiri and Rao (1995). All the other similar arguments throughout this dissertation can be derived in the same way.

By neglecting all the terms with lower order, ADM REML and ADM ML methods produce the same g_{3i} terms, which are given by

$$\begin{aligned} g_{3i\mathrm{AR}}(A) &= g_{3i\mathrm{AM}}(A) &(3.10)\\ &= \frac{b_i^4 D_i^2}{(b_i^2 A + D_i)^3} \mathrm{E}[(\hat{A} - A)^2] &(3.11)\\ &= \frac{b_i^4 D_i^2}{(b_i^2 A + D_i)^3} \frac{2}{\mathrm{tr}(\Sigma^{-2})}. &(3.12) \end{aligned}$$

See the Proof section of this chapter for the detailed computations. Note that the above g_{3i} terms are identical to the ones for the REML and ML estimators. Denoting the common g_{3i} term for these four likelihood methods by $g_{3i\mathrm{L}}(A)$, we have

$$\begin{aligned} g_{3i\mathrm{L}}(A) &= g_{3i\mathrm{AR}}(A) = g_{3i\mathrm{AM}}(A) = g_{3i\mathrm{RE}}(A) = g_{3i\mathrm{ML}}(A)\\ &= \frac{b_i^4 D_i^2}{(b_i^2 A + D_i)^3} \frac{2}{\mathrm{tr}(\Sigma^{-2})}. &(3.13) \end{aligned}$$

The $g_{3i}(A)$ terms for the Prasad-Rao estimator and the Fay-Herriot estimator of A are given by:

$$\begin{aligned} g_{3i\mathrm{PR}}(A) &= \frac{b_i^4 D_i^2}{(D_i + b_i^2 A)^3} \frac{2\mathrm{tr}(\Sigma^2)}{m^2}; &(3.14)\\ g_{3i\mathrm{FH}}(A) &= \frac{b_i^4 D_i^2}{(D_i + b_i^2 A)^3} \frac{2m}{(\mathrm{tr}(\Sigma^{-1}))^2}. &(3.15) \end{aligned}$$

Note that all these $g_{3i}(A)$ terms have order $O(m^{-1})$. As pointed out by Datta, Rao and Smith (2004), comparing (3.13), (3.14) and (3.15), we have $g_{3iL}(A) \leq g_{3i\mathrm{FH}}(A) \leq g_{3i\mathrm{PR}}(A)$. Equality holds if and only if the b_i's and D_i's are all equal.

Thus the second-order MSPE approximation of $\hat{\theta}_i$ is given by

$$\text{MSPE}[\hat{\theta}_i(Y;\hat{A})] = g_{1i}(A) + g_{2i}(A) + g_{3i}(A) + o(m^{-1}). \quad (3.16)$$

3.2 Estimation of Mean Squared Prediction Error

Note that the second-order approximation to the mean squared prediction error of EBLUP derived in the last section involves the unknown variance component A and thus cannot be used to assess the uncertainty of EBLUP for a given data set. However, this second-order approximation is useful in obtaining a second-order unbiased estimator of the MSPE of EBLUP.

First, we shall follow Datta and Lahiri (2000) to obtain the order of the biases of $g_{1i}(\hat{A})$, $g_{2i}(\hat{A})$, and $g_{3i}(\hat{A})$ up to order $O(m^{-1})$. We shall then use the results in correcting the biases up to order $O(m^{-1})$ and obtaining the second-order unbiased estimator of MSPE.

An application of the Taylor series expansion of $g_{1i}(\hat{A})$ around A yields

$$g_{1i}(\hat{A}) - g_{1i}(A) = \frac{b_i^2 D_i^2}{(b_i^2 A + D_i)^2}(\hat{A} - A) - \frac{b_i^4 D_i^2}{(b_i^2 A + D_i)^3}(\hat{A} - A)^2 + o((\hat{A} - A)^2).$$

Since $\text{E}(\hat{A} - A)^2 = O(m^{-1})$ (see Chapter 2), we have

$$\begin{aligned} \text{E}[g_{1i}(\hat{A})] &= g_{1i}(A) + \frac{b_i^2 D_i^2}{(b_i^2 A + D_i)^2}\text{E}(\hat{A} - A) - \frac{b_i^4 D_i^2}{(b_i^2 A + D_i)^3}\text{E}(\hat{A} - A)^2 + o(m^{-1}), \\ &= g_{1i}(A) + \frac{b_i^2 D_i^2}{(b_i^2 A + D_i)^2}\text{Bias}(\hat{A}) - g_{3i}(A) + o(m^{-1}). \quad (3.17) \end{aligned}$$

Since both $\frac{\partial g_{2i}(A)}{\partial A}$ and $\frac{\partial g_{3i}(A)}{\partial A}$ have order $O(m^{-1})$, and $\hat{A} - A = O_p(m^{-1/2})$, we obtain the following using Taylor series expansions of $g_{2i}(\hat{A})$ and $g_{3i}(\hat{A})$ around A:

$$\begin{aligned} \mathrm{E}[g_{2i}(\hat{A})] &= g_{2i}(A) + o(m^{-1}), & (3.18) \\ \mathrm{E}[g_{3i}(\hat{A})] &= g_{3i}(A) + o(m^{-1}). & (3.19) \end{aligned}$$

Thus we obtain the following second-order unbiased estimator of the mean squared error of the EBLUP $\hat{\theta}_i$:

$$\mathrm{mspe}(\hat{\theta}_i) = g_{1i}(\hat{A}) + g_{2i}(\hat{A}) + 2g_{3i}(\hat{A}) - \frac{b_i^2 D_i^2}{(b_i^2 A + D_i)^2}\mathrm{Bias}(\hat{A}). \qquad (3.20)$$

Hence, with the bias terms: $\mathrm{Bias}_{\mathrm{AR}}(A)$ (2.38) and $\mathrm{Bias}_{\mathrm{AM}}(A)$ (2.39) derived in Theorem 2.3, we have

$$\begin{aligned} \mathrm{mspe}(\hat{\theta}_i^{\mathrm{AR}}) &= g_{1i}(\hat{A}^{\mathrm{AR}}) + g_{2i}(\hat{A}^{\mathrm{AR}}) + 2g_{3i\mathrm{AR}}(\hat{A}^{\mathrm{AR}}) - \frac{b_i^2 D_i^2}{(b_i^2 \hat{A}^{\mathrm{AR}} + D_i)^2}\mathrm{Bias}_{\mathrm{AR}}(\hat{A}^{\mathrm{AR}}); \\ \mathrm{mspe}(\hat{\theta}_i^{\mathrm{AM}}) &= g_{1i}(\hat{A}^{\mathrm{AM}}) + g_{2i}(\hat{A}^{\mathrm{AM}}) + 2g_{3i\mathrm{AM}}(\hat{A}^{\mathrm{AM}}) \\ &\quad - \frac{b_i^2 D_i^2}{(b_i^2 \hat{A}^{\mathrm{AM}} + D_i)^2}\mathrm{Bias}_{\mathrm{AM}}(\hat{A}^{\mathrm{AM}}). \end{aligned}$$

Second-order unbiased estimators of MSPE of EBLUP are available when A is estimated by the Prasad-Rao (PR), REML, ML and the Fay-Herriot (FH) estimators (see Prasad and Rao, 1990; Datta and Lahiri, 2000; and Datta, Rao and Smith, 2005). Since for the Prasad-Rao and REML estimators of A, $\mathrm{Bias}_{\mathrm{PR}}(\hat{A}^{\mathrm{PR}}) = o(m^{-1})$

and $\text{Bias}_{\text{RE}}(\hat{A}^{\text{RE}}) = o(m^{-1})$, we have

$$\text{mspe}(\hat{\theta}_i^{\text{PR}}) = g_{1i}(\hat{A}^{\text{PR}}) + g_{2i}(\hat{A}^{\text{PR}}) + 2g_{3i\text{PR}}(\hat{A}^{\text{PR}}); \tag{3.21}$$

$$\text{mspe}(\hat{\theta}_i^{\text{RE}}) = g_{1i}(\hat{A}^{\text{RE}}) + g_{2i}(\hat{A}^{\text{RE}}) + 2g_{3i\text{RE}}(\hat{A}^{\text{RE}}); \tag{3.22}$$

For the maximum likelihood estimator $\hat{A}^{\text{ML}}$, Datta and Lahiri (2000) showed that

$$\text{Bias}_{\text{ML}}(\hat{A}^{\text{ML}}) = \text{E}[\hat{A}^{\text{ML}}] - A = \frac{\text{tr}(P - \Sigma^{-1})}{\text{tr}(\Sigma^{-2})} + o(m^{-1}). \tag{3.23}$$

For the Fay-Herriot estimator $\hat{A}^{\text{FH}}$, Datta, Rao and Smith (2005) showed that

$$\text{Bias}_{\text{FH}}(\hat{A}^{\text{FH}}) = \text{E}[\hat{A}^{\text{FH}}] - A = \frac{2[m\text{tr}(\Sigma^{-2}) - (\text{tr}(\Sigma^{-1}))^2]}{(\text{tr}(\Sigma^{-1}))^3} + o(m^{-1}). \tag{3.24}$$

Hence

$$\begin{aligned}
\text{mspe}(\hat{\theta}_i^{\text{ML}}) &= g_{1i}(\hat{A}^{\text{ML}}) + g_{2i}(\hat{A}^{\text{ML}}) + 2g_{3i\text{ML}}(\hat{A}^{\text{ML}}) - \frac{b_i^2 D_i^2}{(b_i^2 \hat{A}^{\text{ML}} + D_i)^2}\text{Bias}_{\text{ML}}(\hat{A}^{\text{ML}}); \\
\text{mspe}(\hat{\theta}_i^{\text{FH}}) &= g_{1i}(\hat{A}^{\text{FH}}) + g_{2i}(\hat{A}^{\text{FH}}) + 2g_{3i\text{FH}}(\hat{A}^{\text{FH}}) - \frac{b_i^2 D_i^2}{(b_i^2 \hat{A}^{\text{FH}} + D_i)^2}\text{Bias}_{\text{FH}}(\hat{A}^{\text{FH}}).
\end{aligned}$$

3.3 Monte Carlo Simulation Study

We use the simulation setting of Chapter 2 in order to study the small sample performances of our proposed second-order unbiased MSPE estimators. We simulate the true MSPE of the EBLUP estimators $\hat{\theta}_i(Y; \hat{A})$ using $\hat{A}^{\text{PR}}$, $\hat{A}^{\text{FH}}$, $\hat{A}^{\text{RE}}$, $\hat{A}^{\text{ML}}$, $\hat{A}^{\text{AR}}$ and $\hat{A}^{\text{AM}}$. Let $Y_i^{(s)}$ and $\theta_i^{(s)}$ be the simulated data and the true mean of area i for the sth simulation, , $i = 1, \cdots, m;\ s = 1, \cdots, S = 10,000$. Let $\hat{A}^{(s)}$ be the value of $\hat{A}$ in sth simulation. Then the simulated value of the true MSPE of $\hat{\theta}_i(Y; \hat{A})$ is given by $S^{-1} \sum_{s=1}^{S} [\hat{\theta}_i(Y^{(s)}; \hat{A}^{(s)}) - \theta_i^{(s)}]^2$.

In Table 3.1, we report the group average of the simulated true MSPE of $\hat{\theta}_i(Y; \hat{A})$. For pattern 1, different estimators of A do not seem to have much impact on the MSPE of the EBLUP. However, when the variability of D_i across areas increases as in patterns 2 and 3, the Prasad-Rao estimator of A leads to larger MSPE than the other methods.

Table 3.2 reports the percent relative biases (RB) for each MSPE estimator. The RB of an MSPE estimator is calculated as RB = [E(MSPE estimator) - simulated true MSPE]/(simulated true MSPE), where E(MSPE estimator) is the Monte Carlo expectation obtained by taking the average of MSPE estimates over $N = 10,000$ simulations. In this table, we report seven simulated MSPE estimators: six of them associated with six estimating methods of A, and one naive MSPE estimator using REML estimator, denoted by "RE-N".

The naive estimator of MSPE has a tendency to unduly underestimate. For pattern 1, all MSPE estimators are comparable and reduce the underestimation of the naive estimator. The MSPE estimator of EBLUP that uses the Prasad-Rao estimator of A tends to overestimate the MSPE of EBLUP, especially when the sampling variances D_i's are highly unbalanced. For example, the overestimation could exceed 700% for pattern 3. For patterns 2 and 3, the Fay-Herriot, REML, and ML methods could lead to underestimation. Of course, the second-order MSPE estimator that uses ADM method improves on the naive method. The ADM based

MSPE estimators lead to $1-2\%$ overestimation. Considering the underestimation of the MSPE will create trouble in prediction interval construction, the proposed ADM methods are more favorable than the other methods.

3.4 Derivation of g_{3i}

To derive g_{3i}s for the ADM REML and ADM ML estimators, we use the following lemma. See Srivastava and Tiwari (1976) for the proof.

Lemma 3.1 *Let $U \sim N(0, \Sigma)$. Then for symmetric matrices A, B and C,*

1. $\mathrm{E}[(U'AU)(U'BU)(U'CU)] = 8\mathrm{tr}(A\Sigma B\Sigma C\Sigma) + 2\{\mathrm{tr}(A\Sigma B\Sigma)\mathrm{tr}(C\Sigma)$
$+\mathrm{tr}(A\Sigma C\Sigma)\mathrm{tr}(B\Sigma) + \mathrm{tr}(B\Sigma C\Sigma)\mathrm{tr}(A\Sigma)\} + \mathrm{tr}(A\Sigma)\mathrm{tr}(B\Sigma)\mathrm{tr}(C\Sigma);$
2. $\mathrm{E}[(U'AU)(U'BU)] = 2\mathrm{tr}(A\Sigma B\Sigma) + \mathrm{tr}(A\Sigma)\mathrm{tr}(B\Sigma).$

3.4.1 Derivation of $g_{3i\mathrm{AR}}$

Let $u_i = y_i - x_i'\beta$ and $u = Y - X\beta = Zv + e$. Define G_i be a $m \times m$ matrix, with the (i,i) element being 1 and all the other elements being 0. Then we have $u_i^2 = u'G_iu$. Using $\hat{A}^{\mathrm{AR}} - A = \frac{y'PRPy - \mathrm{tr}(PR) + 2/A}{\mathrm{tr}((PR)^2) + 2/A^2} + r_{\mathrm{R}}$ (see 2.29), and the fact that $y'PRPy = u'PRPu$ and $P\Sigma P = P$, we have

$$
\begin{aligned}
& \mathrm{E}[(\hat{A}^{\mathrm{AR}} - A)^2(y_i - x_i'\beta)^2] \\
= \quad & \frac{1}{(\mathrm{tr}(PR)^2 + 2/A^2)^2}\mathrm{E}[(u'PRPu - \mathrm{tr}(PR) + 2/A)^2 u'G_iu] \\
= \quad & \frac{1}{(\mathrm{tr}(PR)^2 + 2/A^2)^2}\mathrm{E}[u'PRPuu'PRPuu'G_iu + (\mathrm{tr}(PR) - 2/A)^2 u'G_iu \\
& -2(\mathrm{tr}(PR) - 2/A)u'PRPuu'G_iu]
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{(\mathrm{tr}(PR)^2+2/A^2)^2}[8\mathrm{tr}(PRP\Sigma PRP\Sigma G_i\Sigma)+2\{\mathrm{tr}(PRP\Sigma PRP\Sigma)\mathrm{tr}(G_i\Sigma) \\
&\quad +2\mathrm{tr}(PRP\Sigma G_i\Sigma)\mathrm{tr}(PRP\Sigma)\}+(\mathrm{tr}(PRP\Sigma))^2\mathrm{tr}(G_i\Sigma) \\
&\quad +(\mathrm{tr}(PR)-2/A)^2\mathrm{tr}(G_i\Sigma)-2(\mathrm{tr}(PR)-2/A)\{2\mathrm{tr}(PRP\Sigma G_i\Sigma) \\
&\quad +\mathrm{tr}(PRP\Sigma)\mathrm{tr}(G_i\Sigma)\}] \\
&= \frac{1}{(\mathrm{tr}(PR)^2+2/A^2)^2}[8\mathrm{tr}(PRPRP\Sigma G_i\Sigma)+2\mathrm{tr}(PR)^2\mathrm{tr}(G_i\Sigma) \\
&\quad +4\mathrm{tr}(PRP\Sigma G_i\Sigma)\mathrm{tr}(PR)+(\mathrm{tr}(PR))^2\mathrm{tr}(G_i\Sigma)+(\mathrm{tr}(PR)-2/A)^2\mathrm{tr}(G_i\Sigma) \\
&\quad -4(\mathrm{tr}(PR)-2/A)\mathrm{tr}(PRPG_i\Sigma)-2(\mathrm{tr}(PR)-2/A)\mathrm{tr}(PR)\mathrm{tr}(G_i\Sigma) \\
&= \frac{1}{(\mathrm{tr}(PR)^2+2/A^2)^2}[8\mathrm{tr}(PRPRP\Sigma G_i\Sigma)+2\mathrm{tr}(PR)^2\mathrm{tr}(G_i\Sigma) \\
&\quad 8/A\mathrm{tr}(PRPG_i\Sigma)+\{\mathrm{tr}(PR)-\mathrm{tr}(PR)+2/A\}^2\mathrm{tr}(G_i\Sigma)] \\
&= \frac{2(\mathrm{tr}(PR)^2+2/A^2)\mathrm{tr}(G_i\Sigma)}{(\mathrm{tr}(PR)^2+2/A^2)^2}+o(m^{-1}) \\
&= \frac{2}{(\mathrm{tr}(PR)^2+2/A^2)^2}(b_i^2A+D_i)+o(m^{-1}) \\
&\doteq \mathrm{E}(\hat{A}^{\mathrm{AR}}-A)^2(b_i^2A+D_i).
\end{aligned}
$$

Thus

$$
\begin{aligned}
g_{3i\mathrm{AR}}(A) &= \frac{b_i^4D_i^2}{(b_i^2A+D_i)^3}\mathrm{E}[(\hat{A}^{\mathrm{AR}}-A)^2] \\
&= \frac{b_i^4D_i^2}{(b_i^2A+D_i)^3}\frac{2}{\mathrm{tr}(\Sigma^{-2})}.
\end{aligned}
$$

3.4.2 Derivation of $g_{3i\text{AM}}$

Using $\hat{A}^{\text{AR}} - A = \frac{y'PRPy - \text{tr}(\Sigma^{-1}R) + 2/A}{\text{tr}(\Sigma^{-2}R^2) + 2/A^2} + r_{\text{M}}$ (see 2.36), we have

$$
\begin{aligned}
& \text{E}[(\hat{A}^{\text{AM}} - A)^2 (y_i - x_i'\beta)^2] \\
= & \frac{1}{\triangle^2} \text{E}[(u'PRPu - \text{tr}(PR) + 2/A)^2 u'G_i u] \\
= & \frac{1}{\triangle^2} \text{E}[u'PRPuu'PRPuu'G_i u + (\text{tr}(\Sigma^{-1}R) - 2/A)^2 u'G_i u \\
& -2(\text{tr}(\Sigma^{-1}R) - 2/A) u'PRPuu'G_i u] \\
= & \frac{1}{\triangle^2} [8\text{tr}(PRP\Sigma PRP\Sigma G_i \Sigma) + 2\{\text{tr}(PRP\Sigma PRP\Sigma)\text{tr}(G_i\Sigma) \\
& +2\text{tr}(PRP\Sigma G_i \Sigma)\text{tr}(PRP\Sigma)\} + (\text{tr}(PRP\Sigma))^2 \text{tr}(G_i\Sigma) \\
& +(\text{tr}(\Sigma^{-1}R) - 2/A)^2 \text{tr}(G_i\Sigma) - 2(\text{tr}(\Sigma^{-1}R) - 2/A)\{2\text{tr}(PRP\Sigma G_i\Sigma) \\
& +\text{tr}(PRP\Sigma)\text{tr}(G_i\Sigma)\}] \\
= & \frac{1}{\triangle^2} [8\text{tr}(PRPRP\Sigma G_i\Sigma) + 2\text{tr}(PR)^2 \text{tr}(G_i\Sigma) \\
& +4\text{tr}(PRP\Sigma G_i\Sigma)\text{tr}(PR) + (\text{tr}(PR))^2 \text{tr}(G_i\Sigma) + (\text{tr}(\Sigma^{-1}R) - 2/A)^2 \text{tr}(G_i\Sigma) \\
& -4(\text{tr}(\Sigma^{-1}R) - 2/A)\text{tr}(PRPG_i\Sigma) - 2(\text{tr}(\Sigma^{-1}R) - 2/A)\text{tr}(PR)\text{tr}(G_i\Sigma) \\
= & \frac{1}{\triangle^2} [8\text{tr}(PRPRP\Sigma G_i\Sigma) + 2\text{tr}(PR)^2 \text{tr}(G_i\Sigma) \\
& 8/A\text{tr}(PRPG_i\Sigma) + \{\text{tr}(PR) - \text{tr}(\Sigma^{-1}R) + 2/A\}^2 \text{tr}(G_i\Sigma)] + o(m^{-1}) \\
= & \frac{2(\text{tr}(\Sigma^{-1}R)^2 + 2/A^2)\text{tr}(G_i\Sigma)}{(\text{tr}(\Sigma^{-1}R)^2 + 2/A^2)^2} + o(m^{-1}) \\
= & \frac{2}{(\text{tr}(\Sigma^{-1}R)^2 + 2/A^2)^2} (b_i^2 A + D_i) + o(m^{-1}) \\
\doteq & \text{E}(\hat{A}^{\text{AM}} - A)^2 (b_i^2 A + D_i),
\end{aligned}
$$

where $\triangle = \text{tr}(\Sigma^{-2}R^2) + 2/A^2$.

Table 3.1: Simulated MSPE of $\hat{\theta}_i(Y;\hat{A})$

PATTERN 1						
	PR	FH	RE	ML	AR	AM
G1	0.46	0.46	0.46	0.47	0.46	0.45
G2	0.41	0.41	0.42	0.42	0.41	0.41
G3	0.37	0.37	0.37	0.38	0.37	0.36
G4	0.31	0.31	0.31	0.31	0.30	0.30
G5	0.25	0.25	0.25	0.25	0.24	0.24
PATTERN 2						
	PR	FH	RE	ML	AR	AM
G1	0.79	0.77	0.77	0.77	0.78	0.77
G2	0.44	0.42	0.42	0.43	0.42	0.41
G3	0.39	0.38	0.38	0.38	0.37	0.37
G4	0.33	0.32	0.31	0.32	0.31	0.31
G5	0.20	0.18	0.18	0.18	0.18	0.18
PATTERN 3						
	PR	FH	RE	ML	AR	AM
G1	0.99	0.93	0.91	0.91	0.93	0.92
G2	0.48	0.42	0.42	0.43	0.41	0.41
G3	0.44	0.38	0.37	0.38	0.37	0.37
G4	0.39	0.32	0.32	0.33	0.31	0.31
G5	0.16	0.10	0.10	0.10	0.09	0.09

Thus

$$\begin{aligned} g_{3i\mathrm{AM}}(A) &= \frac{b_i^4 D_i^2}{(b_i^2 A + D_i)^3} \mathrm{E}[(\hat{A}^{\mathrm{AM}} - A)^2] \\ &= \frac{b_i^4 D_i^2}{(b_i^2 A + D_i)^3} \frac{2}{\mathrm{tr}(\Sigma^{-2})}. \end{aligned}$$

Table 3.2: Percent Average Relative Bias of MSPE estimators of $\hat{\theta}_i(Y;\hat{A})$

PATTERN 1							
	PR	FH	RE	ML	AR	AM	RE-N
G1	-0.1	-0.8	-1.2	-1.7	1.1	1.2	-14.1
G2	1.0	0.4	0.1	-0.2	1.7	1.9	-13.1
G3	-0.1	-0.6	-0.7	-0.9	0.7	0.8	-13.8
G4	1.9	1.6	1.8	1.8	2.7	2.9	-11.5
G5	2.6	2.4	2.8	3.2	2.4	2.8	-10.5
PATTERN 2							
	PR	FH	RE	ML	AR	AM	RE-N
G1	-0.8	-2.9	-3.1	-4.6	1.4	0.7	-14.0
G2	7.7	-0.4	-0.9	-2.0	1.8	1.8	-14.2
G3	7.9	-1.3	-1.6	-2.6	0.7	0.8	-14.9
G4	13.0	0.9	0.9	0.0	2.8	2.9	-12.5
G5	34.5	3.5	3.9	4.0	2.5	2.9	-9.4
PATTERN 3							
	PR	FH	RE	ML	AR	AM	RE-N
G1	1.8	-3.8	-1.8	-4.2	3.1	1.9	-9.8
G2	50.6	-3.5	-2.2	-5.0	1.1	1.0	-14.4
G3	61.9	-5.0	-3.6	-6.3	-0.3	-0.5	-15.4
G4	87.4	-3.4	-1.2	-4.1	1.7	1.7	-13.0
G5	726.4	-0.2	1.9	-4.3	0.5	0.8	-7.6

Chapter 4

Parametric Bootstrap Prediction Interval

Point prediction using the empirical best linear unbiased prediction (EBLUP) and the associated mean square prediction error (MSPE) estimation have been discussed extensively in the small area literature. But little advancement has been made in interval prediction problems. Prediction intervals are useful in small area studies in many ways. For example, prediction intervals may help establish if different counties have similar resources and needs, or if different ethnic or other sub-population groups are equally exposed to a particular disease.

In the small area context, prediction intervals are often produced using the standard EBLUP $\pm z_{\alpha/2}\sqrt{\text{mspe}}$ rule, where mspe is an estimate of the true MSPE of the EBLUP and $z_{\alpha/2}$ is the upper $100(1-\alpha/2)\%$ point of the standard normal distribution. These prediction intervals are asymptotically correct, in the sense that the coverage probability converges to $1-\alpha$ for large sample size m. However, they are not efficient in the sense they have either under-coverage or over-coverage problem for small m depending on the particular choice of the MSPE estimator. In statistical terms, the coverage error of such interval is of order $O(m^{-1})$. This is not accurate enough for most small area applications because of small m. See Jiang and Lahiri (2006) for a review of different prediction interval methods.

Chatterjee, Lahiri and Li (2006) proposed a parametric bootstrap method to obtain a prediction interval directly from the bootstrap histogram. Their method is based on the general linear mixed model using ordinary least square estimator of β and REML/ML estimator of variance components. In this dissertation, we address the parametric bootstrap method for the general Fay-Herriot model using the weighted least squares estimator of β and ADM estimator of variance components. The coverage accuracy of this new prediction interval is $O(m^{-3/2})$.

4.1 Parametric Bootstrap Prediction Interval

Consider the following two-level general Fay-Herriot model:

General Fay-Herriot Model:

Level 1 (sampling model): $y_i|\theta_i \overset{ind}{\sim} N(\theta_i, D_i),\ i = 1, \cdots, m;$

Level 2 (linking model): $\theta_i|A \overset{ind}{\sim} N(x_i'\beta, b_i^2 A),\ i = 1, \cdots, m.$

We are interested in obtaining a parametric bootstrap prediction interval for the small area mean θ_i. The conditional distribution of θ_i given y_i is:

$$\theta_i|y_i \sim N(\mu_i, \sigma_i^2), i = 1, \cdots, m. \tag{4.1}$$

where $\mu_i = B_i y_i + (1 - B_i)x_i'\beta$ and $\sigma_i^2 = b_i^2 A D_i/(b_i^2 A + D_i) = b_i^2 A\tilde{D}_i/(A + \tilde{D}_i)$. Here $B_i = \tilde{D}_i/(A + \tilde{D}_i)$ and $\tilde{D}_i = D_i/b_i^2$.

Note that μ_i is the best predictor of θ_i. In this chapter, we use ADM method to estimate A and the following weighted least squares estimator with estimated A

to estimate β:

$$\hat{\beta} = (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}Y \tag{4.2}$$

$$= \beta + (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}(Zv + e) \tag{4.3}$$

where $\hat{\Sigma} = \text{diag}(b_i^2\hat{A} + D_i, i = 1, \cdots, m)$.

By replacing ψ by $\hat{\psi}$, we obtain the estimates $\hat{\mu}_i$ and $\hat{\sigma}_i^2$ of the posterior mean and variance of θ_i. A prediction interval of θ_i can be constructed based on the distribution of $\hat{\sigma}_i^{-1}(\theta_i - \hat{\mu}_i)$ which we denote as $\mathcal{L}_i$. In this paper, we provide an accurate approximation to $\mathcal{L}_i$ using a parametric bootstrap method.

Let

$$Y_i^* = x_i\hat{\beta} + b_i v_i^* + e_i^*, \quad i = 1, \cdots, m$$

where $v_i^* \sim N(0, \hat{A})$ and $e_i^* \sim N(0, D_i)$ are independent of one another. Using the same techniques used to obtain $\hat{\beta}$ and $\hat{A}$, one can obtain $\hat{\beta}^*$ and $\hat{A}^*$ from Y^*. $\hat{\mu}_i^*$ and $\hat{\sigma}_i^*$. The distribution of

$$\hat{\sigma}_i^{-1*}(\theta_i^* - \hat{\mu}_i^*)$$

conditional on the data Y, is the parametric bootstrap approximation $\mathcal{L}_i^*$ of $\mathcal{L}_i$.

Define $h_{ii} = x_i'(X'\Sigma^{-1}X)^{-1}x_i$. Our parametric bootstrap confidence interval is given in the following theorem.

Theorem 4.1 *Assume that* $\sup_{i\geq 1} h_{ii} = O(m^{-1})$. *Then, for a preassigned* $\alpha \in (0, 1)$

and arbitrary $i = 1, \cdots, m$, *let* q_1 *and* q_2 *be real numbers such that*

$$\mathcal{L}_i^*(q_2) - \mathcal{L}_i^*(q_1) = 1 - \alpha.$$

We have

$$\mathrm{P}[\hat{\mu}_i - q_1\hat{\sigma}_i \leq \theta_i \leq \hat{\mu}_i + q_2\hat{\sigma}_i] = 1 - \alpha + O(m^{-3/2}).$$

4.2 Proof of Theorem 4.1.

We establish this result by obtaining an asymptotic expansion of $\mathcal{L}_i$. A similar expansion holds for $\mathcal{L}_i^*$ leading to the result. In this proof, C, with or without suffix, denotes a generic constant.

Let $\phi(\cdot)$ $[\Phi(\cdot)]$ be the standard Normal probability density (cumulative distribution) function. Let ϕ' and ϕ'' denote the first and second derivative of $\phi(\cdot)$. Thus for $x \in R$, we have

$$\phi'(x) = x\phi(x), \qquad \phi''(x) = (x^2 - 1)\phi(x).$$

For any arbitrary $i = 1, \cdots, m$, define

$$Q(q, Y) = \sigma_i^{-1}\{\hat{\mu}_i - \mu_i + q(\hat{\sigma}_i - \sigma_i)\}.$$

Then for any $q \in R$, we have

$$\begin{aligned}
\mathcal{L}_i(q) &= \mathrm{P}(\hat{\sigma}_i^{-1}(\theta_i - \hat{\mu}_i) \le q) \\
&= \mathrm{P}(\theta \le \hat{\mu}_i + q\hat{\sigma}_i) \\
&= \mathrm{E}\Big[\mathrm{P}\Big(\sigma_i^{-1}(\theta_i - \mu_i) \le q + \sigma_i^{-1}(\hat{\mu}_i - \mu_i + q(\hat{\sigma}_i - \sigma_i))\Big)|Y\Big] \\
&= \mathrm{E}\Big[\Phi(q + Q(q, Y))\Big] \\
&= \Phi(q) + \phi(q)\mathrm{E}Q + \frac{1}{2}\phi'(q)\mathrm{E}Q^2 + \frac{1}{2}\mathrm{E}\Big[\int_q^{q+Q}(q + Q - x)^2\phi''(x)dx\Big] \\
&= \Phi(q) + \phi(q)\mathrm{E}Q - \frac{1}{2}q\phi(q)\mathrm{E}Q^2 + \frac{1}{2}\mathrm{E}\Big[\int_q^{q+Q}(x^2 - 1)(q + Q - x)^2\phi(x)dx\Big] \\
&= \Phi(q) + \phi(q)T_1(q) - \frac{1}{2}q\phi(q)T_2(q) + T_3(q).
\end{aligned}$$

Notice that for $x \in (q, q+Q)$, we have $0 \le |q+Q-x| \le |Q|$ and $(x^2-1)\phi(x) \le 2\phi(\sqrt{3})$. Thus, we have

$$\begin{aligned}
& \mathrm{E}\int_q^{q+Q}(q + Q - x)^2(x^2 - 1)\phi(x)dx \\
\le\ & \mathrm{E}\int_q^{q+Q}|(q + Q - x)^2||(x^2 - 1)\phi(x)|dx \\
\le\ & \mathrm{E}Q^2\int_q^{q+Q}2\phi(\sqrt{3})dx \\
\le\ & C\mathrm{E}|Q|^3.
\end{aligned}$$

We shall show that that $\mathrm{E}Q^8 = O(m^{-4})$. Using Lyapunov's Inequality, we have $T_3(q) = O(m^{-3/2})$. We now simplify the expression for $Q(q, Y)$.

First, note that

$$
\begin{aligned}
\hat{\mu}_i - \mu_i &= \frac{\hat{A}}{\hat{A}+\tilde{D}_i} y_i + \frac{\tilde{D}_i}{\hat{A}+\tilde{D}_i} x_i'\hat{\beta} - \frac{A}{A+\tilde{D}_i} y_i - \frac{\tilde{D}_i}{A+\tilde{D}_i} x_i'\beta \\
&= \Big(\frac{\hat{A}}{\hat{A}+\tilde{D}_i} + \frac{\tilde{D}_i}{\hat{A}+\tilde{D}_i} - \frac{A}{A+\tilde{D}_i} - \frac{\tilde{D}_i}{A+\tilde{D}_i}\Big) x_i'\beta \\
&\quad + \frac{\tilde{D}_i}{\hat{A}+\tilde{D}_i} x_i'(X'X)^{-1}X'(v+e) + \Big(\frac{\hat{A}}{\hat{A}+\tilde{D}_i} - \frac{A}{A+\tilde{D}_i}\Big)(v_i+e_i) \\
&= \frac{\tilde{D}_i}{\hat{A}+\tilde{D}_i} x_i'(X'X)^{-1}X'(v+e) + \Big(\frac{\hat{A}}{\hat{A}+\tilde{D}_i} - \frac{A}{A+\tilde{D}_i}\Big)(v_i+e_i) \ a.s. \\
&= \frac{\tilde{D}_i}{A+\tilde{D}_i} x_i'(X'X)^{-1}X'(v+e) \\
&\quad + \Big(\frac{\hat{A}}{\hat{A}+\tilde{D}_i} - \frac{A}{A+\tilde{D}_i}\Big)(d_i' - x_i'(X'X)^{-1}X')(v+e) \ a.s.,
\end{aligned}
$$

where d_i is a column vector with the i^{th} element unity and the rest zeros and a.s. means almost surely.

In view of the above, we can write

$$
Q(q,Y) = Q_1 + Q_2(Y) + Q_3(Y) + Q_4(q,Y),
$$

where

$$
\begin{aligned}
Q_1 &= \sigma_i^{-1} \frac{\tilde{D}_i}{A+\tilde{D}_i} x_i'(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}(Zv+e), \\
Q_2(Y) &= \sigma_i^{-1} \Big[\frac{\tilde{D}_i}{\hat{A}+\tilde{D}_i} x_i'(X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}(Zv+e) \\
&\quad - \frac{\tilde{D}_i}{A+\tilde{D}_i} x_i'(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}(Zv+e) \Big], \\
Q_3(Y) &= \sigma_i^{-1} \Big(\frac{\hat{A}}{\hat{A}+\tilde{D}_i} - \frac{A}{A+\tilde{D}_i} \Big)(b_i v_i + e_i), \\
Q_4(q,Y) &= q\sigma_i^{-1}(\hat{\sigma}_i - \sigma_i).
\end{aligned}
$$

In order to prove Theorem 4.1, we need to show that $\mathrm{E}Q = O(m^{-1})$, $\mathrm{E}Q^2 = O(m^{-1})$ and $\mathrm{E}Q^8 = O(m^{-4})$, or equivalently, $\mathrm{E}Q_i = O(m^{-1})$, $\mathrm{E}Q_i^2 = O(m^{-1})$ and $\mathrm{E}Q_i^8 = O(m^{-4})$, for $i = 1, \cdots, 4$.

Obviously, $\mathrm{E}Q_1 = 0$. Using the assumption $h_{ii} = O(m^{-1})$, we have

$$\begin{aligned}
\mathrm{E}Q_1^2 &= \frac{\tilde{D}_i}{b_i^2 A(A+\tilde{D}_i)} x_i'(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\mathrm{E}((v+e)(v+e)')\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}x_i \\
&= \frac{\tilde{D}_i}{b_i^2 A(A+\tilde{D}_i)} x_i'(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}x_i \\
&= \frac{\tilde{D}_i}{b_i^2 A(A+\tilde{D}_i)} \tilde{h}_{ii} \\
&= O(m^{-1}).
\end{aligned}$$

Define $\tilde{h}_{ii} = x_i'(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-2}X(X'\Sigma^{-1}X)^{-1}x_i$. Under the assumption $h_{ii} = O(m^{-1})$, it is easily to see that $\tilde{h}_{ii} = O(m^{-1})$. By Cauchy-Schwarz inequality we have $\tilde{h}_{ij} \le \tilde{h}_{ii}^{1/2}\tilde{h}_{jj}^{1/2} = O(m^{-1})$. Then for $P_X = X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}$, we have

$$\begin{aligned}
\mathrm{E}Q_1^8 &= \left(\frac{\tilde{D}_i}{\sigma_i(A+\tilde{D}_i)}\right)^8 \mathrm{E}(x_i'(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}(v+e))^8 \\
&= \left(\frac{\tilde{D}_i}{\sigma_i(A+\tilde{D}_i)}\right)^8 \mathrm{E}(d_i'P_XP_X(v+e))^8 \\
&\le \left(\frac{\tilde{D}_i}{\sigma_i(A+\tilde{D}_i)}\right)^8 \mathrm{E}(d_i'P_XP_X'd_i\ (v+e)'P_XP_X'(v+e))^4 \\
&= \left(\frac{\tilde{D}_i}{\sigma_i(A+\tilde{D}_i)}\right)^8 \mathrm{E}(\sum_{i=1}^m\sum_{j=1}^m (v_i+e_i)(v_j+e_j)\tilde{h}_{ij})^4 \\
&\le Cm^{-8}\mathrm{E}(\sum_{i=1}^m (v_i+e_i)(v_i+e_i))^4 \\
&= O(m^{-4}),
\end{aligned}$$

where d_i is a column vector with the i^{th} element unity and the rest zeros.

The term $Q_2(Y)$ is considerably more complicated than Q_1. We first break down this quantity in terms of more tractable variables and remainder terms.

Note that

$$\begin{aligned}
Q_2(Y) &= \sigma_i^{-1}\Big[\frac{\tilde{D}_i}{\hat{A}+\tilde{D}_i}x_i'(X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}(Zv+e) \\
&\quad -\frac{\tilde{D}_i}{A+\tilde{D}_i}x_i'(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}(Zv+e)\Big], \\
&= \sigma_i^{-1}\Big[\Big(\frac{\tilde{D}_i}{\hat{A}+\tilde{D}_i}-\frac{\tilde{D}_i}{A+\tilde{D}_i}\Big) \\
&\quad x_i'\Big((X'\hat{\Sigma}^{-1}X)^{-1}-(X'\Sigma^{-1}X)^{-1}\Big)X'\Big(\hat{\Sigma}^{-1}-\Sigma^{-1}\Big)(Zv+e) \\
&\quad +\Big(\frac{\tilde{D}_i}{\hat{A}+\tilde{D}_i}-\frac{\tilde{D}_i}{A+\tilde{D}_i}\Big)x_i'\Big((X'\hat{\Sigma}^{-1}X)^{-1}-(X'\Sigma^{-1}X)^{-1}\Big)X'\Sigma^{-1}(Zv+e) \\
&\quad +\Big(\frac{\tilde{D}_i}{\hat{A}+\tilde{D}_i}-\frac{\tilde{D}_i}{A+\tilde{D}_i}\Big)x_i'(X'\Sigma^{-1}X)^{-1}X'\Big(\hat{\Sigma}^{-1}-\Sigma^{-1}\Big)(Zv+e) \\
&\quad +\Big(\frac{\tilde{D}_i}{\hat{A}+\tilde{D}_i}-\frac{\tilde{D}_i}{A+\tilde{D}_i}\Big)x_i'(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}(Zv+e) \\
&\quad +\frac{\tilde{D}_i}{A+\tilde{D}_i}x_i'\Big((X'\hat{\Sigma}^{-1}X)^{-1}-(X'\Sigma^{-1}X)^{-1}\Big)X'\Big(\hat{\Sigma}^{-1}-\Sigma^{-1}\Big)(Zv+e) \\
&\quad +\frac{\tilde{D}_i}{A+\tilde{D}_i}x_i'\Big((X'\hat{\Sigma}^{-1}X)^{-1}-(X'\Sigma^{-1}X)^{-1}\Big)X'\Sigma^{-1}(Zv+e) \\
&\quad +\frac{\tilde{D}_i}{A+\tilde{D}_i}x_i'(X'\Sigma^{-1}X)^{-1}X'\Big(\hat{\Sigma}^{-1}-\Sigma^{-1}\Big)(Zv+e)\Big].
\end{aligned}$$

Using the Taylor series expansion, we have

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

and

$$
\begin{aligned}
& \hat{\Sigma}^{-1} - \Sigma^{-1} \\
= \;& \mathrm{diag}\{\frac{1}{\hat{A} + \tilde{D}_i}\}_{i=1}^m - \mathrm{diag}\{\frac{1}{A + \tilde{D}_i}\}_{i=1}^m \\
= \;& -(\hat{A} - A)\mathrm{diag}\{\frac{1}{(A + \tilde{D}_i)^2}\}_{i=1}^m + (\hat{A} - A)^2\mathrm{diag}\{\frac{1}{(A + \tilde{D}_i)^3}\}_{i=1}^m \\
& -(\hat{A} - A)^3\mathrm{diag}\{\frac{1}{(A^* + \tilde{D}_i)^4}\}_{i=1}^m.
\end{aligned}
$$

Using the regularity condition $(X'\Sigma^{-1}X)^{-1} = O(m^{-1})$, we have

$$
(X'\hat{\Sigma}^{-1}X)^{-1} - (X'\Sigma^{-1}X)^{-1} = O(m^{-1}).
$$

Before analyzing Q_2 and Q_3, we evaluate the moments of $(\hat{A}^{\mathrm{ADM}} - A)$ first. From Chapter 2, we know

$$
\mathrm{E}(\hat{A}^{\mathrm{ADM}} - A) = O(m^{-1}). \tag{4.4}
$$

Using Lemma 5.1 of Das, Jiang and Rao (2004) and the fact that $y'P^2y = u'P^2u$ $(u = v + e)$, we can show that for any $g \geq 2$, $\mathrm{E}|y'P^2y - \mathrm{tr}(P)|^g \leq c||P^2||_2^g = O(m^{g/2})$. Then we have for any $g \geq 2$, $\mathrm{E}|\hat{A}^{\mathrm{ADM}} - A|^g = O(m^{-g/2})$.

Now we examine the term $Q_2(Y)$. Using the Taylor expansion, we first have

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

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

Note that

$$\begin{aligned}
Q_2(Y) &= \sigma_i^{-1}\Big((\hat{A}-A)\frac{\tilde{D}_i}{(A+\tilde{D}_i)^2}(d_i'-x_i'(X'X)^{-1}X')(v+e) \\
&\quad -(\hat{A}-A)^2\frac{\tilde{D}_i}{(A+\tilde{D}_i)^3}(d_i'-x_i'(X'X)^{-1}X')(v+e) \\
&\quad +(\hat{A}-A)^3\frac{\tilde{D}_i}{(A^*+\tilde{D}_i)^4}(d_i'-x_i'(X'X)^{-1}X')(v+e)\Big) \\
&= L_1+L_2+L_3;
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{E}L_1 &= C\mathrm{E}[(\hat{A}-A)(d_i'-x_i'(X'X)^{-1}X')(v+e)] \\
&= O(m^{-1})\mathrm{E}(\mu'P^2\mu(d_i'-x_i'(X'X)^{-1}X')\mu)+r \\
&= O(m^{-1})\mathrm{E}(\mu'P^2\mu(\mu_i-\sum_{j=1}^{m}h_{ij}\mu_j)+r \\
&= O(m^{-1}).
\end{aligned}$$

Using Holder inequality, we have

$$\begin{aligned}
\mathrm{E}L_2 &= C\mathrm{E}[(\hat{A}-A)^2(d_i'-x_i'(X'X)^{-1}X')(v+e)] \\
&\leq C\sqrt{\mathrm{E}(\hat{A}-A)^4\mathrm{E}((d_i'-x_i'(X'X)^{-1}X')\mu)^2} \\
&= C\sqrt{O(m^{-2})\mathrm{E}(\mu_i-\sum_{j=1}^{m}h_{ij}\mu_j)^2} \\
&= O(m^{-1}).
\end{aligned}$$

Using a similar argument, we can show that $\mathrm{E}L_3 = O(m^{-3/2})$, $\mathrm{E}L_1^2 = O(m^{-1})$, $\mathrm{E}L_1^8 = O(m^{-4})$, $\mathrm{E}L_2^2 = O(m^{-1})$, $\mathrm{E}L_2^8 = O(m^{-1})$ and that all the other terms have lower orders. Thus, we establish that $\mathrm{E}Q_2 = O(m^{-1})$, $\mathrm{E}Q_2^2 = O(m^{-1})$ and $\mathrm{E}Q_2^8 = O(m^{-4})$.

Finally, we examine the term $Q_3(q, Y)$. Letting $W = \sigma_i^{-2}\left(\hat{\sigma}_i^2 - \sigma_i^2\right)$, we have

$$
\begin{aligned}
Q_3(q, Y) &= q\sigma_i^{-1}\left(\hat{\sigma}_i - \sigma_i\right) \\
&= q\left(\sigma_i^{-1}\hat{\sigma}_i - 1\right) \\
&= q\left[\left\{\sigma_i^{-2}\left(\hat{\sigma}_i^2 - \sigma_i^2\right) + 1\right\}^{1/2} - 1\right] \\
&= q\left[\{W + 1\}^{1/2} - 1\right], \\
&= q\left[W/2 - W^2/8 + r_n\right].
\end{aligned}
$$

The last line can be justified by Taylor series expansion of $(1+W)^{1/2}$, the remainder being $r_n = O(W^3)$.

Now, we simplify W by the Taylor expansion.

$$
\begin{aligned}
W &= \sigma_i^{-2}\left(\hat{\sigma}_i^2 - \sigma_i^2\right) \\
&= \sigma_i^{-2}\left(\frac{\hat{A}D_i}{\hat{A} + D_i} - \frac{AD_i}{A + D_i}\right) \\
&= \sigma_i^{-2}\left((\hat{A} - A)\frac{D_i^2}{(A + D_i)^2} - (\hat{A} - A)^2\frac{D_i^2}{(A + D_i)^3} + (\hat{A} - A)^3\frac{D_i^2}{(A^* + D_i)^4}\right).
\end{aligned}
$$

Using the moment properties of $\hat{A}^{\mathrm{ADM}} - A$ derived in Chapter 3, we have $\mathrm{E}W = O(m^{-1})$, $\mathrm{E}W^2 = O(m^{-1})$, and $\mathrm{E}W^8 = O(m^{-4})$. Furthermore, $\mathrm{E}Q_3 = O(m^{-1})$, $\mathrm{E}Q_3^2 = O(m^{-1})$, and $\mathrm{E}Q_3^8 = O(m^{-4})$.

The above arguments support the final expansion:

$$\mathcal{L}_i(q) = \Phi(q) + m^{-1}\gamma(q, \beta, A) + O(m^{-3/2}),$$

where $\gamma(\cdot, \cdot, \cdot)$ is a smooth function of order $O(1)$.

A similar representation holds for $\mathcal{L}_i^*(q)$ with $\hat{\beta}$ and $\hat{A}$ in place of β and A respectively. Thus we have

$$\begin{aligned} 1 - \alpha &= \mathcal{L}_i^*(q_2) - \mathcal{L}_i^*(q_1) \\ &= \Phi(q_2) - \Phi(q_1) + m^{-1}\gamma(q_2, \hat{\beta}, \hat{A}) - m^{-1}\gamma(q_1, \hat{\beta}, \hat{A}) + O(m^{-3/2}), \end{aligned}$$

which yields

$$\Phi(q_2) - \Phi(q_1) = 1 - \alpha - (m^{-1}\gamma(q_2, \hat{\beta}, \hat{A}) - m^{-1}\gamma(q_1, \hat{\beta}, \hat{A})) + O(m^{-3/2}).$$

Finally, we have

$$\begin{aligned} & \mathrm{P}[\hat{\mu}_i - q_1\hat{\sigma}_i \le \theta_i \le \hat{\mu}_i + q_2\hat{\sigma}_i] \\ = \; & \mathcal{L}_i(q_2) - \mathcal{L}_i(q_1) \\ = \; & \Phi(q_2) - \Phi(q_1) + m^{-1}\gamma(q_2, \beta, A) - m^{-1}\gamma(q_1, \beta, A) + O(m^{-3/2}) \\ = \; & 1 - \alpha + (m^{-1}\gamma(q_2, \beta, A) - m^{-1}\gamma(q_1, \beta, A)) - (m^{-1}\gamma(q_2, \hat{\beta}, \hat{A}) \\ & -m^{-1}\gamma(q_1, \hat{\beta}, \hat{A})) + O(m^{-3/2}) \\ = \; & 1 - \alpha + O(m^{-3/2}). \end{aligned}$$

■

4.3 A Monte Carlo Simulation

We use the simulation setting of Chapter 2 and Chapter 3 and consider the following three patterns of D_i:

1. Pattern 1: $D_i = (0.7, 0.6, 0.5, 0.4, 0.3)$;
2. Pattern 2: $D_i = (2.0, 0.6, 0.5, 0.4, 0.2)$;
3. Pattern 3: $D_i = (4.0, 0.6, 0.5, 0.4, 0.1)$.

We compare seven prediction intervals of θ_i using coverage probabilities and average lengths. Different prediction intervals include the Cox's empirical Bayes prediction interval with $\hat{A}^{\text{RE}}$ (labeled as Cox-RE), three traditional prediction intervals of type EBLUP $\pm\, 1.96\sqrt{\text{mspe}}$, where mspe is the estimator of the MSPE of EBLUP based on $\hat{A}^{\text{PR}}$ (labelled PR) , $\hat{A}^{\text{FH}}$ (labeled FH) and $\hat{A}^{\text{RE}}$ (labeled RE) estimators of A , and three prediction intervals based on the proposed parametric bootstrap methods using $\hat{A}^{\text{AR}}$ (labeled PB-AR), $\hat{A}^{\text{AM}}$ (labeled PB-AM) and $\hat{A}^{\text{RE}}$ (labeled PB-RE) estimators of A.

All the results are based on $10,000$ simulation runs. Table 4.1-4.2 reports the simulated coverage probabilities and average lengths of seven different prediction intervals with nominal coverage 0.95. The Cox prediction method consistently undercover, which is due to its usage of the underestimated MSPE estimator. The parametric bootstrap prediction interval method with REML estimator of A consistently overcover. That is perhaps because the frequent zero estimate of A by REML method. In the proposed parametric bootstrap method, the estimator of A occurs in the denominator of the approximated pivot, then zero estimates create extremely large values of the pivot. The cut-off points resulting from such pivot values tend to be extraordinarily large, which results in the large length of the prediction interval and over-coverage.

Now let us look at the other five prediction intervals. In pattern 1, which has almost balanced sampling variances, the five prediction intervals are almost identical. They have perfect coverage probabilities and lengths. However, the situation changes in the Pattern 2 and 3. For those unbalanced sampling error cases, the traditional Taylor prediction intervals with Fay-Herriot estimator and REML estimators always have undercoverage problem. The traditional Taylor prediction interval with the Prasad-Rao estimator always has overcoverage problem. It is generally seen that an increase in the sampling variances results in even more poorly performance of the traditional intervals. In contrast, the performances of our parametric bootstrap methods with ADM estimators remain stable over all of three different patterns and always close to the target nominal level.

Table 4.1: Average Coverage of Different Intervals (Nominal Coverage=0.95)

	FH	PR	RE	Cox-RE	PB-RE	PB-AR	PB-AM
PATTERN 1							
	FH	PR	RE	Cox-RE	PB-RE	PB-AR	PB-AM
G1	93.7	94.0	93.6	90.0	97.3	94.6	94.2
G2	94.2	94.5	94.2	90.1	97.5	94.5	94.6
G3	94.4	94.6	94.4	90.6	97.3	94.4	94.4
G4	94.5	94.7	94.6	90.8	97.2	94.6	94.4
G5	95.0	95.1	95.0	91.2	97.0	94.7	94.6
PATTERN 2							
	FH	PR	RE	Cox-RE	PB-RE	PB-AR	PB-AM
G1	91.7	92.6	92.1	87.9	97.6	94.3	94.5
G2	93.6	95.6	93.8	89.8	97.3	94.6	94.4
G3	93.9	95.7	94.0	89.9	97.1	94.5	94.4
G4	94.5	96.3	94.6	90.1	97.0	94.4	94.6
G5	95.2	96.4	95.2	91.3	96.7	94.3	94.4
PATTERN 3							
	FH	PR	RE	Cox-RE	PB-RE	PB-AR	PB-AM
G1	89.6	90.7	90.8	88.1	97.7	94.4	94.2
G2	91.6	98.0	93.3	90.0	97.0	94.3	94.5
G3	92.1	98.1	93.6	90.5	96.7	94.7	94.5
G4	92.5	98.1	93.7	90.7	96.9	94.5	94.4
G5	95.3	97.6	95.3	93.0	96.2	94.6	94.8

Table 4.2: Average Length of Different Intervals (Nominal Coverage=0.95)

	FH	PR	RE	Cox-RE	PB-RE	PB-AR	PB-AM
PATTERN 1							
	FH	PR	RE	Cox-RE	PB-RE	PB-AR	PB-AM
G1	2.63	2.64	2.63	2.37	3.44	2.65	2.65
G2	2.51	2.52	2.51	2.27	3.25	2.52	2.51
G3	2.37	2.38	2.37	2.14	3.02	2.36	2.35
G4	2.20	2.21	2.20	1.99	2.75	2.16	2.16
G5	1.98	1.99	1.98	1.80	2.41	1.93	1.93
PATTERN 2							
	FH	PR	RE	Cox-RE	PB-RE	PB-AR	PB-AM
G1	3.32	3.40	3.33	2.98	4.67	3.55	3.56
G2	2.52	2.68	2.51	2.25	3.30	2.52	2.53
G3	2.38	2.55	2.37	2.13	3.07	2.36	2.37
G4	2.20	2.39	2.20	1.97	2.79	2.17	2.17
G5	1.69	1.92	1.68	1.53	1.99	1.62	1.63
PATTERN 3							
	FH	PR	RE	Cox-RE	PB-RE	PB-AR	PB-AM
G1	3.55	3.76	3.59	3.31	5.03	4.01	4.00
G2	2.46	3.30	2.49	2.26	3.20	2.53	2.53
G3	2.32	3.23	2.35	2.14	2.98	2.36	2.37
G4	2.15	3.15	2.17	1.99	2.72	2.17	2.18
G5	1.25	2.89	1.22	1.15	1.36	1.19	1.19

Chapter 5

SAIPE Data Analysis

The Small Area Income and Poverty Estimation (SAIPE) project is an ongoing project of the United States Census Bureau to estimate the counts of poor school-age children by state, county, and ultimately school district. The primary source of the data comes from the Current Population Survey (CPS), an important nationwide complex survey conducted by the Census Bureau. The Census Bureau employs an EBLUP method that combines the CPS data with income data available from the Internal Revenue Service (IRS), food stamp data, and Census residuals (see National Research Council, 2000, U.S. Census Bureau, 2005 for details).

In this Chapter, we analyze the SAIPE state level data using the Fay-Herriot model. We first compare various variance estimation methods and point out the zero estimates produced by the currently used methods. In contrast, the ADM estimators proposed in this dissertation always yield positive variance estimates across year 1989 - 1997. We also compare the performances of shrinkage estimators and EBLUP estimators using different variance estimation methods. With the theoretical support presented in the Chapter 3 and 4, we demonstrate the MSPE estimators and parametric prediction intervals using ADM estimators for SAIPE data. Several criteria are employed to describe the application of our ADM estimators.

5.1 Various Estimations of A

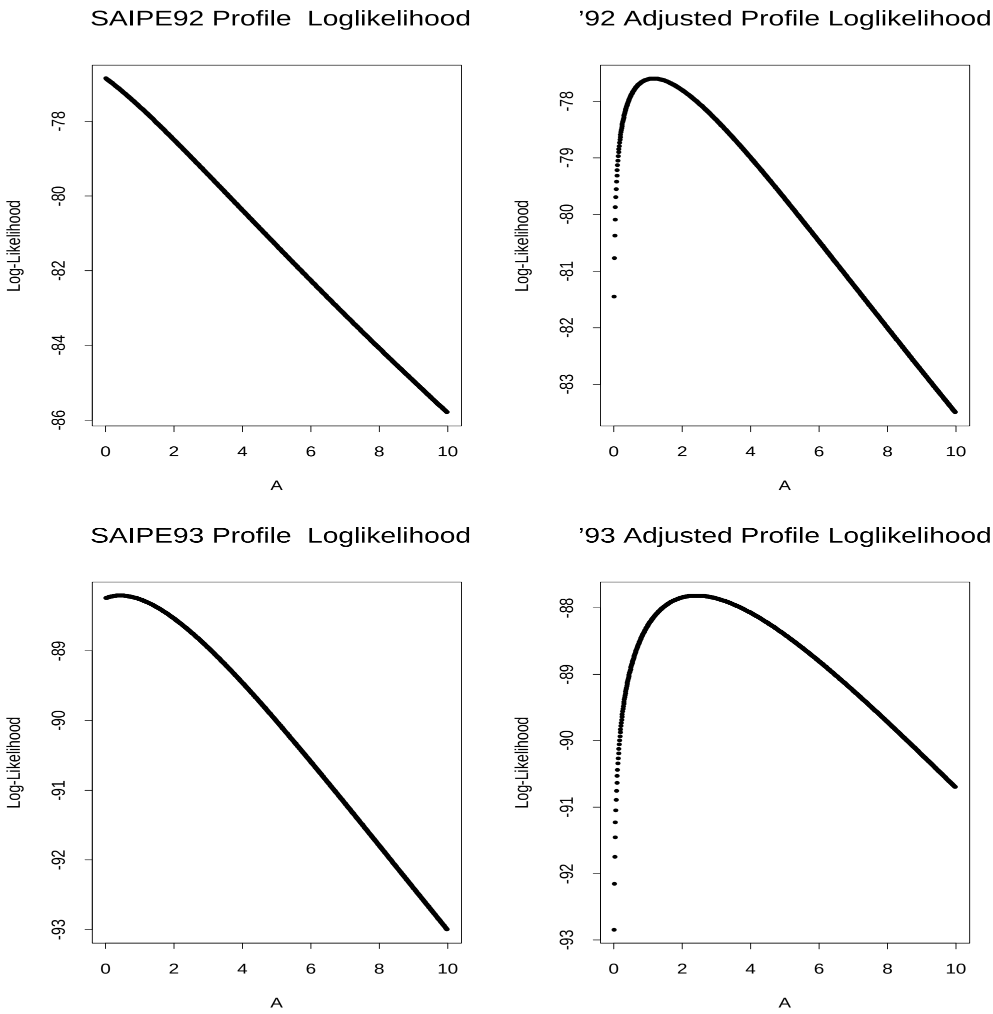
Table 5.1 displays the estimates of A using the Prasad-Rao simple method-of-moments (denoted by PR), the Fay-Herriot method-of-moments (denoted by FH), REML (denoted by RE), maximum likelihood (denoted by ML), ADM REML (denoted by AR) and ADM ML (denoted by AM) methods for five years, from 1989 to 1993. Bell (1999) produced the same number for REML and ML estimates. It is interesting to note that all the methods except the ADM methods are subject to zero estimate. As noted earlier zero estimate of A causes problem in EBLUP methodology.

Table 5.1: Different Variance Estimation Methods

Year	PR	FH	RE	ML	AR	AM
1989	0.00	0.06	0.00	0.00	1.83	1.15
1990	1.13	0.85	0.00	0.00	2.31	1.42
1991	0.00	0.00	0.00	0.00	1.67	1.14
1992	0.00	0.00	0.00	0.00	1.74	1.17
1993	5.85	2.67	1.70	0.43	3.61	2.42

Let us now illustrate the ML and ADM ML methods using the SAIPE data 1992 and 1993. Note that the maximum likelihood estimate of A for the years 1992 is zero. In Figure 5.1, we draw the profile log-likelihood function of A for year 1992. We can readily see that the function attains its maximum on the boundary which gives the zero ML estimate. On the other hand, the ADM ML method maximizes an adjusted profile likelihood which is just the product of the profile likelihood and A. Thus, the multiplier A pushes the maximum point of profile likelihood to the right. Figure 5.1 shows that the ADM ML estimates always fall into the positive half axis.

Figure 5.1: SAIPE 92 and 93 Log-Likelihood



For the year 1993, maximum likelihood estimate of A is 0.43. In this case also, the multiplier A pushes the maximum point of profile likelihood to the right yielding a larger ADM ML estimate (2.42). As described in Chapter 2, the overestimation of the ADM methods can be used to make up for the underestimation of the shrinkage estimator B_i due to its own convexity.

Next, we compute the shrinkage estimators of B_i. We select the year 1992 and 1993 to compare the impact of the REML, maximum likelihood and two ADM estimators of A on B_i. Recall D_i is the sampling error, $\hat{B}_i = D_i/(\hat{A} + D_i)$ and $\hat{\theta}_i^{\text{EBLUP}} = (1 - \hat{B}_i)y_i + \hat{B}_i x_i'\hat{\beta}$. Figure 5.2 displays B_i versus states, where states are sorted in increasing order of the D_i's.

Since $\hat{A}^{\text{RE}} = \hat{A}^{\text{ML}} = 0$ for 1992, we have $\hat{B}_i = 1$ for all the states. While the ADM estimates of A yield large $\hat{B}_i$ for the states with large D_i and small $\hat{B}_i$ for the states with small D_i. When D_i is small, the direct estimate y_i is reasonable. Thus, in this situation a sensible EBLUP method should put more weight to y_i than the synthetic regression part. In this sense, the REML and ML does not provide a sensible method; but the ADM's do. For year 1993, all the methods produce none-zero estimates of A, and yield large $\hat{B}_i$ for the states with large D_i and small $\hat{B}_i$ for the states with small D_i. However, ADM estimates of B are much smaller than REML and ML estimates. ADM methods put more weight to the direct part and are more conservative.

5.2 MSPE Estimates Using ADM Methods

In Chapter 3, we derived the estimator of MSPE for $\hat{\theta}_i^{\text{EBLUP}}$ when the ADM methods are used to estimate A. In this section, we use the formula of Chapter 3 to obtain MSPE estimates for all the states in the year 1992 and 1993.

Table 5.2-5.3 compares the coefficient of variance(CV) of the EBLUP that uses the ADM methods with the direct method. For the direct method, the CV is given by: CV direct $= \sqrt{D_i}/y_i$. The CV's for the ADM REML and ADM ML methods are given by:

$$\text{CV AR} = \sqrt{\text{mspe}_i^{\text{AR}}}/[\hat{\theta}_i^{\text{EBLUP}}(\hat{A}^{\text{AR}})];$$

$$\text{CV AM} = \sqrt{\text{mspe}_i^{\text{AM}}}/[\hat{\theta}_i^{\text{EBLUP}}(\hat{A}^{\text{AM}})].$$

The gain in CV due to the use of EBLUP over the direct method is given by CV ADM Gain = (CV direct − CV ADM)/CV direct × 100%. The results show that both the EBLUP methods have about 60% gain over the direct method in term of the CV.

5.3 Parametric Bootstrap Prediction Interval

In this section, we compare the parametric bootstrap prediction intervals discussed in Chapter 4 with the Cox prediction interval that uses the Fay-Herriot method of estimating A. For this purpose, we use the SAIPE data for the year 1993 and use the Fay-Herriot method of estimating A in obtaining the Cox empirical Bayes prediction interval. In Chapter 4, we have seen that the Cox prediction interval is prone to undercover due to unduly narrow prediction interval. In Table 5.4, we report the inflations in the length that result from the use of reliable parametric bootstrap prediction interval that uses the ADM methods. In this table, we arrange the states in increasing order of the sampling variances D_i. We define the inflation as:

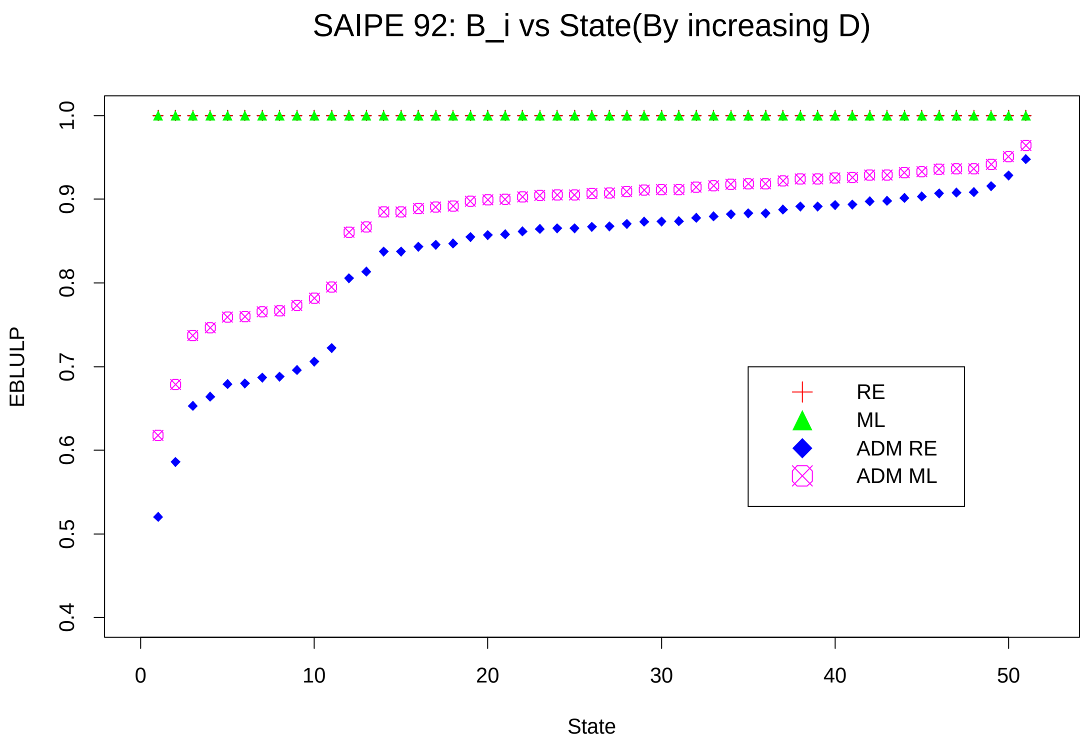
$$\text{ADM Length Inflation} = (\text{Length of ADM} - \text{Length of Cox})/\text{Length of Cox} \times 100\%.$$

Table 5.2: SAIPE 92 Percent Gain in CV of EBLUP over the direct estimator

State	CV Direct	CV AR % Gain	CV AM % Gain	State	CV Direct	CV AR % Gain	CV AM % Gain
CA	0.07	12.2	8.2	IN	0.29	78.1	80.1
NY	0.07	23.2	21.4	TN	0.20	68.8	70.1
NJ	0.14	44.5	46.1	ME	0.23	66.7	68.2
PA	0.13	48.6	50.7	SC	0.13	59.1	60.4
IL	0.10	42.0	43.8	MS	0.12	47.5	47.9
MI	0.11	46.7	48.5	NM	0.13	60.8	61.8
OH	0.11	44.8	46.3	MD	0.22	69.0	71.2
MA	0.11	36.7	37.5	WA	0.26	70.7	72.3
TX	0.09	44.3	45.4	MT	0.25	77.9	79.7
FL	0.10	48.3	50.4	CT	0.22	62.5	64.3
NC	0.09	38.2	39.1	OK	0.20	78.0	80.0
NE	0.19	50.6	52.3	AL	0.17	72.0	73.5
IA	0.23	64.5	66.8	AR	0.20	71.7	72.9
WI	0.26	68.1	69.9	RI	0.21	72.2	74.2
ID	0.16	45.3	46.7	MO	0.21	77.3	79.4
KS	0.26	74.3	76.7	KY	0.18	76.1	77.7
WY	0.28	67.0	68.9	WV	0.13	54.4	55.1
CO	0.29	76.7	78.9	ND	0.32	76.3	77.6
UT	0.29	63.0	64.4	SD	0.26	71.8	72.9
VA	0.23	75.6	78.1	AZ	0.19	77.0	78.8
MN	0.20	62.7	65.1	GA	0.15	75.3	77.5
NH	0.37	68.2	70.0	DE	0.37	83.3	84.9
OR	0.31	66.1	67.3	VT	0.36	74.7	75.8
NV	0.19	62.3	64.5	LA	0.15	68.4	69.3
AK	0.35	71.9	73.6	DC	0.18	70.7	71.7
HI	0.22	51.5	52.4				

Table 5.4 shows that the parametric bootstrap prediction intervals with ADM methods have longer length than the Cox interval - this is consistent with our theory and simulations presented in Chapter 4.

Figure 5.2: SAIPE 92 and 93 B_i versus State



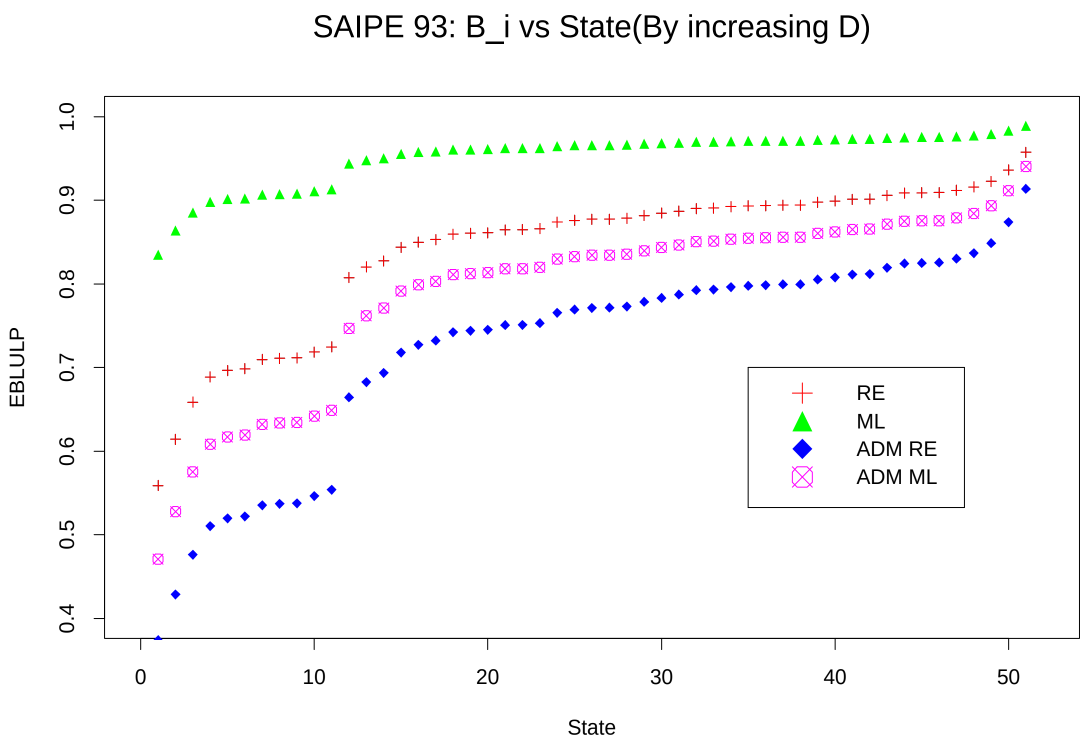


Table 5.3: SAIPE 93 Percent Gain in CV of EBLUP over the direct estimator

State	CV Direct	CV AR % Gain	CV AM % Gain	State	CV Direct	CV AR % Gain	CV AM % Gain
CA	0.06	5.0	0.6	ME	0.20	47.9	49.5
NY	0.07	16.5	14.7	AK	0.37	59.7	62.2
NJ	0.11	18.1	16.9	NM	0.20	63.3	65.1
PA	0.11	28.3	28.9	MT	0.26	60.7	63.2
MA	0.13	27.3	27.9	MS	0.12	39.7	40.3
IL	0.11	30.3	31.5	TN	0.13	38.0	38.8
TX	0.09	31.8	32.7	HI	0.29	48.9	50.4
MI	0.09	23.6	23.5	CT	0.25	51.2	53.2
OH	0.12	31.1	32.0	OK	0.17	55.7	58.1
FL	0.08	20.9	20.3	ND	0.39	66.7	68.7
NC	0.12	28.5	29.0	WA	0.34	64.3	66.4
NE	0.21	31.9	32.8	MD	0.30	57.8	60.0
IA	0.27	50.6	53.2	AL	0.19	59.9	61.7
ID	0.21	36.7	38.4	RI	0.19	51.5	53.6
WI	0.20	38.8	40.2	AR	0.16	49.8	51.5
WY	0.31	50.3	52.6	SD	0.26	55.2	57.0
CO	0.29	57.1	59.8	GA	0.24	68.0	70.2
KS	0.20	46.9	49.2	MO	0.21	59.3	61.4
UT	0.23	33.7	34.9	KY	0.16	58.8	60.8
MN	0.27	50.7	53.0	WV	0.15	51.2	52.5
NV	0.26	49.9	52.4	AZ	0.19	60.8	63.1
VA	0.29	62.7	65.4	DE	0.34	64.4	66.5
NH	0.24	27.5	27.9	VT	0.35	63.0	64.6
IN	0.33	65.7	68.0	LA	0.13	51.2	52.5
SC	0.13	39.4	40.5	DC	0.13	46.0	47.2
OR	0.25	46.3	48.1				

Table 5.4: SAIPE 93 Percent Length Inflation for the Parametric Bootstrap Prediction Intervals with ADM estimators over the Cox Prediction Interval

State	Cox FH Length	PB AR % Inflation	PB AM % Inflation	State	Cox FH Length	PB AR % Inflation	PB AM % Inflation
CA	3.82	30.9	32.8	ME	4.79	46.7	42.6
NY	4.01	36.5	27.9	AK	4.79	57.3	43.0
NJ	4.15	19.9	22.3	NM	4.80	56.5	51.2
PA	4.24	31.4	18.2	MT	4.81	46.3	41.3
MA	4.27	35.0	35.0	MS	4.82	78.1	83.1
IL	4.27	27.9	24.4	TN	4.83	57.7	54.0
TX	4.31	42.2	29.1	HI	4.83	51.6	57.9
MI	4.31	31.9	29.2	CT	4.83	51.2	54.8
OH	4.31	29.1	26.1	OK	4.83	38.8	32.1
FL	4.34	34.6	32.0	ND	4.83	51.4	45.2
NC	4.35	33.2	36.0	WA	4.84	48.3	47.8
NE	4.60	42.8	29.2	MD	4.84	39.3	34.0
IA	4.63	35.4	28.5	AL	4.85	52.1	39.5
ID	4.65	48.2	43.8	RI	4.85	46.0	43.2
WI	4.70	47.6	43.5	AR	4.85	54.5	48.7
WY	4.71	42.9	41.4	SD	4.86	52.3	56.6
CO	4.72	37.6	33.6	GA	4.87	45.3	31.3
KS	4.74	31.7	23.2	MO	4.88	35.4	28.4
UT	4.74	54.3	48.4	KY	4.88	38.6	35.8
MN	4.75	43.9	33.8	WV	4.88	88.6	82.9
NV	4.76	44.3	30.5	AZ	4.88	52.4	52.8
VA	4.76	35.7	23.8	DE	4.89	35.8	34.4
NH	4.76	53.1	38.4	VT	4.91	62.1	54.4
IN	4.78	38.1	24.3	LA	4.95	61.4	61.3
SC	4.79	50.6	47.4	DC	5.00	83.3	73.2
OR	4.79	56.0	59.7				

Chapter 6

Future Research

For the Fay-Herriot model, zero estimate of the variance component A causes severe problem in the EBLUP-based inference for small area mean. In this dissertation, we have illustrated the use of ADM-based methods as a possible solution to this problem. The ADM methods have not been developed for linear mixed models with more than one variance components. As an immediate future research, we would like to explore the ADM-based methods for the nested error regression model described in Chapter 1.

Note that the nested error regression model induces the following area level model:

$$\bar{y}_i = \bar{x}_i^T \beta + v_i + \bar{e}_i, i = 1, \cdots, m,$$

where area specific random effects $\{v_i\}$ and sampling errors $\bar{e}_i$ are independently distributed with $v_i \overset{ind}{\sim} N(0, \sigma_v^2)$, $\bar{e}_i \overset{ind}{\sim} N(0, \sigma_e^2 k_i)$, and $k_i =$. Unlike the Fay-Herriot model, here the sampling variances $D_i = D_i(\sigma_e^2)$ are unknown and involves an additional variance component estimation. For this model, the weighted least squares estimator of β depends on two variance components. Note that the standard method-of-moments estimator of σ_e^2 is the within area mean square which is strictly positive and consistent. However, the estimation of σ_v^2 is problematic. We can apply the same ADM methods developed in this dissertation treating the above model as a

Fay-Herriot model with $D_i = \hat{\sigma}_e^2 k_i$. We shall investigate the theoretical properties of this proposed method. The extension of this dissertation to the general linear mixed model with multiple variance components is a highly non-trivial problem that we will consider for future research.

We used normality in developing different methods in the dissertation. Certain robustness properties of the simple method-of-moments against non-normality of the random effects are well-known in the literature. Does such robustness extend to the ADM-based methods? This is certainly a topic of interest and we plan to explore this as a future research.

Chapter A

SAS Code for Simulation and Data Analysis

```
-------Chapter 2: Comparison of A, B, Theta
                              for different methods-------

Simulation setting is exact same as Datta, Rao and Smith(2005).
Model: y_i = v_i + e_i; i=1, ... , 15,
       v_i ~ (0,1) exponential (0,1)
       e_i ~ (0,D_i) --Double Exponential
       1) D_i=(0.7 0.6 0.5 0.4 0.3)
       2) D_i=(2.0 0.6 0.5 0.4 0.2)
       3) D_i=(4.0 0.6 0.5 0.4 0.1)

Monte Carlo times = 10,000
*/

options nodate nonumber  formdlim="-";

data _null_;
   %let seed_nv=-22123;
   %let seed_ne=-22123445;

   ***Distribution of e*****;
   %let edist="N";
   *%let edist="Exp";

   %let A=2;
   %let beta=0;
   %let m=15; *num of small areas;
   %let p=1;  *num of linear parameters;
   %let times=10000; *num. of simulation times;
run;

proc iml;
    ************Fay-Herriot Method***************;
    start F_fh(Ahat, y) GLOBAL(d, x);
```

```
      sigma=diag(Ahat+d);
      yhat=x*inv(t(x)*inv(sigma)*x)*t(x)*inv(sigma)*y;
      f=sum( (y-yhat)##2/(Ahat+d));
      return (f);
finish F_fh;

start G_fh(Ahat, y) GLOBAL(d, x);
      sigma=diag(Ahat+d);
      yhat=x*inv(t(x)*inv(sigma)*x)*t(x)*inv(sigma)*y;
      g= -sum ( ((y-yhat)/(Ahat+d))##2 );
      return (g);
finish G_fh;

*******ADM  REML Method************;
start F_admR(Ahat, y) GLOBAL(d, x);
      sigma=diag(Ahat+d);
      P=inv(sigma) - inv(sigma)*x * inv(t(x)*
         inv(sigma)*x)*t(x)*inv(sigma);
      f=0.5*(t(y)*P*P*y - trace(P)) +1/Ahat;
      return (f);
finish F_admR;

start G_admR(Ahat, y) GLOBAL(d, x);
      sigma=diag(Ahat+d);
      P=inv(sigma) - inv(sigma)*x * inv(t(x)*
        inv(sigma)*x)*t(x)*inv(sigma);
      g= -trace(P*P)/2 -1/Ahat**2; *score method;
      return (g);
finish G_admR;

*******REML Method************;
start F_reml(Ahat, y) GLOBAL(d, x);
      sigma=diag(Ahat+d);
      P=inv(sigma) - inv(sigma)*x * inv(t(x)*
        inv(sigma)*x)*t(x)*inv(sigma);
      f=0.5*(t(y)*P*P*y - trace(P)) ;
      return (f);
finish F_reml;

start G_reml(Ahat, y) GLOBAL(d, x);
      sigma=diag(Ahat+d);
      P=inv(sigma) - inv(sigma)*x * inv(t(x)*
        inv(sigma)*x)*t(x)*inv(sigma);
      g= - trace(P*P)/2 ; *score method;
      return (g);
```

```
    finish G_reml;

    *********ADM  ML Method ********************;

    start F_admM(Ahat, y) GLOBAL(d, x);
          sigma=diag(Ahat+d);
          P=inv(sigma) - inv(sigma)*x * inv(t(x)*
            inv(sigma)*x)*t(x)*inv(sigma);
          f=0.5*(t(y)*P*P*y - trace(inv(sigma))) +1/Ahat;
          return (f);
    finish F_admM;

    start G_admM(Ahat, y) GLOBAL(d, x);
          sigma=diag(Ahat+d);
          P=inv(sigma) - inv(sigma)*x * inv(t(x)*
            inv(sigma)*x)*t(x)*inv(sigma);
          g= -trace(P*P) + trace(inv(sigma*sigma))/2 -1/Ahat**2;
          return (g);
    finish G_admM;

    *******ML Method************;
    start F_ml(Ahat, y) GLOBAL(d, x);
          sigma=diag(Ahat+d);
          P=inv(sigma) - inv(sigma)*x * inv(t(x)*
            inv(sigma)*x)*t(x)*inv(sigma);
          f=0.5*(t(y)*P*P*y - trace(inv(sigma))) ;
          return (f);
    finish F_ml;

    start G_ml(Ahat, y) GLOBAL(d, x);
          sigma=diag(Ahat+d);
          P=inv(sigma) - inv(sigma)*x * inv(t(x)*
            inv(sigma)*x)*t(x)*inv(sigma);
          g= - trace(P*P) + trace(inv(sigma*sigma))/2 ;
          return (g);
    finish G_ml;

******************************Generate Data Parameters*******;

    /*3 types of D*/
    mm=&m/5;

    dtype=j(&m,7,0);

    dtype[1:mm, 1]=repeat(0.7,mm,1);
```

```
dtype[mm+1:mm*2, 1]=repeat(0.6,mm,1);
dtype[mm*2+1:mm*3, 1]=repeat(0.5,mm,1);
dtype[mm*3+1:mm*4, 1]=repeat(0.4,mm,1);
dtype[mm*4+1:mm*5, 1]=repeat(0.3,mm,1);

dtype[1:mm, 2]=repeat(2.0,mm,1);
dtype[mm+1:mm*2, 2]=repeat(0.6,mm,1);
dtype[mm*2+1:mm*3, 2]=repeat(0.5,mm,1);
dtype[mm*3+1:mm*4, 2]=repeat(0.4,mm,1);
dtype[mm*4+1:mm*5, 2]=repeat(0.2,mm,1);

dtype[1:mm, 3]=repeat(4.0,mm,1);
dtype[mm+1:mm*2, 3]=repeat(0.6,mm,1);
dtype[mm*2+1:mm*3, 3]=repeat(0.5,mm,1);
dtype[mm*3+1:mm*4, 3]=repeat(0.4,mm,1);
dtype[mm*4+1:mm*5, 3]=repeat(0.1,mm,1);

do tt=3 to 3;
    d=dtype[,tt];
    A=&A;
    B=repeat(D/(D+A), 1 ,6);
    *******************Begin simulation******************;
    A1= j(1,6, 0); A2= j(1,6, 0); A3= j(1,6, 0);
    B1= j(&m,6, 0); B2= j(&m,6, 0); B3= j(&m,6, 0);
    theta1= j(&m,6, 0); theta2= j(&m,6, 0); theta3= j(&m,6, 0);
    cc_fh=0;  cc_aR=0;  cc_am=0;
    cc_pr=0;  cc_re=0;  cc_ml=0;

    do j=1 to &times;

       x=j(&m,1,1);
       theta=x*&beta + rannor(j(&m,1,&seed_nv))*sqrt(A);
       e=rannor(j(&m,1,&seed_ne))#sqrt(d);
       y=theta+e;

       *********Estimation of A************************;

       *************Prasad-Rao estimator*********************;
       beta_ols=inv(t(x)*x)*t(x)*y;
       h=vecdiag(x*inv(t(x)*x)*t(x));
       A_pr=(t(y-x*beta_ols)*(y -x*beta_ols)-
            sum(d#(1-h)))/(&m-&p);
       if  A_pr<0 then
           do;
           cc_pr=cc_pr+1; A_pr= 0;
```

```
    end;

**************Fay-Herriot Estimator*******************;
A_fh=(A_pr<>0.5); diff=10;
do while  ((diff>0.0001)&(diff<10E4));
   Anew=A_fh+(&m-&p-f_fh(A_fh, y))/g_fh(A_fh, y);
   diff=abs((Anew-A_fh)/A_fh);
   A_fh=Anew;
end;

if A_fh< 0 then
   do;
   A_fh=0;  cc_fh=cc_fh+1;
   end;

   *print A_fh;
**************ADM_REML Estimator***********************;
A_aR=(A_pr<>0.5); diff=10; i=0;
do while  ((diff>0.0001)&(diff<10E4)&(i<40));
   Anew=A_aR - f_admR(A_aR, y)/g_admR(A_aR, y);
   diff=abs((Anew-A_aR)/A_aR);
   A_aR=Anew;
   i=i+1;
end;

if (A_aR< 0) | (i>39) then
   do;
   A_aR=0;  cc_aR=cc_aR+1;
   end;

**************ADM_ML Estimator*************************;
A_aM=(A_pr<>0.5); diff=10; i=0;
do while  ((diff>0.0001)&(diff<10E4)&(i<40));
   Anew=A_aM - f_admM(A_aM, y)/g_admM(A_aM, y);
   diff=abs((Anew-A_aM)/A_aM);
   A_aM=Anew;
    i=i+1;
end;

if (A_aM< 0) | (i>39) then
    do;
    A_aM=0;  cc_aM=cc_aM+1;
    end;

*print A_adm;
```

```
**************REML Estimator****************************;
A_re=(A_pr<>0.5);
diff=10; i=0;
do while  ((diff>0.0001)&(diff<10E4)&(i<40));
   Anew=A_re - f_reml(A_re, y)/g_reml(A_re, y);
   diff=abs((Anew-A_re)/A_re);
   A_re=Anew;
   i=i+1;
end;

if (A_re< 0)|(i>39) then
    do;
    A_re=0;  cc_re=cc_re+1;
    end;

**************ML Estimator****************************;
A_ml=(A_pr<>0.5); diff=10; i=0;
do while  ((diff>0.0001)&(diff<10E4)&(i<40)&(A_ml>0));
   Anew=A_ml - f_ml(A_ml,y)/g_ml(A_ml,y);
   diff=abs((Anew-A_ml)/A_ml);
   A_ml=Anew;
   i=i+1;
end;

if (A_ml< 0)|(i>39) then
   do;
   A_ml= 0;cc_ml=cc_ml+1;
   end;

Ahat=j(1, 6,0);
Ahat[,1]=A_pr;
Ahat[,2]=A_fh;
Ahat[,3]=A_re;
Ahat[,4]=A_ml;
Ahat[,5]=A_ar;
Ahat[,6]=A_am;

sigma_PR=diag(A_PR+d);
sigma_FH=diag(A_FH+d);
sigma_RE=diag(A_RE+d);
sigma_ml=diag(A_ml+d);
sigma_AR=diag(A_AR+d);
sigma_AM=diag(A_AM+d);
```

```
   betahat=j(&p, 6,0);
   betahat[,1]=inv(t(x)*inv(sigma_PR)*x)*t(x)*inv(sigma_PR)*y;
   betahat[,2]=inv(t(x)*inv(sigma_FH)*x)*t(x)*inv(sigma_FH)*y;
   betahat[,3]=inv(t(x)*inv(sigma_RE)*x)*t(x)*inv(sigma_RE)*y;
   betahat[,4]=inv(t(x)*inv(sigma_ML)*x)*t(x)*inv(sigma_ML)*y;
   betahat[,5]=inv(t(x)*inv(sigma_AR)*x)*t(x)*inv(sigma_AR)*y;
   betahat[,6]=inv(t(x)*inv(sigma_AM)*x)*t(x)*inv(sigma_AM)*y;

   Bhat=repeat(d, 1, 6)/(repeat(d, 1, 6)+repeat(Ahat, &m,1));
   thetahat=(Bhat)#repeat(y,1,6) +(1- Bhat)#(x*betahat);
   A1=A1 + Ahat;
   A2=A2 + Ahat##2;
   A3=A3 + abs(Ahat - A );

   B1=B1 + (Bhat );
   B2=B2 + Bhat##2;
   B3=B3 + abs(Bhat - B );

   thetat=repeat(theta,1,6);
   theta1=theta1 + (thetahat );
   theta2=theta2 + thetahat##2;
   theta3=theta3 + abs(thetahat- thetat );
end;

Bias_A=round( A1/&times - A, 0.001);
RB_A=round(Bias_A/A, 0.001);
CV_A=round(sqrt( (A2 - A1##2/&times)/(&times-1)**2 )
     *&times/A1, 0.001);
ARE_A=round(A3/(&times*A), 0.001);
MSE_A=round(A2/&times + A**2 - 2*A*A1/&times, 0.001);

Bias_B=B1/&times - B;
RB_B=Bias_B/B;
CV_B=sqrt( (B2 - B1##2/&times)/(&times-1)**2 )*&times/B1;
ARE_B=B3/(&times*B);
MSE_B=B2/&times + B - 2*B #B1/&times;

Bias_T=theta1/&times - thetat;
RB_T=Bias_T/thetat;
CV_T=sqrt( (theta2 - theta1##2/&times)/(&times-1)**2 )
     *&times/theta1;
ARE_T=theta3/(&times*thetat);
MSE_T=theta2/&times + thetat - 2*theta #theta1/&times;

gBias_B=j(5,6,0); gRB_B=j(5,6,0); gCV_B=j(5,6,0);
```

```
        gARE_B=j(5,6,0); gMSE_B=j(5,6,0);

        gBias_T=j(5,6,0); gRB_T=j(5,6,0); gCV_T=j(5,6,0);
        gARE_T=j(5,6,0); gMSE_T=j(5,6,0);


        mm=&m/5;
        do i=1 to 5;
            do j=1 to 6;
                gBias_B[i,j]=round(sum(Bias_B[((i-1)*mm+1):i*mm,j]),
                             0.001);
                gRB_B[i,j]=round(sum(RB_B[((i-1)*mm+1):i*mm,j]),
                             0.001)*100;
                gCV_B[i,j]=round(sum(CV_B[((i-1)*mm+1):i*mm,j]),
                             0.001);
                gARE_B[i,j]=round(sum(ARE_B[((i-1)*mm+1):i*mm,j]),
                             0.001)*100;
                gMSE_B[i,j]=round(sum(MSE_B[((i-1)*mm+1):i*mm,j]),
                             0.001);

                gBias_T[i,j]=round(sum(Bias_T[((i-1)*mm+1):i*mm,j]),
                             0.001);
                gRB_T[i,j]=round(sum(RB_T[((i-1)*mm+1):i*mm,j]),
                             0.001)*100;
                gCV_T[i,j]=round(sum(CV_T[((i-1)*mm+1):i*mm,j]),
                             0.001);
                gARE_T[i,j]=round(sum(ARE_T[((i-1)*mm+1):i*mm,j]),
                             0.001)*100;
                gMSE_T[i,j]=round(sum(MSE_T[((i-1)*mm+1):i*mm,j]),
                             0.001);
            end;
        end;

        title "DRS, A=&A, D: type tt, edist=&edist;  m=&m, N=&times";
        PRINT TT;
        print cc_pr cc_fh cc_re cc_ml cc_ar cc_am;
        *print Bias_A RB_A CV_A ARE_A MSE_A;
        *print gBias_B gRB_B gCV_B gARE_B gMSE_B;
        print  RB_A  ARE_A MSE_A;
        print  gRB_B  gARE_B gMSE_B;
        *print gBias_T gRB_T gCV_T gARE_T gMSE_T;
    end;
quit;

-------Chapter 3: Comparison of MSE of theta for different methods--
```

```
options nodate nonumber  formdlim="-";

data _null_;
   %let seed_nv=22123;
   %let seed_ne=22123445;

    ***Distribution of e*****;
    %let edist="N";
    *%let edist="Exp";

   %let A=1;
   %let beta=0;
   %let m=15; *num of small areas;
   %let p=1;  *num of linear parameters;
   %let times=10000;          *num. of simulation times;
run;
title " A=&A, eist=&edist, N=&times ";

proc iml;
    ************Fay-Herriot Method***************;
    start F_fh(Ahat, y) GLOBAL(d, x);
          sigma=diag(Ahat+d);
          yhat=x*inv(t(x)*inv(sigma)*x)*t(x)*inv(sigma)*y;
          f=sum( (y-yhat)##2/(Ahat+d));
          return (f);
    finish F_fh;

    start G_fh(Ahat, y) GLOBAL(d, x);
          sigma=diag(Ahat+d);
          yhat=x*inv(t(x)*inv(sigma)*x)*t(x)*inv(sigma)*y;
          g= -sum ( ((y-yhat)/(Ahat+d))##2 );
          return (g);
    finish G_fh;

    *******ADM  REML Method************;
    start F_admR(Ahat, y) GLOBAL(d, x);
          sigma=diag(Ahat+d);
          P=inv(sigma) - inv(sigma)*x * inv(t(x)*
             inv(sigma)*x)*t(x)*inv(sigma);
          f=0.5*(t(y)*P*P*y - trace(P)) +1/Ahat;
          return (f);
    finish F_admR;

    start G_admR(Ahat, y) GLOBAL(d, x);
```

```
      sigma=diag(Ahat+d);
      P=inv(sigma) - inv(sigma)*x * inv(t(x)*
        inv(sigma)*x)*t(x)*inv(sigma);
      g= -trace(P*P)/2 -1/Ahat**2; *score method;
      return (g);
finish G_admR;

*******REML Method************;
start F_reml(Ahat, y) GLOBAL(d, x);
      sigma=diag(Ahat+d);
      P=inv(sigma) - inv(sigma)*x * inv(t(x)*
        inv(sigma)*x)*t(x)*inv(sigma);
      f=0.5*(t(y)*P*P*y - trace(P)) ;
      return (f);
finish F_reml;

start G_reml(Ahat, y) GLOBAL(d, x);
      sigma=diag(Ahat+d);
      P=inv(sigma) - inv(sigma)*x * inv(t(x)*
        inv(sigma)*x)*t(x)*inv(sigma);
      g= - trace(P*P)/2 ; *score method;
      return (g);
finish G_reml;

*********ADM  ML Method ********************;

start F_admM(Ahat, y) GLOBAL(d, x);
      sigma=diag(Ahat+d);
      P=inv(sigma) - inv(sigma)*x * inv(t(x)*
        inv(sigma)*x)*t(x)*inv(sigma);
      f=0.5*(t(y)*P*P*y - trace(inv(sigma))) +1/Ahat;
      return (f);
finish F_admM;

start G_admM(Ahat, y) GLOBAL(d, x);
      sigma=diag(Ahat+d);
      P=inv(sigma) - inv(sigma)*x * inv(t(x)*
        inv(sigma)*x)*t(x)*inv(sigma);
      g= -trace(P*P) + trace(inv(sigma*sigma))/2 -1/Ahat**2;
      return (g);
finish G_admM;

*******ML Method************;
start F_ml(Ahat, y) GLOBAL(d, x);
      sigma=diag(Ahat+d);
```

```
      P=inv(sigma) - inv(sigma)*x * inv(t(x)*
        inv(sigma)*x)*t(x)*inv(sigma);
      f=0.5*(t(y)*P*P*y - trace(inv(sigma))) ;
      return (f);
finish F_ml;

start G_ml(Ahat, y) GLOBAL(d, x);
      sigma=diag(Ahat+d);
      P=inv(sigma) - inv(sigma)*x * inv(t(x)*
        inv(sigma)*x)*t(x)*inv(sigma);
      g= - trace(P*P) + trace(inv(sigma*sigma))/2 ;
      return (g);
finish G_ml;

******************Generate Data Parameters***************;

/*3 types of D*/
mm=&m/5;

dtype=j(&m,7,0);

dtype[1:mm, 1]=repeat(0.7,mm,1);
dtype[mm+1:mm*2, 1]=repeat(0.6,mm,1);
dtype[mm*2+1:mm*3, 1]=repeat(0.5,mm,1);
dtype[mm*3+1:mm*4, 1]=repeat(0.4,mm,1);
dtype[mm*4+1:mm*5, 1]=repeat(0.3,mm,1);

dtype[1:mm, 2]=repeat(2.0,mm,1);
dtype[mm+1:mm*2, 2]=repeat(0.6,mm,1);
dtype[mm*2+1:mm*3, 2]=repeat(0.5,mm,1);
dtype[mm*3+1:mm*4, 2]=repeat(0.4,mm,1);
dtype[mm*4+1:mm*5, 2]=repeat(0.2,mm,1);

dtype[1:mm, 3]=repeat(4.0,mm,1);
dtype[mm+1:mm*2, 3]=repeat(0.6,mm,1);
dtype[mm*2+1:mm*3, 3]=repeat(0.5,mm,1);
dtype[mm*3+1:mm*4, 3]=repeat(0.4,mm,1);
dtype[mm*4+1:mm*5, 3]=repeat(0.1,mm,1);

*print dtype;
do  tt=1 to 3;
    d=dtype[,tt];
    A=&A;
    *******************Begin simulation********************;
    theta1= j(&m,7, 0); theta2= j(&m,7, 0); amse=j(&m, 7,0);
```

```
cc_fh=0;  cc_aR=0;  cc_am=0;
cc_pr=0;  cc_re=0;  cc_ml=0;

do j=1 to &times;

   x=j(&m,1,1);
   theta=x*&beta + rannor(j(&m,1,&seed_nv))*sqrt(A);
   e=rannor(j(&m,1,&seed_ne))#sqrt(d);
   y=theta+e;

   *********Estimation of A*************************;

   *************Prasad-Rao estimator********************;
   beta_ols=inv(t(x)*x)*t(x)*y;
   h=vecdiag(x*inv(t(x)*x)*t(x));
   A_pr=(t(y-x*beta_ols)*(y-x*beta_ols)-sum(d#(1-h)))
        /(&m-&p);
   if  A_pr<0 then
       do;
       cc_pr=cc_pr+1; A_pr= 0;
       end;

   **************Fay-Herriot Estimator*******************;
   A_fh=(A_pr<>0.5); diff=10;
   do while  ((diff>0.0001)&(diff<10E4));
      Anew=A_fh+(&m-&p-f_fh(A_fh, y))/g_fh(A_fh, y);
      diff=abs((Anew-A_fh)/A_fh);
      A_fh=Anew;
   end;

   if A_fh< 0 then
      do;
      A_fh=0;  cc_fh=cc_fh+1;
      end;

      *print A_fh;
   **************ADM_REML Estimator************************;
   A_aR=(A_pr<>0.5); diff=10; i=0;
   do while  ((diff>0.0001)&(diff<10E4)&(i<40));
      Anew=A_aR - f_admR(A_aR, y)/g_admR(A_aR, y);
      diff=abs((Anew-A_aR)/A_aR);
      A_aR=Anew;
      i=i+1;
   end;
```

```
if (A_aR< 0) | (i>39) then
   do;
   A_aR=0;  cc_aR=cc_aR+1;
   end;

**************ADM_ML Estimator*************************;
A_aM=(A_pr<>0.5); diff=10; i=0;
do while  ((diff>0.0001)&(diff<10E4)&(i<40));
   Anew=A_aM - f_admM(A_aM, y)/g_admM(A_aM, y);
   diff=abs((Anew-A_aM)/A_aM);
   A_aM=Anew;
    i=i+1;
end;

if (A_aM< 0) | (i>39) then
    do;
    A_aM=0;  cc_aM=cc_aM+1;
    end;

*print A_adm;

**************REML Estimator*************************;
A_re=(A_pr<>0.5);
diff=10; i=0;
do while  ((diff>0.0001)&(diff<10E4)&(i<40));
   Anew=A_re - f_reml(A_re, y)/g_reml(A_re, y);
   diff=abs((Anew-A_re)/A_re);
   A_re=Anew;
   i=i+1;
end;

if (A_re< 0)|(i>39) then
    do;
    A_re=0;  cc_re=cc_re+1;
    end;

**************ML Estimator***************************;
A_ml=(A_pr<>0.5); diff=10; i=0;
do while  ((diff>0.0001)&(diff<10E4)&(i<40)&(A_ml>0));
   Anew=A_ml - f_ml(A_ml,y)/g_ml(A_ml,y);
   diff=abs((Anew-A_ml)/A_ml);
   A_ml=Anew;
   i=i+1;
end;
```

```
if (A_ml< 0)|(i>39) then
   do;
   A_ml= 0;cc_ml=cc_ml+1;
   end;

******************THETA AND MSE*************************;
****theta_pr***;
sigma_pr=diag(A_pr+d);
beta_pr=inv(t(x)*inv(sigma_pr)*x)*t(x)*inv(sigma_pr)*y;
B_pr=d/(d+A_pr);
theta_eb_pr=(1-B_pr)#y + B_pr#(x*beta_pr);

g1_pr=d#(1-B_pr);
g2_pr=B_pr##2#vecdiag(x*inv(t(x)*inv(Sigma_pr)*x)*t(x));
g3_pr= d##2 # (A_pr+D)##(-3) * 2* sum((A_pr+d)##2)/&m**2;

mse_pr = g1_pr + g2_pr + 2*g3_pr;

****theta_fh***;
sigma_fh=diag(A_fh+d);
beta_fh=inv(t(x)*inv(sigma_fh)*x)*t(x)*inv(sigma_fh)*y;
B_fh=d/(d+A_fh);
theta_eb_fh=(1-B_fh)#y + B_fh#(x*beta_fh);

g1_fh = d#(1-B_fh);
g2_fh = B_fh##2 # vecdiag(x*inv(t(x)*
     inv(Sigma_fh)*x)*t(x));
g3_fh = 2*&m * d##2 # (A_fh+D)##(-3) *
     (sum(1/(A_fh+d)))**(-2);
 *see Datta, Rao and Smith(2005);
g4_fh = 2*B_fh##2 # ( &m*trace( inv(Sigma_fh)**2)
       - (trace(Sigma_fh**(-1)))**2 )
         /( trace(Sigma_fh**(-1)))**3;

mse_fh = g1_fh + g2_fh + 2*g3_fh - g4_fh;


****theta_adm re***;
sigma_ar=diag(A_ar+d);
beta_ar=inv(t(x)*inv(sigma_ar)*x)*t(x)*inv(sigma_ar)*y;
B_ar=d/(d+A_ar);
theta_eb_ar=(1-B_ar)#y + B_ar#(x*beta_ar);

g1_ar=d#(1-B_ar);
g2_ar=B_ar##2#vecdiag(x*inv(t(x)*inv(Sigma_ar)*x)*t(x));
```

```
g3_ar= d##2 # (A_ar+D)##(-3)*2/trace( inv(Sigma_ar)**2);
if A_ar=0 then
   g4_ar=0;
else
   g4_ar = B_ar##2 *( 2/A_ar )/trace( inv(Sigma_ar)**2);

mse_ar = g1_ar + g2_ar + 2*g3_ar - g4_ar;

****theta_adm ml***;
sigma_am=diag(A_am+d);
beta_am=inv(t(x)*inv(sigma_am)*x)*t(x)*inv(sigma_am)*y;
W_am= inv(sigma_am)* x * inv(t(x)*inv(sigma_am)*x)
      *t(x)*inv(sigma_am);
B_am=d/(d+A_am);
theta_eb_am=(1-B_am)#y + B_am #(x*beta_am);

g1_am=d#(1-B_am);
g2_am=B_am##2#vecdiag(x*inv(t(x)*inv(Sigma_am)*x)*t(x));
g3_am= d##2 # (A_am+D)##(-3)*2/trace(inv(Sigma_am)**2 );
if A_am=0 then
   g4_am=0;
else  g4_am = B_am##2 *( trace(-W_am) + 2/A_am )
    / trace( inv(Sigma_am)**2 );


mse_am = g1_am + g2_am + 2*g3_am - g4_am;
mse_amn = g1_am + g2_am;

****theta_reml***;
sigma_re=diag(A_re+d);
beta_re=inv(t(x)*inv(sigma_re)*x)*t(x)*inv(sigma_re)*y;
B_re=d/(d+A_re);
theta_eb_re=(1-B_re)#y + B_re#(x*beta_re);

g1_re = d#(1-B_re);
g2_re = B_re##2#vecdiag(x*inv(t(x)*inv(Sigma_re)*x)*t(x));
g3_re = d##2#(A_re+D)##(-3)*2/trace(inv(Sigma_re)**2);

mse_re = g1_re + g2_re + 2*g3_re;
mse_ren = g1_re + g2_re;

****theta_ml***;
sigma_ml=diag(A_ml+d);
beta_ml=inv(t(x)*inv(sigma_ml)*x)*t(x)*inv(sigma_ml)*y;
W_ml= inv(sigma_ml)* x * inv(t(x)*inv(sigma_ml)*x)*t(x)
```

```
          *inv(sigma_ml);
    B_ml=d/(d+A_ml);
    theta_eb_ml=(1-B_ml)#y + B_ml#(x*beta_ml);

    g1_ml=d#(1-B_ml);
    g2_ml=B_ml##2#vecdiag(x*inv(t(x)*inv(Sigma_ml)*x)*t(x));
    g3_ml=d##2 # (A_ml+D)##(-3)*2/trace(inv(Sigma_ml)**2);
    g4_ml=B_ml##2 *trace(-W_ml)/ trace( inv(Sigma_ml)**2);

    mse_ml = g1_ml + g2_ml + 2*g3_ml - g4_ml;
    *****************************************************;
    thetahat=j(&m, 7,0);
    thetahat[,1]=theta_eb_pr;
    thetahat[,2]=theta_eb_fh;
    thetahat[,3]=theta_eb_re;
    thetahat[,4]=theta_eb_ml;
    thetahat[,5]=theta_eb_ar;
    thetahat[,6]=theta_eb_am;
    thetahat[,7]=theta_eb_re;

    mse=j(&m, 7,0);
    mse[,1]=mse_pr;
    mse[,2]=mse_fh;
    mse[,3]=mse_re;
    mse[,4]=mse_ml;
    mse[,5]=mse_ar;
    mse[,6]=mse_am;
    mse[,7]=mse_ren;

    thetat=repeat(theta,1,7);
    theta1=theta1 + (thetahat - thetat );
    theta2=theta2 + (thetahat- thetat )##2;

    amse=amse+mse;

end;

Bias_T=theta1/&times;
MSE_T=theta2/&times;
amse=amse/&times;
ramse=(amse - mse_t)/mse_t;

gBias_B=j(5,7,0); gRB_B=j(5,7,0); gCV_B=j(5,7,0);
gARE_B=j(5,7,0); gMSE_B=j(5,7,0);
```

```
        gBias_T=j(5,7,0); gamse=j(5,7,0); gramse=j(5,7,0);
        gMSE_t=j(5,7,0);


        mm=&m/5;
        do i=1 to 5;
            do j=1 to 7;
                gBias_T[i,j]=round(sum(Bias_T[((i-1)*mm+1):i*mm,j]
                            /3),0.001);
                gamse[i,j]=round( sum(amse[((i-1)*mm+1):i*mm,j]/3),
                            0.001);
                gramse[i,j]=round( sum(ramse[((i-1)*mm+1):i*mm,j]/3),
                            0.001);
                gMSE_T[i,j]=round( sum(MSE_T[((i-1)*mm+1):i*mm,j]/3),
                            0.001);
            end;
        end;

        PRINT TT;

        print cc_pr cc_fh cc_re cc_ml cc_ar cc_am;
        print gBias_T  gMSE_T gamse gramse;
    end;
quit;



--------Chapter 4 Comparison of Different Prediction Intervals---
options nodate nonumber  formdlim="-";

data _null_;
   %let seed_nv=-22123;
   %let seed_ne=-22123445;
    %let seed_bv=-22132423;
   %let seed_be=-22178945;

    ***Distribution of e*****;
    %let edist="N";
    *%let edist="Exp";

    %let alpha=.05;
   %let beta=0;
    %let A=1;
    %let k=1; *increase variability of A D;
   %let m=15; *num of small areas;
```

```
   %let p=1;  *num of linear parameters;
   %let type=1; * Pattern D;

   %let Bt_times=1000;   *num. of bootstrap times;
   %let times=10000;          *num. of simulation times;

run;

title "DRS-Normal, A=&A, REML D: type &type; m=&m, N=&times,
B=&bt_times ";


proc iml;
    ************Fay-Herriot Method***************;
    start F_fh(Ahat, y) GLOBAL(d, x);
          sigma=diag(Ahat+d);
          yhat=x*inv(t(x)*inv(sigma)*x)*t(x)*inv(sigma)*y;
          f=sum( (y-yhat)##2/(Ahat+d));
          return (f);
    finish F_fh;

    start G_fh(Ahat, y) GLOBAL(d, x);
          sigma=diag(Ahat+d);
          yhat=x*inv(t(x)*inv(sigma)*x)*t(x)*inv(sigma)*y;
          g= -sum ( ((y-yhat)/(Ahat+d))##2 );
          return (g);
    finish G_fh;

    *******ADM  REML Method************;
    start F_admR(Ahat, y) GLOBAL(d, x);
          sigma=diag(Ahat+d);
          P=inv(sigma) - inv(sigma)*x * inv(t(x)*
             inv(sigma)*x)*t(x)*inv(sigma);
          f=0.5*(t(y)*P*P*y - trace(P)) +1/Ahat;
          return (f);
    finish F_admR;

    start G_admR(Ahat, y) GLOBAL(d, x);
          sigma=diag(Ahat+d);
          P=inv(sigma) - inv(sigma)*x * inv(t(x)*
            inv(sigma)*x)*t(x)*inv(sigma);
          g= -trace(P*P)/2 -1/Ahat**2; *score method;
          return (g);
    finish G_admR;
```

```
*******REML Method************;
start F_reml(Ahat, y) GLOBAL(d, x);
      sigma=diag(Ahat+d);
      P=inv(sigma) - inv(sigma)*x * inv(t(x)*
        inv(sigma)*x)*t(x)*inv(sigma);
      f=0.5*(t(y)*P*P*y - trace(P)) ;
      return (f);
finish F_reml;

start G_reml(Ahat, y) GLOBAL(d, x);
      sigma=diag(Ahat+d);
      P=inv(sigma) - inv(sigma)*x * inv(t(x)*
        inv(sigma)*x)*t(x)*inv(sigma);
      g= - trace(P*P)/2 ; *score method;
      return (g);
finish G_reml;

*********ADM  ML Method *********************;

start F_admM(Ahat, y) GLOBAL(d, x);
      sigma=diag(Ahat+d);
      P=inv(sigma) - inv(sigma)*x * inv(t(x)*
        inv(sigma)*x)*t(x)*inv(sigma);
      f=0.5*(t(y)*P*P*y - trace(inv(sigma))) +1/Ahat;
      return (f);
finish F_admM;

start G_admM(Ahat, y) GLOBAL(d, x);
      sigma=diag(Ahat+d);
      P=inv(sigma) - inv(sigma)*x * inv(t(x)*
        inv(sigma)*x)*t(x)*inv(sigma);
      g= -trace(P*P) + trace(inv(sigma*sigma))/2 -1/Ahat**2;
      return (g);
finish G_admM;

*******ML Method************;
start F_ml(Ahat, y) GLOBAL(d, x);
      sigma=diag(Ahat+d);
      P=inv(sigma) - inv(sigma)*x * inv(t(x)*
        inv(sigma)*x)*t(x)*inv(sigma);
      f=0.5*(t(y)*P*P*y - trace(inv(sigma))) ;
      return (f);
finish F_ml;

start G_ml(Ahat, y) GLOBAL(d, x);
```

```
      sigma=diag(Ahat+d);
      P=inv(sigma) - inv(sigma)*x * inv(t(x)*
        inv(sigma)*x)*t(x)*inv(sigma);
      g= - trace(P*P) + trace(inv(sigma*sigma))/2 ;
      return (g);
finish G_ml;

********************Generate Data Parameters*******;
/*3 types of D*/
mm=&m/5;

dtype=j(&m,7,0);

dtype[1:mm, 1]=repeat(0.7,mm,1);
dtype[mm+1:mm*2, 1]=repeat(0.6,mm,1);
dtype[mm*2+1:mm*3, 1]=repeat(0.5,mm,1);
dtype[mm*3+1:mm*4, 1]=repeat(0.4,mm,1);
dtype[mm*4+1:mm*5, 1]=repeat(0.3,mm,1);

dtype[1:mm, 2]=repeat(2.0,mm,1);
dtype[mm+1:mm*2, 2]=repeat(0.6,mm,1);
dtype[mm*2+1:mm*3, 2]=repeat(0.5,mm,1);
dtype[mm*3+1:mm*4, 2]=repeat(0.4,mm,1);
dtype[mm*4+1:mm*5, 2]=repeat(0.2,mm,1);

dtype[1:mm, 3]=repeat(4.0,mm,1);
dtype[mm+1:mm*2, 3]=repeat(0.6,mm,1);
dtype[mm*2+1:mm*3, 3]=repeat(0.5,mm,1);
dtype[mm*3+1:mm*4, 3]=repeat(0.4,mm,1);
dtype[mm*4+1:mm*5, 3]=repeat(0.1,mm,1);

do kk=3 to 3 ;

    d=dtype[,kk];
    A=&A;
    ********************Begin simulation******************;
    cc_re=0; cc_pr=0;   cc_fh=0;   cc_am=0;
    coverage= j(&m,6, 0); length= j(&m,6, 0);

    do j=1 to &times;

        x=j(&m,1,1);
        theta=x*&beta + rannor(j(&m,1,&seed_nv))*sqrt(A);
        e=rannor(j(&m,1,&seed_ne))#sqrt(d);
        y=theta+e;
```

```
*********Estimation of A************************;
*************Prasad-Rao estimator******************;
beta_ols=inv(t(x)*x)*t(x)*y;
h=vecdiag(x*inv(t(x)*x)*t(x));
A_pr=(t(y-x*beta_ols)*(y -x*beta_ols)-sum(d#(1-h)))
        /(&m-&p);
if  A_pr<0 then
do;
    cc_pr=cc_pr+1;    A_pr= 0;
end;

**************Fay-Herriot Estimator****************;
A_fh=(A_pr<>0.5); diff=10;
do while  ((diff>0.0001)&(diff<10E4));
    Anew=A_fh+(&m-&p-f_fh(A_fh, y))/g_fh(A_fh, y);
    diff=abs((Anew-A_fh)/A_fh);
    A_fh=Anew;
end;

if A_fh< 0 then
    do;
       A_fh=0;  cc_fh=cc_fh+1;
    end;
**************ADM_ML Estimator*********************;
A_aM=(A_pr<>0.5); diff=10; i=0;
do while  ((diff>0.0001)&(diff<10E4)&(i<40));
    Anew=A_aM - f_admM(A_aM, y)/g_admM(A_aM, y);
    diff=abs((Anew-A_aM)/A_aM);
    A_aM=Anew;
    i=i+1;
end;

if (A_aM< 0) | (i>39) then
    do;
       A_aM=0;  cc_aM=cc_aM+1;
    end;
**************REML Estimator************************;
A_re=(A_pr<>0.5);
diff=10; i=0;
do while  ((diff>0.0001)&(diff<10E4)&(i<40));
    Anew=A_re - f_re(A_re, y)/g_re(A_re, y);
    diff=abs((Anew-A_re)/A_re);
    A_re=Anew;
    i=i+1;
end;
```

```
if (A_re< 0)|(i>39) then
    do;
       A_re=0;  cc_re=cc_re+1;
    end;


******************THETA AND MSE*********************;
****theta_pr***;
sigma_pr=diag(A_pr+d);
beta_pr=inv(t(x)*inv(sigma_pr)*x)*t(x)*inv(sigma_pr)*y;
B_pr=d/(d+A_pr);
theta_eb_pr=(1-B_pr)#y + B_pr#(x*beta_pr);

g1_pr=d#(1-B_pr);
g2_pr=B_pr##2#vecdiag(x*inv(t(x)*inv(Sigma_pr)*x)*t(x));
g3_pr= d##2 # (A_pr+D)##(-3) * 2* sum((A_pr+d)##2)/&m**2;

mse_pr = g1_pr + g2_pr + 2*g3_pr;
CI_pr=j(&m,2,0);
CI_pr[,1]=theta_eb_pr - 1.96#sqrt(mse_pr);
CI_pr[,2]=theta_eb_pr + 1.96#sqrt(mse_pr);

****theta_fh***;
sigma_fh=diag(A_fh+d);
beta_fh=inv(t(x)*inv(sigma_fh)*x)*t(x)*inv(sigma_fh)*y;
B_fh=d/(d+A_fh);
theta_eb_fh=(1-B_fh)#y + B_fh#(x*beta_fh);

g1_fh = d#(1-B_fh);
g2_fh = B_fh##2#vecdiag(x*inv(t(x)*
        inv(Sigma_fh)*x)*t(x));
g3_fh = 2*&m*d##2#(A_fh+D)##(-3)*(sum(1/(A_fh+d)))**(-2);
 *see Datta, Rao and Smith(2005);
g4_fh = 2*B_fh##2 # ( &m*trace( inv(Sigma_fh)**2)
         - (trace(Sigma_fh**(-1)))**2 )
        /( trace(Sigma_fh**(-1)))**3;

mse_fh = g1_fh + g2_fh + 2*g3_fh - g4_fh;
if sum(mse_fh<0)>0 then
    do;
        mse_fh2 = g1_fh + g2_fh + 2*g3_fh;
        mse_fh[loc(mse_fh<0),]=mse_fh2[loc(mse_fh<0),];
    end;
```

```
CI_fh=j(&m,2,0);
CI_fh[,1]=theta_eb_fh - 1.96#sqrt(mse_fh);
CI_fh[,2]=theta_eb_fh + 1.96#sqrt(mse_fh);

****theta_adm ml***;
sigma_am=diag(A_am+d);
beta_am=inv(t(x)*inv(sigma_am)*x)*t(x)*inv(sigma_am)*y;
W_am= inv(sigma_am)* x * inv(t(x)*inv(sigma_am)*x)*t(x)
        *inv(sigma_am);
B_am=d/(d+A_am);
theta_eb_am=(1-B_am)#y + B_am #(x*beta_am);

g1_am=d#(1-B_am);
g2_am=B_am##2#vecdiag(x*inv(t(x)*inv(Sigma_am)*x)*t(x));
g3_am=d##2#(A_am+D)##(-3)*2/trace(inv(Sigma_am)**2);
if A_am=0 then
     g4_am=0;
else  g4_am = B_am##2 *( trace(-W_am) + 2/A_am )
            / trace( inv(Sigma_am)**2 );

mse_am = g1_am + g2_am + 2*g3_am - g4_am;
if sum(mse_am<0)>0 then
    do;
        mse_am2 = g1_am + g2_am + 2*g3_am;
        mse_am[loc(mse_am<0),]=mse_am2[loc(mse_am<0),];
    end;
mse_amn = g1_am;

CI_am=j(&m,2,0);
CI_am[,1]=theta_eb_am - 1.96#sqrt(mse_am);
CI_am[,2]=theta_eb_am + 1.96#sqrt(mse_am);

CI_amn=j(&m,2,0);
CI_amn[,1]=theta_eb_am - 1.96#sqrt(mse_amn);
CI_amn[,2]=theta_eb_am + 1.96#sqrt(mse_amn);

****theta_reml***;
sigma_re=diag(A_re+d);
beta_re=inv(t(x)*inv(sigma_re)*x)*t(x)*inv(sigma_re)*y;
B_re=d/(d+A_re);
theta_eb_re=(1-B_re)#y + B_re#(x*beta_re);

g1_re=d#(1-B_re);
g2_re=B_re##2#vecdiag(x*inv(t(x)*inv(Sigma_re)*x)*t(x));
g3_re= d##2#(A_re+D)##(-3)*2/trace( inv(Sigma_re)**2 );
```

```
mse_re = g1_re + g2_re + 2*g3_re;

CI_re=j(&m,2,0);
CI_re[,1]=theta_eb_re - 1.96#sqrt(mse_re);
CI_re[,2]=theta_eb_re + 1.96#sqrt(mse_re);

**********************BOOTSTRAP METHOD*****************;
pivot_cllg=j(&bt_times,&m,0);

do btimes=1 to &bt_times;

    btheta_g=beta_am+rannor(j(&m,1,&seed_bv))*sqrt(A_am);
    be=rannor(j(&m,1,&seed_be))#sqrt(d);
    by_g=btheta_g+be;

    Ahatstar_g=.5; diff=10; i=0;
    do while  ((diff>0.00001)&(diff<10E4)&(i<40));
        Anewstar=Ahatstar_g -f_admm(Ahatstar_g,by_g)
                /g_admm(Ahatstar_g,by_g);
        diff=abs(Anewstar-Ahatstar_g);
        Ahatstar_g=Anewstar;
        i=i+1;
    end;

    if (Ahatstar_g < 0)|(i>39)then Ahatstar_g=0.01;

    *print Ahatstar;
    bsigma_g=diag(Ahatstar_g+d);
    betastar_g=inv(t(x)*inv(bsigma_g)*x)*t(x)
                *inv(bsigma_g)*by_g;

    Bhatstar_g=d/(d+Ahatstar_g);

    theta_ebstar_g=(1-Bhatstar_g)#by_g +
                    Bhatstar_g#(x*betastar_g);

    g1star_g=d#(1-Bhatstar_g);

    pivot_cllg[btimes,]=t((btheta_g - theta_ebstar_g)
                        /sqrt(g1star_g));

end;
```

```
    *print pivot_cllg;
    *********************PB_cllgls****************;
    n=&bt_times; ns=(n*&alpha-1);
    ts=j(&m,3,0); tt=j(ns,3,0);
    do i=1 to &m;
        temp=pivot_cllg[,i];bymean=pivot_cllg[,i];
        bymean[rank(bymean),]=temp;
        frac=n*&alpha/2;
        tt[,1]=bymean[1:ns,]; tt[,2]=bymean[n-ns+1:n,];
        tt[,3]=tt[,2]-tt[,1];
        ss=tt[><,3];
        ts[i,1]=tt[loc(tt[,3]=ss),1];
        ts[i,2]=tt[loc(tt[,3]=ss),2];
    end;

    *print A_am ts;
    *print g1_am;
    CI_cllgs=j(&m,2,0);
    CI_cllgs[,1]=theta_eb_am+ts[,1]#sqrt(g1_am);
    CI_cllgs[,2]=theta_eb_am+ts[,2]#sqrt(g1_am);

    *print ci_cllgs;

    *****************COVERAGE AND LENGTH*******************;
    coverage[,1]=coverage[,1]+
                ((CI_amn[,1]<theta)&(CI_amn[,2]>theta));
    coverage[,2]=coverage[,2]+
                ((CI_fh[,1]<theta) & (CI_fh[,2]>theta));
    coverage[,3]=coverage[,3]+
                ((CI_pr[,1]<theta) & (CI_pr[,2]>theta));
    coverage[,4]=coverage[,4]+
                ((CI_re[,1]<theta) & (CI_re[,2]>theta));
    coverage[,5]=coverage[,5]+
                ((CI_am[,1]<theta) & (CI_am[,2]>theta));
    coverage[,6]=coverage[,6]+
                ((CI_cllgs[,1]<theta)&(CI_cllgs[,2]>theta));

    length[,1]=length[,1]+(CI_amn[,2]-CI_amn[,1]);
    length[,2]=length[,2]+(CI_fh[,2]-CI_fh[,1]);
    length[,3]=length[,3]+(CI_pr[,2]-CI_pr[,1]);
    length[,4]=length[,4]+(CI_re[,2]-CI_re[,1]);
    length[,5]=length[,5]+(CI_am[,2]-CI_am[,1]);
    length[,6]=length[,6]+(CI_cllgs[,2]-CI_cllgs[,1]);

end;
```

```
        gcoverage=j(5,6,0);
        glength=j(5,6,0);

        do i=1 to 5;
            do j=1 to 6;
                gcoverage[i,j]=sum(coverage[((i-1)*mm+1):i*mm,j]);
                glength[i,j]=sum(length[((i-1)*mm+1):i*mm,j]);
            end;
        end;
        nn=&times*mm;
        gcoverage=round(gcoverage*100/nn, 0.1);
        glength=round(glength/nn,0.01);

        print kk;
        print gcoverage glength;
    end;
quit;

-----------Chapter 5 SAIPE Data Analysis---------------------------

----Estimators of A---------

PROC IMPORT OUT= WORK.SAIPE
            DATAFILE= "C:\Documents and Settings\Eileen\My Documents
            \My research\Dissertation\Bill Bell\saipe.txt"
            DBMS=TAB REPLACE;
     GETNAMES=YES;
     DATAROW=2;
RUN;

options nodate nonumber  formdlim="-";

data _null_;
    %let year=199;
    %let m=51;   * num. of small areas;
    %let p=5;   *num. of covariates;

    %let seed_bv=-22132423;
    %let seed_be=-22178945;

    %let alpha=.05;   *Norminal value for the CI;
    %let A=1.5;   * True value of A;
    %let beta=0;
```

```
    %let Bt_times=1000;   *num. of bootstrap times;
run;

************from SAIPE97 data, get beta, X, D_i, A***********;

data saipe&year;
set saipe;
x0=1;
x1=irsprchld;
x2=irsnf0_64;
x3= fsrate;
x4=cenrsd;
if year=&year;
d=ve;
dinv=1/ve;
state=f1;
keep cps x0-x4 d dinv state;
run;
proc sort data=saipe&year out=saipe&year;
by D;
run;
/*proc print data=saipe;
var d;
run;*/
/*
proc mixed data=saipe&year method=ml noprofile;
        class state;
        id state cps;
        model cps=x0 x1 x2 x3 x4/noint ddfm=kr ;
        random state;
        parms (1) (1)/hold=2;
        weight dinv;
        *ods output SolutionF= Mixedfix  CovParms=sigmave2 ;
run;
*/
title" SAIPE &YEAR  data analysis";
proc iml;
    start F_fh(Ahat, y) GLOBAL(d, x);
          sigma=diag(Ahat+d);
          yhat=x*inv(t(x)*inv(sigma)*x)*t(x)*inv(sigma)*y;
          f=sum( (y-yhat)##2/(Ahat+d));
          return (f);
    finish F_fh;

    start G_fh(Ahat, y) GLOBAL(d, x);
```

```
            sigma=diag(Ahat+d);
            yhat=x*inv(t(x)*inv(sigma)*x)*t(x)*inv(sigma)*y;
            g= -sum ( ((y-yhat)/(Ahat+d))##2 );
            return (g);
finish G_fh;
*******ADM  REML Method************;
start F_admR(Ahat, y) GLOBAL(d, x);
            sigma=diag(Ahat+d);
            P=inv(sigma) - inv(sigma)*x * inv(t(x)*
               inv(sigma)*x)*t(x)*inv(sigma);
            f=0.5*(t(y)*P*P*y - trace(P)) +1/Ahat;
            return (f);
finish F_admR;

start G_admR(Ahat, y) GLOBAL(d, x);
            sigma=diag(Ahat+d);
            P=inv(sigma) - inv(sigma)*x * inv(t(x)*
              inv(sigma)*x)*t(x)*inv(sigma);
            g= -trace(P*P)/2 -1/Ahat**2; *score method;
            return (g);
finish G_admR;

*******REML Method************;
start F_reml(Ahat, y) GLOBAL(d, x);
            sigma=diag(Ahat+d);
            P=inv(sigma) - inv(sigma)*x * inv(t(x)*
               inv(sigma)*x)*t(x)*inv(sigma);
            f=0.5*(t(y)*P*P*y - trace(P)) ;
            return (f);
finish F_reml;

start G_reml(Ahat, y) GLOBAL(d, x);
            sigma=diag(Ahat+d);
            P=inv(sigma) - inv(sigma)*x * inv(t(x)*
             inv(sigma)*x)*t(x)*inv(sigma);
            g= - trace(P*P)/2 ; *score method;
            return (g);
finish G_reml;

*********ADM  ML Method ********************;
start F_admM(Ahat, y) GLOBAL(d, x);
            sigma=diag(Ahat+d);
            P=inv(sigma) - inv(sigma)*x * inv(t(x)*
              inv(sigma)*x)*t(x)*inv(sigma);
            f=0.5*(t(y)*P*P*y - trace(inv(sigma))) +1/Ahat;
```

```
            return (f);
finish F_admM;

start G_admM(Ahat, y) GLOBAL(d, x);
            sigma=diag(Ahat+d);
            P=inv(sigma) - inv(sigma)*x * inv(t(x)*
              inv(sigma)*x)*t(x)*inv(sigma);
            g= -trace(P*P) + trace(inv(sigma*sigma))/2
               -1/Ahat**2;
            return (g);
finish G_admM;

*******ML Method************;
start F_ml(Ahat, y) GLOBAL(d, x);
            sigma=diag(Ahat+d);
            P=inv(sigma) - inv(sigma)*x * inv(t(x)*
               inv(sigma)*x)*t(x)*inv(sigma);
            f=0.5*(t(y)*P*P*y - trace(inv(sigma))) ;
            return (f);
finish F_ml;

start G_ml(Ahat, y) GLOBAL(d, x);
            sigma=diag(Ahat+d);
            P=inv(sigma) - inv(sigma)*x * inv(t(x)*
               inv(sigma)*x)*t(x)*inv(sigma);
            g= - trace(P*P) + trace(inv(sigma*sigma))/2 ;
            return (g);
finish G_ml;
***********Generate Data Parameters**************;
use saipe&year;
read all var { x0 x1 x2 x3 x4 } into x;
read all var{d} into d;
read all var{cps} into y;
close;

beta_ols=inv(t(x)*x)*t(x)*y;

************Prasad-Rao estimator**********************;
beta_ols=inv(t(x)*x)*t(x)*y;
h=vecdiag(x*inv(t(x)*x)*t(x));
A_pr=(t(y -x*beta_ols)*(y-x*beta_ols)-
        sum(d#(1-h)))/(&m-&p);
if  A_pr<0 then
    do;
      A_pr= 0;
```

```
    end;

**************Fay-Herriot Estimator***************;
A_fh=(0.5); diff=10;
do while  ((diff>0.0001)&(diff<10E4));
   Anew=A_fh+(&m-&p-f_fh(A_fh, y))/g_fh(A_fh, y);
   diff=abs((Anew-A_fh)/A_fh);
   A_fh=Anew;
end;

if A_fh< 0 then
    do;
        A_fh=0;
    end;

*print A_fh;
**************ADM_REML Estimator******************;
A_aR=(0.5); diff=10; i=0;
do while  ((diff>0.0001)&(diff<10E4)&(i<40));
    Anew=A_aR - f_admR(A_aR, y)/g_admR(A_aR, y);
    diff=abs((Anew-A_aR)/A_aR);
    A_aR=Anew;
    i=i+1;
end;

if (A_aR< 0) | (i>39) then
    do;
       A_aR=0;
    end;

**************ADM_ML Estimator*********************;
A_aM=(0.5); diff=10; i=0;
do while  ((diff>0.0001)&(diff<10E4)&(i<40));
     Anew=A_aM - f_admM(A_aM, y)/g_admM(A_aM, y);
     diff=abs((Anew-A_aM)/A_aM);
     A_aM=Anew;
     i=i+1;
end;

if (A_aM< 0) | (i>39) then
    do;
      A_aM=0;
    end;

*print A_adm;
```

```
**************REML Estimator*********************;
A_re=(0.5);
diff=10; i=0;
do while  ((diff>0.0001)&(diff<10E4)&(i<40));
     Anew=A_re - f_reml(A_re, y)/g_reml(A_re, y);
     diff=abs((Anew-A_re)/A_re);
     A_re=Anew;
     i=i+1;
end;

if (A_re< 0)|(i>39) then
    do;
     A_re=0;
    end;

**************ML Estimator***************************;
A_ml=(0.5); diff=10; i=0;
do while  ((diff>0.0001)&(diff<10E4)&(i<40)&(A_ml>0));
    Anew=A_ml - f_ml(A_ml,y)/g_ml(A_ml,y);
    diff=abs((Anew-A_ml)/A_ml);
    A_ml=Anew;
    i=i+1;
end;

if (A_ml< 0)|(i>39) then
    do;
       A_ml= 0;
    end;

****theta_fh***;
sigma_fh=diag(A_fh+d);
beta_fh=inv(t(x)*inv(sigma_fh)*x)*t(x)*inv(sigma_fh)*y;
B_fh=d/(d+A_fh);
theta_eb_fh=(1-B_fh)#y + B_fh#(x*beta_fh);

g1_fh = d#(1-B_fh);
g2_fh = B_fh##2 # vecdiag(x*inv(t(x)*inv(Sigma_fh)*x)*t(x));
g3_fh = 2*&m*d##2#(A_fh+D)##(-3)*(sum(1/(A_fh+d)))**(-2);
g4_fh = 2*B_fh##2 # ( &m*trace( inv(Sigma_fh)**2) -
 (trace(Sigma_fh**(-1)))**2 )/( trace(Sigma_fh**(-1)))**3;

mse_fh = g1_fh + g2_fh + 2*g3_fh - g4_fh;

****theta_adm re***;
sigma_ar=diag(A_ar+d);
```

```
beta_ar=inv(t(x)*inv(sigma_ar)*x)*t(x)*inv(sigma_ar)*y;
B_ar=d/(d+A_ar);
theta_eb_ar=(1-B_ar)#y + B_ar#(x*beta_ar);

g1_ar=d#(1-B_ar);
g2_ar=B_ar##2#vecdiag(x*inv(t(x)*inv(Sigma_ar)*x)*t(x));
g3_ar= d##2 # (A_ar+D)##(-3) * 2/trace( inv(Sigma_ar)**2);
if A_ar=0 then
   g4_ar=0;
else
   g4_ar = B_ar##2 *( 2/A_ar )/trace( inv(Sigma_ar)**2);

mse_ar = g1_ar + g2_ar + 2*g3_ar - g4_ar;
msep_ar=j(&m, 3,0);
msep_ar[,1]=(g1_ar+g3_ar - g4_ar)/mse_ar;
msep_ar[,2]=g2_ar/mse_ar;
msep_ar[,3]=g3_ar/mse_ar;

msep_ar=round(msep_ar, 0.01)*100;
*msep_ar[,4]=round(mse_ar,0.01);

****theta_adm ml***;
sigma_am=diag(A_am+d);
beta_am=inv(t(x)*inv(sigma_am)*x)*t(x)*inv(sigma_am)*y;
W_am= inv(sigma_am)* x * inv(t(x)*inv(sigma_am)*x)
      *t(x)*inv(sigma_am);
B_am=d/(d+A_am);
theta_eb_am=(1-B_am)#y + B_am #(x*beta_am);

g1_am=d#(1-B_am);
g2_am=B_am##2#vecdiag(x*inv(t(x)*inv(Sigma_am)*x)*t(x));
g3_am= d##2 # (A_am+D)##(-3)*2/trace(inv(Sigma_am)**2);
if A_am=0 then
   g4_am=0;
else  g4_am = B_am##2 *( trace(-W_am) + 2/A_am )
              / trace( inv(Sigma_am)**2 );

mse_am = g1_am + g2_am + 2*g3_am - g4_am;
mse_amn = g1_am ;

msep_am=j(&m, 3,0);
msep_am[,1]=(g1_am+g3_am - g4_am)/mse_am;
msep_am[,2]=g2_am/mse_am;
msep_am[,3]=g3_am/mse_am;
```

```
msep_am=round(msep_am, 0.01)*100;
*msep_am[,4]=round(mse_am,0.01);
* print  msep_ar msep_am;

****theta_reml***;
sigma_re=diag(A_re+d);
beta_re=inv(t(x)*inv(sigma_re)*x)*t(x)*inv(sigma_re)*y;
B_re=d/(d+A_re);
theta_eb_re=(1-B_re)#y + B_re#(x*beta_re);

g1_re = d#(1-B_re);
g2_re = B_re##2 # vecdiag(x*inv(t(x)*
        inv(Sigma_re)*x)*t(x));
g3_re = d##2# A_re+D)##(-3)*2/trace(inv(Sigma_re)**2);

mse_re = g1_re + g2_re + 2*g3_re;

****theta_ml***;
sigma_ml=diag(A_ml+d);
beta_ml=inv(t(x)*inv(sigma_ml)*x)*t(x)*inv(sigma_ml)*y;
W_ml= inv(sigma_ml)* x * inv(t(x)*inv(sigma_ml)*x)
      *t(x)*inv(sigma_ml);
B_ml=d/(d+A_ml);
theta_eb_ml=(1-B_ml)#y + B_ml#(x*beta_ml);

g1_ml = d#(1-B_ml);
g2_ml = B_ml##2#vecdiag(x*inv(t(x)*inv(Sigma_ml)*x)*t(x));
g3_ml = d##2#(A_ml+D)##(-3) * 2 /trace(inv(Sigma_ml)**2);
g4_ml = B_ml##2 *trace(-W_ml)/ trace(inv(Sigma_ml)**2);

mse_ml = g1_ml + g2_ml + 2*g3_ml - g4_ml;

**********************BOOTSTRAP METHOD AM***************;
betahat=beta_AR; Ahat=a_AR;
pivot_cllg=j(&bt_times,&m,0);

do btimes=1 to &bt_times;

   btheta=x*betahat + rannor(j(&m,1,&seed_bv))*sqrt(ahat);
   be=rannor(j(&m,1,&seed_be))#sqrt(d);
   by=btheta+be;

   Ahatstar=Ahat; diff=10; i=0;
   do while  ((diff>0.00001)&(diff<10E4)& (i<40));
       Anewstar=Ahatstar - f_AR(Ahatstar,by)
```

```
                /g_AR(Ahatstar,by);
        diff=abs(Anewstar-Ahatstar);
        Ahatstar=Anewstar;
        i=i+1;
    end;

    if (Ahatstar<0)|(i >39) then   Ahatstar=0.01;

    bsigma=diag(Ahatstar+d);
    betastar_g=inv(t(x)*inv(bsigma)*x)*t(x)*inv(bsigma)*by;
    Bhatstar=d/(d+Ahatstar);
    theta_ebstar=(1-Bhatstar)#by + Bhatstar#(x*betastar_g);
    g1star=d#(1-Bhatstar);

    pivot_cllg[btimes,]=t((btheta-theta_ebstar)/sqrt(g1star));

end;

**********************PB_cllgls*****************;
n=&bt_times; ns=(n*&alpha-1);
ts_AR=j(&m,3,0); tt=j(ns,3,0);
do i=1 to &m;
   temp=pivot_cllg[,i];bymean=pivot_cllg[,i];
   bymean[rank(bymean),]=temp;
   frac=n*&alpha/2;
   tt[,1]=bymean[1:ns,]; tt[,2]=bymean[n-ns+1:n,];
   tt[,3]=tt[,2]-tt[,1];
   ss=tt[><,3];
   ts_AR[i,1]=tt[loc(tt[,3]=ss),1];
   ts_AR[i,2]=tt[loc(tt[,3]=ss),2];
 end;

 *********************BOOTSTRAP METHOD ADM ML*********;
 betahat=beta_AM; Ahat=a_AM;
 pivot_cllg=j(&bt_times,&m,0);

 do btimes=1 to &bt_times;

    btheta=x*betahat + rannor(j(&m,1,&seed_bv))*sqrt(ahat);
    be=rannor(j(&m,1,&seed_be))#sqrt(d);
    by=btheta+be;

    Ahatstar=Ahat; diff=10; i=0;
    do while  ((diff>0.00001)&(diff<10E4)& (i<40));
        Anewstar=Ahatstar - f_AM(Ahatstar,by)/g_AM(Ahatstar,by);
```

```
        diff=abs(Anewstar-Ahatstar);
        Ahatstar=Anewstar;
        i=i+1;
    end;

    if (Ahatstar<0)|(i >39) then   Ahatstar=2/(&m-3);

    bsigma=diag(Ahatstar+d);
    betastar_g=inv(t(x)*inv(bsigma)*x)*t(x)*inv(bsigma)*by;
    Bhatstar=d/(d+Ahatstar);
    theta_ebstar=(1-Bhatstar)#by + Bhatstar#(x*betastar_g);
    g1star=d#(1-Bhatstar);

    pivot_cllg[btimes,]=t((btheta - theta_ebstar)/sqrt(g1star));

end;

*********************PB_cllgls*****************;
n=&bt_times; ns=(n*&alpha-1);
ts_AM=j(&m,3,0); tt=j(ns,3,0);
do i=1 to &m;
   temp=pivot_cllg[,i];bymean=pivot_cllg[,i];
   bymean[rank(bymean),]=temp;
   frac=n*&alpha/2;
   tt[,1]=bymean[1:ns,]; tt[,2]=bymean[n-ns+1:n,];
   tt[,3]=tt[,2]-tt[,1];
   ss=tt[><,3];
   ts_AM[i,1]=tt[loc(tt[,3]=ss),1];
   ts_AM[i,2]=tt[loc(tt[,3]=ss),2];
end;

B=j(&m,4 ,0);
B[,1]=B_re;
B[,2]=B_ml;
B[,3]=B_ar;
B[,4]=B_am;

Theta_eb=j(&m,4 ,0);
Theta_eb[,1]=Theta_eb_re;
Theta_eb[,2]=Theta_eb_ml;
Theta_eb[,3]=Theta_eb_ar;
Theta_eb[,4]=Theta_eb_am;

CI=j(&m,16 ,0);
CI[,1]=theta_eb_ar - 1.96#sqrt(g1_ar);
```

```
CI[,2]=theta_eb_ar+ 1.96#sqrt(g1_ar);
CI[,3]=theta_eb_ar + ts_ar[,1]#sqrt(g1_ar);
CI[,4]=theta_eb_ar + ts_ar[,2]#sqrt(g1_ar);
CI[,5]=theta_eb_am - 1.96#sqrt(g1_am);
CI[,6]=theta_eb_am+ 1.96#sqrt(g1_am);
*CI[,7]=theta_eb_am + ts_am[,1]#sqrt(g1_am);
*CI[,8]=theta_eb_am + ts_am[,2]#sqrt(g1_am);

length_cox_re=2*1.96#sqrt(g1_re);
*length_cox_am=2*1.96#sqrt(g1_am);
length_pb_ar=(ts_ar[,2]-ts_ar[,1])#sqrt(g1_ar);
length_pb_am=(ts_am[,2]-ts_am[,1])#sqrt(g1_am);

inflation_pb_ar =round( (length_pb_ar - length_cox_re)
                 *100/length_cox_re,0.1);
inflation_pb_am =round( (length_pb_am - length_cox_re)
                 *100/length_cox_re,0.1);

length_cox_re=round(length_cox_re,0.01);
*length_cox_am=round(length_cox_am,0.01);

print length_cox_re  inflation_pb_ar  inflation_pb_am;

********CV GAINS**********************;
CV_d= sqrt(d)/y;
CV_ar=sqrt(mse_ar)/theta_eb_ar;
CV_am=sqrt(mse_am)/theta_eb_am;

Gain_ar=round((cv_d-cv_ar)/cv_d*100, 0.1);
Gain_am=round((cv_d-cv_am)/cv_d*100, 0.1);

cv_d=round(cv_d,0.01);

print  A_Fh A_re A_ml A_ar A_am;
print y theta_eb_ar;
print cv_d gain_ar gain_am;
print y  ss;
print d  B ;
print  d  theta_eb ;


create B from B
[colname={ 'RE', 'ML','AR','AM'}];
append from B;
```

```
    create theta_eb from theta_eb
    [colname={ 'RE', 'ML','AR','AM'}];
    append from theta_eb;

quit;
```

Bibliography

[1] Battese, G.E., Harter, R.M. and Fuller, W.A. (1988), "An error-components model for prediction of county crop areas using survey and satellite data," *Journal of the American Statistical Association*, 80, 28-36.

[2] Bell, William R. (1999) "Accounting for uncertainty about variances in small area estimation," *Bulletin of the International Statistical Institute*, 52nd Session, Helsinki, 1999.

[3] Chatterjee, A. Lahiri, P., and Li, H. (2006), "Parametric bootstrap approximation to the distribution of EBLUP, and related prediction intervals in linear mixed models,"*Annals of Statisitcs*, in press.

[4] Chen, S. (2001), *Empirical Best Prediction and Hierarchical Bayes Methods in Small-Area Estimation,* Ph.D. Dissertation, Department of Mathematics and Statistics, University of Nebraska-Lincoln.

[5] Christiansen, C. L. and Morris, C. N. (1997), "Hierarchical Poisson regression modeling," *Journal of the American Statistical Association* 92, 618-632.

[6] Cox, D.R. (1975), "Prediction intervals and empirical Bayes confidence intervals," in *Perspectives in Probability and Statistics,* papers in honor of M.S. Bartlett. Ed: J. Gani, Academic Press, 47-55.

[7] Das, K., Jiang, J. and Rao, J. N. K. (2004), "Mean squared error of empirical predictor," *Annals of Statisitcs* 32, 818-840.

[8] Datta, G.S., and Lahiri, P. (2000), "Second order approximation to the mean squared error of EBLUP in small-area estimation problems," tentatively accepted in *Statistica Sinica,* 10, 613-627.

[9] Datta, G.S., Rao, J.N.K. and Smith, D.D. (2005), "On measuring the variability of small area estimators under a basic area level model," *Biometrika*, 92, 183-196.

[10] Fay, R. E., and Herriot, R. A. (1979), "Estimates of income for small places: an application of James-Stein procedure to Census data, " *Journal of the American Statistical Association*, 74, 269-277.

[11] Henderson, C.R. (1975), "Best linear unbiased estimation and prediction under a selection model," *Biometrics* 31, 423-447.

[12] Jiang, J. (1996), "REML estimation: asymptotic behavior and related topics," *Annals Statistics* 24, 255-286.

[13] Jiang, J. (2007), *Linear and generalized linear mixed models and their applications,* Springer.

[14] Jiang, J., and Lahiri, P. (2006),"Mixed model prediction and small area estimation (with discussions)," *Test*, 15, 1, 1-96.

[15] Kackar, R. N. and Harville, D. A. (1984), "Approximations for standard errors of estimators of fixed and random effects in mixed linear models," *Journal of the American Statistical Association* 79, 853-862.

[16] Lahiri, P., and Rao, J.N.K. (1995), "Robust estimation of mean squared error of small area estimators," *Journal of the American Statistical Association* 90, 758-766.

[17] Li, Huilin (2006), "An application of parametric bootstrap method in small area estimation problem," *2006 Proceedings of the American Statistical Association.*

[18] Miller, J. J. (1977), "Asymptotic properties of maximum likelihood estimates in the mixed model of analysis of variance, "*Annals Statistics* 5, 746-762.

[19] Morris, C.N. (1988), "Approximating posterior distributioins and posterior moments," In J.M. Bernardo, M.H. DeGroot, D. V. Lindley, and A.F.M. Smith, eds., Bayesian Statistics, Vol. 3, pp. 327-344. Oxford University Press, Oxford.

[20] National Research Council (2000), *Small-area estimates of school-age children in poverty: evaluation of current methodology,* C.F. Citro and G. Kalton (Eds.) Committee on National Statistics, Washington, DC: National Acedemy Press.

[21] Patterson, H. D. and Thompson, R. (1971), "Recovery of interblock information when block sizes are unequal," *Biometrika* 58, 545-554.

[22] Prasad, N.G.N., and Rao, J.N.K. (1990), "The Estimation of mean squared errors of small area estimators," *Journal of the American Statistical Association*, 85, 163-171.

[23] Srivastava, V.K., and Tiwari, R.(1976), "Evaluation of expectation of products of stochastic matrics, " Scandinavian Journal of Statistics, 3, 135-138.

[24] Rao, J. N. K. (2003), *Small Area Estimation*, Wiley, New York.

[25] Tang, R. (2002), *Fitting and evaluating certain two-Level hierarchical models,* Ph.D. Thesis, Dept. of Statistics, Harvard University.

[26] U.S. Census Bureau (2005), "Small area income and poverty estimates: 2003 state level estimation details," Housing and Household Economic Statistics Division, Washington, DC, available at http://www.census.gov/hhes/www/saipe/techdoc/methods/state/03statemod .html.

[27] Wolter, K. (1985), *Introduction to variance estimation*, New York: Springer-Verlag.