Worcester Polytechnic Institute

Digital WPI

2010-04-20

Bayesian Predictive Inference and Multivariate Benchmarking for Small Area Means

Ma. Criselda Santos Toto
Worcester Polytechnic Institute

Follow this and additional works at: https://digitalcommons.wpi.edu/etd-dissertations

Bayesian Predictive Inference and Multivariate Benchmarking for Small Area Means

A Dissertation

Submitted to the Faculty

of

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Doctor of Philosophy

in

Mathematical Sciences

by

Ma. Criselda Santos Toto

April 16, 2010

APPROVED:

Balgobin Nandram, Ph. D.
Dissertation Advisor
Department of Mathematical Sciences
Worcester Polytechnic Institute

Jayson Wilbur, Ph. D.
Instrumentation Laboratory

Domokos Vermes, Ph. D.
Department of Mathematical Sciences
Worcester Polytechnic Institute

Myron Katzoff, Ph. D.
Office of Research and Methodology
National Center for Health Statistics

Zheyang Wu, Ph. D.
Department of Mathematical Sciences
Worcester Polytechnic Institute

Jai Won Choi
Department of Biostatistics
Medical College of Georgia

Abstract

Direct survey estimates for small areas are likely to yield unacceptably large standard errors due to the small sample sizes in the areas. This makes it necessary to use models to "borrow strength" from related areas to find more reliable estimate for a given area or, simultaneously, for several areas. For instance, in many applications, data on related multiple characteristics and auxiliary variables are available. Thus, multivariate modeling of related characteristics with multiple regression can be implemented.

However, while model-based small area estimates are very useful, one potential difficulty with such estimates when models are used is that the combined estimate from all small areas does not usually match the value of the single estimate on the large area. Benchmarking is done by applying a constraint to ensure that the "total" of the small areas matches the "grand total". Benchmarking can help to prevent model failure, an important issue in small area estimation. It can also lead to improved prediction for most areas because of the information incorporated in the sample space due to the additional constraint. We describe both the univariate and multivariate Bayesian nested error regression models and develop a Bayesian predictive inference with a benchmarking constraint to estimate the finite population means of small areas. Our models are unique in the sense that the benchmarking constraint involves unit-level sampling weights and the prior distribution for the covariance of the area effects follows a specific structure.

We use Markov chain Monte Carlo procedures to fit our models. Specifically, we use Gibbs sampling to fit the multivariate model; our univariate benchmarking only needs random samples. We use two datasets, namely the crop data (corn and soybeans) from the LANDSAT and Enumerative survey and the NHANES III data (body mass index and bone mineral density), to illustrate our results. We also conduct a simulation study to assess frequentist properties of our models.

Acknowledgments

I extend my heartfelt gratitude, appreciation, and indebtedness to Jesus Christ, Our Lord and Saviour, and to the following persons, who in one way or the other were material in making this thesis possible.

To my advisor, Balgobin Nandram, for his full support and guidance, unending patience and kindness, and for all the time he sacrificed for me.

To my parents and sisters, whose support and love are endless, even from thousands of miles away.

To all the members of my dissertation committee for the time and effort they spent in reading and in evaluating my dissertation, for their presence at my dissertation defense, and for the all valuable comments and suggestions they gave to improve my dissertation.

To all the professors I have taken classes with and worked with as a TA, who have been a big part of my graduate experience at WPI, for all of the theoretical and practical lessons I learned and for all the help and encouragement that you have given me along the way.

To all my relatives, colleagues, and friends, both in the Philippines and in the United States, for the support and for boosting my confidence in accomplishing this thesis. Without their emotional, moral, and spiritual support, especially during the toughest times of my life, this thesis would not have been completed.

Lastly, to Andy, for all the love, care, help, support, and inspiration you have given me. Your time on Earth might be over... but I will carry on the mission you promise to live by. Your memories will always be remembered. You will always have a special place in my heart. ♡

Contents

Chapter 1

Introduction

This dissertation is focused on implementing a Bayesian predictive inference to estimate the finite population means of small domains with benchmarking. We use a Bayesian nested-error regression model with an internal benchmarking constraint that incorporates unit-level sampling weight.

In this chapter, we introduce the concept of benchmarking and discuss the reasons for implementing it in small area estimation. In Section 1.1, we discuss the issues in model-based small area estimation and how benchmarking would help resolve some of the issues. We introduce our univariate nonbenchmarking Bayesian model in Section 1.2. Moreover, we present detailed discussions of the different studies related to our work in Section 1.3. Lastly, we will introduce in Section 1.4 the form of the benchmarking constraint we use in our study. Our benchmarking constraint incorporates unit-level survey or sampling weights, so we also give a brief discussion of that in Section 1.4. We would also introduce in Section 1.5 the two applications on which we apply our results, as well as discuss some of the literature related to our study.

This dissertation has three additional chapters following the introduction. In Chapter 2, we describe our benchmarking Bayesian model for the univariate case. Here we introduce the Bayesian version of the Battese, Harter, and Fuller (1988) nested-error regression model. We present in Chapter 2 how we incorporate the constraint into the regression model. We use the resulting univariate Bayesian benchmarking model to estimate the finite population means of small areas by means of Bayesian predictive inference procedures. We will present the results of our univariate Bayesian models us-

ing the crop data (Battese, Harter, and Fuller (1988)) and the body mass index (BMI) and bone mineral density (BMD) data from NHANES III (1996). We also present a simulation study drawing data similar to the corn hectares data to further compare the univariate nonbenchmarking (NBM) and univariate benchmarking (BM) models.

In Chapter 3, we extend the univariate results of Chapter 2 to the multivariate setting. We discuss the Bayesian version of the multivariate nested-error regression model of Fuller and Harter (1987). There is a practical significance to this extension of the benchmarking model to the multivariate case because in many small area problems, data on related multiple characteristics and auxiliary variables are available. For instance, the two crops, corn and soybean, in the survey and satellite data would be more appropriately benchmarked simultaneously, so as the two health variables, BMI and BMD, in the NHANES III data.

Finally, in Chapter 4 we discuss our methodological contributions, summarize our results, present concluding remarks, and discuss the different problems we want to do in the future that are related to this research.

1.1   Benchmarking in Small Areas

Small area estimation has gained popularity in recent years due to the growing demand for reliable small area statistics from both the public and the private sectors worldwide. For instance, in many countries, including United States and Canada, the government use small area statistics in the apportionment of funds, as well as in regional and city planning. On the other hand, the private sectors have demands for small area statistics because the policy making of many businesses and industries relies on local socio-economic conditions. One major problem in small area estimation is that the sample sizes for small areas are typically too small. This is because these sample sizes depend on the overall sample size in a survey. But the overall sample size is usually determined to provide specific accuracy at a much higher level of aggregation than that of small area. For instance, sample survey data are typically used to derive reliable estimators of totals and means for large areas or domains. Thus, the usual direct survey estimators for a small area, based on data only from the small units in the area, are likely to yield

unacceptably large standard errors due to the unwarranted small size of the sample in the area. Thus, making the direct estimates from the small areas unreliable. In order to provide reliable estimates, a common practice in small area estimation is to use appropriate models that "borrow strength" from neighboring or related small areas, or other correlated variables (via multivariate approach) and relevant covariate information available from other sources, such as administrative records. This concern about the unreliability of small area estimates led to the development of various model-based methods in small-area estimation; see Rao (2003) for a complete discussion of small area estimation, and Ghosh and Rao (1994) for an appraisal of some of these methods.

Appropriate models are often used to produce reliable small area estimates by "borrowing strength" from the ensemble. For instance, since there are many applications with multiple characteristics and auxiliary variables in small area problems, a multivariate modeling of related characteristics using a multiple regression model can be implemented (Fay, 1987). The success of such modeling rests on the strength of dependence among these characteristics. Multivariate analysis can help increase precision by using data from other characteristics. For instance, in the LANDSAT and Enumerative Survey data (see Battese, Harter, and Fuller (1988)), the yield of corn may be related to the yield of soybean, synergistically or antagonistically. This can also be true with the body mass index (BMI) and bone mineral density (BMD) from NHANES III (1996). The model-based small area estimates can differ widely from the direct estimates of the individual areas, especially for areas with very low sample sizes. Thus, while model-based small area estimates are very useful, one potential difficulty with such estimates is that when aggregated, the combined estimate of all the small areas may be quite different from the corresponding direct estimate on the large area, the latter being usually believed to be quite reliable. For example, sample surveys are usually designed to achieve specified inferential accuracy at a higher level of aggregation. Thus, the samples from each small area might not yield accurate or reliable estimates.

In the government and public sector, an overall agreement with the direct estimates at an aggregate level may sometimes be politically necessary to convince the legislators of the utility of small area estimates. One way to ensure that the combined model-based estimates matches the direct design-based estimate is by implementing benchmarking, which is done by applying a constraint, internally or externally, to ensure

that the "total" of the small areas matches the "grand total." Internal benchmarking occurs when the pre-specified estimator can be a weighted average of the direct small area estimators, and external benchmarking occurs when the pre-specified estimator is obtained from external sources, such as a different survey census, or other administrative records. Through benchmarking, the model-based estimates are modified in order to get the same aggregate estimate for the larger area. We are forcing the combined small area estimates to match the direct estimate of the large area obtained when the small areas are collapsed into a single area. Model failure is an important issue in small area estimation. The problem of non-overall agreement between combined small area estimates and the overall estimate from using model-based estimates can be a more severe issue if the model fails. This can cause serious problems since most often, there is no real way to check for validity of the assumed model. Benchmarking can help to prevent this model failure, because the implementation of benchmarking corrects for some bias. Thus, it provides some guard against model failure. Benchmarking also shifts the small area estimators to accommodate the benchmark constraint. In doing so, it can provide some increase in the precision of the small area estimators of the finite population means or totals because the sample space is reduced by the constraint. This makes the benchmarking technique desirable to practitioners of model-assisted small area estimation.

1.2 Basic Problem

Assume that there are ℓ small areas. Let N_i be the population size of the ith area of a finite population $(i = 1, \ldots, \ell)$ and $N = \sum_{i=1}^{\ell} N_i$ be the total population size.

Let $\{y_{ij}, i = 1, 2, \ldots, \ell, j = 1, 2, \ldots, N_i\}$ denote a finite population where y_{ij} is the response associated with the jth unit within the ith small area. Assume that $y_{ij}, \ i = 1, \ldots, \ell, \ j = 1, \ldots, n_i$, are observed from the finite population, where n_i is the size of the sample from the ith area $(i = 1, \ldots, \ell)$ and $n = \sum_{i=1}^{\ell} n_i$ is the total sample size. We assume N_i and n_i are fixed known quantities. We write the $n \times 1$ vector of observed values as $\underset{\sim}{y}_s = \left(\underset{\sim}{y}'_{s_1}, \ldots, \underset{\sim}{y}'_{s_\ell} \right)'$ with $\underset{\sim}{y}_{s_i} = (y_{i1}, \ldots, y_{in_i})'$ for $i = 1, 2, \ldots, \ell$.

Our main goal is to make inference about the *finite population mean of the* i^{th}

area, $\overline{Y}_i = N_i^{-1} \sum_{j=1}^{N_i} y_{ij}, \; i = 1, \ldots, \ell,$ based on the sample data $\underset{\sim}{y}_s$ under the nested-error regression model of Battese, Harter, and Fuller (1988). To predict the finite population mean, we need to predict the nonsample values in the population. Our approach is to use Bayesian predictive inference to generate values of the finite population means from the small areas based on the sample data. We predict the values of the nonsample data from its posterior predictive distribution.

The nested-error regression model of Battese, Harter, and Fuller (1988) for finite population means accommodates covariates at the unit level. Assume $\underset{\sim}{x}_{ij}$ is the $p \times 1$ vector of covariates associated with $y_{ij}, \; i = 1, 2, \ldots, \ell, j = 1, 2, \ldots, N_i$. Note that each $\underset{\sim}{x}_{ij}$ contains an intercept. Let X_s be the $n \times p$ matrix of covariates for the sample data $\underset{\sim}{y}_s$. Similarly, let X_{ns} be the $(N - n) \times p$ matrix of covariates for the nonsample data $\underset{\sim}{y}_{ns}$. Also, denote by $X_{ns_{(N)}}$ the $(N - 1 - n) \times p$ matrix of covariates for the nonsampled observations without y_{ℓ, N_ℓ}. Note that these matrices of covariates are assumed to be full-rank.

The population model of the Battese, Harter, and Fuller (1988) nested-error regression model for finite population means is given by

$$y_{ij} = \underset{\sim}{x}_{ij}' \underset{\sim}{\beta} + \nu_i + e_{ij},$$

$i = 1, 2, \ldots, \ell, j = 1, 2, \ldots, N_i,$ where $\underset{\sim}{\beta} = (\beta_0, \ldots, \beta_{p-1})'$ is the $p \times 1$ vector of regression parameters. Let $\underset{\sim}{\nu} = (\nu_1, \ldots, \nu_\ell)'$ be the $\ell \times 1$ vector of random effects. These area effects are assumed to be independent and each follows a normal distribution. That is,

$$\nu_i \overset{iid}{\sim} \text{Normal} \left\{ 0, \left(\frac{\rho}{1 - \rho} \right) \sigma^2 \right\}, i = 1, 2, \ldots, \ell$$

and independent of the unit errors e_{ij}, which are also assumed to follow

$$e_{ij} \overset{ind}{\sim} \text{Normal} \left\{ 0, \sigma^2 \right\}.$$

We define ρ as the intra-class correlation within areas and is the same for each area. We assume that $0 < \rho < 1$. The introduction of ρ in the model is a novel idea in this study.

5

Using the prior

$$p\left(\underset{\sim}{\beta}, \sigma^2, \rho\right) \propto \frac{1}{\sigma^2}, \quad \sigma^2 > 0,$$

the Bayesian nonbenchmarking model (NBM) is given by

$$y_{ij}|\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2 \overset{ind}{\sim} \text{Normal}\left\{\underset{\sim}{x}'_{ij}\underset{\sim}{\beta} + \nu_i, \sigma^2\right\}, \quad i = 1, \ldots, \ell, \quad j = 1, \ldots, N_i, \quad (1.1)$$

$$\nu_i|\rho, \sigma^2 \overset{iid}{\sim} \text{Normal}\left\{0, \left(\frac{\rho}{1-\rho}\right)\sigma^2\right\}, \quad 0 < \rho < 1, \quad (1.2)$$

$$p\left(\underset{\sim}{\beta}, \sigma^2, \rho\right) \propto \frac{1}{\sigma^2}, \quad \sigma^2 > 0. \quad (1.3)$$

Essentially, the model specified by (1.1) and (1.2) is a reparameterization of the Battese, Harter, and Fuller (BHF) model, and in fact, it is equivalent to the BHF model. To see this, suppose we replace (1.2) by

$$\nu_i|\sigma^2, \delta^2 \overset{iid}{\sim} \text{Normal}\left\{0, \delta^2\right\},$$

then $\rho = \dfrac{\delta^2}{\delta^2 + \sigma^2}$ is the intra-class correlation of the y_{ij} within each area. This is also the same as

$$y_{ij} = \underset{\sim}{x}'_{ij}\underset{\sim}{\beta} + \nu_i + e_{ij},$$

where $e_{ij} \overset{ind}{\sim} \text{Normal}\{0, \sigma^2\}$ and independently $\nu_i \overset{iid}{\sim} \text{Normal}\{0, \delta^2\}$, as in the BHF model. However, there is a minor difference which is ρ must be strictly less than 1 in (1.2). This reparameterization is helpful to prove propriety of the joint posterior density, and it also helps to simplify the computation.

Using Bayes' theorem in (1.1) - (1.3), the joint posterior density of the nonbenchmarking model is

$$\pi\left(\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2, \rho|\underset{\sim}{y}_s\right) \propto \left(\frac{1}{\sigma^2}\right)^{1+(n+\ell)/2} \left(\frac{1-\rho}{\rho}\right)^{\ell/2}$$

$$\times \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{\ell}\left(\frac{n_i}{\lambda_i}\left(\nu_i - \lambda_i\left(\overline{y}_{s_i} - \overline{\underset{\sim}{x}}'_{s_i}\underset{\sim}{\beta}\right)\right)^2 + \lambda_i\left(\frac{1-\rho}{\rho}\right)\left(\overline{y}_{s_i} - \overline{\underset{\sim}{x}}'_{s_i}\underset{\sim}{\beta}\right)^2\right.\right.$$

$$\left.\left.+ \sum_{j=1}^{n_i}\left[\left(y_{ij} - \overline{y}_{s_i}\right) - \left(\underset{\sim}{x}_{ij} - \overline{\underset{\sim}{x}}_{s_i}\right)'\underset{\sim}{\beta}\right]^2\right)\right]\right\},$$

where $\overline{y}_{s_i} = \dfrac{1}{n_i} \displaystyle\sum_{j=1}^{n_i} y_{ij},\ \overline{x}_{s_i} = \dfrac{1}{n_i} \displaystyle\sum_{j=1}^{n_i} \underset{\sim}{x}_{ij},$ and $\lambda_i = \dfrac{\rho n_i}{\rho n_i + (1 - \rho)}$ for $i = 1, 2, \ldots, \ell$. Assuming that $0 < \rho < 1$ and the design matrix is full rank, the joint posterior density $\pi\left(\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2, \rho | \underset{\sim}{y}_s\right)$ is proper; see Nandram, Toto, and Choi (2010) for a similar proof.

Based on the observed values $\underset{\sim}{y}_s$, we want to make inference for the finite population mean $\overline{Y}_i$. Let $\overline{x}_{ns_i} = \dfrac{1}{N_i - n_i} \displaystyle\sum_{j=n_i+1}^{N_i} \underset{\sim}{x}_{ij}$ and define the sampling fraction $f_i = \dfrac{n_i}{N_i}, i = 1, 2, \ldots, \ell$. To be able to make inference for $\overline{Y}_i\ (i = 1, \ldots, \ell)$, given the sample observations $\underset{\sim}{y}_s$, a possible approach is to sample from the distribution $\overline{Y}_i | \underset{\sim}{y}_s$. The distribution of $\overline{Y}_i | \underset{\sim}{y}_s$ is

$$p\left(\overline{Y}_i | \underset{\sim}{y}_s\right) = \int p\left(\overline{Y}_i | \underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2, \underset{\sim}{y}_s\right) \pi\left(\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2 | \underset{\sim}{y}_s\right) d\underset{\sim}{\nu} d\underset{\sim}{\beta} d\sigma^2. \tag{1.4}$$

This is called Bayesian predictive inference. Under the nonbenchmarking Bayesian model (NBM), $\overline{Y}_i$ given $\underset{\sim}{y}_s, \underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2$, for $i = 1, \ldots, \ell$, has the following distribution

$$\overline{Y}_i | \underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2, \underset{\sim}{y}_s \sim \text{Normal}\left\{f_i \overline{y}_{s_i} + (1 - f_i) \overline{x}_{ns_i}' \underset{\sim}{\beta} + (1 - f_i) \nu_i, \dfrac{\sigma^2}{N_i} (1 - f_i)\right\}. \tag{1.5}$$

Notice that the mean of the distribution of $\overline{Y}_i | \underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2, \underset{\sim}{y}_s$ depends on information both from the sample data and nonsample data from the ith area, and are independent from the distribution of the finite population means of the i^{*th} areas, $i^* = 1, 2, \ldots, \ell, i^* \neq i$.

1.3 Literature Review

In this section we mention some of the existing benchmarking literature for small area estimation. Benchmarking in small area estimation has been an active area of research in recent years, with the use of Bayesian analysis becoming a more popular method in executing the procedure.

Pfeffermann and Tiller (2006) used state-space models to benchmark small area estimators to aggregates of survey direct estimators within a group of areas. Their method is implemented to improve the procedure of the US Bureau of Labor Statistics

(BLS), which produces monthly employment and unemployment estimates for different census divisions and states. The work of Pfeffermann and Tiller (2006) is an example of internal benchmarking, where their benchmark constraint depended on a weighted direct estimator of the small area means. Using hierarchical Bayes (HB) approach, You, Rao, and Dick (2004) studied benchmarked estimators for small area means based on unmatched sampling and linking area level models proposed by You and Rao (2002b). Their method is evaluated using the 1991 Canadian census under-coverage estimation. You, Rao, and Dick (2004) used external benchmarking to develop their hierarchical Bayes (HB) estimators for area level unmatched sampling and linking models.

Hillmer and Trabelsi (1987) developed a statistical model-based approach to the benchmarking problem in the study of time series data. Benchmarking is done when data from a monthly sample survey are combined with data from an annual census for the purpose of improving the survey estimates. Previous authors have used numerical analysis techniques to derive methods to perform benchmarking. This article formulates the benchmarking problem in a statistical framework and uses times series methods to derive a solution. This solution is based in part upon the statistical properties of the time series being benchmarked and upon the properties of the survey errors associated with that time series. The article makes use of the theory of signal extraction that has been derived for non stationary time series. Two common types of benchmarking problems are studied in greater detail. The results of the theory derived in the article are illustrated by an example.

Wang, Fuller, and Qu (2008) gave a characterization of the "best" linear unbiased predictor (BLUP) for small area means under an area level model

$$Y_i = \underset{\sim}{x}_i' \beta + b_i + e_i, \quad i = 1, 2, \dots, n,$$

that satisfies a benchmarking constraint involving the small area predictors $\hat{y}_i$

$$\sum_{i=1}^{n} \omega_i \hat{y}_i = \sum_{i=1}^{n} \omega_i Y_i,$$

where ω_i are the area-level sampling weights such that $\sum_{i=1}^{n} \omega_i Y_i$ is a design consistent

8

estimator of the total (or the mean). They aim at minimizing the 'loss function' criterion

$$Q\left(\hat{y}^a\right) = \sum_{i=1}^{n} \varphi_i \mathrm{E}(\hat{y}^a - y_i)^2,$$

where $\varphi_i, i = 1, 2, \ldots, n$ are a chosen set of positive weights, while satisfying the benchmarking constraint above. In fact, all linear unbiased predictors satisfy this 'loss function' criterion. In benchmarking, this characterization led to a unifying view of some BLUP based predictors (e.g., Battese, Harter, and Fuller (1988), Isaki, Tsay, and Fuller (2000), Pfeffermann and Barnard (1991)). Wang, Fuller, and Qu (2008) also presented an alternative way of imposing the benchmarking constraint. In this approach, the BLUP estimator would have a self-calibrated property. They used an approach in which the weights are included in an augmented model, that is, we have

$$\underset{\sim}{Y} = \underset{\sim}{X}_1 \underset{\sim}{\beta} + \underset{\sim}{b} + \underset{\sim}{e}$$

where $\underset{\sim}{X}_1 = (\underset{\sim}{X}, \underset{\sim}{\omega})$ and $\underset{\sim}{\omega} = (\omega_1, \omega_2, \ldots, \omega_n)'$. This is used to obtain the best linear unbiased predictor (BLUP) or Empirical BLUP (EBLUP). With the weights $\underset{\sim}{\omega}$ in the model, their adjustment met the benchmarking criterion above, which has smaller adjustment for the model than without the weights. Their proposed self-calibrated augmented model reduces bias both at the overall and small area level. However, their model does not predict finite population means, and so their benchmarking constraint is different from ours. We wish to benchmark the finite population mean.

You and Rao (2003, 2002a) developed self-calibrated estimators for small area means under the unit level nested error regression model

$$y_{ij} = \underset{\sim}{x}'_{ij} \underset{\sim}{\beta} + v_i + e_{ij}, \quad j = 1, 2, \ldots, n_i, \, i = 1, 2, \ldots, \ell.$$

You and Rao (2003) benchmarked the Battese, Harter, and Fuller (1988) regression models using survey weights. They proposed three pseudo-hierarchical Bayes estimators for the small area means based on survey weights. The use of survey weights in their models ensured design-consistency of their estimators. Like You and Rao (2003, 2002a), we also combine a unit-level benchmarking constraint with a nested error regression model. They also illustrated the resulting survey-weighted univariate model using the crop data. You and Rao (2003) proposed a pseudo-HB estimator obtained from a two-

step estimation process involving both area and unit level models, while You and Rao (2002a) used a pseudo-EBLUP procedure to find the estimator. The pseudo-HB (or similarly, pseudo-EBLUP) estimates depend on the survey weights and satisfy design consistency. An advantage of their proposed estimator is that it automatically satisfies the benchmarking property when the estimators $\hat{\theta}_{iw}$ are aggregated over i, if the weights $\tilde{\omega}_{ij}$ are calibrated to agree with the known population total N_i of the ith area, that is,

$$\sum_{i=1}^{n_i} \tilde{\omega}_{ij} = \tilde{\omega}_{i\cdot} = N_i.$$

This means $\sum_{i=1}^{n_i} N_i\hat{\theta}_i$ equals the direct survey regression estimator $\hat{Y}_{\omega} + \left(X - \hat{X}'_{\omega}\underset{\sim}{\beta}_w \right)$, of the overall total Y, where

$$\hat{Y}_{\omega} = \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \tilde{\omega}_{ij} y_{ij},$$

is the direct estimator of the overall total Y. They used the Gibbs sampler to obtain samples of the estimates of the regression coefficients and error-term variance parameters. However, because of the awkwardness of the joint posterior density, it was difficult for You and Rao (2003, 2002a) to perform complete Bayesian analyses. Our work is different from the work of You and Rao because we discuss a full Bayesian predictive inference of the finite population means of small areas. Without transforming the data, we show in Chapter 2 how to do a full Bayesian analysis to estimate the finite population means of small areas; we also show this for multivariate outcomes in Chapter 3.

Datta, Ghosh, Steorts, and Maples (2009) discussed several Bayesian benchmarked estimators for area-level models. That is, they found the benchmarked Bayes estimator of $\hat{\underset{\sim}{\theta}} = \left(\hat{\theta}_1, \ldots, \hat{\theta}_{\ell} \right)$ of the small area means vector $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_{\ell})$ such that

$$\sum_{i=1}^{\ell} \omega_i \hat{\theta} = t,$$

where t is either pre-specified from some other source or $t = \sum_{i=1}^{\ell} \omega_i \hat{\theta}_i$, and ω_i are the weights attached to the direct estimators. They implemented two sets of benchmarking constraints, one with respect to some weighted mean and another with respect to both

weighted mean and weighted variability. Datta, Ghosh, Steorts, and Maples (2009) developed a class of constrained empirical Bayes estimators for area-level models using either internal or external benchmarking. They applied their results to produce model-based estimates of the number of poor school-aged children at the state, county, and school district levels. Their Bayes estimators are obtained by minimizing the posterior expectation of the weighted squared error loss

$$\sum_{i=1}^{\ell} \phi_i E\left[(\theta_i - e_i)^2 | \hat{\underline{\theta}}\right]$$

with respect to the e_i's satisfying $\overline{e}_{\omega} = \sum_{i=1}^{\ell} \omega_i e_i = t$. Moreover, the ϕ_i's may be the same as the w_i, but this does not necessarily have to be the case. They mentioned that these ϕ_i may be regarded as weights for a multiple-objective decision process, that is, each specific weight is relevant only to the decision-make for the corresponding small area, who may not be concerned with the weights related to decision-makers in other small areas. One advantage of their Bayesian approach is its ability to adjust to any general Bayes estimator, linear or non-linear. This is a decision-theoretic approach to benchmark small areas with only area-level data. It is interesting that they also showed that the standard raking procedure arises as a special case from their procedure. Our approach is different from the work of these authors because they do not make inference directly about the finite population mean of an area.

Nandram, Toto, and Choi (2010) did a full Bayesian analysis of the finite population mean of each area using the Scott-Smith superpopulation model (Scott and Smith, 1969)

$$y_{ij} \overset{iid}{\sim} \text{Normal}\left\{\theta + \nu_i, \sigma^2\right\}, j = 1, \ldots, N_i, i = 1, \ldots, \ell,$$

where θ is an overall effect and the ν_i are area effects. Note that the nested-error regression model of Battese, Harter, and Fuller (1988) used in this study is an extension of the Scott-Smith model to accommodate covariates at the unit level. Similar to the goal of our study, Nandram, Toto, and Choi (2010) wanted to benchmark the estimators of the $\bar{Y}_i$, $i = 1, \ldots, \ell$. In this case, Nandram, Toto, and Choi (2010) allowed a practitioner to specify the finite population mean for all areas combined; one can estimate the finite population mean for all areas combined and use that as the specified value ignoring the

variability. They included the benchmark constraint

$$\sum_{i=1}^{\ell} \sum_{j=1}^{N_i} y_{ij} = aN,$$

where it is assumed that a is a fixed known quantity. If a is not known, one would need to fit an independent model that does not distinguish the small areas to get a "precise" value of a. Our approach is similar to Nandram, Toto, and Choi (2010), but we are doing internal benchmarking, which is considerably more difficult.

Lastly, Datta, Day, and Maiti (1998) discussed a multivariate model for Bayesian small area estimation. They implemented Bayes prediction of small area mean vectors using the multivariate nested error regression model of Fuller and Harter (1987). Datta, Day, and Maiti (1998) developed a multivariate hierarchical Bayes approach to predict the small area mean vectors based on the multivariate nested error regression model

$$\underset{\sim}{Y}_{ij} = \underset{\sim}{x}_{ij} B + \underset{\sim}{u}_{ij}, \qquad (1.6)$$

where $\underset{\sim}{Y}_{ij}$ is an $q-$dimensional row vector of observations on the variables for which small area estimates are desired, $\underset{\sim}{x}_{ij}$ is a $p-$dimensional row vector of auxiliary variables for which the small area totals (or means) are known, B is a $q \times p$ matrix of coefficients, and $\underset{\sim}{u}_{ij}$ are the $q-$dimensional row vector of errors. Our multivariate benchmarking Bayesian model in Chapter 3 is similar to this model. Note that in (1.6),

$$\underset{\sim}{u}_{ij} = \underset{\sim}{v}_i + \underset{\sim}{e}_{ij},$$

$$\underset{\sim}{v}_i \sim \text{Normal}\left(\underset{\sim}{0}, \Sigma_{vv}\right),$$

$$\underset{\sim}{e}_{ij} \sim \text{Normal}\left(\underset{\sim}{0}, \Sigma_{ee}\right),$$

where $\underset{\sim}{v}_i$ is independent of $\underset{\sim}{e}_{ij}$ for all i, j, k. They were interested in predicting

$$\underset{\sim}{\gamma}_i = N_i^{-1} \sum_{j=1}^{N_i} \underset{\sim}{Y}_{ij},$$

the finite population mean for the ith domain, $i = 1, 2, \ldots, \ell$. Datta, Day, and Maiti (1998) also applied their model to estimate the finite population means of the crop (corn

and soybeans) data (Battese, Harter, and Fuller, 1988). They used Gibbs sampling to generate their Monte Carlo samples using the full conditional distributions from the joint posterior distribution, which is the same approach we implement in our numerical calculation. However, like Battese, Harter, and Fuller (1988), Datta, Day, and Maiti (1998) did not incorporate any means of benchmarking in their hierarchical Bayesian model. Datta, Day, and Maiti (1998) observed some efficiency gains in the use of the multivariate model over its univariate counterpart. One of the main difference of our model with Datta, Day, and Maiti (1998) is the form of the joint prior imposed in their hierarchical Bayes model. They used an improper prior of the form

$$\pi\left(\underset{\sim}{\beta}, \Sigma_{ee}, \Sigma_{vv}\right) \propto |\Sigma_{vv}|^{-a_v/2} |\Sigma_{ee}|^{-a_e/2}.$$

This type of prior belongs to the family of inverse-Wishart distributions. Gelman (2006) discussed the problems with the inverse-gamma family of "noninformative" prior distribution for the variance of the area effects in the two-level normal data. The inverse-gamma(ϵ, ϵ) model is one of the commonly used improper prior of the scalar area effects variance. This prior does not have any proper limiting posterior distribution. As a result, posterior inferences are sensitive to ϵ. It cannot be comfortably set to a low value such as 0.001. This problem can generalize in the inverse-Wishart family of prior distributions. Despite the use of these improper inverse-Wishart priors, Datta, Day, and Maiti (1998) obtain a set of necessary and sufficient conditions for the propriety of the posterior distributions corresponding to this class of improper priors on the components of variance matrices. As a difference with the work of Datta, Day, and Maiti (1998), we use a more specific form of prior for the distribution of the area effects $\underset{\sim}{\nu}$ to alleviate some of the issues that Gelman (2006) brought up in the use of the "noninformative" prior distributions for variance components. Our prior distribution is more specific that it makes the numerical computations easier, but it also led to a more complex joint posterior distribution. To compare the performance of the multivariate approach with the usual univariate approach they analyzed the survey and satellite data of Battese, Harter, and Fuller (1988). Their simulations show that the multivariate approach may result in substantial improvement over its univariate counterpart.

1.4   The Benchmarking Constraint

Many surveys use differential probabilities of selection (e.g., probability proportional to size) and sampling weights are used to correct for other sources of bias as well. Sampling weights are used to adjust for oversampling of certain populations; sampling weight adjustment for noncoverage and nonresponse. Generally, models used in small area estimation do not make use of the unit level survey weights. For instance, the Third National Health and Nutrition Examination Survey (NHANES III, 1996), has a large oversampling of young children, older persons, black persons, and Mexican-Americans because there are many nonrespondents among children and adolescents, and blacks and Mexican-Americans are selected with higher probability. To account for this complex survey design, including oversampling and survey non-response, sampling weights were assigned to each sampled individual; see NHANES III (1996) for more details. To achieve design consistency, it is essential that sample weights be incorporated into the model. Otherwise, it is likely that users will misinterpret the results. The sampling weights are used to produce estimates of population means and standard errors of the mean.

For $i = 1, \ldots, \ell, \ j = 1, \ldots, n_i$, denote the sampling weight corresponding to the sampled unit y_{ij} as ω_{ij}. Note that the sample weights for the nonsampled units are not generally available for secondary data analysts. So, the jth sampled unit within the ith area must represent ω_{ij} units in the entire population including itself. Thus, for $i = 1, \ldots, \ell, \ \sum_{j=1}^{n_i} \omega_{ij} \cong N_i$. The sampling weights can be $\omega_{ij} = N/n, \ i = 1, \ldots, \ell, \ j = 1, \ldots, n_i$ (e.g., proportional allocation), or $\omega_{ij} = N_i/n_i$ for $i = 1, \ldots, \ell, \ j = 1, \ldots, n_i$ (e.g., simple random sampling within areas).

To estimate the small area mean $\overline{Y}_i$, we want to determine a standard predictor $\hat{\overline{Y}}_i, \ i = 1, 2, \ldots, \ell$. A practitioner might want to use design-consistent estimator to predict $\overline{Y}_i$. Given fixed known sampling weights $\omega_{ij}, \ i = 1, \ldots, \ell, \ j = 1, \ldots, n_i$, a direct design-based estimator of the small area mean $\overline{Y}_i$ that is popular among survey practitioners is the Horvitz-Thompson estimator,

$$\hat{\overline{y}}_i = \sum_{j=1}^{n_i} \frac{\omega_{ij} y_{ij}}{N_i}.$$

If the design is simple random sampling within areas, that is, $\omega_{ij} = N_i/n_i$, then a design-consistent estimator of $\overline{Y}_i$ is the ith sample mean $\overline{y}_i$.

But, as mentioned earlier, the problem with the use of direct design-based estimators is that the sample sizes for the small areas might be too small and these standard survey estimators would yield unacceptably large standard error. This enforces the use of model-based estimators. However, model-based estimators of the small area means based on unit level model, such as a nested error regression model, do not use the unit level survey weights, ω_{ij}, attached to y_{ij}. Ignoring the sampling weights when using a unit level model can lead to a small area estimator that is not design-consistent. One way to achieve design consistency when using model-based estimators is by incorporating the design-consistent estimator in the model as a benchmarking constraint.

For the area level models, a common form of the benchmarking constraint is,

$$\sum_{i=1}^{\ell} \omega_i \hat{\overline{y}}_i = \sum_{i=1}^{\ell} \omega_i \overline{y}_i, \tag{1.7}$$

where ω_i is the area level sampling weights, $\hat{\overline{y}}_i$ are the small area predictors (i.e., model-based), and $\overline{y}_i$ are the direct area estimators (i.e., design-based). Wang, Fuller, and Qu (2008) used (1.7) as the restriction satisfied by their proposed EBLUP estimator for small area mean. In an alternative way of estimating the small area means that satisfy (1.7), they included the fixed survey weights ω_i in the constraint as part of the covariates in their augmented area level model.

It is reasonable to expect $\sum_{i=1}^{\ell} \sum_{j=1}^{N_i} y_{ij} = \sum_{i=1}^{\ell} N_i \hat{\overline{y}}_i$. Thus, because $\sum_{j=1}^{n_i} \omega_{ij} = N_i$, a possible benchmarking constraint is given by

$$\sum_{i=1}^{\ell} \sum_{j=1}^{N_i} y_{ij} = \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \omega_{ij} y_{ij}. \tag{1.8}$$

Note that both sides of (1.8) are random, thereby providing an internal benchmarking. For instance, in the left-hand side of (1.8), y_{ij}, $i = 1, \ldots, \ell$, $j = 1, \ldots, n_i$, are observed, and $y_{ij}, i = 1, \ldots, \ell$, $j = n_i + 1, \ldots, N_i$, are to be predicted using Bayesian predictive inference. If a simple random sampling design is assumed, one possible form of the sampling weight ω_{ij} is given by $\omega_{ij} = N/n$, for $i = 1, \ldots, \ell; j = 1, \ldots, n_i$. Thus, the

benchmarking constraint in (1.8) simplifies to

$$\frac{1}{N}\sum_{i=1}^{\ell}\sum_{j=1}^{N_i} y_{ij} = \frac{1}{n}\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} y_{ij}. \tag{1.9}$$

Note that (1.9) can be written as

$$\sum_{i=1}^{\ell} \frac{N_i \overline{Y}_i}{N} = \sum_{i=1}^{\ell} \frac{n_i \overline{y}_i}{n}.$$

Nandram and Toto (2010) used (1.9) as the benchmarking constraint in the nested-error regression model with unequal sampling variances.

However, note that $\overline{y}_s$ is not the only choice for an overall direct estimator of $\overline{Y}$. A practitioner could use other values as the direct estimator of $\overline{Y}$. For instance, one can use the constraint $\sum_{i=1}^{\ell}\sum_{j=1}^{N_i} y_{ij} = Na$, where a is a fixed known constant to be specified by the practitioner. The value of the benchmark constraint can be specified either by using a direct estimate of the finite population mean of the large area based on some auxiliary data or by determining a model that would give a value of a with relatively small standard error; for example, Nandram, Toto, and Choi (2010) used a fixed and known value of a. However, we focus on the benchmark constraints, where both sides of the equation are random. This avoids the uncertainty in the specification of a, and in fact, $\overline{y}_s$ is a design consistent estimator of $\overline{Y}$.

In Chapter 3, we incorporate q benchmarking constraints

$$\sum_{i=1}^{\ell}\sum_{j=1}^{N_i} y_{ijk} = \sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \omega_{ij} y_{ijk}, \; k = 1, 2, \ldots, q. \tag{1.10}$$

Once again in these q constraints, the ω_{ij} is the sampling weight of the jth unit in the ith population. Because the sampling weight are unit-based, the sampling weights are the same for all q benchmarking equations.

1.5   Applications

We apply our results to two illustrative examples. Our first example involves crop data based on the survey and satellite data conducted by the United States Department of Agriculture (USDA). The second example is from one of the large-scale national studies of the US government, the National Health and Nutrition Examination Survey (NHANES III), which was conducted by the Center for Disease Control, National Center for Health Statistics (CDC-NCHS). These applications show the diversity of the usage of the results of our study.

1.5.1   Corn and Soybeans Hectares Data

Knowledge of the area under different crops is important to the U.S. Department of Agriculture. Different surveys have been designed to estimate crop areas for large regions, like crop-reporting districts, individual states, and the whole United States. However, prediction of crop areas for small domains such as counties has been one of the problems of the National Agriculture Statistics Service (NASS) because of the lack of enough available data from farm surveys with small areas. In recent years, the United States Department of Agriculture (USDA) has been investigating the use of LANDSAT satellite data in association with farm-level survey observations, both to improve its estimates of crop area for crop-reporting districts and to develop estimates for individual counties (Battese, Harter, and Fuller, 1988).

Most of the corn grown in the United States comes from the Corn Belt, which consists of ten major states; Iowa is the largest corn-producing state in the U.S.. Iowa is also among the major states for soybean production. Producers in the U.S. feed most of the corn crop to cattle, hogs, sheep and poultry. The rest is used for processed food, industrial products such as cornstarch and plastic, renewable energy and ethanol. Corn is the U.S. largest crop, both in volume and value. Iowa has produced the largest corn crop over the current decade. In an average year Iowa produces more corn than most countries (e.g., three times as much corn as Argentina). Soybean and corn are grown in rotation in Iowa. Soybean is a healthy and rich source of protein for both animals and humans (e.g., tofu). Nearly all soybeans are processed into oils, many industrial products such as lubricants, solvents, cleaners and paints. Soybeans are also used for animal feed, biodiesel, cleaning product and candles. Thus, it is important to study the

production of corn and soybean in Iowa; in fact, corn and soybeans are the main products of the "Breadbasket" or "Grain Belt" of the United States. The "Breadbasket," also known as the Granary, of the country is a region which, because of richness of soil and/or advantageous climate, produces an agricultural surplus which is considered vital for the country as a whole. "Grain Belt" is an informal name for the prairie-region states across the Midwest that produces a substantial amount of the world's grain and soybean. This region includes the state of Iowa, from which our data is based upon. The National Agricultural Statistics Service (NASS) has an enormous amount of data, but these data are highly confidential and are kept under strict surveillance.

In this application, we are interested in predicting the mean crop (corn and soybeans) acreage for a number counties (small area) in north central Iowa using the enumerative survey and satellite data (see Battese, Harter, and Fuller, 1988). Corn and soybeans are important crops grown in the United States and they are of enormous support to the U.S. economy. Our variable of interest y_{ij} is the number of hectares of corn (or soybeans) in the jth segment of the ith county. This study used a simple random sampling within county (i.e., no weighting for individual sampled unit); we have taken the survey weight for sampled unit j in area i as $\omega_{ij} = N/n$, for all $i = 1, \ldots, \ell, j = 1, \ldots, n_i$. A strong linear relationship between the reported hectares of corn (soybeans) and the number of pixels of corn (soybean). Thus, we use a linear regression model to describe the relationship between the survey and satellite data and apply it to predict the mean crop area per segment in the sample counties. Table 1.1 presents the sampled crop data from the 1978 June Enumerative survey and LANDSAT satellite data. Table 1.1 is taken from Battese, Harter, and Fuller (1988). Table 1.1 provides information from $n = 37$ sampled segments (units) in the $\ell = 12$ counties. Observe that the sample size n_i $(i = 1, \ldots, \ell)$ within each county ranges from 1 to 6, while the total number of segments N_i (population size) within each county ranges from 395 to 965. The population mean pixels of corn and soybeans were also provided in the data.

1.5.2   Body Mass Index and Bone Mineral Density Data

The National Health and Nutrition Examination Survey (NHANES III) is one of the surveys used to assess an aspect of health of the United States population. The general structure of the NHANES III sample design is the same as that of the previous NHANES. Each of the surveys used a stratified, multi-stage probability design. The

NHANES III sample represented the total civilian, noninstitutionalized population, two months of age or over, in the 50 states and the District of Columbia in the United States. Two of the variables in this survey are body mass index (BMI) and bone mineral density (BMD), which are used as measures to diagnose obesity and osteoporosis respectively.

Obesity is one of the leading public health concern in the United States. Obesity is associated with increased health-care costs, reduced quality of life, and increased risk for premature death. Body mass index (BMI) is calculated from a person's weight and height. It provides a reasonable indicator of body fatness and weight categories that may lead to health problems. "Overweight" and "obese" are defined by the person's body mass index (BMI). An adult with a BMI between 25 and 30 is classified as overweight, while an adult with a BMI of 30 or greater is classified as obese. People who are overweight are at increased risk for diabetes, heart disease, stroke, high blood pressure, arthritis, and cancer. Reports in the last two decades showed that the prevalence of overweight and obesity in the United States remains high. There is little indication that the prevalence is decreasing in any subgroup of the population. Given the overall high prevalence of obesity, effective policies and environmental strategies that promote healthy eating and physical activity are needed for all populations and geographic areas, but particularly for those populations and areas disproportionately affected by obesity. This makes the design of the NHANES III and the study of obesity considerably interesting. See Nandram and Choi (2005, 2010) for extensive analysis of the BMI data.

On the other hand, osteoporosis is a disease that causes bones to become thin and weak, often resulting in fractures (broken bones). A broken bone can interfere with one's daily activities and can have serious consequences. For instance, once an elderly person suffered a fracture, it can negatively impact the standard of living and may lead to permanent disability. It has been found that the lower the bone mass, the greater the risk for fracture. Low bone mass, often called osteopenia, is a condition in which the bone density or bone thickness is lower than the average bone density of healthy adults of the same gender. The development of osteoporosis or osteopenia is a major concern as it is estimated that as many as four out of every ten women over the age of 50 in the USA will suffer a bone fracture, to the wrist, spine, or hip in their lifetime. Bone mineral density (BMD) is used to characterize osteoporosis. Women who suffer from osteopenia have a 1.8 fold increase in the likelihood of suffering a bone fracture

compared to a non-osteopenia person, whilst this increases to four times the risk if a person has osteoporosis. Research also showed that depression is a possible risk factor for low BMD. The World Health Organization (WHO) uses T-scores to define normal bone mass, low bone mass (or osteopenia), and osteoporosis. If a person has a BMD T-score between -1 and -2.5 then the person is diagnosed as having osteopenia, while a patient has a full blow osteoporosis if the BMD T-score is -2.5 or less. It is estimated that the risk of suffering a bone fracture doubles with every decrease in BMD T-score of 1.

In this example, we study the body mass index (BMI) and bone mineral density (BMD) of the adult domains in the state of New York. We define the adult domains as the Mexican-American gender domains of the population over 20 years of age in five counties of the state of New York with over $500,000$ people. We obtain the sample data from the Third National Health and Nutrition Examination Survey, (NHANES III, 1996). We have five counties and two gender groups (male, female). Hence, we get $\ell = 5 \times 2 = 10$ adult domains, representing the small areas for this example. We model BMI and BMD both as a linear function of age, although this is a very rough approximation for the age group. As in the crop data example, we note that there is a very large difference between the sample sizes and the population sizes. The sample size $n_i\ (i = 1, \ldots, \ell)$ within each domain ranges from 1 to 8, while the population size N_i within each youth domain ranges from 6 to 150. The information about the population sizes of the adult domains were obtained from the United States Decennial Census (2000). We used the information from the Census to also calculate the population total age by multiplying the population size with the median age of the different age groups. Table 1.2 gives the data for BMI and BMD. We include in Table 1.2 the information on the survey weights provided by NHANES III.

Table 1.1: Data for the crop data from Enumerative Survey and LANDSAT

County	Number of segments		Reported Hectares		Number of Pixel		Mean Pixel	
	Sample	County	Corn	Soybeans	Corn	Soybeans	Corn	Soybeans
1	1	545	165.78	8.09	374	55	295.29	189.70
2	1	566	96.32	106.03	209	218	300.40	196.65
3	1	394	76.08	103.60	253	250	289.60	205.28
4	2	424	185.35	6.47	432	96	290.74	220.22
			116.43	63.82	367	178		
5	3	564	162.08	43.50	361	137	318.21	188.06
			152.04	71.43	288	206		
			161.75	42.49	369	165		
6	3	570	92.88	105.26	206	218	257.17	247.13
			149.94	76.49	316	221		
			64.75	174.34	145	338		
7	3	402	127.07	95.67	355	128	291.77	185.37
			133.55	76.57	295	147		
			77.70	93.48	223	204		
8	3	567	206.39	37.84	459	77	301.26	221.36
			108.33	131.12	290	217		
			118.17	124.44	307	258		
9	4	687	99.96	144.15	252	303	262.17	247.09
			140.43	103.60	293	221		
			98.95	88.59	206	222		
			131.04	115.58	302	274		
10	5	569	114.12	99.15	313	190	314.28	198.66
			100.60	124.56	246	270		
			127.88	110.88	353	172		
			116.90	109.14	271	228		
			87.41	143.66	237	297		
11	5	965	93.48	91.05	221	167	298.65	204.61
			121.00	132.33	369	191		
			109.91	143.14	343	249		
			122.66	104.13	342	182		
			104.21	118.57	294	179		
12	6	556	88.59	102.59	220	262	325.99	177.05
			88.59	29.46	340	87		
			165.35	69.28	355	160		
			104.00	99.15	261	221		
			88.63	143.66	187	345		
			153.70	94.49	350	190		

See Battese, Harter, and Fuller (1988).

Table 1.2: Data for the adult domains in the state of New York from NHANES III and Census 2000

Domain ℓ	Sample Size, n	Population Size, N	BMI	BMD	Age	Survey Weights, ω
1	1	6	19.3	0.965	30	2178.11
2	1	112	26.8	0.983	39	4400.50
3	2	96	22.4	1.003	28	2913.42
			24.2	0.920	44	3061.85
4	8	150	20.1	0.653	20	1970.84
			22.0	0.971	20	1970.84
			25.1	0.900	20	1970.84
			22.5	0.880	22	1970.84
			21.4	0.806	26	1970.84
			24.6	1.097	29	1996.15
			25.4	1.206	34	1996.15
			26.4	0.809	35	2097.44
5	4	39	28.5	1.003	23	2320.99
			21.9	1.400	20	2274.38
			20.3	0.916	23	2513.96
			21.3	1.032	28	2274.38
6	2	6	21.3	0.727	24	3524.08
			24.1	0.948	30	1872.93
7	2	118	24.5	0.934	20	3609.89
			28.8	0.792	34	3609.89
8	5	101	23.6	0.834	23	2389.98
			19.5	0.856	26	2389.98
			20.0	0.726	28	2396.26
			19.3	0.646	34	2396.26
9	3	142	24.5	0.822	36	2322.29
			34.9	0.923	23	1952.89
			25.3	0.875	35	1952.89
			23.2	0.609	48	1605.64
10	2	37	23.1	0.865	28	1865.75
			29.1	0.890	29	2062.29

See NHANES III (1996).

Chapter 2

Univariate Benchmarking

In this chapter, we discuss the univariate Bayesian benchmarking model. We have introduced our Bayesian version of the Battese, Harter, and Fuller (1988) nested-error regression model. With the benchmarking constraint, our model would include both covariates and survey weights at the unit-level. We show in this chapter how to incorporate the constraint into the model. Given the sampled data, we use Bayesian predictive inference to estimate the finite population means of small areas. We analyze the two examples and perform some simulation studies to characterize the resulting model. Let us begin with some basic notations we will use.

We denote by $\underset{\sim}{1}_m$ a column vector of size m with each of its elements being unity, the identity matrix of size m by I_m and the $r \times s$ matrix with each of its elements being unity by J_s^r. If $r = s$, we write J_r. Also, let Z_s be the $n \times \ell$ block diagonal matrix with $\underset{\sim}{1}_{n_i}$ on the ith diagonal, $i = 1, 2, \ldots, \ell$, Z_{ns} be the $(N-n) \times \ell$ block diagonal matrix with $\underset{\sim}{1}_{N_i - n_i}$ on the ith diagonal, $i = 1, 2, \ldots, \ell$, and $Z_{ns_{(N)}}$ as the $(N-1-n) \times \ell$ corresponding block diagonal matrix with $\underset{\sim}{1}_{N_i - n_i}$ on the ith diagonal entry, $i = 1, 2, \ldots, \ell - 1$, and $\underset{\sim}{1}_{N_\ell - n_\ell - 1}$ on the ℓth diagonal entry.

We add benchmarking constraint (1.8) onto our Bayesian version of the nested error regression model of Battese, Harter, and Fuller (1988).

2.1 Univariate Nested-Error Regression Model

The Bayesian benchmarking model (BM) is

$$y_{ij}|\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2 \overset{ind}{\sim} \text{Normal}\left\{x_{ij}'\underset{\sim}{\beta} + \nu_i, \sigma^2\right\}, \qquad \sum_{i=1}^{\ell}\sum_{j=1}^{N_i} y_{ij} = \sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\omega_{ij}y_{ij}, \qquad (2.1)$$

$$\nu_i|\sigma^2, \rho \overset{iid}{\sim} \text{Normal}\left\{0, \left(\frac{\rho}{1-\rho}\right)\sigma^2\right\}, \qquad 0 < \rho < 1, \qquad (2.2)$$

$$p\left(\underset{\sim}{\beta}, \sigma^2, \rho\right) \propto \frac{1}{\sigma^2}, \qquad \sigma^2 > 0. \qquad (2.3)$$

Conditional on the benchmarking constraint, we derive the distribution of the data. Let $\underset{\sim}{y}_{(N)} = \left(\underset{\sim}{y}_s', \underset{\sim}{y}_{ns_{(N)}}'\right)'$, where $\underset{\sim}{y}_{ns_{(N)}}$ is the vector of $N - n - 1$ unobserved y_{ij} values excluding the N_ℓ^{th} unit y_{ℓ,N_ℓ} from the ℓth area. That is, $\underset{\sim}{y}_{ns_{(N)}} = \left(\underset{\sim}{y}_{ns_1}', \ldots, \underset{\sim}{y}_{ns_\ell(N)}'\right)'$ with $\underset{\sim}{y}_{ns_i} = (y_{i,n_i+1}, \ldots, y_{i,N_i})'$ for $i = 1, 2, \ldots, \ell - 1$, and $\underset{\sim}{y}_{ns_\ell(N)} = (y_{\ell,n_\ell+1}, \ldots, y_{\ell,N_\ell-1})'$. Also, anticipating a transformation of $y_{\ell N_\ell}$, let $\phi = \sum_{i=1}^{\ell}\sum_{j=1}^{N_i} y_{ij} - \sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\omega_{ij}y_{ij}$. Furthermore, define $\mu_{ij} = x_{ij}'\underset{\sim}{\beta} + \nu_i$, for $i = 1, \ldots, \ell, \quad j = 1, \ldots, N_i$, and write $\underset{\sim}{\mu} = \left(\underset{\sim}{\mu}_{(N)}', \mu_{\ell,N_\ell}\right)'$, where $\underset{\sim}{\mu}_{(N)} = \left(\underset{\sim}{\mu}_s', \underset{\sim}{\mu}_{ns(N)}'\right)'$. We write the $n \times 1$ vector of sampling weights as $\underset{\sim}{\omega} = (\omega_{11}, \ldots, \omega_{\ell n_\ell})$. Lastly, write $\underset{\sim}{\omega}^* = \left(\underset{\sim}{\omega} - \underset{\sim}{1}_n\right)$, $\underset{\sim}{a} = \left(\underset{\sim}{\omega}^{*'}, -\underset{\sim}{1}_{N-n-1}'\right)'$, and

$$D = \left(\frac{1}{\underset{\sim}{\omega}^{*'}\underset{\sim}{\omega}^* + \underset{\sim}{\omega}^{*'}\underset{\sim}{1}_n}\right)\begin{bmatrix} (\underset{\sim}{\omega}^*\underset{\sim}{\omega}^{*'}) & -\left(\underset{\sim}{\omega}^*\underset{\sim}{1}_{(N-1-n)}'\right) \\ -\left(\underset{\sim}{1}_{(N-1-n)}\underset{\sim}{\omega}^{*'}\right) & J_{(N-1-n)} \end{bmatrix}.$$

Note that we can also write the matrix D in terms of the vector $\underset{\sim}{a}$ as follows

$$D = \frac{\underset{\sim}{a}\underset{\sim}{a}'}{1 + \underset{\sim}{a}'\underset{\sim}{a}}.$$

Lemma 2.1. Under the benchmarking model, conditional on the benchmarking constraint,

$$\underset{\sim}{y}_{(N)}|\underset{\sim}{\mu}, \sigma^2, \phi = 0 \sim \text{Normal}\left\{\underset{\sim}{\mu}_{(N)} + \left(\frac{1}{\underset{\sim}{\omega}^{*'}\underset{\sim}{\omega}^* + \underset{\sim}{\omega}^{*'}\underset{\sim}{1}_n}\right)\left(\mu_{\ell,N_\ell} - \underset{\sim}{a}'\underset{\sim}{\mu}_{(N)}\right)\underset{\sim}{a}, \sigma^2\left(I_{N-1} - D\right)\right\},$$

and

$$y_{\ell,N_\ell} = \underset{\sim}{\omega}^{*'}\underset{\sim}{y}_s - \underset{\sim}{1}_{N-n-1}'\underset{\sim}{y}_{ns_{(N)}}.$$

Note that the **multivariate normal distribution** is a generalization of the one-dimensional (univariate) normal distribution to higher dimension. A random vector $X = (X_1, X_2, \ldots, X_k)$ has a multivariate normal distribution if the support of X is the entire space R^k and there exist a $k-$vector μ and a positive-definite $k \times k$ matrix Σ such that the probability density function of X can be expressed as

$$f_X(x) = \frac{1}{2\pi^{k/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x - \mu)'\Sigma^{-1}(x - \mu)\right\},$$

where $|\Sigma|$ is the determinant of Σ. This expression reduces to the density of the univariate normal distribution if Σ is a scalar (i.e., a 1×1 matrix).

Proof.

First, consider the slightly simpler notation in which $Z_k|\alpha_k, \sigma^2 \overset{ind}{\sim} \text{Normal}(\alpha_k, \sigma^2)$, $k = 1, 2, \ldots, N$, subject to the constraint $\sum_{k=1}^{N} z_k = \sum_{k=1}^{n} \omega_k z_k$. Make the transformation $\phi = \sum_{k=1}^{N} z_k - \sum_{k=1}^{n} \omega_k z_k$ with $Z_1, \ldots, Z_{N-1}$ untransformed. We can show that the jacobian is 1, and the joint distribution is given by

$$p\left(z_1, \ldots, z_{N-1}, \phi | \alpha, \sigma^2\right) =$$

$$\left(\frac{1}{2\pi\sigma^2}\right)^{N/2} \exp\left(-\frac{1}{2\sigma^2}\left\{\sum_{k=1}^{N-1}(z_k - \alpha_k)^2 + \left[\left(\phi - \sum_{k=1}^{N-1} z_k + \sum_{k=1}^{n} \omega_k z_k\right) - \alpha_N\right]^2\right\}\right),$$

where $z_N = \phi - \sum_{k=1}^{N-1} z_k + \sum_{k=1}^{n} \omega_k z_k$. But since our distribution is subject to the constraint,

$$\sum_{k=1}^{N} z_k = \sum_{k=1}^{n} \omega_k z_k, \phi = z_N + \sum_{k=1}^{N-1} z_k - \sum_{k=1}^{n} \omega_k z_k = 0. \text{ Equivalently, } z_N = \sum_{k=1}^{n} \omega_k z_k - \sum_{k=1}^{N-1} z_k.$$

Therefore, the joint density is $p\left(z_1, \ldots, z_{N-1}, \phi = 0 | \alpha, \sigma^2\right)$. There is some simplicity when we normalize this density over $(z_1, \ldots, z_{N-1})$ to get $p\left(z_1, \ldots, z_{N-1} | \phi = 0, \alpha, \sigma^2\right)$, which is the joint density we use. Hence, conditional on the benchmarking constraint, we have the distribution

$$Z_{(N)}|\alpha, \sigma^2, \phi = 0 \sim \text{Normal}\left\{\alpha_{(N)} + \left(\frac{1}{\omega^{*'}\omega}\right)\left(\alpha_N - \alpha'\alpha_{(N)}\right)\alpha, \sigma^2\left(I_{N-1} - D\right)\right\},$$

where $\underset{\sim}{\alpha}_{(N)} = (\alpha_1, \ldots, \alpha_{N-1})'$. Now, match the original vector $\underset{\sim}{y} = \left(\underset{\sim}{y_s}', \underset{\sim}{y_{ns}}'\right)'$ with $(z_1, \ldots, z_n, z_{n+1}, \ldots, z_N)'$, as well as $\underset{\sim}{\mu}$ with $\underset{\sim}{\alpha} = \left(\underset{\sim}{\alpha}_{(N)}, \alpha_N\right)$, to obtain the desired result.

∎

Lemma 2.1 is a simple result that incorporates the benchmark constraint into the distribution of the data given the parameters. From the resulting density in Lemma 2.1, the addition of the benchmark constraint into the model causes the y_{ij}'s to be correlated. It is interesting that the sample values are negatively correlated and nonsample values are also negatively correlated but a sample value is positively correlated with a nonsample value; this will reduce the variance of the sum of the nonsample values. Theorem 2.4 has the conditional distribution of $\underset{\sim}{y}_{ns_{(N)}}$ given $\underset{\sim}{y}_s$ and other parameters. Note that y_{ℓ, N_ℓ} inherits its distributional properties from $\underset{\sim}{y}_{(N)}|\mu, \sigma^2, \phi = 0$. Henceforth, we drop the notation $\phi = 0$, but it must be understood that all distributions are conditional on $\phi = 0$.

Remark 2.2. Applying the marginal distribution property of the multivariate normal density in Lemma 2.1, we find that $\underset{\sim}{y}_s|\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2$ follows a Normal distribution with

$$\mathrm{E}\left[\underset{\sim}{y}_s|\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2\right] = C_x\underset{\sim}{\beta} + C_z\underset{\sim}{\nu} \text{ and } \mathrm{Var}\left[\underset{\sim}{y}_s|\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2\right] = \sigma^2\left[I_n - W\right],$$

where

$$W = \frac{\underset{\sim}{\omega}^*\underset{\sim}{\omega}^{*\prime}}{\underset{\sim}{\omega}^{*\prime}\underset{\sim}{\omega}^* + \underset{\sim}{\omega}^{*\prime}1_n},$$

$$C_x = \left[I_n - W\right]X_s + \left(\frac{\underset{\sim}{\omega}^*}{\underset{\sim}{\omega}^{*\prime}\underset{\sim}{\omega}^* + \underset{\sim}{\omega}^{*\prime}1_n}\right)\left(1'_{N-n}X_{ns}\right),$$

and

$$C_z = \left[I_n - W\right]Z_s + \left(\frac{\underset{\sim}{\omega}^*}{\underset{\sim}{\omega}^{*\prime}\underset{\sim}{\omega}^* + \underset{\sim}{\omega}^{*\prime}1_n}\right)\left(1'_{N-n}Z_{ns}\right).$$

From Remark 2.2, the likelihood function is given by

$$p\left(\underset{\sim}{y}_s|\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2\right) \propto$$

$$\left(\frac{1}{\sigma^2}\right)^{n/2}\exp\left\{-\frac{1}{2\sigma^2}\left[\left(\underset{\sim}{y}_s - \left[C_x\underset{\sim}{\beta} + C_z\underset{\sim}{\nu}\right]\right)'\left[I_n - W\right]^{-1}\left(\underset{\sim}{y}_s - \left[C_x\underset{\sim}{\beta} + C_z\underset{\sim}{\nu}\right]\right)\right]\right\}.$$

Now, applying Bayes' theorem on $p\left(\underset{\sim}{y}_s|\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2\right)$ with the distribution of $\underset{\sim}{\nu}|\sigma^2, \rho$ in (2.2),

and prior distribution given in (2.3), we get the joint posterior density

$$\pi\left(\underset{\sim}{\nu},\underset{\sim}{\beta},\sigma^2,\rho|y_s\right) \propto \left(\frac{1}{\sigma^2}\right)^{1+(n+\ell)/2}\left(\frac{1-\rho}{\rho}\right)^{\ell/2}$$

$$\times \exp\left\{-\frac{1}{2\sigma^2}\left(\frac{1-\rho}{\rho}\right)\left[\left(\underset{\sim}{\nu}-B^{-1}C_z'A\left(y_s-C_x\underset{\sim}{\beta}\right)\right)'B\left(\underset{\sim}{\nu}-B^{-1}C_z'A\left(y_s-C_x\underset{\sim}{\beta}\right)\right)\right]\right\}$$

$$\times \exp\left\{-\frac{1}{2\sigma^2}\left(\frac{1-\rho}{\rho}\right)\left[\left(y_s-C_x\underset{\sim}{\beta}\right)'\left(A^{-1}+C_zC_z'\right)^{-1}\left(y_s-C_x\underset{\sim}{\beta}\right)\right]\right\},$$

where $A=\left[\left(\frac{1-\rho}{\rho}\right)(I_n-W)\right]^{-1}$ and $B=I_\ell+C_z'AC_z$.

Lemma 2.3. Assuming that $0<\rho<1$ and the design matrix is full rank, under the benchmarking model the joint posterior density $\pi\left(\underset{\sim}{\nu},\underset{\sim}{\beta},\sigma^2,\rho|y_s\right)$ is proper.

Lemma 2.3 shows that the addition of the benchmarking constraint to the model does not affect the propriety of the joint posterior distribution. To prove Lemma 2.3, we use the multiplication rule

$$\pi\left(\underset{\sim}{\nu},\underset{\sim}{\beta},\sigma^2,\rho|y_s\right) = \pi_1\left(\underset{\sim}{\nu}|\underset{\sim}{\beta},\sigma^2,\rho,y_s\right)\pi_2\left(\underset{\sim}{\beta}|\sigma^2,\rho,y_s\right)\pi_3\left(\sigma^2|\rho,y_s\right)\pi_4\left(\rho|y_s\right),$$

and show that $\pi_1\left(\underset{\sim}{\nu}|\underset{\sim}{\beta},\sigma^2,\rho,y_s\right)$, $\pi_2\left(\underset{\sim}{\beta}|\sigma^2,\rho,y_s\right)$, $\pi_3\left(\sigma^2|\rho,y_s\right)$, and $\pi_4\left(\rho|y_s\right)$ are all proper densities.

Proof.
Recall that the joint posterior density is

$$\pi\left(\underset{\sim}{\nu},\underset{\sim}{\beta},\sigma^2,\rho|y_s\right) \propto \left(\frac{1}{\sigma^2}\right)^{1+(n+\ell)/2}\left(\frac{1-\rho}{\rho}\right)^{\ell/2}$$

$$\times \exp\left\{-\frac{1}{2\sigma^2}\left(\frac{1-\rho}{\rho}\right)\left[\left(\underset{\sim}{\nu}-B^{-1}C_z'A\left(y_s-C_x\underset{\sim}{\beta}\right)\right)'B\left(\underset{\sim}{\nu}-B^{-1}C_z'A\left(y_s-C_x\underset{\sim}{\beta}\right)\right)\right]\right\}$$

$$\times \exp\left\{-\frac{1}{2\sigma^2}\left(\frac{1-\rho}{\rho}\right)\left[\left(y_s-C_x\underset{\sim}{\beta}\right)'\left(A^{-1}+C_zC_z'\right)^{-1}\left(y_s-C_x\underset{\sim}{\beta}\right)\right]\right\}.$$

Taking $0<\rho<1$, it is easy to show that $A=\left[\left(\frac{1-\rho}{\rho}\right)(I_n-W)\right]^{-1}$ is positive definite.

Therefore, $B = I_\ell + C_z' A C_z$ is positive definite. First, it can be shown that

$$\nu | \beta, \sigma^2, \rho, y_s \sim \text{Normal}\left(B^{-1} C_z' A \left(y_s - C_x \beta \right), \sigma^2 \left(\frac{\rho}{1-\rho} \right) B^{-1} \right). \qquad (2.4)$$

Integrating out ν from $\pi \left(\nu, \beta, \sigma^2, \rho | y_s \right)$, we get

$$\pi \left(\beta, \sigma^2, \rho | y_s \right) \propto \left(\frac{1}{\sigma^2} \right)^{1+n/2} \left(\frac{1}{|B|} \right)^{1/2} \exp\left\{ -\frac{1}{2\sigma^2} \left(\frac{1-\rho}{\rho} \right) \left[\left(\beta - \widehat{\beta} \right)' V \left(\beta - \widehat{\beta} \right) \right] \right\}$$

$$\times \exp\left\{ -\frac{1}{2\sigma^2} \left(\frac{1-\rho}{\rho} \right) \left[y_s' \left(D - D C_x V^{-1} C_x' D \right) y_s \right] \right\},$$

and it can be shown that

$$\beta | \sigma^2, \rho, y_s \sim \text{Normal}\left\{ \widehat{\beta}, \sigma^2 \left(\frac{\rho}{1-\rho} \right) V^{-1} \right\}, \qquad (2.5)$$

where $\widehat{\beta} = V^{-1} C_x' D y_s$, $D = \left(A^{-1} + C_z C_z' \right)^{-1}$, and $V = C_x' D C_x$. Then, integrating out β from $\pi \left(\beta, \sigma^2, \rho | y_s \right)$, we have

$$\pi \left(\sigma^2, \rho | y_s \right) \propto \left(\frac{1}{\sigma^2} \right)^{1+(n-p)/2} \left(\frac{\rho}{1-\rho} \right)^{p/2} \left(\frac{1}{|V|} \right)^{1/2} \left(\frac{1}{|B|} \right)^{1/2}$$

$$\times \exp\left\{ -\frac{1}{2\sigma^2} \left(\frac{1-\rho}{\rho} \right) \left[y_s' \left(D - D C_x V^{-1} C_x' D \right) y_s \right] \right\},$$

and it can be shown that

$$\sigma^{-2} | \rho, y_s \sim \text{Gamma}\left(\frac{n-p}{2}, \frac{G}{2} \right), \qquad (2.6)$$

where $G = \left(\frac{1-\rho}{\rho} \right) \left[y_s' \left(D - D C_x V^{-1} C_x' D \right) y_s \right]$.

Finally, integrating out σ^2 from $\pi \left(\sigma^2, \rho | y_s \right)$, we get

$$\pi_4 \left(\rho | y_s \right) \propto \left(\frac{\rho}{1-\rho} \right)^{n/2} (|V|)^{-1/2} (|B|)^{-1/2} \left[y_s' \left(D - D C_x V^{-1} C_x' D \right) y_s \right]^{-\left(\frac{n-p}{2} \right)}. \qquad (2.7)$$

To complete the proof, we showed that $\pi_4 \left(\rho | y_s \right)$ is proper for all ρ, $0 < \rho < 1$. Because A is a function of ρ, the matrices B and D are functions of ρ. The distribution of $\pi_4 \left(\rho | y_s \right)$

is proper as long as ρ is set to a value strictly between 0 and 1, say $0.0001 < \rho < 0.9999$. ■

To be able to make inference for $\overline{Y}_i$ $(i = 1, \ldots, \ell)$, given the sample observations y_s, a possible approach is to sample from the distribution $\overline{Y}_i | y_s$. Theorem 2.4 below gives us the Bayesian predictive distribution of $\overline{Y}_i | y_s$.

Theorem 2.4. The distribution of $\overline{Y}_i | y_s$ is

$$p\left(\overline{Y}_i | y_s\right) = \int p\left(\overline{Y}_i | \nu, \beta, \sigma^2, y_s\right) \pi\left(\nu, \beta, \sigma^2 | y_s\right) d\nu d\beta d\sigma^2, \qquad (2.8)$$

where under the benchmarking model,

$$\overline{Y}_i | \nu, \beta, \sigma^2, y_s$$

$$\sim \text{Normal}\left\{ f_i \overline{y}_{s_i} + (1 - f_i) \overline{x}_{ns_i}{}'\beta + (1 - f_i) \nu_i + A_i, \frac{\sigma^2}{N_i}(1 - f_i) V_i \right\}, \qquad (2.9)$$

where $A_i = \left(\dfrac{1 - f_i}{N - n}\right)\left[\omega^{*\prime} y_s - \left(1'_{N-n} X_{ns}\right)\beta - \left(1'_{N-n} Z_{ns}\right)\nu\right]$ and $V_i = 1 - \left(\dfrac{N_i - n_i}{N - n}\right), i = 1, \ldots, \ell.$

Proof.

Note that $\overline{Y}_i$ can be written as

$$\overline{Y}_i = \frac{1}{N_i}\left[n_i \overline{y}_{s_i} + (N_i - n_i) \overline{y}_{ns_i}\right] = \frac{1}{N_i}\left(1'_{n_i} y_{s_i} + 1'_{N_i - n_i} y_{ns_i}\right), \ i = 1, 2, \ldots, \ell.$$

Hence, to prove Theorem 2.4, we need the distribution of y_s and y_{ns}. Using the conditional distribution property of the multivariate normal density on Lemma 2.1, we find that $y_{ns_{(N)}} | \nu, \beta, \sigma^2, y_s$ also follows a normal distribution with

$$\text{E}\left[y_{ns_{(N)}} | \nu, \beta, \sigma^2, y_s\right] =$$

$$\left(\frac{1}{N - n}\right) 1_{N-1-n} \omega^{*\prime} y_s + \left[X_{ns_{(N)}} - \left(\frac{1}{N - n}\right) J_{N-n}^{N-1-n} X_{ns}\right]\beta$$

$$+ \left[Z_{ns_{(N)}} - \left(\frac{1}{N - n}\right) J_{N-n}^{N-1-n} Z_{ns}\right]\nu,$$

and
$$\text{Var}\left[\underset{\sim}{y}_{ns_{(N)}}|\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2, \underset{\sim}{y}_s\right] = \sigma^2\left[I_{N-1-n} - \left(\frac{1}{N-n}\right)J_{N-1-n}\right].$$

The distribution of $\overline{Y}_i|\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2, \underset{\sim}{y}_s$ follows from these results and the following expression for $\overline{Y}_\ell$ from the benchmarking constraint $\overline{Y}_\ell = \frac{1}{N_\ell}\left(\underset{\sim}{1}'_{n_\ell}\underset{\sim}{y}_{s\ell} + \underset{\sim}{\omega}^{*'}\underset{\sim}{y}_s - \sum_{i=1}^{\ell-1}\underset{\sim}{1}'_{N_i-n_i}\underset{\sim}{y}_{ns_i}\right)$.

Note that the distribution of $\overline{Y}_i|\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2, \underset{\sim}{y}_s$ does not depend on ρ. From Lemma 2.3, $\pi_4\left(\rho|\underset{\sim}{y}_s\right)$ is proper. Thus, $\pi\left(\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2|\underset{\sim}{y}_s\right) = \int \pi\left(\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2, \rho|\underset{\sim}{y}_s\right)\mathrm{d}\rho$ is proper. $\blacksquare$

Thus, we can compare the two distributions of $\overline{Y}_i|\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2, \underset{\sim}{y}_s$ without the benchmarking constraint in (1.5) and with the benchmarking constraint in (2.9) to observe the effects of the benchmarking constraint on the distribution of the finite population mean. First, we see an overall adjustment $A_i, i = 1, 2, \ldots, \ell$, to the expected value in the distribution of $\overline{Y}_i|\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2, \underset{\sim}{y}_s$ under the benchmarking model, with respect to the nonbenchmarking model. Note that unlike Nandram, Toto, and Choi (2010), where the overall adjustment is constant throughout all areas, here the adjustment A_i is different for each small area, since f_i depends on the sample and population sizes in each area; covariates are not included in Nandram, Toto, and Choi (2010). Furthermore, observe that in the benchmarking model, the variance of the finite population mean in each small area is reduced by different amounts $V_i, i = 1, 2, \ldots, \ell$.

2.2 Posterior Inference

We use Bayesian predictive procedure to make posterior inference about $\overline{Y}_i|\underset{\sim}{y}_s$. Our approach is to combine samples from the posterior distribution $\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2, \rho|\underset{\sim}{y}_s$, and the distribution of $\overline{Y}_i|\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2, \underset{\sim}{y}_s$ via Theorem 2.4.

The proof of propriety of the joint posterior density provides a prescription of how to draw samples from the posterior density. Note that by the multiplication rule,

$$\pi\left(\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2, \rho|\underset{\sim}{y}_s\right) = \pi_1\left(\underset{\sim}{\nu}|\underset{\sim}{\beta}, \sigma^2, \rho, \underset{\sim}{y}_s\right)\pi_2\left(\underset{\sim}{\beta}|\sigma^2, \rho, \underset{\sim}{y}_s\right)\pi_3\left(\sigma^2|\rho, \underset{\sim}{y}_s\right)\pi_4\left(\rho|\underset{\sim}{y}_s\right).$$

Since the conditional densities $\pi_1\left(\underset{\sim}{\nu}|\underset{\sim}{\beta}, \sigma^2, \rho, \underset{\sim}{y}_s\right),\ \pi_2\left(\underset{\sim}{\beta}|\sigma^2, \rho, \underset{\sim}{y}_s\right)$ and $\pi_3\left(\sigma^2|\rho, \underset{\sim}{y}_s\right)$ are known distributions [see (2.4), (2.5) and (2.6)], we can draw samples from $\pi\left(\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2, \rho|\underset{\sim}{y}_s\right)$ using the composition method. We would first draw samples from $\rho|\underset{\sim}{y}_s$ [see (2.7)]using a

grid method. With each sample of ρ, we draw a sample from $\sigma^{-2}|\rho, \underset{\sim}{y}_s$. Then, with the resulting pair of values of ρ and σ^2, we draw a sample from $\underset{\sim}{\beta}|\sigma^2, \rho, \underset{\sim}{y}_s$. Also, with the sampled values of ρ, σ^2, and $\underset{\sim}{\beta}$, draw values of $\underset{\sim}{\nu}$ from $\underset{\sim}{\nu}|\underset{\sim}{\beta}, \sigma^2, \rho, \underset{\sim}{y}_s$. Once the parameters are obtained, we use them to draw values of the finite population means for each of the areas from the distribution of $\overline{Y}_i|\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2, \underset{\sim}{y}_s$; see Theorem 2.4. We apply this algorithm to draw samples in both the nonbenchmarking and benchmarking models.

To facilitate inference, we generate $M = 10,000$ iterates from the sampling method. We use 200 grids in $(0, 1)$ to approximate $\pi_4\left(\rho|\underset{\sim}{y}_s\right)$. This is a reasonable approximation because the width of each grid is 0.005. It is convenient to use the grid method because ρ is bounded in the interval $(0, 1)$, and the function $\pi_4\left(\rho|\underset{\sim}{y}_s\right)$ is easy to compute for each ρ. It is interesting that we do not need to use Markov chain Monte Carlo (MCMC) to fit the nested error regression model, with or without benchmarking.

2.3 Applications

In this section, we present two illustrative examples to compare the benchmarking model with its nonbenchmarking counterpart. We summarize the results using the posterior mean (PM), posterior standard deviation (PSD), and the 95% credible interval (Int) of the finite population mean of each area. In our first example, we apply our model to estimate the corn and soybeans hectares from the 1978 Enumerative survey and satellite data. We model the number of hectares of corn (or soybeans) y_{ij} using the nested error regression model of Battese, Harter, and Fuller (1988) with two covariates ($p = 3$). That is, $\underset{\sim}{x}_{ij} = (1, x_{ij1}, x_{ij2})'$, where x_{ij1} is the number of pixels of corn, and x_{ij2} is the number of pixels of soybeans. Information from $n = 37$ sampled segments (units) in the $\ell = 12$ counties are available. The sample size n_i $(i = 1, \ldots, \ell)$ within each county ranges from 1 to 6, while the total number of segments N_i (population size) within each county ranges from 395 to 965. The population mean pixels of corn and soybeans were also provided in the data.

In our second example, we estimate the body mass index (BMI) and bone mineral density (BMD) of the adult domains (gender groups of Mexican American population over 20 years). Note that for the enumerative survey and satellite data the design is self-weighting within areas, but for the BMI and BMD, there are differential probabilities of selection (i.e., varying sampling weights). See Section 1.5 to learn more about the

background of these two examples.

2.3.1   Corn and Soybeans Hectares Data

Table 2.1 shows that the posterior means from the two models are mostly the same, with a slight shifting in the posterior means from NBM to BM. These estimates from NBM are similar to those in Battese, Harter, and Fuller (1988). Notice too that there is a slight increase in precision in BM, relative to NBM. In Table 2.1, the posterior standard deviations (PSD) in BM for most counties in both crops (corn and soybeans) are smaller than the PSDs in NBM; with the exception of the PSD of County 12 for soybeans in the NBM. These decreases in the PSDs are also reflected in the 95% credible intervals in Table 2.1. However, notice that the gains in precision observed are not significantly large. There are even few counties or domains where the NBM has smaller posterior SD than BM. If the values of n_i were more comparable to N_i, larger gains in precision in the benchmarking model could be expected. From Table 2.1, the means for the corn (soybeans) data are $N^{-1} \sum_{i=1}^{\ell} N_i \overline{y}_i = 120.32(95.35) = N^{-1} \sum_{i=1}^{\ell} N_i PM_{i(BM)}$, but $N^{-1} \sum_{i=1}^{\ell} N_i PM_{i(NBM)} = 119.67(96.87)$. That is, the value of the overall posterior mean from BM is equal to the sample mean of the data, but that of NBM is not equal. This observation reflects the benchmarking constraint incorporated into BM. To further assess the performance of BM over NBM, we looked at the distance between the direct estimators $\overline{y}_i$ and the PM_i from both NBM and BM. As a summary measure, we computed the shrinkage $SHR = \sqrt{\sum_{i=1}^{\ell} (PM_i - \overline{y}_i)^2}$. For the corn data, $SHR_{(NBM)} = 81.05$ and $SHR_{(BM)} = 80.84$. On the other hand, for the soybeans data, $SHR_{(NBM)} = 92.81$ and $SHR_{(BM)} = 91.03$. For both corn and soybeans data, we find that BM has a slightly lower shrinkage than NBM, indicating that BM is slightly more design-consistent than NBM. We use box plots to show the general features in Figure 2.1.

Table 2.1: Comparison of NBM and BM using posterior means (PM) and the posterior standard deviations (PSD) and 95% credible intervals (Int) for the finite population mean $\left(\overline{Y}\right)$ by county (area) of the corn and soybeans hectares data

				Nonbenchmarking			Benchmarking		
				Posterior		Credible	Posterior		Credible
County	Sample Size	Population Size	Direct Estimate	Mean	SD	Interval	Mean	SD	Interval
				Corn					
1	1	545	165.78	123.47	9.32	(105.72,143.26)	124.04	8.32	(108.40,142.34)
2	1	566	96.32	124.20	9.28	(106.42,143.61)	124.89	8.22	(109.09,142.29)
3	1	394	76.08	110.95	10.04	(88.68,128.66)	111.68	9.34	(90.38,127.25)
4	2	424	150.89	114.16	8.37	(96.38,130.14)	114.74	7.83	(97.84,129.33)
5	3	564	158.62	138.82	8.37	(123.69,156.10)	139.31	7.95	(125.88,156.42)
6	3	570	102.52	109.78	7.66	(95.60,126.05)	110.48	6.83	(98.04,124.89)
7	3	402	112.77	116.05	7.20	(101.02,129.52)	116.51	6.72	(102.67,129.35)
8	3	567	144.30	122.90	7.20	(108.76,137.60)	123.46	6.49	(110.69,136.80)
9	4	687	117.59	112.07	7.00	(98.78,126.74)	112.74	6.33	(101.27,126.04)
10	5	569	109.38	123.99	6.25	(111.29,136.04)	124.55	5.91	(112.29,135.69)
11	5	965	110.25	111.71	6.96	(97.67,124.48)	112.36	6.47	(98.79,123.43)
12	6	556	114.81	131.25	5.92	(119.35,142.97)	131.69	5.69	(120.05,143.14)
				Soybeans					
1	1	545	8.09	78.76	11.27	(55.99,100.120)	77.31	10.33	(56.75,97.29)
2	1	566	106.03	94.34	10.92	(72.74,115.87)	92.77	9.98	(73.61,112.99)
3	1	394	103.60	87.71	10.70	(66.48,108.55)	86.20	10.19	(65.32,105.61)
4	2	424	35.15	82.04	10.09	(62.09,101.88)	80.66	9.39	(62.03,99.08)
5	3	564	52.47	67.15	7.93	(51.97,83.16)	65.46	7.51	(50.53,80.26)
6	3	570	118.70	113.83	7.34	(99.71,128.24)	112.32	6.91	(98.91,125.91)
7	3	402	88.57	97.23	7.63	(82.15,112.27)	95.95	7.41	(81.48,110.43)
8	3	567	97.80	111.93	7.60	(96.72,126.95)	110.41	7.20	(96.08,124.70)
9	4	687	112.98	110.06	6.54	(97.18,123.09)	108.39	6.16	(96.46,120.66)
10	5	569	117.48	100.36	6.13	(88.15,112.39)	98.97	5.98	(86.97,110.69)
11	5	965	117.84	118.28	6.48	(105.02,130.66)	116.73	6.07	(104.26,128.30)
12	6	556	89.77	75.04	**5.65**	(63.85,85.98)	73.53	**5.75**	(62.33,84.94)

2.3.2 Body Mass Index and Bone Mineral Density Data

The Third National Health and Nutrition Examination Survey, (NHANES III, 1996), has a complex sampling design and there are survey weights. The sample weights are calibrated such that $\sum_{j=1}^{n_i} \omega_{ij} = N_i, \ i = 1, \ldots, \ell$. To see the importance of taking into consideration the sampling weights in small area estimation, we will fit the benchmarking model to the BMI and BMD data without survey weights. that is, we implement the use of the benchmarking constraint in 1.9. This model will be called the self-weighting benchmarking model (SBM). Thus, we will apply to the BMI and BMD data the three models (NBM, SBM, BM). Note that while BM contains survey weights, SBM and NBM do not.

Table 2.2 shows that the posterior means (PM) of the benchmarking models

(both BM and SBM) are mostly smaller than the posterior means from NBM, shifting the distribution down with reference to NBM. Under BM, the benchmarking constraint implies that $\hat{\bar{Y}} = N^{-1} \sum_{i=1}^{10} \sum_{j=1}^{n_i} \omega_{ij} y_{ij}$. On the other hand, under SBM, we have

$\hat{\bar{Y}} = n^{-1} \sum_{i=1}^{10} \sum_{j=1}^{n_i} y_{ij}$. From our calculations, we find that $N^{-1} \sum_{i=1}^{10} \sum_{j=1}^{n_i} \omega_{ij} y_{ij} = 24.992 =$

$N^{-1} \sum_{i=1}^{10} N_i PM_{i(BM)}$. Similarly, $n^{-1} \sum_{i=1}^{10} \sum_{j=1}^{n_i} y_{ij} = 23.780 = N^{-1} \sum_{i=1}^{\ell} N_i PM_{i(SBM)}$, but

$N^{-1} \sum_{i=1}^{10} N_i PM_{i(NBM)} = 24.330$. The two benchmarking models are closer to each other than the nonbenchmarking model. But BM, which is the more appropriate model for the BMI data, has the smaller shrinkage of $SHR_{(BM)} = 5.606$, while $SHR_{(SBM)} = 5.667$. NBM has the greatest shrinkage with $SHR_{(NBM)} = 6.482$. For the BMD data, we find that $N^{-1} \sum_{i=1}^{10} \sum_{j=1}^{n_i} \omega_{ij} y_{ij} = 0.895 = N^{-1} \sum_{i=1}^{10} N_i PM_{i(BM)}$. Similarly, $n^{-1} \sum_{i=1}^{10} \sum_{j=1}^{n_i} y_{ij} = 0.903 = N^{-1} \sum_{i=1}^{10} N_i PM_{i(SBM)}$, but $N^{-1} \sum_{i=1}^{10} N_i PM_{i(NBM)} = 0.878$. For the shrinkage, the most appropriate model for BMD, the benchmarking model with unequal survey weights, has a shrinkage of $SHR_{(BM)} = 0.102$, while $SHR_{(SBM)} = 0.029$. This implies that the shrinkage of the model with equal survey weights is smaller than the benchmarking with unequal weights. NBM again has the greatest shrinkage with $SHR_{(NBM)} = 0.258$. However, overall, in this example, the two benchmarking models are closer to each other than the nonbenchmarking model. The PSDs of all three models are very close to each other. However, we observe that the PSDs from BM are mostly smaller than those from NBM and SBM, which reflects the adjustments from the survey weights in the constraint. As the survey weights are not incorporated in an informative manner (i.e., non ignorable design) in the model, there may not be gains in precision. We use box plots to show the general features in Figure 2.2. Figure 2.2 shows that the posterior means (PM) of the benchmarking models (both BM and SBM) are mostly smaller than the posterior means from NBM, shifting the distribution down with reference to NBM. Moreover, we see that the length of the whiskers of the boxplots for each domain is shorter for the BM and SBM models compared to that of NBM.

In summary, the two examples showed us the diversity of the applications of the univariate benchmarking model presented in this section. From the results, we

observe the effects of the benchmarking constraint. As our theory shows, we observe that the model-based estimates from the small areas match the chosen direct estimator for the single large area when these small areas are collapsed. We also observe that in most cases there is a gain in precision in using the benchmarking models, relative to the nonbenchmarking model, although the gains are not so much significant. This is perhaps due to the samples being relatively too small compared to the population size. Now, we will compare NBM and BM further using a simulation study, and we will discuss effects of benchmarking on biases.

Table 2.2: Comparison of the nonbenchmarking (NBM), self-weighted benchmarking (SBM) and survey-weighted benchmarking (BM) models using posterior means (PM), the posterior standard deviations (PSD) and the 95% credible intervals for the finite population mean $(\overline{Y})$ of the BMI and BMD data for the adult domains, gender domains of Mexican American population over 20 years, in the state of New York

				Nonbenchmarking			Self-Weighting Benchmarking			Benchmarking		
				Posterior		Credible	Posterior		Credible	Posterior		Credible
i	n	N	Direct Estimate	Mean	SD	Interval	Mean	SD	Interval	Mean	SD	Interval
— Body Mass Index
1	1	6	19.30	22.42	**2.09**	(18.13,26.40)	21.92	2.04	(17.55,25.62)	22.80	**2.15**	(18.21,26.74)
2	1	112	26.80	24.68	1.82	(21.21,28.54)	24.12	1.51	(21.21,27.46)	25.55	1.59	(22.50,29.04)
3	2	96	23.30	23.77	1.52	(20.61,26.73)	23.23	1.35	(20.28,25.75)	24.34	1.41	(21.36,27.06)
4	8	150	23.44	23.78	1.12	(21.54,26.00)	23.24	0.96	(21.19,24.99)	24.33	1.00	(22.25,26.24)
5	4	39	23.00	23.61	1.42	(20.71,26.40)	23.09	1.29	(20.32,25.44)	24.12	1.34	(21.27,26.60)
6	2	6	22.70	23.43	1.64	(20.16,26.58)	23.01	1.56	(19.83,25.96)	23.84	1.62	(20.56,26.99)
7	2	118	26.65	25.10	1.66	(22.03,26.58)	24.49	1.32	(22.05,27.41)	25.93	1.38	(23.38,28.86)
8	5	101	21.38	22.64	1.36	(19.87,25.20)	22.11	1.30	(19.39,24.31)	23.09	1.34	(20.38,25.41)
9	3	142	27.80	25.74	1.60	(22.85,29.08)	25.18	1.31	(22.97,28.07)	26.48	1.35	(24.11,29.38)
10	2	37	26.10	24.92	1.63	(21.82,28.37)	24.42	1.45	(21.62,27.46)	25.57	1.48	(22.73,28.66)
— Bone Mineral Density
1	1	6	0.965	0.900	0.09	(0.732,1.078)	0.921	0.08	(0.759,1.095)	0.914	0.08	(0.750,1.087)
2	1	112	0.983	0.914	0.08	(0.765,1.084)	0.941	0.07	(0.812,1.092)	0.935	0.06	(0.812,1.091)
3	2	96	0.962	0.915	0.07	(0.789,1.066)	0.940	0.06	(0.826,1.074)	0.931	0.06	(0.818,1.066)
4	8	150	0.927	0.890	0.05	(0.791,0.992)	0.913	0.04	(0.834,0.997)	0.904	0.04	(0.820,0.988)
5	4	39	1.088	0.971	0.07	(0.838,1.120)	0.992	0.06	(0.881,1.132)	0.984	0.06	(0.873,1.115)
6	2	6	0.838	0.836	**0.07**	(0.691,0.978)	0.852	0.07	(0.712,0.988)	0.846	**0.07**	(0.704,0.981)
7	2	118	0.863	0.864	0.07	(0.724,1.000)	0.889	0.06	(0.766,1.003)	0.886	0.06	(0.769,0.997)
8	5	101	0.776	0.813	0.06	(0.688,0.926)	0.836	0.05	(0.725,0.929)	0.826	0.05	(0.712,0.920)
9	3	142	0.802	0.849	0.06	(0.721,0.970)	0.874	0.05	(0.760,0.970)	0.867	0.05	(0.753,0.964)
10	2	37	0.878	0.871	0.07	(0.731,1.008)	0.893	0.06	(0.763,1.018)	0.886	0.06	(0.752,1.011)

2.4 Simulation Study

We perform a simulation study to further compare the nonbenchmarking and the benchmarking models. For simplicity, we assume that the design is self-weighting. We present four data generating models. We simulate the corn hectares data in the crop example in Section 3.5. To facilitate the simulation, we generate $M = 1,000$ datasets from each of the four generating models.

2.4.1 Design

We used four simulation models to generate data for this simulation. The first model is a homogeneous model, which ignores the small areas,

$$y_{ij} \overset{ind}{\sim} \text{Normal}\left\{ \underset{\sim}{x}'_{ij}\beta, \sigma^2 \right\}, \quad i = 1, 2, \ldots, \ell, j = 1, 2, \ldots, N_i. \tag{2.10}$$

The second model is the nested error regression model,

$$y_{ij} \overset{ind}{\sim} \text{Normal}\left\{ \underset{\sim}{x}'_{ij}\beta + \nu_i, \sigma^2 \right\}, \quad i = 1, 2, \ldots, \ell, j = 1, 2, \ldots, N_i; \tag{2.11}$$

$$\nu_i \overset{iid}{\sim} \text{Normal}\left\{ 0, \left(\frac{\rho}{1 - \rho} \right) \sigma^2 \right\}, \ 0 < \rho < 1, \quad i = 1, 2, \ldots, \ell.$$

Note that model (2.11) is simply model (2.10) with an addition of the area effect ν_i, for each area i, $i = 1, 2, \ldots, \ell$ [i.e., if $\nu_i = 0$, $i = 1, \ldots, \ell$, (2.11) becomes (2.10)].

The third model is the benchmarking model described in Section 2.1. Note that the third model is the second model in (2.11) with the benchmarking constraint. We use the results from Lemma 2.1, Remark 2.2, and Theorem 2.4 to generate the values of y. To do this, we first generate the sampled y values, $\underset{\sim}{y}_s$, from the distribution in Remark 2.2. Then, we use these generated values of $\underset{\sim}{y}_s$ to find the values of the nonsampled y's from the results in Lemma 2.1 and the proof of Theorem 2.2.

Note that the distributions of $\underset{\sim}{y}_s | \underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2$ and $\underset{\sim}{y}_{ns_{(N)}} | \underset{\sim}{y}_s, \underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2$ have similar forms and both can be written in a more general form as

$$\underset{\sim}{Y} \sim \text{Normal}\left\{ \underset{\sim}{\delta}, \sigma^2 \left(I - aJ \right) \right\}, \ 0 < a < 1. \tag{2.12}$$

Observe that from (2.12), we can write

$$\underset{\sim}{T} = \frac{1}{\sigma} \left(\underset{\sim}{Y} - \underset{\sim}{\delta} \right) \sim \text{Normal} \left\{ \underset{\sim}{0}, (I - aJ) \right\}. \tag{2.13}$$

So, if we can generate values of $\underset{\sim}{t}$ from (2.13), we can use these values to get the corresponding values of $\underset{\sim}{y}$. Using properties of multivariate normal and some algebraic manipulation in (2.13), we find that

$$T_1 \sim \text{Normal} \left(0, 1 - a \right) \tag{2.14}$$

and

$$T_k | \underset{\sim}{t}_{k-1} \sim \text{Normal} \left\{ - \left(\frac{a}{1 - a(k-1)} \right) \sum_{i=1}^{k-1} t_i, \frac{1 - ak}{1 - a(k-1)} \right\}, \tag{2.15}$$

where $\underset{\sim}{t}_{k-1} = (t_1, t_2, \ldots, t_{k-1})$, $k = 2, 3, \ldots$. So, to generate values of $\underset{\sim}{y}$, we generate values of $\underset{\sim}{t}$ via (2.14) and (2.15), and solve for the corresponding values of $\underset{\sim}{y}$ using the relation $\underset{\sim}{y} = \sigma \underset{\sim}{t} + \underset{\sim}{\delta}$. We apply this technique to generate both the sampled and nonsampled values of y. Then, we use Lemma 2.1 to get the value of the last observation, y_{ℓ, N_ℓ}.

Since the first three models are either the homogeneous model, the NBM or the BM, the results might favor either one of the models. Thus, it is sensible to generate data from a different model. In this fourth model, we use the Parzen-Rosenblatt estimator

$$\hat{f}(y | x_1, x_2) = \sum_{k=1}^{n} w_k \frac{1}{h_{opt}^{(3)}} \Phi \left(\frac{y - y_k}{h_{opt}^{(3)}} \right), w_k = \frac{\prod_{j=1}^{2} \Phi \left(\frac{x_j - x_{jk}}{h_{opt}^{(j)}} \right)}{\sum_{k=1}^{n} \prod_{j=1}^{2} \Phi \left(\frac{x_j - x_{jk}}{h_{opt}^{(j)}} \right)},$$

where $h_{opt}^{(1)}$, $h_{opt}^{(2)}$, and $h_{opt}^{(3)}$ are the window widths for x_1, x_2, and y, respectively, with $h_{opt} = \frac{1.06}{n^{1/5}} \min \left\{ \text{STD}, \frac{\text{IQR}}{1.34} \right\}$; the values of STD and IQR are based on all observed data for the respective variables. We use the following algorithm to generate the values of y:

(1) Draw an integer value k $(1 \le k \le n)$ with corresponding probability w_k of being chosen.

(2) Based on the drawn k in step (1), draw the value of y_{ij} from $y \sim \text{Normal}\left(y_k, h_{opt}^{(3)2}\right)$.

We also need to generate the values of the covariates for the nonsampled units. The approach used in the method is similar to our benchmarking technique. Again, we use the Parzen-Rosenblatt kernel density estimator $f(x) = \dfrac{1}{n}\sum\limits_{j=1}^{n}\dfrac{1}{h_{opt}}\Phi\left(\dfrac{x - x_j^*}{h_{opt}}\right)$. The distribution of $\underset{\sim}{X}_{ns(N)}$ given the sample covariate vector $\underset{\sim}{x}_s = (x_1, \ldots, x_n)$ is

$$\underset{\sim}{X}_{ns(N)}|\underset{\sim}{x}_s \sim \text{Normal}\left\{\left[\frac{t}{N} - \frac{n}{N-n}\left(\overline{x}_s - \frac{t}{N}\right)\right]1_{N-n-1}, h_{opt}^2\left[I_{N-n-1} - \frac{J_{N-n-1}}{N-n}\right]\right\}.$$

Note that the distribution of $\underset{\sim}{X}_{ns(N)}|\underset{\sim}{x}_s$ also takes the general form we stated in (2.12). Hence, we can also generate the values of $X_{n+1}, \ldots, X_N$ given $x_1, \ldots, x_n$, by using the technique mentioned above to generate the y values.

In our study we simulate the corn acreage data. We first generate the two covariates, namely, the number of corn pixels and the number of soybeans pixels, for all the N_i units in each of the $i = 1, \ldots, \ell = 12$ counties (area) using the above method for generating covariates. Based on the $n = 37$ samples from the crop data, we got $h_{opt(corn)} = 32.99$ and $h_{opt(soybeans)} = 24.20$. We use these constants and the sampled covariates to generate all the values of the covariates that would match the information given in the crop data. The number of sampled units and population units generated for each data is based on the sample and population sizes given in the corn data. Once all the values of the two covariates have been generated for the population units, we fix these values and use them to generate the values of y from all four generating data models for y. We use the posterior means from the conditional distributions of ρ, σ^2, and $\underset{\sim}{\beta} = (\beta_0, \beta_1, \beta_2)$ obtained by fitting the original corn data as the fixed values of the parameters ρ, σ^2, and $\underset{\sim}{\beta} = (\beta_0, \beta_1, \beta_2)$ to generate the simulated data. The area effects, $\nu_i, i = 1, 2, \ldots, \ell$, are simulated from its distribution given in the nested error regression model conditional on the observed values of ρ and σ^2 in the corn data. To make the comparison, we run both NBM and BM in the simulated datasets from the four models. Again we use $M = 10,000$ iterates in the model fitting.

2.4.2   Results

To compare the nonbenchmarking and benchmarking models, for each area and simulation run, we compute the accuracy ratios $R_{acc} = \left(PM - \overline{Y}\right)/\overline{Y}$, and the shrink-

age ratios $R_{shr} = (PM - \overline{y}_s)/\overline{y}_s$, where PM is the posterior mean for each area, $\overline{Y}$ is the true finite population mean of each area, and $\overline{y}_s$ is the direct estimate and the sample mean of each area. In Table 2.3, we present a comparison of the distribution of the R_{acc} and R_{shr} in BM and NBM in all four scenarios. Observe that the medians are very close to zero in all areas and the interquartile ranges are very small. These results indicate that the point estimators of the finite population means are very close to the true values and that there is not much shrinkage or bias using our estimator.

As in the examples above, we also calculated the posterior mean (PM), posterior standard deviation (PSD) and the 95% credible interval (Int) for each area under each model fitting of all the simulated datasets. The results are summarized in Table 2.4 for the first two models, the homogeneous and the nonbenchmarking models while the results for the benchmarking and nonparametric modelsand are given in Table 2.5. We present the average posterior mean, the average posterior standard deviation, the average width of the 95% credible intervals, and the coverage probabilities for each area. Note that the coverage probability (C) is calculated as the proportion of 95% credible intervals containing the true finite population mean $\overline{Y}$ for each area over the 1,000 simulated data. It is possible to compute these proportions since we know the value of the true finite population mean $\overline{Y}$ for each area from the simulated data. The average of the true values of the finite population mean for each area are also presented in Table 2.4 and Table 2.5. We also computed the 'root posterior mean squared error' for the kth data set, $RMSE_k = \sqrt{PMSE_k} = \sqrt{(PM_k - \overline{Y}_k)^2 + PSD_k^2}$ for each area. We summarize these values by presenting the average RMSE over the 1,000 data sets and its corresponding standard error for each area.

The results show that the average posterior means are relatively the same for both BM and NBM. Moreover, note that the average PSDs from BM are smaller than NBM in all four cases, exhibiting a better precision in BM. In terms of the interval estimations, the coverage probabilities for both NBM and BM are reasonably close to the nominal value of 95%, with those from BM closer to the nominal value in most areas in all four cases. (The closer coverage to the 95% nominal values are highlighted.) For the mean squared error, the RMSEs are generally smaller under BM. Observe that the few cases where BM's RMSEs are not smaller than those of NBM, the coverage probabilities of NBM are also better. The standard errors of the RMSEs offset these extreme differences.

Thus, these results are in concordance with what we have seen in the illustrative

examples. Both models give roughly the same accuracy and shrinkage. The posterior means are similar but with BM always giving smaller posterior standard deviations. This is natural because the benchmarking constraint reduces the sample space. Lastly, the mostly smaller posterior mean squared error from BM shows that the BM is performing better in estimating the finite population mean than NBM. Overall, we saw that in all four models, BM shows better estimation than NBM, with a few non-significant exceptions.

2.5　Concluding Remarks

In this chapter, we presented the univariate Bayesian benchmarking model based on the nested-error regression model of Battese, Harter, and Fuller (1988). We used Bayesian predictive inference to estimate the finite population means of small areas; see Toto and Nandram (2010a) for the journal article written based on the results of this chapter. We discussed both the theoretical and numerical effects of adding the benchmarking constraint into the Bayesian model. From the results of the examples as well as the mathematical calculation, we found that there were adjustments in the posterior mean of the finite population mean after the benchmarking constraint was included in the model. We also observed that the benchmarking model with the appropriate weights produces the most precise results. There is a definite gain in precision due to the inclusion of the benchmarking constraint into the model. However, we do not see too much improvement, because our sample sizes are relatively too small compared to the population sizes.

In the next chapter, we will extend the univariate Bayesian benchmarking model to the multivariate Bayesian benchmarking model to accommodate multivariate outcomes. That is, we discuss the Bayesian regression model for any number of responses, $q \geq 2$. There is a practical significance to the extension of the benchmarking model to the multivariate case because in many small area problems, data on related multiple characteristics and auxiliary variables are available. For instance, the two crops, corn and soybean, in the survey and satellite data of Battese, Harter, and Fuller (1988) would be more appropriately analyzed simultaneously because these two are correlated. Similarly, the body mass index (BMI) and bone mineral density (BMD) from NHANES III (1996) are also correlated characteristics that would be better analyzed at the same time.

We simultaneously model these examples in Chapter 3. We want to find out if the possible correlation between the responses would add significant contribution to the results. The work of Datta, Day, and Maiti (1998) presented a hierarchical Bayes prediction of the small area mean vector using the multivariate nested-error regression model of Fuller and Harter (1987). They compared the performance of the multivariate approach with the usual univariate approach and found that the multivariate approach may result in substantial improvement over its univariate counterpart through some data analysis

and simulations. Our model is also adapted from the multivariate nested-error regression model of Fuller and Harter (1987). However, the work of Datta, Day, and Maiti (1998) does not include benchmarking in their model, so we can only compare it to our non-benchmarking model. Like that of Datta, Day, and Maiti (1998), we will also compare the performance of the multivariate approach with the univariate approach both under nonbenchmarking and benchmarking.

Figure 2.1: Box plots of the distribution of the finite population mean $\left(\overline{Y}\right)$ by county (area) of the corn and soybeans acreage data

Distribution of Small Area Means for Corn Acreage

Distribution of Small Area Means for Soybeans Acreage
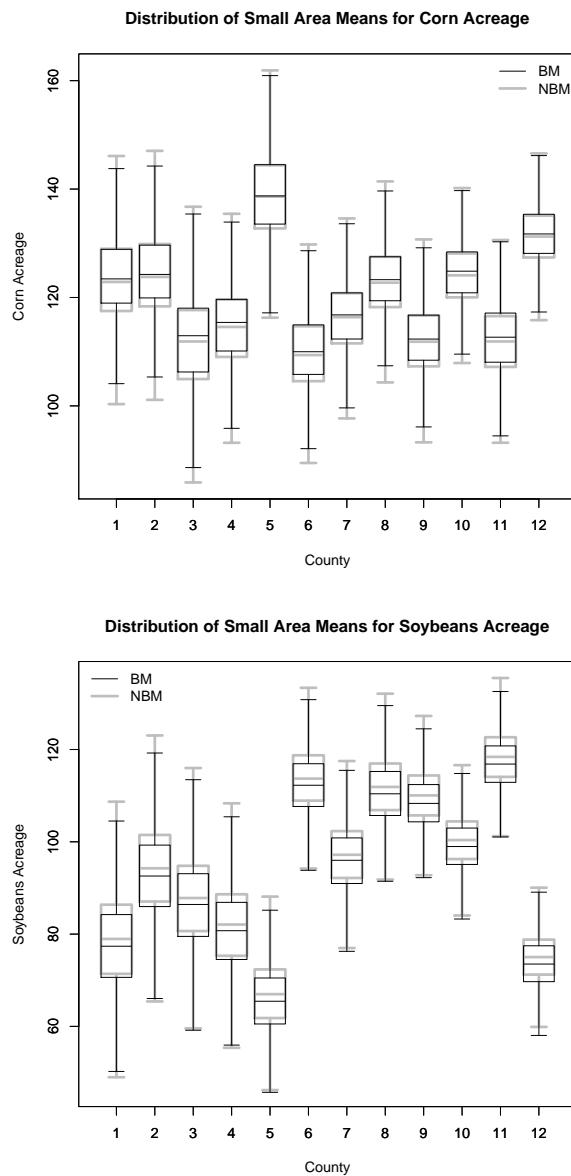
Note: There is a consistent upward shift in the center of the distributions for the corn acreage. On the other hand, there is a downward shift in the center of the distributions for the soybeans acreage. Most of the whiskers of the box plots in both corn and soybeans acreage results for the BM are shorter compared to the NBM boxplots, indicating an increase in the precision in most of the counties.

44

Figure 2.2: Boxplots of the distribution of the finite population mean $\left(\overline{Y}\right)$ by domain of the BMI and BMD data

Distribution of Small Area Means for BMI

Distribution of Small Area Means for BMD
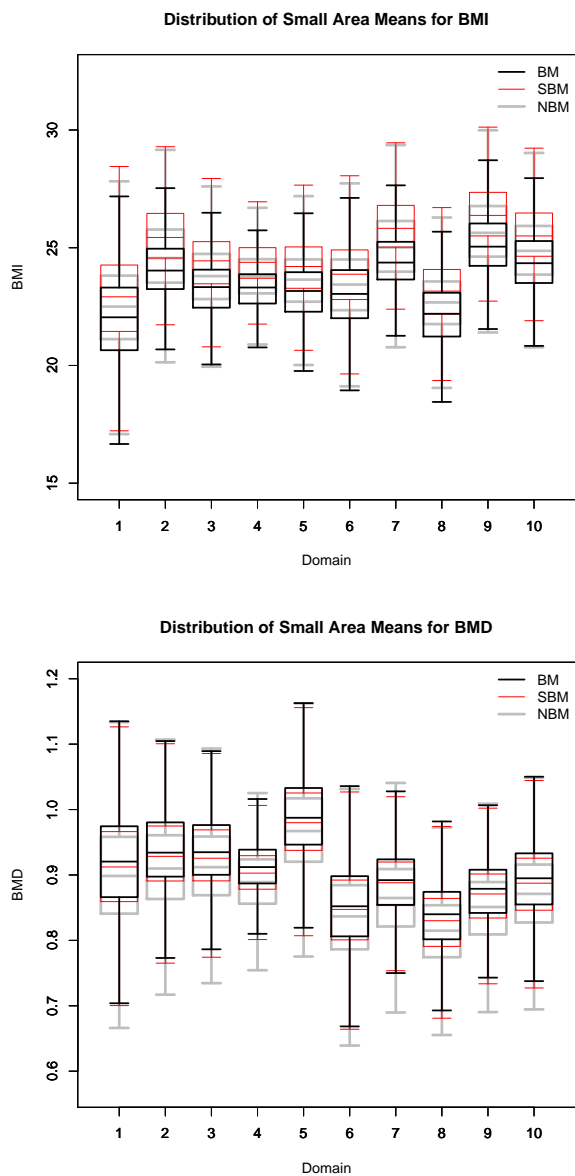
Note: There is a shift in the center of the distribution of the finite population means for BMI and BMD. For the BMI data, notice the downward shift in the center of the distributions. On the other hand, there is an upward shift in the center of the distributions for the BMD data. Notice also the shorter whiskers of most of the BM and SBM boxplots compared to the NBM boxplots, indicating an increase in the precision in most of the domains.

Table 2.3: Comparison of the distribution of the accuracy ratio R_{acc} and the shrinkage ratio R_{shr} in the nonbenchmarking and the benchmarking models for estimating the finite population mean

| | R_{acc} | | | | | | R_{shr} | | | | | |
| | NonBenchmarking | | | Benchmarking | | | NonBenchmarking | | | Benchmarking | | |
i	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
	Data generated using homogeneous model											
1	-0.042	-0.006	0.025	-0.034	0	0.031	-0.092	-0.021	0.072	-0.089	-0.015	0.083
2	-0.036	-0.001	0.036	-0.028	0.005	0.041	-0.081	0.004	0.105	-0.077	0.012	0.117
3	-0.019	0.022	0.056	-0.013	0.028	0.062	-0.057	0.024	0.125	-0.053	0.031	0.133
4	-0.037	0.003	0.038	-0.028	0.01	0.044	-0.046	0.002	0.062	-0.042	0.012	0.072
5	-0.064	-0.032	0.000	-0.059	-0.027	0.007	-0.059	-0.032	0	-0.054	-0.025	0.007
6	0.025	0.076	0.117	0.035	0.083	0.125	0.026	0.068	0.117	0.033	0.079	0.126
7	-0.024	0.013	0.046	-0.016	0.018	0.054	-0.021	0.014	0.051	-0.013	0.022	0.057
8	-0.04	-0.002	0.033	-0.033	0.003	0.04	-0.033	-0.002	0.037	-0.026	0.006	0.044
9	0.013	0.060	0.103	0.021	0.066	0.109	0.023	0.052	0.085	0.029	0.06	0.091
10	-0.053	-0.020	0.017	-0.047	-0.014	0.026	-0.038	-0.02	0.003	-0.032	-0.013	0.01
11	-0.038	0.000	0.029	-0.032	0.006	0.036	-0.025	-0.003	0.02	-0.019	0.004	0.028
12	-0.062	-0.034	0.002	-0.059	-0.028	0.011	-0.047	-0.029	-0.01	-0.041	-0.023	-0.004
	Data generated using Nonbenchmarking model											
1	-0.042	0.007	0.052	-0.030	0.019	0.061	-0.062	0.008	0.102	-0.051	0.02	0.116
2	-0.088	-0.044	-0.002	-0.075	-0.036	0.007	-0.1	-0.043	0.024	-0.093	-0.034	0.039
3	-0.096	-0.057	-0.006	-0.085	-0.046	0.003	-0.106	-0.052	0.009	-0.099	-0.041	0.023
4	-0.074	-0.032	0.013	-0.064	-0.023	0.022	-0.071	-0.03	0.015	-0.06	-0.018	0.027
5	-0.061	-0.016	0.025	-0.050	-0.005	0.037	-0.036	-0.014	0.01	-0.027	-0.002	0.022
6	-0.003	0.0490	0.097	0.010	0.061	0.109	0.013	0.045	0.081	0.023	0.059	0.095
7	-0.043	0.000	0.043	-0.033	0.012	0.054	-0.029	-0.006	0.021	-0.018	0.007	0.034
8	-0.016	0.034	0.089	0.001	0.046	0.103	0.006	0.038	0.072	0.019	0.05	0.088
9	0.036	0.102	0.166	0.05	0.118	0.178	0.07	0.099	0.13	0.089	0.116	0.146
10	-0.064	-0.027	0.012	-0.057	-0.017	0.024	-0.039	-0.024	-0.008	-0.03	-0.013	0.005
11	-0.027	0.016	0.056	-0.015	0.027	0.068	-0.001	0.014	0.031	0.009	0.027	0.045
12	-0.078	-0.041	-0.001	-0.071	-0.031	0.009	-0.05	-0.036	-0.021	-0.04	-0.026	-0.01
	Data generated using Benchmarking model											
1	-0.014	0.047	0.111	0.026	0.081	0.142	-0.027	0.049	0.145	0.004	0.084	0.191
2	-0.044	0.019	0.071	-0.009	0.05	0.102	-0.045	0.015	0.096	-0.021	0.051	0.142
3	0.005	0.068	0.135	0.042	0.102	0.162	-0.002	0.066	0.149	0.027	0.104	0.191
4	-0.058	-0.002	0.045	-0.026	0.028	0.075	-0.04	-0.004	0.04	-0.011	0.028	0.074
5	-0.047	-0.001	0.053	-0.017	0.029	0.084	-0.016	0.003	0.024	0.012	0.036	0.06
6	0.022	0.106	0.177	0.067	0.146	0.214	0.065	0.1	0.143	0.105	0.143	0.186
7	-0.059	-0.013	0.03	-0.03	0.015	0.059	-0.029	-0.011	0.009	0	0.02	0.041
8	-0.023	0.028	0.085	0.008	0.062	0.118	0	0.026	0.057	0.033	0.061	0.093
9	-0.015	0.038	0.09	0.021	0.073	0.123	0.012	0.028	0.051	0.047	0.065	0.088
10	-0.069	-0.029	0.005	-0.047	-0.004	0.033	-0.043	-0.03	-0.019	-0.02	-0.004	0.01
11	-0.045	-0.008	0.033	-0.015	0.021	0.064	-0.012	0	0.012	0.016	0.032	0.046
12	-0.074	-0.039	-0.003	-0.056	-0.016	0.025	-0.052	-0.041	-0.027	-0.032	-0.016	-0.001
	Data generated using nonparametric model											
1	-0.05	-0.019	0.025	-0.043	-0.014	0.028	-0.105	0.059	0.129	-0.103	0.065	0.137
2	-0.04	-0.004	0.04	-0.034	0.003	0.043	-0.109	0.031	0.117	-0.108	0.036	0.125
3	-0.01	0.020	0.063	-0.004	0.027	0.066	-0.044	0.072	0.145	-0.044	0.078	0.155
4	-0.036	0.005	0.049	-0.03	0.01	0.052	-0.052	0.029	0.103	-0.048	0.037	0.11
5	-0.074	-0.034	0.014	-0.068	-0.029	0.019	-0.069	-0.026	0.029	-0.065	-0.02	0.035
6	0.019	0.048	0.091	0.025	0.055	0.097	0.021	0.065	0.107	0.024	0.073	0.116
7	-0.036	0.007	0.046	-0.032	0.014	0.051	-0.038	0.008	0.054	-0.033	0.012	0.062
8	-0.045	-0.007	0.037	-0.04	-0.002	0.042	-0.057	-0.008	0.046	-0.053	-0.002	0.052
9	0.009	0.040	0.08	0.014	0.045	0.086	0.019	0.054	0.086	0.024	0.061	0.093
10	-0.063	-0.024	0.015	-0.059	-0.02	0.021	-0.05	-0.021	0.006	-0.046	-0.016	0.013
11	-0.032	0.004	0.042	-0.025	0.008	0.048	-0.022	0.006	0.037	-0.018	0.013	0.044
12	-0.067	-0.034	0.011	-0.064	-0.03	0.018	-0.052	-0.026	0.002	-0.046	-0.021	0.006

Table 2.4: Comparison of the Average Posterior Mean (PM), Average Posterior Standard Deviation (PSD), Average Width of the 95% credible intervals (width), Coverage Probabilities (C), and average Root 'Posterior' MSE (and its standard error) of the finite population means for each area over the $1,000$ simulation runs for the nonbenchmarking (NBM) and the benchmarking (BM) models

				Nonbenchmarking					Benchmarking				
i	n	N	FPM $\overline{Y}$	Posterior Mean	SD	Interval width	C	RMSE(SE)	Posterior Mean	PSD	Interval width	C	RMSE(SE)
						Data generated using homogeneous model							
1	1	545	122.65	121.76	11.43	45.97	1	13.55(4.05)	122.58	10.23	41.37	1	12.34 (3.60)
2	1	566	121.8	121.86	11.36	45.68	1	13.61(3.79)	122.7	10.18	41.23	1	12.49 (3.52)
3	1	395	118.41	120.53	11.26	45.16	1	13.83(4.09)	121.34	10.38	41.86	**0.99**	13.17 (3.90)
4	2	424	119.77	119.73	10.19	40.59	0.99	12.75 (3.62)	120.57	9.39	37.64	0.99	12.05 (3.48)
5	3	564	129.24	125.36	8.74	34.6	0.96	11.97 (3.29)	126.19	8.05	32.03	**0.95**	11.21 (3.08)
6	3	570	106.72	114.13	9.16	36.07	**0.89**	14.06 (3.69)	114.97	8.34	32.89	0.84	13.94 (3.93)
7	3	402	118.73	120.02	8.72	34.58	0.98	11.24 (2.86)	120.86	8.17	32.52	**0.96**	11.01 (3.02)
8	3	567	121.23	120.74	8.83	35.04	0.99	11.35 (2.75)	121.58	8.01	32.04	**0.97**	10.65 (2.63)
9	4	687	108.22	114.2	8.33	32.79	**0.91**	12.52 (2.93)	115.03	7.55	29.75	0.85	12.36 (3.16)
10	5	569	128.24	126.13	7.59	29.95	**0.95**	10.57 (2.40)	126.95	7.25	28.68	0.94	10.44 (2.65)
11	5	965	122.21	121.73	7.39	29.25	0.97	10.10 (2.64)	122.56	6.51	25.9	**0.94**	9.41 (2.60)
12	6	566	133.26	129.51	7.47	29.34	**0.93**	10.76 (2.34)	130.34	7.31	28.67	0.92	10.72 (2.52)
						Data generated using Nonbenchmarking model							
1	1	545	119.56	119.92	13.76	54.86	1	16.59 (5.47)	121.33	12.49	49.9	1	15.32 (4.56)
2	1	566	130.68	125.14	13.74	54.68	0.99	17.46 (5.45)	126.57	12.53	49.92	0.99	15.87 (4.58)
3	1	395	132.89	126.06	13.57	53.89	0.98	17.94 (5.72)	127.5	12.75	50.67	0.98	16.69 (5.03)
4	2	424	128.38	124.31	11.87	47.02	0.98	15.60 (5.98)	125.69	11.16	44.29	**0.97**	14.64 (5.28)
5	3	564	126.36	124.33	9.85	38.97	0.98	13.09 (3.63)	125.75	9.2	36.42	**0.97**	12.48 (3.57)
6	3	570	108.43	113.3	10.11	39.87	**0.94**	14.14 (3.83)	114.72	9.38	37	0.92	14.03 (4.13)
7	3	402	122.43	122.5	9.87	38.98	0.98	12.92 (3.51)	123.91	9.42	37.35	**0.97**	12.75 (3.76)
8	3	567	110.08	113.8	10.18	40.15	0.97	13.94 (3.83)	115.21	9.47	37.38	**0.94**	13.69 (3.99)
9	4	687	93.5	102.79	9.91	38.8	**0.86**	16.27 (5.59)	104.21	9.27	36.18	0.79	16.64 (5.98)
10	5	569	133.39	130.11	8.49	33.42	0.94	11.98 (2.92)	131.53	8.3	32.63	0.94	11.80 (3.08)
11	5	965	115.66	117.23	8.23	32.51	0.97	11.12 (2.55)	118.65	7.41	29.25	**0.94**	10.71 (2.65)
12	6	566	142.61	137.15	8.42	33.12	0.9	12.91 (3.45)	138.56	8.42	33	0.9	12.81 (3.45)

47

Table 2.5: Comparison of the Average Posterior Mean (PM), Average Posterior Standard Deviation (PSD), Average Width of the 95% credible intervals (width), Coverage Probabilities (C), and average Root 'Posterior' MSE (and its standard error) of the finite population means for each area over the $1,000$ simulation runs for the nonbenchmarking (NBM) and the benchmarking (BM) models

				Nonbenchmarking					Benchmarking				
			FPM	Posterior		Interval	C	RMSE(SE)	Posterior		Interval	C	RMSE(SE)
i	n	N	$\overline{Y}$	Mean	SD	width			Mean	PSD	width		
						Data generated using Benchmarking model							
1	1	545	110.37	115.56	15.32	60.63	0.99	19.57 (7.03)	119.32	13.97	55.31	**0.96**	19.53 (7.74)
2	1	566	117.55	119.3	15.09	59.66	1	18.57 (6.00)	123.07	13.81	54.83	**0.98**	17.95 (6.19)
3	1	395	107.66	114.7	14.91	58.94	0.98	19.91 (7.04)	118.44	14.02	55.56	**0.93**	20.68 (8.41)
4	2	424	120.02	119.65	12.7	50.11	0.99	16.24 (5.45)	123.38	12.02	47.52	**0.97**	15.95 (6.12)
5	3	564	119.74	120.11	10.4	40.98	0.97	14.08 (4.49)	123.84	9.78	38.57	0.93	14.22 (5.29)
6	3	570	91.07	99.95	11.31	44.45	**0.89**	17.72 (6.40)	103.69	10.5	41.21	0.78	19.23 (7.91)
7	3	402	125.71	124	10.41	41.04	0.97	13.61 (3.52)	127.68	10.07	39.71	0.97	13.47 (3.84)
8	3	567	110.27	113.31	10.7	42.13	**0.97**	14.40 (4.09)	117.03	9.96	39.28	0.92	14.99 (5.08)
9	4	687	106.98	110.81	9.48	37.3	**0.95**	13.20 (3.59)	114.54	8.84	34.83	0.88	14.27 (4.78)
10	5	569	141.95	137.66	8.92	35.11	0.93	12.83 (3.51)	141.38	8.99	35.32	**0.94**	12.58 (3.43)
11	5	965	121.34	120.77	8.51	33.56	0.96	11.44 (2.78)	124.49	7.78	30.7	0.94	11.30 (3.33)
12	6	566	154.55	148.71	8.89	35.04	0.91	13.60 (4.07)	152.41	9.19	36.17	**0.93**	13.22 (3.71)
						Data generated using nonparametric model							
1	1	545	109.55	108.67	11.24	45.42	1	13.24 (3.19)	109.24	9.97	40.63	1	12.02 (2.92)
2	1	566	109.1	109.39	11.2	45.27	1	13.14 (3.29)	109.97	9.95	40.56	1	12.01 (3.14)
3	1	395	105.04	108.44	11.09	44.64	0.99	13.60 (4.74)	109	10.13	41.29	**0.98**	12.90 (4.74)
4	2	424	106.2	107.11	10.2	40.72	0.99	12.64 (3.51)	107.7	9.3	37.52	**0.98**	11.91 (3.51)
5	3	564	114.55	111.47	8.92	35.43	**0.95**	12.39 (3.73)	112.05	8.12	32.42	0.93	11.68 (3.67)
6	3	570	98.77	104.24	9.04	35.87	0.97	12.12 (3.11)	104.83	8.15	32.5	**0.94**	11.68 (3.15)
7	3	402	107.63	108.43	8.83	35.13	0.99	11.11 (2.43)	109	8.2	32.85	**0.98**	10.75 (2.62)
8	3	567	109.4	109.22	8.98	35.76	0.98	11.55 (3.12)	109.8	8.1	32.47	**0.97**	10.87 (3.31)
9	4	687	99.31	103.74	8.36	33.04	0.97	11.12 (2.65)	104.31	7.48	29.77	**0.94**	10.67 (2.71)
10	5	569	114.79	112.51	7.82	30.87	**0.93**	11.20 (3.22)	113.09	7.35	29.23	0.91	10.96 (3.51)
11	5	965	107.33	107.96	7.58	30.03	0.99	9.83 (2.10)	108.55	6.57	26.27	**0.96**	9.09 (2.11)
12	6	566	117.04	113.88	7.59	29.94	**0.91**	11.15 (2.90)	114.45	7.32	28.94	0.89	11.15 (3.33)

Chapter 3

Multivariate Benchmarking

In this chapter, we extend the Bayesian benchmarking model discussed in Chapter 2 for any finite number of responses, $q \geq 2$. The extension of the benchmarking model to the multivariate case is interesting because in many small area problems, data on related multiple characteristics and auxiliary variables are available. For instance, the two crops, corn and soybean, in the survey and satellite data of Battese, Harter, and Fuller (1988) would be more appropriately analyzed simultaneously. Similarly, the body mass index (BMI) and bone mineral density (BMD) from NHANES III (1996) are also correlated characteristics that would be better analyzed at the same time. Datta, Day, and Maiti (1998) presented a hierarchical Bayes prediction of the small area mean vector also using the multivariate nested-error regression model of Fuller and Harter (1987). We will compare our results with the results of Datta, Day, and Maiti (1998) for the nonbenchmarking case. We will also compare the performance of the multivariate approach with the univariate approach with and without the benchmarking constraint. The work of Datta, Day, and Maiti (1998) did not incorporate a benchmarking constraint in their model, but they found that the nonbenchmarking multivariate approach may result in substantial improvement over its univariate counterpart. We present some data analysis and comparison of results for both the crop and health data.

We first describe the notation used for the Cholesky decomposition.

Let

$$
P = \begin{bmatrix} 1 & \phi_{12} & \cdots & \phi_{1q} \\ \phi_{21} & 1 & \cdots & \phi_{2q} \\ & & \ddots & \\ \phi_{q1} & \phi_{q2} & \cdots & 1 \end{bmatrix}
$$

be a positive definite matrix. Then **Cholesky decomposition of** P, denoted by $\mathrm{Chol}\,(P)$, is given by

$$
\mathrm{Chol}(P) = \begin{bmatrix} 1 & \phi_{12} & \cdots & \phi_{1q} \\ 0 & (1-\phi_{12}^2)^{\frac{1}{2}} & \cdots & \dfrac{\phi_{2q} - \phi_{12}\phi_{1q}}{(1-\phi_{12}^2)^{\frac{1}{2}}} \\ & & \ddots & \\ 0 & 0 & \cdots & \left(1 - \displaystyle\sum_{k=1}^{q-1} c_{kq}^2\right)^{\frac{1}{2}} \end{bmatrix},
$$

where $c_{11} = 1$, and $c_{1i} = \phi_{1i}$, $i = 2, \ldots, q$. Moreover, for $i = 2, \ldots, q-1$,

$$
c_{ij} = 0, j = 1, 2, \ldots, i - 1; \qquad\qquad c_{ii} = \left(1 - \sum_{k=1}^{i-1} c_{ki}^2\right)^{\frac{1}{2}};
$$

$$
c_{ij} = \frac{\phi_{ij} - \displaystyle\sum_{k=1}^{i-1} c_{ki}c_{kj}}{c_{ii}}, j = i + 1, \ldots, q; \quad c_{qq} = \left(1 - \sum_{k=1}^{q-1} c_{kq}^2\right)^{\frac{1}{2}}.
$$

With most, if not all of our applications in this dissertation, being bivariate vectors ($q = 2$), we describe the covariance matrix structure for the bivariate nested-error regression model. Thus, for $q = 2$, we will obtain the matrices G and P, given by

$$
G = \begin{bmatrix} \left(\dfrac{\gamma_1}{1 - \gamma_1}\right)^{\frac{1}{2}} & 0 \\ 0 & \left(\dfrac{\gamma_2}{1 - \gamma_2}\right)^{\frac{1}{2}} \end{bmatrix} \quad \text{and} \quad P = \begin{bmatrix} 1 & \phi \\ \phi & 1 \end{bmatrix}.
$$

Thus, applying the Cholesky decomposition on P, we get

$$
\mathrm{Chol}(P) = \begin{bmatrix} 1 & \phi \\ 0 & (1 - \phi^2)^{\frac{1}{2}} \end{bmatrix}.
$$

Hence, we will write

$$\Gamma = \text{Chol}(P)G = \begin{bmatrix} \left(\dfrac{\gamma_1}{1-\gamma_1}\right)^{\frac{1}{2}} & \phi\left(\dfrac{\gamma_2}{1-\gamma_2}\right)^{\frac{1}{2}} \\[2ex] 0 & (1-\phi^2)^{\frac{1}{2}}\left(\dfrac{\gamma_2}{1-\gamma_2}\right)^{\frac{1}{2}} \end{bmatrix},$$

and

$$\Gamma'\Gamma = G'PG = \begin{bmatrix} \left(\dfrac{\gamma_1}{1-\gamma_1}\right) & \phi\left(\dfrac{\gamma_1}{1-\gamma_1}\right)^{\frac{1}{2}}\left(\dfrac{\gamma_2}{1-\gamma_2}\right)^{\frac{1}{2}} \\[2ex] \phi\left(\dfrac{\gamma_1}{1-\gamma_1}\right)^{\frac{1}{2}}\left(\dfrac{\gamma_2}{1-\gamma_2}\right)^{\frac{1}{2}} & \left(\dfrac{\gamma_2}{1-\gamma_2}\right) \end{bmatrix}.$$

3.1 Multivariate Nested-Error Regression Model

Fuller and Harter (1987) introduced the multivariate nested-error regression model with components-of-variance error structure. This model is the multivariate version of the model of Battese, Harter, and Fuller (1988). The multivariate nested-error regression model is given by

$$
\underset{\sim}{y}_{ij} = \begin{pmatrix} y_{ij1} \\ y_{ij2} \\ \vdots \\ y_{ijq} \end{pmatrix} \overset{ind}{\sim} \text{Normal}\left\{ \begin{pmatrix} x'_{ij1}\underset{\sim}{\beta}_1 + \nu_{i1} \\ x'_{ij2}\underset{\sim}{\beta}_2 + \nu_{i2} \\ \vdots \\ x'_{ijq}\underset{\sim}{\beta}_q + \nu_{iq} \end{pmatrix}, \Sigma \right\} \tag{3.1}
$$

$$= \text{Normal}\left\{ X'_{ij}\underset{\sim}{\beta} + \underset{\sim}{\nu}_i, \Sigma \right\},$$

where $X_{ij} = \begin{bmatrix} \underset{\sim}{x}_{ij1} & \cdots & \underset{\sim}{0}_p \\ \vdots & \ddots & \vdots \\ \underset{\sim}{0}_p & \cdots & \underset{\sim}{x}_{ijq} \end{bmatrix}$, $\underset{\sim}{\nu}_i = (\nu_{i1}, \nu_{i2}, \ldots, \nu_{iq})'$, for $i = 1, \ldots, \ell$, $\underset{\sim}{\beta} = \left(\underset{\sim}{\beta}_1', \ldots, \underset{\sim}{\beta}_q'\right)'$,

with $\underset{\sim}{\beta}_k = (\beta_{k1}, \ldots, \beta_{kp})'$, for $k = 1, \ldots, q$, and Σ is a symmetric, positive-definite matrix.

Suppose that n_i samples are given for each $i = 1, 2, \ldots, \ell$. Since each unit y_{ij}, $i = 1, 2, \ldots, \ell$, $j = 1, 2, \ldots, n_i$, are independent, the sample data vector $\underset{\sim}{y}_s = \left(\underset{\sim}{y}'_{11}, \ldots, \underset{\sim}{y}'_{\ell n_\ell}\right)'$

has the following distribution function

$$
\begin{aligned}
p\left(\underset{\sim}{y_s}|\underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma\right) \quad \propto \quad & \prod_{i=1}^{\ell}\prod_{j=1}^{n_i}\left(\frac{1}{2\pi}\right)^{1/2}\left(\frac{1}{|\Sigma|}\right)^{1/2} \\
\times \quad & \exp\left\{-\frac{1}{2}\left(\underset{\sim}{y_{ij}} - X'_{ij}\underset{\sim}{\beta} - \underset{\sim}{\nu_i}\right)'\Sigma^{-1}\left(\underset{\sim}{y_{ij}} - X'_{ij}\underset{\sim}{\beta} - \underset{\sim}{\nu_i}\right)\right\}. \quad (3.2)
\end{aligned}
$$

In Chapter 2, we introduced a novel structure on the distribution of the area effect ν_i. In this chapter, we extend that structure for $q \geq 2$. That is, for $i = 1, 2, \ldots, \ell$, the area effects vector $\underset{\sim}{\nu_i}$ has distribution

$$
\underset{\sim}{\nu_i} = (\nu_{i1}, \nu_{i2}, \ldots, \nu_{iq})' \overset{ind}{\sim} \text{Normal}\left\{\underset{\sim}{0}_q, \Gamma'\Sigma\Gamma\right\},
$$

where $\Gamma = \{\text{chol}(P)\}G$ is an upper-triangular matrix such that $\Gamma'\Gamma = G'PG$ with

$$
G = \begin{bmatrix} \left(\dfrac{\gamma_1}{1-\gamma_1}\right)^{1/2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \left(\dfrac{\gamma_q}{1-\gamma_q}\right)^{1/2} \end{bmatrix}, \quad P = \begin{bmatrix} 1 & \phi_{12} & \cdots & \phi_{1q} \\ \phi_{21} & 1 & \cdots & \phi_{2q} \\ & & \ddots & \\ \phi_{q1} & \phi_{q2} & \cdots & 1 \end{bmatrix},
$$

with $0 < \gamma_k < 1$, $\quad k = 1, \ldots, q$; and $\phi_{ij} = \phi_{ji}$, $\quad -1 < \phi_{ij} < 1$, $\quad i, j = 1, \ldots, q$.

Since each $\underset{\sim}{\nu_i}, i = 1, \ldots, \ell$, are independent, the distribution function of $\underset{\sim}{\nu} = (\nu'_1, \ldots, \nu'_\ell)'$ is given by

$$
\pi\left(\underset{\sim}{\nu}\right) \propto \prod_{i=1}^{\ell}\left(\frac{1}{2\pi}\right)^{1/2}\left(\frac{1}{|\Gamma'\Sigma\Gamma|}\right)^{1/2}\exp\left\{-\frac{1}{2}\underset{\sim}{\nu_i}(\Gamma'\Sigma\Gamma)^{-1}\underset{\sim}{\nu_i}\right\}. \quad (3.3)
$$

3.2 The Nonbenchmarking Bayesian Model

In this section, we introduce the nonbenchmarking multivariate model. This is an extension of the nonbenchmarking Bayesian model discussed in Chapter 2. This model can also be considered as a Bayesian version of the multivariate nested-error regression model discussed by Fuller and Harter (1987). This is similar to the multivariate Bayesian model of Datta, Day, and Maiti (1998). The difference between the model of Datta, Day, and Maiti (1998) and our nonbenchmarking model lies in the prior specification in the model. Datta, Day, and Maiti (1998) imposed inverse-Wishart priors on both covariance matrices in the model, whereas, in our model, we applied reparameteri-

zation, so that we do not have to put priors on one of the covariance structures. This is one of the novel ideas that we are contributing to this study. We will elaborate on this in the next paragraphs.

Since our approach is to use Bayesian methodology, we need to put prior distributions on the following model parameters, $\underset{\sim}{\beta}, \Sigma, \underset{\sim}{\gamma} = (\gamma_1, \ldots, \gamma_q)$, and $\underset{\sim}{\phi} = (\phi_{12}, \ldots, \phi_{q-1,q})$, in our model.

We assume the following prior distribution on the parameters $\underset{\sim}{\beta}, \Sigma, \underset{\sim}{\gamma}, \underset{\sim}{\phi}$

$$\pi\left(\underset{\sim}{\beta}\right) \propto 1, ;$$

$$\pi\left(\Sigma\right) \propto \left(\frac{1}{|\Sigma|^{(q+1)/2}}\right) \quad \text{(Jeffrey's prior)};$$

$$\gamma_k \overset{iid}{\sim} \text{Uniform}\left(0, 1\right), k = 1, \ldots, q-1;$$

$$\phi_{ij} \overset{iid}{\sim} \text{Uniform}\left(-1, 1\right), \quad i = 1, 2, \ldots, q, \, j = 2, \ldots, q, i \neq j.$$

Assuming that the parameters are independent of each other, the joint prior distribution for $\underset{\sim}{\beta}, \Sigma, \Gamma$ is given by

$$\pi\left(\underset{\sim}{\beta}, \Sigma, \Gamma\right) \propto \left(\frac{1}{|\Sigma|^{(q+1)/2}}\right), \tag{3.4}$$

where $\Gamma = \Gamma\left(\underset{\sim}{\gamma}, \underset{\sim}{\phi}\right)$. As a result of the structure of the priors on the model parameters in our model, note that the joint prior distribution is a Wishart-type prior, a degenerate case. Observe that this generalizes the structure of the prior in the univariate problem.

Note that in the bivariate model,

$$\Gamma = \text{Chol}(P)G = \begin{bmatrix} \left(\frac{\gamma_1}{1-\gamma_1}\right)^{\frac{1}{2}} & \phi\left(\frac{\gamma_2}{1-\gamma_2}\right)^{\frac{1}{2}} \\ 0 & (1-\phi^2)^{\frac{1}{2}}\left(\frac{\gamma_2}{1-\gamma_2}\right)^{\frac{1}{2}} \end{bmatrix}.$$

With the structure of the covariance matrix of the area effects $\underset{\sim}{\nu_i}$, we only need to estimate the values of the scalar parameters. In this case, the three scalar parameters $\gamma_1, \gamma_2,$ and ϕ, of the covariance matrix, instead of estimating the covariance matrix as a whole. This makes the numerical calculations easier, at the same time avoids the problems of using the commonly-used inverse-Wishart family of distribution in hierarchical models

(Gelman, 2006). For instance, in our sampling procedure, we simply need to generate samples from Uniform $(0,1)$ to obtain values of the parameters $\gamma_k, k = 1, 2, \ldots, k - 1$, and ϕ_{ij}^*. We use a grid to generate the samples from the Uniform $(0,1)$. Then to generate samples from ϕ_{ij}, and get an appropriate prior value of the ϕ_{ij}'s, which are supposed to be Uniform $(0,1)$, $\quad i = 1, 2, \ldots, q, j = 2, \ldots, q, i \neq j$, we use linear transformation, $\phi_{ij} = 2\phi_{ij}^* - 1$. Thus, $\phi_{ij} \overset{iid}{\sim}$ Uniform $(-1,1)$, for $i = 1, 2, \ldots, q, j = 2, \ldots, q, i \neq j$.

Using Bayes' theorem in (3.2), (3.3), and (3.4), we find that the joint posterior density of $\underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma, \Gamma$ given the sample data $\underset{\sim}{y}_s$ is given by

$$
\pi\left(\nu, \underset{\sim}{\beta}, \Sigma, \Gamma | y_s\right) \propto p\left(\underset{\sim}{y}_s | \underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma, \Gamma\right) \times \pi\left(\underset{\sim}{\nu}\right) \times \pi\left(\underset{\sim}{\beta}, \Sigma, \Gamma\right)
$$

$$
\propto \left[\prod_{i=1}^{\ell}\prod_{j=1}^{n_i}\left(\frac{1}{(2\pi)^{1/2}|\Sigma|^{1/2}}\right) \exp\left\{-\frac{1}{2}\left[\underset{\sim}{y}_{ij} - \left(X_{ij}\underset{\sim}{\beta} + \underset{\sim}{\nu}_i\right)\right]'\Sigma^{-1}\left[\underset{\sim}{y}_{ij} - \left(X_{ij}\underset{\sim}{\beta} + \underset{\sim}{\nu}_i\right)\right]\right\}\right]
$$

$$
\times \left[\left(\prod_{i=1}^{\ell}\left(\frac{1}{(2\pi)^{1/2}|\Gamma'\Sigma\Gamma|^{1/2}}\right) \exp\left\{-\frac{1}{2}\left(\underset{\sim}{\nu}_i - \underset{\sim}{0}_q\right)'(\Gamma'\Sigma\Gamma)^{-1}\left(\underset{\sim}{\nu}_i - \underset{\sim}{0}_q\right)\right\}\right)\right]\left(\frac{1}{|\Sigma|^{(q+1)/2}}\right).
$$

Simplifying the above equation, the joint posterior density of $\underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma, \Gamma$ given the sample data $\underset{\sim}{y}_s$ is

$$
\pi\left(\nu, \underset{\sim}{\beta}, \Sigma, \Gamma | \underset{\sim}{y}_s\right) \propto \left(\frac{1}{|\Gamma|}\right)^{\ell}\left(\frac{1}{|\Sigma|}\right)^{(n+\ell+q+1)/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{\ell}\left(\underset{\sim}{\nu}_i - \underset{\sim}{0}_q\right)'(\Gamma'\Sigma\Gamma)^{-1}\left(\underset{\sim}{\nu}_i - \underset{\sim}{0}_q\right)\right\}
$$

$$
\times \exp\left\{-\frac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\left[\underset{\sim}{y}_{ij} - \left(X_{ij}\underset{\sim}{\beta} + \underset{\sim}{\nu}_i\right)\right]'\Sigma^{-1}\left[\underset{\sim}{y}_{ij} - \left(X_{ij}\underset{\sim}{\beta} + \underset{\sim}{\nu}_i\right)\right]\right\}.
$$

We use Gibbs sampling (see Gelfand and Smith (1990)) to draw samples from the above joint distribution. Hence, we need to know the full conditional distributions of the parameters we want to estimate. From the joint posterior distribution $\underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma, \Gamma | \underset{\sim}{y}_s$, under the nonbenchmarking model, we have the following full conditional distributions of the parameters of the model.

1. For $i = 1, 2, \ldots, \ell$, the full conditional distribution of $\underset{\sim}{\nu}_i | \underset{\sim}{y}_s, \underset{\sim}{\beta}, \Sigma, \Gamma$ is given by

$$
\underset{\sim}{\nu}_i | \underset{\sim}{y}_s, \underset{\sim}{\beta}, \Sigma, \Gamma \overset{ind}{\sim} \text{Normal}\left\{B_i^{-1}\Sigma^{-1}\sum_{j=1}^{n_i}\left(\underset{\sim}{y}_{ij} - X_{ij}'\underset{\sim}{\beta}\right), B_i^{-1}\right\}, \qquad (3.5)
$$

54

where $B_i = n_i\Sigma^{-1} + (\Gamma'\Sigma\Gamma)^{-1}$.

2. The full conditional distribution of $\beta|y_s, \underset{\sim}{\nu}, \Sigma, \Gamma$ is given by

$$\beta|y_s, \underset{\sim}{\nu}, \Sigma, \Gamma \sim \text{Normal}\left\{\widehat{\beta}, D^{-1}\right\}, \tag{3.6}$$

where $\widehat{\beta} = \left(\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} X_{ij}\Sigma^{-1}X_{ij}'\right)^{-1}\left(\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} X_{ij}\Sigma^{-1}\left(y_{ij} - \underset{\sim}{\nu_i}\right)'\right)$ and

$D = \sum_{i=1}^{\ell}\sum_{j=1}^{n_i} X_{ij}\Sigma^{-1}X_{ij}'$.

3. The full conditional distribution of $\Sigma|y_s, \underset{\sim}{\nu}, \beta, \Gamma$ is given by

$$\Sigma|y_s, \underset{\sim}{\nu}, \beta, \Gamma \sim \text{Inverse-Wishart}\,(n + \ell, S) \tag{3.7}$$

where $n + \ell > q - 1$ and S is a positive definite $q \times q$ scale matrix given by

$$S = \sum_{i=1}^{\ell}\left[\left(\Gamma^{-1}\right)'\left(\underset{\sim}{\nu_i}\underset{\sim}{\nu_i'}\right)\left(\Gamma^{-1}\right)\right] + \sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\left(y_{ij} - X_{ij}'\beta - \underset{\sim}{\nu_i}\right)\left(y_{ij} - X_{ij}'\beta - \underset{\sim}{\nu_i}\right)'.$$

The **Inverse-Wishart distribution** is the conjugate prior distribution for the multivariate normal covariance matrix. A degenerate form occurs when $\nu < k$. The probability density function of the inverse-Wishart random $k \times k$ matrix W is

$$p\,(W|S, \nu) = \frac{|S|^{\nu/2}|W|^{-(\nu+k+2)/2}\exp\left\{-\frac{1}{2}\text{trace}\left(SW^{-1}\right)\right\}}{2^{\nu k/2}\Gamma_k\,(nu/2)},$$

where ν is the degrees of freedom, S is a $k \times k$ symmetric positive definite scale matrix, and $\Gamma_k(\cdot)$ is the multivariate gamma function.

4. The full conditional density of $\Gamma|y_s, \underset{\sim}{\nu}, \beta, \Sigma$ is given by

$$\pi\left(\Gamma|y_s, \underset{\sim}{\nu}, \beta, \Sigma\right) \propto \left(\frac{1}{|\Gamma|}\right)^{\ell}\exp\left\{\frac{-1}{2}\text{trace}\left[\left(\Gamma'\Sigma\Gamma\right)^{-1}\left(\sum_{i=1}^{\ell}\underset{\sim}{\nu_i}\underset{\sim}{\nu_i'}\right)\right]\right\} \tag{3.8}$$

55

Now, we describe the distribution of the finite population mean under the non-benchmarking model. For $i = 1, \ldots, \ell$, the $q \times 1$ finite population mean vector for the ith area is defined as

$$\overline{Y}_{\sim i} = \frac{1}{N_i} \sum_{j=1}^{n_i} y_{\sim ij}.$$

Also, denote by $\overline{y}_{\sim s_i} = \left(\overline{y}_{s_{i1}}, \ldots, \overline{y}_{s_{iq}}\right)'$, the $q \times 1$ sample mean vector for the ith area, and by $f_i = n_i/N_i$ the ith sampling fraction. Moreover, $\overline{X}_{s_i} = \text{block diagonal}\left(\overline{x}_{s_{i1}}, \ldots, \overline{x}_{s_{iq}}\right)$ and $\overline{X}_i = \text{block diagonal}\left(\overline{x}_{i1}, \ldots, \overline{x}_{iq}\right)$ are the $qp \times q$ matrices of the means of the p covariates for the ith area, and the ith population, $i = 1, 2, \ldots, \ell$ respectively, where $\overline{x}_{s_{ik}} = \left(\overline{x}_{s_{ik1}}, \ldots, \overline{x}_{s_{ikp}}\right)'$, and $\overline{x}_{ik} = \left(\overline{x}_{ik1}, \ldots, \overline{x}_{ikp}\right)'$, for $k = 1, 2, \ldots, q$.

Theorem 3.1. Using properties of multivariate normal distribution, under the nonbenchmarking model, the distribution of the finite population mean vector, $\overline{Y}_{\sim i}$ given $y_{\sim s}, \nu_{\sim}, \beta_{\sim}, \Sigma$ is

$$\overline{Y}_{\sim i} | y_{\sim s}, \nu_{\sim}, \beta_{\sim}, \Sigma \overset{ind}{\sim} \text{Normal} \left\{ f_i \overline{y}_{\sim s_i} + (1 - f_i)\left(\overline{X}'_{ns_i} \beta_{\sim} + \nu_{\sim i}\right), \left(\frac{1 - f_i}{N_i}\right) \Sigma \right\}. \qquad (3.9)$$

Observe that if $q = 1$, the distribution in (3.9) reduces to (1.5) given in Chapter 1.

Proof.
We can write $\overline{Y}_{\sim i}$ as

$$
\begin{aligned}
\overline{Y}_{\sim i} &= \frac{1}{N_i} \sum_{j=1}^{N_i} y_{\sim ij} = f_i \overline{y}_{\sim s_i} + (1 - f_i)\overline{y}_{\sim ns_i} \\
&= \frac{1}{n_i}\left(1'_{n_i} \otimes I_q\right) y_{\sim s_i} + \frac{1}{N_i - n_i}\left(1'_{N_i - n_i} \otimes I_q\right) y_{\sim ns_i}.
\end{aligned}
$$

The above expression implies that the distribution of the finite population mean vector $\overline{Y}_{\sim i} | y_{\sim s}, \nu_{\sim}, \beta_{\sim}, \Sigma$ depends on the distribution of the sampled units, $y_{\sim s_i}$, and the non-sampled units, $y_{\sim ns_i}$, in the ith area. Using properties of multivariate normal distribution, the nonsampled units, $y_{\sim ns_i}$, are normally distributed. Hence, for $i = 1, 2, \ldots, \ell$, the finite population mean vector $\overline{Y}_{\sim i} | y_{\sim s}, \nu_{\sim}, \beta_{\sim}, \Sigma$ follows a multivariate normal distribution with mean vector

$$E\left[\overline{Y}_{\sim i} | y_{\sim s}, \nu_{\sim}, \beta_{\sim}, \Sigma\right] = f_i \overline{y}_{\sim s_i} + (1 - f_i)\left(\overline{X}'_{ns_i} \beta_{\sim} + \nu_{\sim i}\right),$$

and covariance matrix

$$Var\left[\overline{Y}_i | y_s, \nu, \beta, \Sigma\right] = \left(\frac{1-f_i}{N_i}\right)\Sigma.$$

∎

As mentioned earlier, Datta, Day, and Maiti (1998) presented a multivariate Bayesian nonbenchmarking model for the finite population mean and they used the crop data as the illustration of their model. Datta, Day, and Maiti (1998) also applied Bayesian methodology to estimate the small area mean vector in the Fuller and Harter (1987) regression model. But they used inverse-Wishart priors (see Appendix) on all their covariance matrices. Thus we can compare the results from their model with our nonbenchmarking (NBM) results from both univariate and multivariate Bayesian models. Table 3.1 shows the comparison of the crop data example.

Table 3.1: Comparison of the Univariate Nonbenchmarking model (uNBM), Multivariate Nonbenchmarking (mNBM), and the Datta, Day, and Maiti model (DDM) results for the finite population mean of the crop hectares data for the counties in Iowa.

County	Corn			Soybean			Correlation	
	mNBM	DDM	uNBM	mNBM	DDM	uNBM	mNBM	DDM
i	Posterior Mean$_{(SD)}$							
1	$124.05_{(8.75)}$	$119.98_{(13.90)}$	$123.47_{(9.32)}$	$79.16_{(11.47)}$	$83.69_{(13.47)}$	$78.76_{(11.27)}$	-0.51	0.02
2	$122.08_{(7.98)}$	$130.00_{(14.03)}$	$124.20_{(9.28)}$	$95.03_{(10.89)}$	$98.72_{(13.43)}$	$94.34_{(10.92)}$	-0.55	-0.13
3	$116.26_{(8.66)}$	$98.83_{(13.65)}$	$110.95_{(10.04)}$	$87.09_{(11.76)}$	$78.54_{(12.86)}$	$87.72_{(10.70)}$	-0.45	-0.07
4	$120.19_{(8.92)}$	$99.79_{(12.19)}$	$114.16_{(8.37)}$	$81.23_{(10.98)}$	$81.01_{(12.13)}$	$82.04_{(10.09)}$	-0.35	-0.26
5	$138.66_{(8.20)}$	$133.13_{(10.76)}$	$138.82_{(8.37)}$	$67.68_{(8.48)}$	$75.90_{(9.94)}$	$67.15_{(7.93)}$	-0.44	-0.44
6	$108.14_{(6.58)}$	$111.43_{(10.36)}$	$109.78_{(7.66)}$	$114.10_{(7.54)}$	$118.23_{(9.47)}$	$113.83_{(7.34)}$	-0.36	-0.38
7	$115.03_{(6.45)}$	$120.08_{(9.93)}$	$116.05_{(7.20)}$	$97.13_{(7.60)}$	$94.40_{(9.47)}$	$97.23_{(7.63)}$	-0.41	-0.39
8	$120.94_{(6.83)}$	$126.62_{(10.30)}$	$122.90_{(7.20)}$	$112.07_{(7.30)}$	$111.79_{(9.44)}$	$111.93_{(7.60)}$	-0.42	-0.36
9	$111.23_{(6.23)}$	$111.40_{(9.71)}$	$112.07_{(7.00)}$	$110.47_{(6.69)}$	$114.09_{(9.21)}$	$110.06_{(6.54)}$	-0.45	-0.53
10	$123.45_{(5.19)}$	$126.31_{(8.97)}$	$123.99_{(6.25)}$	$100.44_{(5.76)}$	$97.74_{(8.35)}$	$100.36_{(6.13)}$	-0.45	-0.49
11	$110.94_{(6.99)}$	$116.69_{(9.46)}$	$111.71_{(6.96)}$	$117.50_{(6.71)}$	$111.43_{(8.97)}$	$118.28_{(6.48)}$	-0.41	-0.55
12	$132.12_{(5.43)}$	$129.65_{(8.45)}$	$131.25_{(5.92)}$	$75.28_{(5.99)}$	$75.86_{(7.86)}$	$75.04_{(5.65)}$	-0.42	-0.51

Observe from Table 3.1 that the multivariate nonbenchmarking results have the most improved precision compared to both the Datta, Day, and Maiti (1998) and the univariate nonbenchmarking results. There are some instances in Table 3.1 that our univariate nonbenchmarking yielded some better precision compared to the multivariate results, however, their values are not so far from each other. Also, observe the last

two columns of Table 3.1 in which we see that the posterior correlation between corn and soybeans is more consistent over all the areas for the multivariate nonbenchmarking model than the Datta, Day, and Maiti (1998) correlations; note the correlation in the first county from the Datta, Day, and Maiti (1998) output.

3.3 The Benchmarking Bayesian Model

We now extend the univariate benchmarking model to include multivariate outcomes. We incorporate to the nonbenchmarking model described in (3.2), (3.3), and (3.4), the following q benchmarking constraints

$$\sum_{i=1}^{\ell}\sum_{j=1}^{N_i} y_{ijk} = \sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \omega_{ij} y_{ijk}, k = 1, 2, \ldots, q. \tag{3.10}$$

Recall that ω_{ij} is the sampling weight of the jth unit in the ith area and these weights are the same for all outcomes. As in the univariate case, this type of benchmarking is called internal benchmarking, since we are using a weighted average of the direct estimators.

Incorporating the benchmarking constraints in (3.10) to the nonbenchmarking model, we get the multivariate Bayesian benchmarking model (MBM) is

$$\underset{\sim}{y}_{ij} \overset{ind}{\sim} \text{Normal} \left\{ X'_{ij}\underset{\sim}{\beta} + \underset{\sim}{\nu}_i, \Sigma \right\}, i = 1, 2, \ldots, \ell, j = 1, 2, \ldots, N_i, \tag{3.11}$$

$$\sum_{i=1}^{\ell}\sum_{j=1}^{N_i} y_{ijk} = \sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \omega_{ij} y_{ijk}, \quad k = 1, 2, \ldots, q,$$

$$\underset{\sim}{\nu}_i \overset{ind}{\sim} \text{Normal} \left\{ \underset{\sim}{0}_q, \Gamma'\Sigma\Gamma \right\}, i = 1, 2, \ldots, \ell, \tag{3.12}$$

$$\pi \left(\underset{\sim}{\beta}, \Sigma, \Gamma \right) \propto \frac{1}{|\Sigma|}^{(q+1)/2} \tag{3.13}$$

Let $\underset{\sim}{y}_s = \left(\underset{\sim}{y}'_{s_1}, \ldots, \underset{\sim}{y}_{s_\ell}' \right)'$, where $\underset{\sim}{y}_{s_i} = \left(\underset{\sim}{y}'_{i1}, \ldots, \underset{\sim}{y}'_{in_i} \right)'$. Similarly, $\underset{\sim}{y}_{ns} = \left(\underset{\sim}{y}'_{ns_{(N)}}, \underset{\sim}{y}'_{\ell N_\ell} \right)'$, where $\underset{\sim}{y}_{ns_{(N)}} = \left(\underset{\sim}{y}'_{ns_1}, \ldots, \underset{\sim}{y}_{ns_\ell}' \right)'$, $\underset{\sim}{y}_{ns_i} = \left(\underset{\sim}{y}'_{in_i+1}, \ldots, \underset{\sim}{y}'_{iN_i} \right)'$, for $i = 1, 2, \ldots, \ell - 1$, and $\underset{\sim}{y}_{ns_\ell} = \left(\underset{\sim}{y}'_{\ell n_\ell + 1}, \ldots, \underset{\sim}{y}'_{\ell N_\ell - 1} \right)'$. Also, write $\underset{\sim}{\omega}^* = (\underset{\sim}{\omega} - \underset{\sim}{1}_n)$, $W = \frac{\underset{\sim}{\omega}^* \underset{\sim}{\omega}^{*'}}{\underset{\sim}{\omega}^{*'}\underset{\sim}{\omega}^* + \underset{\sim}{\omega}^{*'}\underset{\sim}{1}_n}$, $\tilde{W} = \frac{\underset{\sim}{\omega}^* \underset{\sim}{1}_{N-n}'}{\underset{\sim}{\omega}^{*'}\underset{\sim}{\omega}^* + \underset{\sim}{\omega}^{*'}\underset{\sim}{1}_n}$. Lastly, we write $X_s = (X_{11}, \ldots, X_{\ell, n_\ell})$, $X_{ns} = (X_{1n_1+1}, \ldots, X_{\ell, N_\ell})$,

Z_s = block diagonal $\left(\underset{\sim}{1}_{n_1}, \ldots, \underset{\sim}{1}_{n_\ell} \right)$, and Z_{ns} = block diagonal $\left(\underset{\sim}{1}_{N_1-n_1}, \ldots, \underset{\sim}{1}_{N_\ell - n_\ell} \right)$.

Conditional on the benchmarking constraint, we find the distribution of the data, both sampled and nonsampled, in the next two lemmas. Before we state our remarks, we define first the Kronecker product, which we use in a lot of our results.

We note that the **direct product**, sometimes called the **Kronecker product** or **Zehfuss product**, of two matrices $\boldsymbol{A}_{p \times q}$ and $\boldsymbol{B}_{m \times n}$ is defined as

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11}\boldsymbol{B} & \cdots & a_{1q}\boldsymbol{B} \\ \vdots & \ddots & \vdots \\ a_{p1}\boldsymbol{B} & \cdots & a_{pq}\boldsymbol{B} \end{bmatrix}. \tag{3.14}$$

Remark 3.2. Under the benchmarking model, conditional on the benchmarking constraint,

$$\underset{\sim}{y}_s | \underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma \sim \text{Normal} \left\{ C_x \underset{\sim}{\beta} + C_z \underset{\sim}{\nu}, (I_n - W) \otimes \Sigma \right\},$$

$$C_x = [(I_n - W) \otimes I_q] X'_s + \left(\tilde{W} \otimes I_q \right) X'_{ns},$$

and

$$C_z = [(I_n - W) \otimes I_q] [Z_s \otimes I_q] + \left[\tilde{W} \otimes I_q \right] [Z_{ns} \otimes I_2].$$

Remark 3.3. Under the benchmarking model, conditional on the benchmarking constraint,

$$\underset{\sim}{y}_{ns} | \underset{\sim}{y}_s, \underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma \sim \text{Normal} \left\{ K_x \underset{\sim}{\beta} + K_z \underset{\sim}{\nu} + K_y \underset{\sim}{y}_s, \left(I_{N-n} - \frac{1}{N-n} J_{N-n} \right) \otimes \Sigma \right\},$$

$$K_x = \left[\left(I_{N-n} - \frac{1}{N-n} J_{N-n} \right) \otimes I_q \right] X'_{ns}, \quad K_y = \left(\frac{1}{N-n} \right) \left(\underset{\sim}{1}_{N-n} \underset{\sim}{\omega}^{*'} \otimes I_q \right), \text{ and}$$

$$K_z = \left[\left(I_{N-n} - \frac{1}{N-n} J_{N-n} \right) \otimes I_q \right] [Z_{ns} \otimes I_q].$$

Using Bayes' theorem in (3.11), (3.12), and (3.13), we find that the joint posterior density of $\underset{\sim}{\nu}, \beta, \Sigma, \Gamma$ given the sample data $\underset{\sim}{y}_s$ under the multivariate benchmarking model is

$$\pi\left(\underset{\sim}{\nu},\underset{\sim}{\beta},\Sigma,\Gamma|y_s\right) \propto p\left(y_s|\underset{\sim}{\nu},\underset{\sim}{\beta},\Sigma\right)\pi(\underset{\sim}{\nu})\pi(\underset{\sim}{\beta},\Sigma,\Gamma)$$

$$\propto \left(\frac{1}{|\Sigma|}\right)^{n/2}\exp\left\{-\frac{1}{2}\left(y_s - C_x\underset{\sim}{\beta} - C_z\underset{\sim}{\nu}\right)'[(I_n - W)\otimes\Sigma]^{-1}\left(y_s - C_x\underset{\sim}{\beta} - C_z\underset{\sim}{\nu}\right)\right\}$$

$$\times \left(\frac{1}{|\Gamma|}\right)^{\ell}\exp\left\{-\frac{1}{2}\underset{\sim}{\nu}'(I_\ell \otimes \Gamma'\Sigma\Gamma)^{-1}\underset{\sim}{\nu}\right\}\left(\frac{1}{|\Sigma|}\right)^{(q+1)/2}.$$

Simplifying the above equation, the joint posterior density of $\underset{\sim}{\nu},\underset{\sim}{\beta},\Sigma,\Gamma$ given the sample data y_s is

$$\pi\left(\underset{\sim}{\nu},\underset{\sim}{\beta},\Sigma,\Gamma|y_s\right) \propto \left(\frac{1}{|\Gamma|}\right)^{\ell}\left(\frac{1}{|\Sigma|}\right)^{(n+\ell+q+1)/2}\exp\left\{-\frac{1}{2}\underset{\sim}{\nu}'(I_\ell \otimes \Gamma'\Sigma\Gamma)^{-1}\underset{\sim}{\nu}\right\}$$

$$\times \exp\left\{-\frac{1}{2}\left(y_s - C_x\underset{\sim}{\beta} - C_z\underset{\sim}{\nu}\right)'[(I_n - W)\otimes\Sigma]^{-1}\left(y_s - C_x\underset{\sim}{\beta} - C_z\underset{\sim}{\nu}\right)\right\}.$$

As in the nonbenchmarking model, we will use Gibbs sampling to generate estimates of the model parameters. From the joint posterior distribution $\underset{\sim}{\nu},\underset{\sim}{\beta},\Sigma,\Gamma|y_s$, under the benchmarking model, we have the following full conditional distributions:

1. The full conditional distribution of $\underset{\sim}{\nu}|y_s,\underset{\sim}{\beta},\Sigma,\Gamma$ is given by

$$\underset{\sim}{\nu}|y_s,\underset{\sim}{\beta},\Sigma,\Gamma \sim \text{Normal}\left\{A^{-1}C_z'[(I_n - W)\otimes\Sigma]^{-1}\left(y_s - C_x\underset{\sim}{\beta}\right), A^{-1}\right\} \qquad (3.15)$$

where $A = [I_\ell \otimes \Gamma'\Sigma\Gamma]^{-1} + C_z'[(I_n - W)\otimes\Sigma]^{-1}C_z$.

2. The full conditional distribution of $\underset{\sim}{\beta}|y_s,\underset{\sim}{\nu},\Sigma,\Gamma$ is given by

$$\underset{\sim}{\beta}|y_s,\underset{\sim}{\nu},\Sigma,\Gamma \sim \text{Normal}\left\{\widehat{\underset{\sim}{\beta}}, \left[C_x'\{(I_n - W)\otimes\Sigma\}^{-1}C_x\right]^{-1}\right\}, \qquad (3.16)$$

where

$$\widehat{\underset{\sim}{\beta}} = \left\{C_x'[(I_n - W)\otimes\Sigma]^{-1}C_x\right\}^{-1}\left\{C_x'[(I_n - W)\otimes\Sigma]^{-1}\left[y_s - C_z\underset{\sim}{\nu}\right]\right\}.$$

3. The full conditional distribution of $\Sigma|y_s,\underset{\sim}{\nu},\underset{\sim}{\beta},\Gamma$ is given by

$$\Sigma|y_s,\underset{\sim}{\nu},\underset{\sim}{\beta},\Gamma \sim \text{Inverse-Wishart}\left(n + \ell, S\right) \qquad (3.17)$$

where $n + \ell > q - 1$ and S is a positive definite $q \times q$ scale matrix given by

$$S = S_1 S_1' + S_2 \left(I_n - W\right)^{-1} S_2',$$

and S_1 is the $q \times \ell$ matrix obtained by unstacking the vector $(I_\ell \otimes \Gamma)^{-1} \underset{\sim}{\nu}$ taking each consecutive q elements of the vector as columns, S_2 is the $q \times n$ matrix obtained by unstacking the vector $\underset{\sim}{y}_s - C_x \beta - C_z \underset{\sim}{\nu}$ taking each consecutive q elements as the columns.

4. The full conditional density of $\Gamma | \underset{\sim}{y}_s, \underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma$ is given by

$$\pi \left(\Gamma | \underset{\sim}{y}_s, \underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma \right) \propto \left(\frac{1}{|\Gamma|} \right)^\ell \exp \left\{ \frac{-1}{2} \text{trace} \left[(\Gamma' \Sigma \Gamma)^{-1} \left(\sum_{i=1}^\ell \underset{\sim}{\nu}_i \underset{\sim}{\nu}_i' \right) \right] \right\}. \tag{3.18}$$

Using similar notations as in the nonbenchmarking model, we describe the distribution of the finite population mean under the benchmarking model.

Theorem 3.4. The Bayesian predictive distribution of $\overline{\underset{\sim}{Y}}_i | \underset{\sim}{y}_s$ is

$$p \left(\overline{\underset{\sim}{Y}}_i | \underset{\sim}{y}_s \right) = \int p \left(\overline{\underset{\sim}{Y}}_i | \underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2, \underset{\sim}{y}_s \right) \pi \left(\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2 | \underset{\sim}{y}_s \right) d\underset{\sim}{\nu} d\underset{\sim}{\beta} d\Sigma. \tag{3.19}$$

Using properties of multivariate normal distribution, the distribution of the finite population mean vector, $\overline{\underset{\sim}{Y}}_i$ given $\underset{\sim}{y}_s, \underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma$ is

$$\overline{\underset{\sim}{Y}}_i | \underset{\sim}{y}_s, \underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma$$
$$\overset{ind}{\sim} \text{Normal} \left\{ f_i \overline{\underset{\sim}{y}}_{s_i} + (1 - f_i) \left(\overline{X}'_{ns_i} \underset{\sim}{\beta} + \underset{\sim}{\nu}_i \right) + \underset{\sim}{A}_i, V_i \left(\frac{1 - f_i}{N_i} \right) \Sigma \right\}, \tag{3.20}$$

where

$$\underset{\sim}{A}_i = \left(\frac{1 - f_i}{N - n} \right) \left[\left(\underset{\sim}{\omega}^{*\prime} \otimes I_q \right) \underset{\sim}{y}_s - \left(N\overline{X} - n\overline{X}_s \right)' \underset{\sim}{\beta} - \left(\underset{\sim}{1}'_{N-n} Z_{ns} \otimes I_q \right) \underset{\sim}{\nu} \right],$$

and

$$V_i = 1 - \left(\frac{N_i - n_i}{N - n} \right).$$

Note that the mean of the distribution of the finite population mean under the benchmarking model has an additional $q \times 1$ vector term $\underset{\sim}{A}_i$ compared to the mean of its nonbenchmarking counterpart. Moreover, there is an adjustment of a scalar factor of V_i to the covariance matrix of the distribution of the finite population mean in the benchmarking model with reference to the nonbenchmarking model. Observe that if we set $q = 1$, (3.20) will be equivalent to distribution of the finite population mean

$\overline{Y}_i, i = 1, 2, \ldots, \ell$ in the benchmarked univariate case (see Equation (2.9) in Section 2.1). Below is the proof of the distribution of $\overline{Y}_i | y_s, \nu, \beta, \Sigma$ under the multivariate benchmarking model given in Theorem 3.4 above.

Proof.

Note that $\overline{Y}_i$ can be written as

$$\begin{aligned}
\overline{Y}_i &= \frac{1}{N_i} \sum_{j=1}^{N_i} y_{ij} = f_i \overline{y}_{s_i} + (1 - f_i) \overline{y}_{ns_i} \\
&= \frac{1}{n_i} \left(1'_{n_i} \otimes I_q \right) y_{s_i} + \frac{1}{N_i - n_i} \left(1'_{N_i - n_i} \otimes I_q \right) y_{ns_i}.
\end{aligned}$$

Hence, to prove Theorem 3.4, we need the distribution of y_s and y_{ns}. Using the conditional distribution property of the multivariate normal density, we find that $y_{ns_{(N)}} | \nu, \beta, \Sigma, y_s$ also follows a normal distribution (see Remark 3.3). The distribution of $\overline{Y}_i | y_s, \nu, \beta, \Sigma$ follows from the distribution of the nonsampled units. Thus, under the benchmarking model, the finite population mean vector $\overline{Y}_i | y_s, \nu, \beta, \Sigma, i = 1, 2, \ldots, \ell$, follows a multivariate normal distribution with mean vector

$$E\left[\overline{Y}_i | y_s, \nu, \beta, \Sigma \right] = f_i \overline{y}_{s_i} + (1 - f_i) \left(\overline{X}'_{ns_i} \beta + \nu_i \right)$$

$$+ \left\{ \left(\frac{1 - f_i}{N - n} \right) \left[\left(\omega^{*'} \otimes I_q \right) y_s - \left(N\overline{X} - n\overline{X}_s \right)' \beta - \left(1'_{N-n} Z_{ns} \otimes I_q \right) \nu \right] \right\},$$

and covariance matrix

$$Var\left[\overline{Y}_i | y_s, \nu, \beta, \Sigma \right] = \left\{ 1 - \left(\frac{N_i - n_i}{N - n} \right) \right\} \left(\frac{1 - f_i}{N_i} \right) \Sigma.$$

∎

3.4 Posterior Inference

As in the univariate case, we use Bayesian predictive procedure to make posterior inference about $\overline{Y}_i | y_s$. Our approach is to combine samples from the posterior distribution $\nu, \beta, \Sigma, \Gamma | y_s$ and the distribution of $\overline{Y}_i | \nu, \beta, \Sigma, y_s$.

We use Gibbs sampling method to draw samples from $\underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma, \Gamma | y_s$, using the full conditional distributions given in Section 3.1 for the multivariate nonbenchmarking Bayesian model and 3.3 for the multivariate benchmarking Bayesian model. Using the obtained values of the model parameters, we draw samples from $\overline{Y}_i | \underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma, y_s$ from the distributions in Theorem 3.1 and Theorem 3.4. We initialize the values of all our model parameters $\underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma, \Gamma$ using some frequentist calculations. In the Gibbs sampling process, we first draw samples from $\Gamma | y_s, \underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma$ using a grid method on its elements $\underset{\sim}{\gamma}$ and ϕ. With each set of samples of these parameters, we form the matrix Γ and use it to draw a sample from $\Sigma | y_s, \underset{\sim}{\nu}, \underset{\sim}{\beta}, \Gamma$. Then, with the new sampled values of Σ and Γ we draw a sample from $\underset{\sim}{\beta} | \underset{\sim}{\nu}, \Sigma, \Gamma, y_s$. Moreover, with the sampled values of Γ, Σ and $\underset{\sim}{\beta}$, draw values of $\underset{\sim}{\nu}$ from $\underset{\sim}{\nu} | \underset{\sim}{\beta}, \Sigma, \Gamma, y_s$. Once the parameters are obtained, we use them to draw values of the finite population means for each of the areas from the distribution of $\overline{Y}_i | \underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma, y_s$. We apply this algorithm to draw samples in both the nonbenchmarking and benchmarking models.

To facilitate inference, we generate $M = 20,000$ iterates for the Gibbs sampling method. We use 5000 burn-in values and thinned our results to 750 Gibbs samples by taking every 20th sampled value. We also use 200 grids in $(0, 1)$ to obtain our values of the $\underset{\sim}{\gamma}$ and $\underset{\sim}{\phi^*}$. Then we use the linear transformation $\phi_{ij} = 2\phi_{ij}^* - 1$ to generate ϕ. We then form the matrix Γ by simple algebra into the matrix

$$\Gamma = \text{Chol}(P)G = \begin{bmatrix} \left(\dfrac{\gamma_1}{1-\gamma_1}\right)^{\frac{1}{2}} & \phi\left(\dfrac{\gamma_2}{1-\gamma_2}\right)^{\frac{1}{2}} \\ 0 & (1-\phi^2)^{\frac{1}{2}}\left(\dfrac{\gamma_2}{1-\gamma_2}\right)^{\frac{1}{2}} \end{bmatrix}.$$

3.5 Applications

In this section, we present two illustrative examples to compare the benchmarking model with its nonbenchmarking counterpart. We summarize the results using the posterior mean (PM), posterior standard deviation (PSD), and the 95% credible interval (Int) of the finite population mean of each area. In our first example, we apply our model to simultaneously estimate the crop data, that is, corn and soybeans hectares, from the 1978 enumerative survey and satellite data. In our second example, we simultaneously estimate the body mass index (BMI) and bone mineral density (BMD) of the Mexican American adult domains (gender-county groups of population over 20 years). These are

the same examples discussed in Chapter 2. We will compare later the results of the multivariate with their univariate counterpart. As in Chapter 2, the enumerative survey and satellite data use equal weights, but for the BMI and BMD, there are differential probabilities of selection (i.e., varying sampling weights).

3.5.1 Corn and Soybeans Hectares Data

Below is the summary of the posterior outcomes of the Gibbs sampling for the multivariate result for the finite population means of the crop data.

Table 3.2: Comparison of the NBM and BM models for the finite population mean of the crop hectares data for the counties in Iowa

Domain	Direct Estimates		Nonbenchmarking Posterior Mean$_{(SD)}$		Corr	Benchmarking Posterior Mean$_{(SD)}$		Corr
i	Corn	Soy	Corn	Soy		Corn	Soy	
1	165.78	8.09	$124.05_{(8.75)}$	$79.16_{(11.47)}$	-0.53	$124.51_{(7.48)}$	$77.96_{(10.12)}$	-0.47
2	96.32	106.03	$122.08_{(7.98)}$	$95.03_{(10.89)}$	-0.44	$122.96_{(7.15)}$	$93.69_{(10.69)}$	-0.49
3	76.08	103.60	$116.26_{(8.66)}$	$87.09_{(11.76)}$	-0.44	$116.05_{(8.62)}$	$85.36_{(11.04)}$	-0.42
4	150.89	35.15	$120.19_{(8.92)}$	$81.23_{(10.98)}$	-0.35	$121.07_{(8.42)}$	$79.41_{(10.69)}$	-0.38
5	158.62	52.47	$138.66_{(8.20)}$	$67.68_{(8.48)}$	-0.49	$139.70_{(7.50)}$	$66.04_{(7.45)}$	-0.42
6	102.52	118.70	$108.14_{(6.58)}$	$114.10_{(7.54)}$	-0.42	$108.89_{(6.14)}$	$112.91_{(7.37)}$	-0.42
7	112.77	88.57	$115.03_{(6.45)}$	$97.13_{(7.60)}$	-0.37	$115.22_{(6.01)}$	$95.55_{(7.17)}$	-0.45
8	144.30	97.80	$120.94_{(6.83)}$	$112.07_{(7.30)}$	-0.38	$121.60_{(5.74)}$	$110.16_{(7.46)}$	-0.39
9	117.60	112.98	$111.23_{(6.23)}$	$110.47_{(6.69)}$	-0.40	$111.80_{(5.83)}$	$108.62_{(6.42)}$	-0.47
10	109.38	117.48	$123.45_{(5.19)}$	$100.44_{(5.76)}$	-0.42	$124.04_{(5.15)}$	$98.90_{(5.82)}$	-0.44
11	110.25	117.84	$110.94_{(6.99)}$	$17.50_{(6.71)}$	-0.44	$111.36_{(6.47)}$	$116.11_{(6.16)}$	-0.33
12	114.81	89.77	$132.12_{(5.43)}$	$75.28_{(5.99)}$	-0.46	$132.77_{(4.98)}$	$73.65_{(5.72)}$	-0.37

As in the univariate results, Table 3.2 shows that the posterior means from the two models are mostly the same, with a slight shifting in the posterior means from NBM to BM. These estimates from NBM are similar to those in Battese, Harter, and Fuller (1988). In Table 3.2, the posterior standard deviations (PSD) in BM for all counties in both crops (corn and soybeans) are smaller than the PSDs in NBM. However, notice that the gains in precision observed are not significantly large, as we have seen in the univariate case. From Table 2.1, the means for the corn (soybean) data are $N^{-1} \sum_{i=1}^{\ell} N_i \overline{y}_i =$

$120.32(95.35) = N^{-1} \sum_{i=1}^{\ell} N_i PM_{i(BM)}$, but $N^{-1} \sum_{i=1}^{\ell} N_i PM_{i(NBM)} = 119.48(96.89)$. Once again, this is the effect of our benchmarking constraint, that is, the value of the overall posterior mean from BM is equal to the sample mean of the data, but that of NBM is not equal. To further assess the performance of BM over NBM, we looked at the distance between the direct estimators $\overline{y}_i$ and the PM_i from both NBM and BM. As a summary measure, we computed the shrinkage $SHR = \sqrt{\sum_{i=1}^{\ell} (PM_i - \overline{y}_i)^2}$. For the corn data, $SHR_{(NBM)} = 81.05$ and $SHR_{(BM)} = 80.84$. On the other hand, for the soybeans data, $SHR_{(NBM)} = 92.81$ and $SHR_{(BM)} = 91.03$. For both corn and soybeans data, we find that BM has a slightly lower shrinkage than NBM, indicating that BM is slightly more design-consistent than NBM. We use box plots to show the general features in Figure 3.1.

3.5.2   Body Mass Index and Bone Mineral Density Data

Table 3.3 presents the posterior results of the Gibbs sampling for the multivariate result for the finite population means of the BMI and BMD data from NHANES III (1996).

Table 3.3: Comparison of the nonbenchmarking and benchmarking models for the finite population mean body mass index (BMI) and finite population mean bone mineral density (BMD) of adult domains (gender domains of Hispanic population over 20 years) in the state of New York

					Nonbenchmarking			Benchmarking		
i	n	N	Direct Estimate		Posterior Mean$_{(SD)}$		Corr	Posterior Mean$_{(SD)}$		Corr
			BMI	BMD	BMI	BMD		BMI	BMD	
1	1	6	19.30	0.97	$22.81_{(1.92)}$	$0.92_{(0.095)}$	0.03	$23.98_{(1.82)}$	$0.87_{(0.061)}$	0.18
2	1	112	26.80	0.98	$24.60_{(1.77)}$	$0.91_{(0.088)}$	0.04	$24.33_{(1.68)}$	$0.91_{(0.051)}$	0.17
3	2	96	23.30	0.96	$23.49_{(1.52)}$	$0.92_{(0.074)}$	0.07	$25.62_{(1.45)}$	$0.90_{(0.045)}$	0.16
4	8	150	23.44	0.93	$23.75_{(1.11)}$	$0.89_{(0.053)}$	0.16	$24.17_{(1.12)}$	$0.87_{(0.041)}$	0.30
5	4	39	23.00	1.09	$23.32_{(1.51)}$	$0.97_{(0.077)}$	0.03	$24.23_{(1.31)}$	$0.88_{(0.048)}$	0.22
6	2	6	22.70	0.84	$23.71_{(1.47)}$	$0.86_{(0.073)}$	0.06	$26.53_{(1.97)}$	$0.89_{(0.065)}$	0.09
7	2	118	26.65	0.86	$25.18_{(1.58)}$	$0.85_{(0.081)}$	0.01	$25.19_{(1.37)}$	$0.90_{(0.043)}$	0.13
8	5	101	21.38	0.78	$23.12_{(1.43)}$	$0.83_{(0.069)}$	0.21	$24.86_{(1.23)}$	$0.92_{(0.043)}$	0.18
9	3	142	27.80	0.80	$25.79_{(1.64)}$	$0.83_{(0.078)}$	0.06	$25.96_{(1.23)}$	$0.89_{(0.043)}$	0.14
10	2	37	26.10	0.88	$24.89_{(1.57)}$	$0.86_{(0.079)}$	0.07	$25.20_{(1.48)}$	$0.90_{(0.050)}$	0.24

Table 3.3 present the multivariate results for the finite population mean BMI and BMD of the Mexican American adult domains. Observe that the tables show that the posterior means (PM) of the benchmarking models shifts the direct estimates, as well as the posterior means from NBM. Under BM, the benchmarking constraint implies that $\hat{\bar{Y}} = N^{-1} \sum_{i=1}^{10} \sum_{j=1}^{n_i} \omega_{ij} y_{ij}$. From our calculations, we find that $N^{-1} \sum_{i=1}^{10} \sum_{j=1}^{n_i} \omega_{ij} y_{ij} = 24.99 = N^{-1} \sum_{i=1}^{10} N_i PM_{i(BM)}$ for the BMI data, while $N^{-1} \sum_{i=1}^{10} \sum_{j=1}^{n_i} \omega_{ij} y_{ij} = 0.90 = N^{-1} \sum_{i=1}^{10} N_i PM_{i(BM)}$ for BMD. However, $N^{-1} \sum_{i=1}^{10} N_i PM_{i(NBM)} = 24.35$ for BMI while $N^{-1} \sum_{i=1}^{10} N_i PM_{i(NBM)} = 0.80$ for the BMD value.

The BMI data has a shrinkage of $SHR_{(BM)} = 5.3049$, while the BMD data has a

shrinkage of $SHR_{(BM)} = 0.0151$. The corresponding NBM shrinkages has for BMI and BMD, respectively are $SHR_{(NBM)} = 5.645$, and $SHR_{(NBM)} = 0.1777$. The PSDs of all models are very close to each other. However, we observe that the PSDs from BM are mostly larger than those from NBM, which reflects the adjustments from the survey weights in the constraint. As the survey weights are not incorporated in an informative manner (i.e., non ignorable design) in the model, there may not be gains in precision. But observe that there is a great increase in precision in most of the benchmarking results for BMD compared to its nonbenchmarking counterpart. This more significant gain in precision might be due to the correlation between BMI and BMD, which we see to be always positive in all areas. This positive correlation is due to the fact that we have modeled them against age, and since BMI and BMD both depend on factors that relate to age, we see a positive correlation between BMI and BMD. Again, we use box plots to show the general features in Figure 3.2.

As a summary, in the two examples we observe the effects of the benchmarking constraint. As our theory shows, we observe that the model-based estimates from the small areas match the chosen direct estimator for the single large area when these small areas are collapsed. We also observe that there is a gain in precision in using the benchmarking models, relative to the nonbenchmarking model, although the gains are not always so significant because the samples are much too small, with a few exceptions.

3.6 Comparison of Univariate and Multivariate Results

In Table 3.4 we present a comparison of the univariate and multivariate results for BMI and BMD. Observe that the multivariate benchmarking results has the most precise result, with a few exceptions in the case of the BMI, where the univariate benchmarking result is better. We can say that most of the gains in precision can be attributed to the incorporation of the benchmarking constraint in the models and the borrowing of strength of the two responses between each other, as well as with the other small areas.

Table 3.4: Comparison of the nonbenchmarking (NBM) and benchmarking (BM) models (univariate and multivariate) for the finite population mean body mass index (BMI) of adult domains (gender domains of Hispanic population over 20 years) in the state of New York

				Univariate		Multivariate	
			Direct	Posterior			
i	n	N	Estimate	Mean$_{(SD)}$			
				NBM	BM	NBM	BM
			Body Mass Index (BMI)				
1	1	6	19.30	$22.42_{(2.09)}$	$22.80_{(2.15)}$	$22.81_{(1.92)}$	$23.98_{(1.82)}$
2	1	112	26.80	$24.68_{(1.82)}$	$25.55_{(1.59)}$	$24.60_{(1.77)}$	$24.33_{(1.68)}$
3	2	96	23.30	$23.77_{(1.52)}$	$24.34_{(1.41)}$	$23.49_{(1.52)}$	$25.62_{(1.45)}$
4	8	150	23.44	$23.78_{(1.12)}$	$24.33_{(1.00)}$	$23.75_{(1.11)}$	$24.17_{(1.12)}$
5	4	39	23.00	$23.61_{(1.42)}$	$24.12_{(1.34)}$	$23.32_{(1.51)}$	$24.23_{(1.31)}$
6	2	6	22.70	$23.43_{(1.64)}$	$23.84_{(1.62)}$	$23.71_{(1.47)}$	$26.53_{(1.97)}$
7	2	118	26.65	$25.10_{(1.66)}$	$25.93_{(1.38)}$	$25.18_{(1.58)}$	$25.19_{(1.37)}$
8	5	101	21.38	$22.64_{(1.36)}$	$23.09_{(1.34)}$	$23.12_{(1.43)}$	$24.86_{(1.23)}$
9	3	142	27.80	$25.74_{(1.60)}$	$26.48_{(1.35)}$	$25.79_{(1.64)}$	$25.96_{(1.23)}$
10	2	37	26.10	$24.92_{(1.63)}$	$25.57_{(1.48)}$	$24.89_{(1.57)}$	$25.20_{(1.48)}$
			Bone Mineral Density (BMD)				
1	1	6	0.97	$0.90_{(0.089)}$	$0.92_{(0.084)}$	$0.92_{(0.095)}$	$0.87_{(0.061)}$
2	1	112	0.98	$0.91_{(0.079)}$	$0.94_{(0.070)}$	$0.91_{(0.088)}$	$0.91_{(0.051)}$
3	2	96	0.96	$0.92_{(0.070)}$	$0.94_{(0.062)}$	$0.92_{(0.074)}$	$0.90_{(0.045)}$
4	8	150	0.93	$0.89_{(0.051)}$	$0.91_{(0.041)}$	$0.89_{(0.053)}$	$0.87_{(0.041)}$
5	4	39	1.09	$0.97_{(0.072)}$	$0.99_{(0.064)}$	$0.97_{(0.077)}$	$0.88_{(0.048)}$
6	2	6	0.84	$0.84_{(0.073)}$	$0.85_{(0.069)}$	$0.86_{(0.073)}$	$0.89_{(0.065)}$
7	2	118	0.86	$0.86_{(0.069)}$	$0.89_{(0.057)}$	$0.85_{(0.081)}$	$0.90_{(0.043)}$
8	5	101	0.78	$0.81_{(0.060)}$	$0.84_{(0.053)}$	$0.83_{(0.069)}$	$0.92_{(0.043)}$
9	3	142	0.80	$0.85_{(0.062)}$	$0.87_{(0.052)}$	$0.83_{(0.078)}$	$0.89_{(0.043)}$
10	2	37	0.88	$0.87_{(0.070)}$	$0.89_{(0.064)}$	$0.86_{(0.079)}$	$0.90_{(0.050)}$

3.7 Concluding Remarks

In this chapter, we presented the multivariate Bayesian benchmarking model. We discuss the Bayesian regression model for any number of responses, $q \geq 2$. There is a practical significance to the extension of the benchmarking model to the multivariate case because in many small area problems, data on related multiple characteristics and auxiliary variables are available. As in the univariate case, we used Bayesian predictive inference to estimate the finite population mean vector of small areas. We discussed both the theoretical and numerical effects of adding the benchmarking constraint into the Bayesian model. From the results of the examples as well as the mathematical calculation, we found that there were adjustment in the posterior mean vector of the finite population mean after the benchmarking constraint was included in the model, as well

as a gain in precision due to the decrease in the covariance of the distribution of the finite population mean vector. However, sometimes, not much improvement is observed, because our sample sizes are relatively too small compared to the population sizes.

We used again the two crops, corn and soybean, in the survey and satellite data of Battese, Harter, and Fuller (1988) and the body mass index (BMI) and bone mineral density (BMD) from NHANES III (1996), both pairs being correlated characteristics as illustration of our multivariate results. We compared also our work with the work of Datta, Day, and Maiti (1998) who presented a super population approach to obtain a hierarchical Bayes prediction of the small area mean vector using the multivariate nested-error regression model of Fuller and Harter (1987). They compared the performance of the multivariate approach with the usual univariate approach and found that the multivariate approach may result in substantial improvement over its univariate counterpart through some data analysis and simulations. Our model is also adapted from the Fuller and Harter (1987) multivariate nested-error regression model. However, the work of Datta, Day, and Maiti (1998) does not include benchmarking in their model, so we can only compare it to our nonbenchmarking model. Like that of Datta, Day, and Maiti (1998), we also compared the performance of the multivariate approach with the univariate approach both under nonbenchmarking and benchmarking. We are currently working on the journal article on the results of this chapter; see Toto and Nandram (2010b).

Figure 3.1: Box plots of the distribution of the finite population mean $\left(\overline{Y}\right)$ by county (area) of the corn and soybeans acreage data
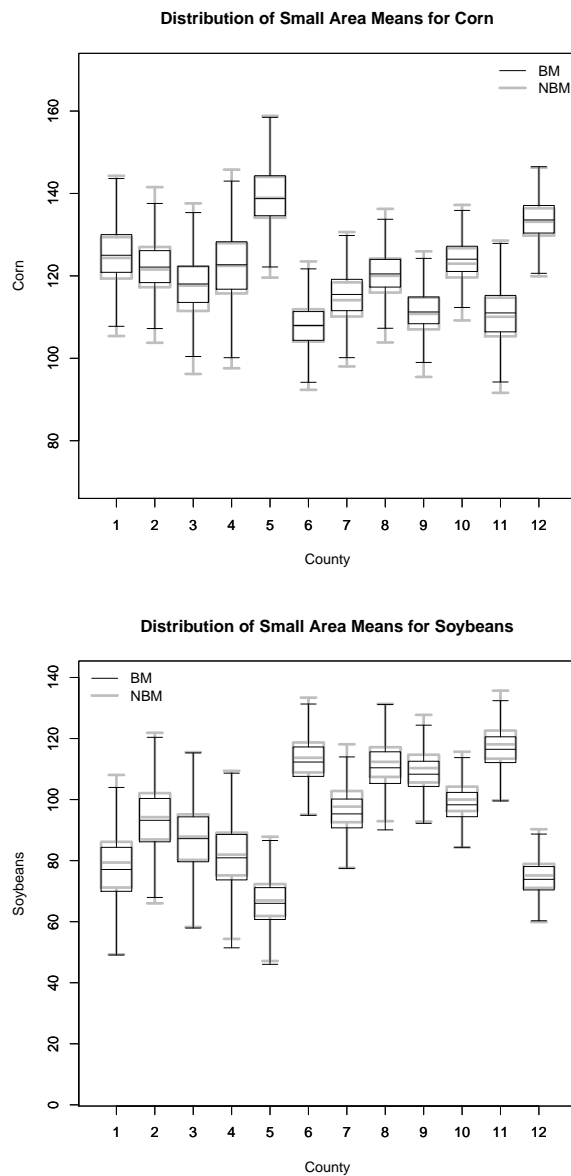
Note: As in the univariate case, we see a consistent upward shift in the center of the distribution of the finite population means for the corn acreage finite population means. For the soybeans data, we observe a consistent downward shift in the center of the distributions. There are minimal difference in the lengths of the whiskers of most of the BM boxplots compared to the NBM boxplots, indicating a slight increase in the precision in most of the domains; see $i = 8$ and $i = 10$ for soybeans, where there the variability in the NBM model is slightly better than the BM results.

Figure 3.2: Boxplots of the distribution of the finite population mean $\left(\overline{Y}\right)$ by domain of the BMI and BMD data

Distribution of Small Area Means for BMI

Distribution of Small Area Means for BMD
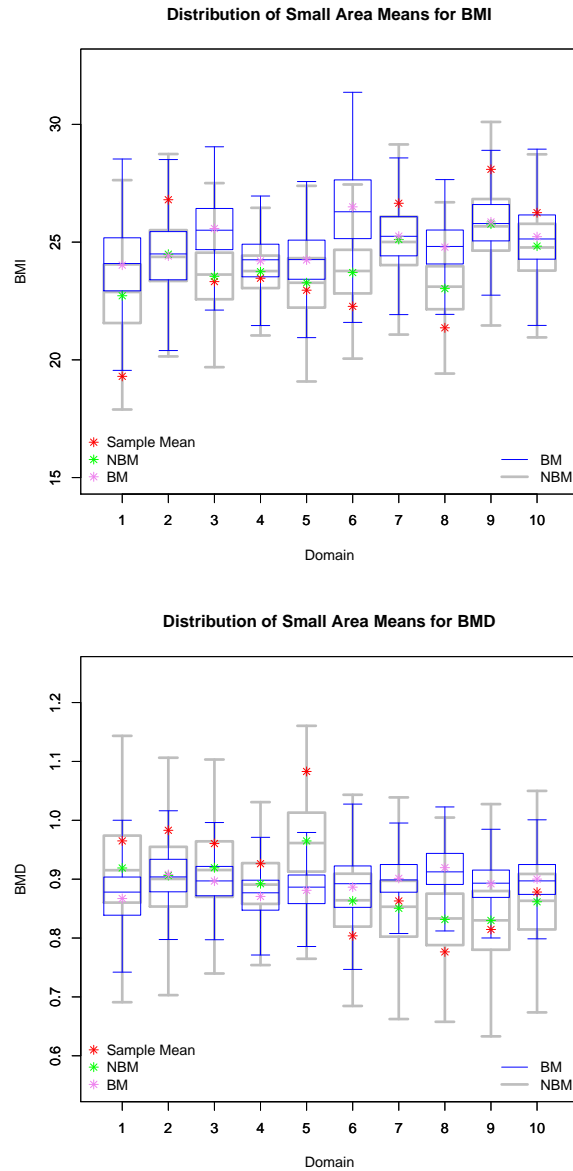
Note: Aside from the boxplots for each of the $\ell = 10$ small areas, we also plot the direct estimates for each areas, the sample means $\overline{y}_i$. We observe a shift in the center of the distribution of the finite population means for BMI and BMD when we compare the center for the BM and the NBM models. For the BMI data, notice that most of the time, there is an upward shift in the center of the distributions of the BM results compared to the NBM model. On the other hand, there is a downward shift in the center of the distributions for the BMD data in the BM model compared to the NBM model. Notice also the shorter whiskers (especially for the BMD results) of most of the BM boxplots compared to the NBM boxplots, indicating an increase in the precision in most of the domains.

71

Chapter 4

Conclusion

In this chapter, we summarize our contributions both in methodology and applications and discuss some future work.

4.1 Methodological Contributions

We developed a Bayesian predictive inference to estimate finite population means under a nested error regression model using internal benchmarking. The benchmarking constraint was constructed using a direct consistent estimator of the overall population mean based on the unit-level data. We incorporated unit level survey weights into the benchmarking constraint with the sampled data.

For the univariate Bayesian benchmarking model, conditional on the benchmarking constraint, we obtained closed forms conditional distribution of the model parameters. The proof of the propriety of the joint posterior distribution showed a simple procedure to obtain samples from the posterior distribution of the parameters. Thus, we do not rely on Markov chain Monte Carlo (MCMC) methods to make inferences; we simply draw random samples from the posterior density of $\rho|\underset{\sim}{y}_s$ and the composition method to obtain samples of the other parameters, and the finite population means. The introduction of the intra-class correlation ρ into the model is one of our novel ideas in this dissertation. Because of the introduction of ρ into the model and the use of unit-level survey weights in the benchmarking constraint are new, both the nonbenchmarking and benchmarking Bayesian model are our own results. We developed Bayesian predictive

distribution for the finite population means.

To get our multivariate Bayesian benchmarking model, we extended the structure of the univariate models to be able to accommodate the multivariate outcomes. First, we generalize the implementation of the intra-class correlation into the nested-error regression model by using a more specific structure of the covariance matrix for the area effects vector, $\underset{\sim}{\nu}_i$, instead of using a simple positive-definite, symmetric matrix for the covariance, as in Datta, Day, and Maiti (1998). This is another novel idea we have in this dissertation. This covariance structure for the area effects gives a direct link to the data and the area effects, however, it also made the model more complex. Due to the complexity of the model, we cannot use the same sampling-based process (i.e., random samples), we used for the univariate model to sample from our multivariate model. We used Gibbs sampling to make inference about the finite population mean vector. Thus, we generated the full conditional distributions of all the model parameters and iterate between these resulting full conditional distributions to generate samples of the parameters, as well as the finite population mean vector. Also, we have not found any other literature on multivariate Bayesian model that incorporates benchmarking. Thus, the introduction of benchmarking in the multivariate Bayesian nested-error regression model is also our original idea.

In both the univariate and multivariate case, we observed important changes in the distribution of the finite population mean when the benchmarking constraint was added to the model. Let us consider the multivariate case, since the univariate case is a special case when $q = 1$. First note that the distribution of the finite population mean vector $\overline{\underset{\sim}{Y}}_i$ given $\underset{\sim}{y}_s, \underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma$ under the nonbenchmarking model is

$$\overline{\underset{\sim}{Y}}_i | \underset{\sim}{y}_s, \underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma \overset{ind}{\sim} \text{Normal} \left\{ f_i \overline{\underset{\sim}{y}}_{s_i} + (1 - f_i) \left(\overline{X}'_{ns_i} \underset{\sim}{\beta} + \underset{\sim}{\nu}_i \right), \left(\frac{1 - f_i}{N_i} \right) \Sigma \right\}.$$

With the addition of the benchmarking constraint into the model, the distribution of the finite population mean vector $\overline{Y}_i$ given y_s, ν, β, Σ under the benchmarking model is

$$\overline{Y}_i | y_s, \nu, \beta, \Sigma$$

$$\overset{ind}{\sim} \text{Normal} \left\{ f_i \overline{y}_{s_i} + (1 - f_i) \left(\overline{X}'_{ns_i} \beta + \nu_i \right) + A_i, V_i \left(\frac{1 - f_i}{N_i} \right) \Sigma \right\},$$

where

$$A_i = \left(\frac{1 - f_i}{N - n} \right) \left[\left(\omega^{*\prime} \otimes I_q \right) y_s - \left(N \overline{X} - n \overline{X}_s \right)' \beta - \left(1'_{N-n} Z_{ns} \otimes I_q \right) \nu \right],$$

and

$$V_i = 1 - \left(\frac{N_i - n_i}{N - n} \right).$$

From these two distributions, we note that there is an additional term $+A_i$ added to the mean vector of the distribution of the finite population mean vector. This term causes a shift in the center of the distribution of the finite population mean due to the benchmarking constraint. Moreover, note that $V_i = 1 - \left(\frac{N_i - n_i}{N - n} \right) < 1$, so the addition of this term in the covariance of the distribution of the finite population mean causes the variances of the finite population mean to be smaller, which implies a possible gain in precision in the distribution of the finite population mean. However, the gain in precision might be offset for some cases when the variability of ν, β, Σ is incorporated. In general though, the gain in precision is small.

4.2   Applications

We applied our nonbenchmarking and benchmarking Bayesian predictive models (both univariate and multivariate) to the corn and soybeans data (LANDSAT satellite data and enumerative survey) and to the body mass index (BMI) and bone mineral density (BMD) data (NHANES III and Census). We found that relative to the nonbenchmarking model, the benchmarking model contributes an improvement in estimating the finite population mean. The benchmarking results showed slightly more precise estimates for most cases with an assurance that the combined estimates from all areas

matches the single direct estimate from the large area when the small areas are collapsed. This is a desirable result for a statistical practitioner of model-based estimation.

From the outcome of our numerical calculations, comparing the nonbenchmarking and benchmarking models, both the univariate and multivariate for the finite population mean, we found that the multivariate benchmarking model gives the most precise estimates of the finite population mean. Moreover, we see that both the univariate and multivariate benchmarking models showed that the combined estimates from all areas match the single direct estimate from the large area when the small areas are collapsed; see Table 4.1 has a summary of both datasets we studied. The values in Table 4.1 are the values of the sample mean $\overline{y} = n^{-1} \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} y_{ij}$ for each of the four dataset and the combined average of the posterior means $\hat{\overline{Y}} = N^{-1} \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \omega_{ij} \mathrm{PM}_{ij}$ from both nonbenchmarking and benchmarking model. Table 4.1 summarizes the effect of benchmarking we mentioned earlier. That is, the combined estimates of the finite population means from the benchmarking model would match the direct estimate of the mean, which the nonbenchmarking model does not achieve.

Table 4.1: Comparison of the nonbenchmarking and benchmarking Bayesian models for Finite Population Means

Variable Name	Direct Estimate	Nonbenchmarking		Benchmarking	
		Univariate	Multivariate	Univariate	Multivariate
Corn	120.32	119.67	119.48	120.32	120.32
Soybeans	95.35	96.87	96.89	95.35	95.35
BMI	24.99	23.78	24.35	24.99	24.99
BMD	0.90	0.88	0.88	0.90	0.90

Relative to the nonbenchmarking model, we observe in the two illustrative examples that there is a shift in the posterior densities of the $\overline{Y}_i$; see Figure 2.1 , Figure 2.2, Figure 3.1 and Figure 3.2. Moreover, the examples showed an increase in the precision of the estimates of the finite population mean in each area under the benchmarking model. This increased precision is reflected in the slightly narrower 95% credible intervals in the benchmarking model and shorter whiskers in the boxplots of the distributions shown in the figures.

Our simulation study also showed that the benchmarking model performs better than the nonbenchmarking model. Both the point estimators of the finite population means from the benchmarking and nonbenchmarking models were found to be very close to the true values of the finite population means. Also, we note that there is not much shrinkage or bias observed in using our estimators compared to the survey direct consistent estimate of the finite population means. The simulation also indicated the gain in precision from using the benchmarking model compared to the nonbenchmarking model. The smaller RMSE from the benchmarking model showed that it does a better job in estimating the finite population mean than the nonbenchmarking model. In most cases the coverage probabilities are better under the benchmarking constraint. Thus, overall we conclude that the benchmarking model contributed an improvement in estimating the finite population means by giving more precise and less biased estimates than the nonbenchmarking model. Of course, this makes the benchmarking approach appealing to survey practitioners. The effects of sampling weights were not included in the simulation study because we simulated the corn data with equal weights for all the samples.

4.3   Future Work

The following problems are next in this research.

1. **Simulation**

   Similar to the univariate model, we also want to further study the properties of the multivariate benchmarking model. Thus, we want to perform some simulation study on the multivariate benchmarking Bayesian model. Once all these are done, we would like to create an R library of our computations so that the statistical community would be able to use our results in doing their statistical research. We also want to include sampling weight in the simulations, both univariate and multivariate.

2. **Propriety of Models**

(a) **Propriety of the Multivariate Nonbenchmarking Bayesian Model**

We believe that the joint posterior density from our multivariate non-benchmarking model, $\pi_{\mathrm{NBM}}\left(\underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma, \Gamma | \underset{\sim}{y}_s\right)$, is proper.

The joint posterior distribution of the multivariate Bayesian model under the nonbenchmarking model is

$$
\pi_{\mathrm{NBM}}\left(\underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma, \Gamma | \underset{\sim}{y}_s\right) \propto \left(\frac{1}{|\Gamma|}\right)^{\ell}\left(\frac{1}{|\Sigma|}\right)^{(n+\ell+q+1)/2} \prod_{i=1}^{\ell} \exp\left\{-\frac{1}{2}\underset{\sim}{\nu}_i{}'(\Gamma'\Sigma\Gamma)^{-1}\underset{\sim}{\nu}_i\right\}
$$

$$
\times \prod_{i=1}^{\ell} \exp\left\{-\frac{1}{2}\sum_{j=1}^{n_i}\left(\underset{\sim}{y}_{ij} - \left[X_{ij}\underset{\sim}{\beta} + \underset{\sim}{\nu}_i\right]\right)'\Sigma^{-1}\left(\underset{\sim}{y}_{ij} - \left[X_{ij}\underset{\sim}{\beta} + \underset{\sim}{\nu}_i\right]\right)\right\}.
$$

Conjecture 4.1. Under the nonbenchmarking model, the joint posterior density $\pi_{\mathrm{NBM}}\left(\underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma, \Gamma | \underset{\sim}{y}_s\right)$ is proper.

See Appendix A for a discussion.

(b) **Propriety of the Multivariate Benchmarking Bayesian Model**

Similarly, we may be able to show that the joint posterior distribution from the multivariate benchmarking Bayesian model is proper. We have a second conjecture.

Conjecture 4.2. Under the benchmarking model, the joint posterior density $\pi_{\mathrm{BM}}\left(\underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma, \Gamma | \underset{\sim}{y}_s\right)$ is proper.

Proof.

To prove that the joint posterior density $\pi_{\mathrm{BM}}\left(\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2, \Gamma | \underset{\sim}{y}_s\right)$, our idea is to implement a strategy which uses the propriety of the nonbenchmarking Bayesian model, if the first conjecture is proved.

The joint posterior distribution of the multivariate Bayesian model under

benchmarking model is

$$\pi_{\text{BM}}\left(\underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma, \Gamma | \underset{\sim}{y}_s\right) \propto \left(\frac{1}{|\Gamma|}\right)^{\ell}\left(\frac{1}{|\Sigma|}\right)^{(n+\ell+q+1)/2} \tag{4.1}$$

$$\times \exp\left\{-\frac{1}{2}\left[\left(\underset{\sim}{y}_s - C_x\underset{\sim}{\beta} - C_z\underset{\sim}{\nu}\right)'[(I_n - W) \otimes \Sigma]^{-1}\left(\underset{\sim}{y}_s - C_x\underset{\sim}{\beta} - C_z\underset{\sim}{\nu}\right)\right]\right\}.$$

We can do this process once we prove that the joint posterior distribution under the nonbenchmarking model is proper. We would attempt to show that the ratio of the joint posterior distribution from the benchmarking model to the joint posterior distribution from the nonbenchmarking model,

$$\frac{\pi_{\text{BM}}\left(\underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma, \Gamma | \underset{\sim}{y}_s\right)}{\pi_{\text{NBM}}\left(\underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma, \Gamma | \underset{\sim}{y}_s\right)},$$

is uniformly bounded.

3. **Transformation**

In the nested error regression model, we assume normality on the responses, but normality may be tenuous. Thus, a transformation (e.g., logarithmic) may be needed. This is true for the data on body mass index; see Nandram and Choi (2005, 2010). However, the benchmarking is done on the original (untransformed) responses. This is a difficult problem. You and Rao (2002b) attempted an approximation for unmatched linking models. Nandram, Toto, and Choi (2010) suggested a different type of approximation using Taylor series expansion. However, it may be possible to perform this analysis without making analytic approximation.

Of course, the variables in the multivariate model might need to be transformed as well, which would make the problem more complex.

4. **Robustification**
Moreover, one can robustify the nested-error regression model and the benchmarking process by using the general skew exponential power distribution instead of the normal error; see, for example, Diciccio and Monti (2004) for an excellent discus-

79

sion of inference in the skew exponential power distribution. This robustification can help to reduce the effect of a possibly misspecified nested error regression model in small area estimation. This is also an extremely difficult problem.

5. **Informative Sampling**

We would like to relate the sampling weights to the responses as in informative sampling; see Pfeffermann, Krieger, and Rinott (1998), Sverchkov and Pfeffermann (2004), Pfeffermann and Sverchkov (2007) and Nandram, Choi, Shen, and Burgos (2006). Another approach can be obtained by improving the Horvitz-Thompson estimator using variable weights (Deville and Sarndal, 1992).

6. **Discrete Outcomes**

We plan to incorporate benchmarking into models with discrete outcomes. For example, the work of Nandram, Toto, and Katzoff (2009) on multiple responses for categorical tables can be extended to include benchmarking.

7. **Application**

Our models are not only applicable to health statistics and agriculture statistics data. We found an application of our results in modeling bacteria data. Prof. Terri A. Camesano of the Chemical Engineering Department of WPI provided a dataset on the properties of the molecules on the surface of the *Escherichia coli (E.coli)* bacteria. For this dataset, we are interested in estimating the finite population mean force of adhesion of the different *E. coli* strains. Each strain is represented by each of the eight slides of *E.coli* strains with different lipopolysaccharide (LPS) structures. Each slide has an area of $25mm^2$ and it is estimated that there are 69 bacteria for each 10 μm^2. These are our small domains in this study. We can also use the different core types as our small areas. Though in this case, instead of eight (8) domains, there are only three core types as domains, namely K12, O157, and O113. If we will be able to estimate the force of adhesion of the different strains of *E.coli*, this can be used to better understand and characterize how the *E.coli* bacteria attach to a surface under different condition and/or phys-

ical properties.

The O-antigens is considered as a highly variable component of the LPS among *E. coli* strains. It is useful for strain identification and assessing virulence. Atomic force microscopy (AFM) is used to characterize the physicochemical properties of the LPS of *E. coli* strains. Moreover, physical properties of the O-antigens, such as length and density, are important for determining binding of bacteria to biomolecules and epithelial cells. However, LPS length and density have not been well studied in characterizing the bacteria. But, LPS length and LPS density were found to be correlated with the force of adhesion. Thus, we can use this relationship to estimate the finite population mean force adhesion of each of the strains of the *E. coli* by using the LPS length and LPS density as the covariates.

Appendix A

Discussion of the Conjectures

To prove that the joint posterior density $\pi\left(\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2, \Gamma | \underset{\sim}{y}_s\right)$ under the nonbenchmarking model is proper, we write the density as

$$\pi_{\text{NBM}}\left(\underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma, \Gamma | \underset{\sim}{y}_s\right) \propto \pi_1\left(\underset{\sim}{\nu} | \underset{\sim}{\beta}, \Sigma, \Gamma, \underset{\sim}{y}_s\right) \pi_2\left(\underset{\sim}{\beta} | \Sigma, \Gamma, \underset{\sim}{y}_s\right) \pi_3\left(\Sigma, \Gamma, | \underset{\sim}{y}_s\right),$$

and attempt to show that each conditional density is proper (as in our univariate proof).

Again, recall that

$$
\begin{aligned}
\pi_{\text{NBM}}\left(\underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma, \Gamma | \underset{\sim}{y}_s\right) \propto{} & \left(\frac{1}{|\Gamma|}\right)^{\ell}\left(\frac{1}{|\Sigma|}\right)^{(n+\ell+q+1)/2} \prod_{i=1}^{\ell} \exp\left\{-\frac{1}{2}\underset{\sim}{\nu_i}'(\Gamma'\Sigma\Gamma)^{-1}\underset{\sim}{\nu_i}\right\} \\
& \times \prod_{i=1}^{\ell} \exp\left\{-\frac{1}{2}\sum_{j=1}^{n_i}\left(\underset{\sim}{y}_{ij} - \left[X_{ij}\underset{\sim}{\beta} + \underset{\sim}{\nu}_i\right]\right)'\Sigma^{-1}\left(\underset{\sim}{y}_{ij} - \left[X_{ij}\underset{\sim}{\beta} + \underset{\sim}{\nu}_i\right]\right)\right\}.
\end{aligned}
$$

After some algebraic manipulation, the joint posterior density is given by

$$
\begin{aligned}
\pi_{\text{NBM}}\left(\underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma, \Gamma | \underset{\sim}{y}_s\right) \propto{} & \left\{\prod_{i=1}^{\ell}\left(|B_i|\right)^{1/2}\right\} \\
& \times \left[\prod_{i=1}^{\ell}\exp\left\{-\frac{1}{2}\left(\underset{\sim}{\nu}_i - B_i^{-1}n_i\Sigma^{-1}\underset{\sim}{b}_i\right)'B_i\left(\underset{\sim}{\nu}_i - B_i^{-1}n_i\Sigma^{-1}\underset{\sim}{b}_i\right)\right\}\right] \quad\quad\text{(A.1)} \\
& \times \left\{\prod_{i=1}^{\ell}\left(|B_i|\right)^{-1/2}\right\}\left(\frac{1}{|\Gamma|}\right)^{\ell}\left(\frac{1}{|\Sigma|}\right)^{(n+\ell+q+1)/2}\left\{\exp\left\{-\frac{1}{2}\sum_{i=1}^{\ell}\left[\underset{\sim}{c}_i - \underset{\sim}{b}_i'\Sigma^{-1}B_i^{-1}\Sigma^{-1}\underset{\sim}{b}_i\right]\right\}\right\},
\end{aligned}
$$

where $B_i = n_i \Sigma^{-1} + (\Gamma' \Sigma \Gamma)^{-1}$, $\underset{\sim}{b}_i = \dfrac{1}{n_i} \sum\limits_{j=1}^{n_i} \left(\underset{\sim}{y}_{ij} - X'_{ij} \underset{\sim}{\beta} \right)$,

and $\underset{\sim}{c}_i = \sum\limits_{j=1}^{n_i} \left(\underset{\sim}{y}_{ij} - X'_{ij} \underset{\sim}{\beta} \right)' \Sigma^{-1} \left(\underset{\sim}{y}_{ij} - X'_{ij} \underset{\sim}{\beta} \right)$ for $i = 1, 2, \ldots, \ell$.

Observe that from (A.1), the conditional posterior density of the small area effects parameters, $\underset{\sim}{\nu}$, is given by

$$\pi_1 \left(\underset{\sim}{\nu} | \underset{\sim}{\beta}, \Sigma, \Gamma, \underset{\sim}{y}_s \right) \propto \prod_{i=1}^{\ell} \pi \left(\underset{\sim}{\nu}_i | \underset{\sim}{\beta}, \Sigma, \Gamma, \underset{\sim}{y}_s \right), \tag{A.2}$$

where

$$\pi \left(\underset{\sim}{\nu}_i | \underset{\sim}{\beta}, \Sigma, \Gamma, \underset{\sim}{y}_s \right) \quad \propto \quad (|B_i|)^{1/2} \exp \left\{ -\frac{1}{2} (\underset{\sim}{\nu}_i - B_i^{-1} n_i \Sigma^{-1} \underset{\sim}{b}_i)' B_i \left(\underset{\sim}{\nu}_i - B_i^{-1} n_i \Sigma^{-1} \underset{\sim}{b}_i \right) \right\}.$$

Hence, given $\underset{\sim}{\beta}, \Sigma, \Gamma$, and $\underset{\sim}{y}_s$, each of the ℓ area effects vector $\underset{\sim}{\nu}_i = (\nu_{i1}, \nu_{i2}, \ldots, \nu_{iq})', i = 1, 2, \ldots, \ell$, follows a Normal distribution with mean vector $E \left[\underset{\sim}{\nu} | \underset{\sim}{\beta}, \Sigma, \Gamma, \underset{\sim}{y}_s \right] = B_i^{-1} n_i \Sigma^{-1} \underset{\sim}{b}_i$ and covariance matrix $\text{Var} \left[\underset{\sim}{\nu} | \underset{\sim}{\beta}, \Sigma, \Gamma, \underset{\sim}{y}_s \right] = B_i^{-1}$. That is,

$$\underset{\sim}{\nu}_i | \underset{\sim}{\beta}, \sigma^2, \underset{\sim}{\Omega}, \underset{\sim}{y}_s \overset{ind}{\sim} \text{Normal} \left\{ B_i^{-1} n_i \Sigma^{-1} \underset{\sim}{b}_i, B_i^{-1} \right\}, \quad i = 1, 2, \ldots, \ell.$$

Now, we can write

$$\pi \left(\underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma, \Gamma | \underset{\sim}{y}_s \right) \propto \pi_1 \left(\underset{\sim}{\nu} | \underset{\sim}{\beta}, \Sigma, \Gamma, \underset{\sim}{y}_s \right) \pi \left(\underset{\sim}{\beta}, \Sigma, \Gamma | \underset{\sim}{y}_s \right).$$

However, integrating out $\underset{\sim}{\nu}$ in the joint distribution $\pi \left(\underset{\sim}{\nu}, \underset{\sim}{\beta}, \Sigma, \Gamma | \underset{\sim}{y}_s \right)$, we get

$$\underset{\sim}{y}_s | \underset{\sim}{\beta}, \Sigma, \Gamma \sim \text{Normal} \left\{ X'_s \underset{\sim}{\beta}, (I_n \otimes \Sigma) + (Z_s \otimes I_q) (I_\ell \otimes \Gamma' \Sigma \Gamma) (Z_s \otimes I_q)' \right\}.$$

Let $R = I_n \otimes \Sigma$, $F = Z_s \otimes I_q$, $G = I_\ell \otimes \Gamma' \Sigma \Gamma$, and $\Omega = R + FGF'$. Then,

$$\underset{\sim}{y}_s | \underset{\sim}{\beta}, \Sigma, \Gamma \sim \text{Normal} \left\{ X'_s \underset{\sim}{\beta}, R + FGF' \right) = \text{Normal} \left(X'_s \underset{\sim}{\beta}, \Omega \right).$$

Hence,

$$\pi \left(\underset{\sim}{y}_s | \underset{\sim}{\beta}, \Sigma, \Gamma \right) \propto \left(\frac{1}{|\Omega|} \right)^{1/2} \exp \left\{ -\frac{1}{2} (\underset{\sim}{y}_s - X'_s \underset{\sim}{\beta})' \Omega^{-1} \left(\underset{\sim}{y}_s - X'_s \underset{\sim}{\beta} \right) \right\}.$$

84

Now, letting $\widehat{\underset{\sim}{\beta}} = \left(X_s\Omega^{-1}X_s'\right)^{-1}\left(X_s\Omega^{-1}\underset{\sim}{y_s}\right)$, we can write $\pi\left(\underset{\sim}{y_s}|\underset{\sim}{\beta}, \Sigma, \Gamma\right)$ as

$$\pi\left(\underset{\sim}{y_s}|\underset{\sim}{\beta}, \Sigma, \Gamma\right) \propto \left(\frac{1}{|\Omega|}\right)^{1/2}$$

$$\times \left[\exp\left\{-\frac{1}{2}\left(\underset{\sim}{y_s} - X_s'\widehat{\underset{\sim}{\beta}}\right)'\Omega^{-1}\left(\underset{\sim}{y_s} - X_s'\widehat{\underset{\sim}{\beta}}\right)\right\}\right]\left[\exp\left\{-\frac{1}{2}\left(\underset{\sim}{\beta} - \widehat{\underset{\sim}{\beta}}\right)'X_s\Omega^{-1}X_s'\left(\underset{\sim}{\beta} - \widehat{\underset{\sim}{\beta}}\right)\right\}\right].$$

Thus, we can write $\pi\left(\underset{\sim}{\beta}, \Sigma, \Gamma|\underset{\sim}{y_s}\right)$ as

$$\pi\left(\underset{\sim}{y_s}|\underset{\sim}{\beta}, \Sigma, \Gamma\right) \propto \left\{\prod_{i=1}^{\ell}(|B_i|)^{-1/2}\right\}\left(\frac{1}{|\Gamma|}\right)^{\ell}\left(\frac{1}{|\Sigma|}\right)^{(n+\ell+q+1)/2}\left(\frac{1}{|\Omega|}\right)^{1/2}$$

$$\times \left[\exp\left\{-\frac{1}{2}\left(\underset{\sim}{y_s} - X_s'\widehat{\underset{\sim}{\beta}}\right)'\Omega^{-1}\left(\underset{\sim}{y_s} - X_s'\widehat{\underset{\sim}{\beta}}\right)\right\}\right]\left[\exp\left\{-\frac{1}{2}\left(\underset{\sim}{\beta} - \widehat{\underset{\sim}{\beta}}\right)'X_s\Omega^{-1}X_s'\left(\underset{\sim}{\beta} - \widehat{\underset{\sim}{\beta}}\right)\right\}\right].$$

Reordering the terms of this product, we get

$$\pi\left(\underset{\sim}{\beta}, \Sigma, \Gamma|\underset{\sim}{y_s}\right) \propto \pi_2\left(\underset{\sim}{\beta}|\Sigma, \Gamma, \underset{\sim}{y_s}\right)\pi\left(\Sigma, \Gamma|\underset{\sim}{y_s}\right)$$

$$\propto \left\{\frac{1}{\det\left[\left(X_s\Omega^{-1}X_s'\right)^{-1}\right]}\right\}^{1/2}\exp\left\{-\frac{1}{2}\left(\underset{\sim}{\beta} - \widehat{\underset{\sim}{\beta}}\right)'X_s\Omega^{-1}X_s'\left(\underset{\sim}{\beta} - \widehat{\underset{\sim}{\beta}}\right)\right\}$$

$$\times \left\{\frac{1}{\det\left[\left(X_s\Omega^{-1}X_s'\right)^{-1}\right]}\right\}^{-1/2}\left\{\prod_{i=1}^{\ell}(|B_i|)^{-1/2}\right\}\left(\frac{1}{|\Gamma|}\right)^{\ell}\left(\frac{1}{|\Sigma|}\right)^{(n+\ell+q+1)/2}\left(\frac{1}{|\Omega|}\right)^{1/2}$$

$$\times \exp\left\{-\frac{1}{2}\left(\underset{\sim}{y_s} - X_s'\widehat{\underset{\sim}{\beta}}\right)'\Omega^{-1}\left(\underset{\sim}{y_s} - X_s'\widehat{\underset{\sim}{\beta}}\right)\right\}.$$

Hence, the vector of regression coefficients $\underset{\sim}{\beta} = \left(\underset{\sim}{\beta}_1', \ldots, \underset{\sim}{\beta}_q'\right)'$, given Σ, Γ, and $\underset{\sim}{y_s}$ follows a Normal distribution with mean vector $E\left[\underset{\sim}{\beta}|\Sigma, \Gamma, \underset{\sim}{y_s}\right] = \widehat{\underset{\sim}{\beta}}$ and covariance matrix $\text{Var}\left[\underset{\sim}{\beta}|\Sigma, \Gamma, \underset{\sim}{y_s}\right] = \left(X_s\Omega^{-1}X_s'\right)^{-1}$. That is,

$$\underset{\sim}{\beta}|\Sigma, \Gamma, \underset{\sim}{y_s} \sim \text{Normal}\left\{\widehat{\underset{\sim}{\beta}}, \left(X_s\Omega^{-1}X_s'\right)^{-1}\right\}.$$

Then, integrating out $\underset{\sim}{\beta}$ from $\pi\left(\underset{\sim}{\beta},\Sigma,\Gamma|\underset{\sim}{y}_s\right)$, we have

$$\pi\left(\Sigma,\Omega|\underset{\sim}{y}_s\right) \propto \left\{\frac{1}{\det\left(X_s\Omega^{-1}X_s'\right)}\right\}^{1/2} \left\{\prod_{i=1}^{\ell}\left(\frac{1}{|B_i|}\right)^{1/2}\right\} \left(\frac{1}{|\Gamma|}\right)^{\ell} \left(\frac{1}{|\Sigma|}\right)^{(n+\ell+q+1)/2} \left(\frac{1}{|\Omega|}\right)^{1/2}$$
$$\times \ \exp\left\{-\frac{1}{2}\left(\underset{\sim}{y}_s - X_s'\widehat{\underset{\sim}{\beta}}\right)'\Omega^{-1}\left(\underset{\sim}{y}_s - X_s'\widehat{\underset{\sim}{\beta}}\right)\right\}.$$

After some algebraic manipulation, $\pi_3\left(\Sigma,\Gamma|\underset{\sim}{y}_s\right)$ is given by

$$\pi_3\left(\Sigma,\Gamma|\underset{\sim}{y}_s\right) \propto \left(\frac{1}{|\Gamma|}\right)^{\ell}\left(\frac{1}{|\Sigma|}\right)^{(n+\ell+q+1)/2}\left(\frac{1}{|\Omega|}\right)^{1/2}$$
$$\times \ \left\{\frac{1}{\det\left(X_s\Omega^{-1}X_s'\right)}\right\}^{1/2}\left\{\prod_{i=1}^{\ell}\left(\frac{1}{|B_i|}\right)^{1/2}\right\}$$
$$\times \ \exp\left\{-\frac{1}{2}\underset{\sim}{y}_s'\left(\Omega^{-1}-\Omega^{-1}\left[X_s'(X_s\Omega^{-1}X_s')^{-1}X_s\right]\Omega^{-1}\right)\underset{\sim}{y}_s\right\}.$$

But note that $\prod_{i=1}^{\ell}|B_i| = \prod_{i=1}^{\ell}\left|n_i\Sigma^{-1}+(\Gamma'\Sigma\Gamma)^{-1}\right| = \left|\left(Z_s'Z_s\otimes\Sigma^{-1}\right)+\left(I_\ell\otimes(\Gamma'\Sigma\Gamma)^{-1}\right)\right|$.
Hence,

$$\pi_3\left(\Sigma,\Gamma|\underset{\sim}{y}_s\right) \propto \left(\frac{1}{|\Gamma|}\right)^{\ell}\left(\frac{1}{|\Sigma|}\right)^{(n+\ell+q+1)/2}\left(\frac{1}{|\Omega|}\right)^{1/2}$$
$$\times \ \left\{\frac{1}{\det\left(X_s\Omega^{-1}X_s'\right)}\right\}^{1/2}\left\{\frac{1}{\det\left[(Z_s'Z_s\otimes\Sigma^{-1})+\left(I_\ell\otimes(\Gamma'\Sigma\Gamma)^{-1}\right)\right]}\right\}^{1/2}$$
$$\times \ \exp\left\{-\frac{1}{2}\underset{\sim}{y}_s'\left(\Omega^{-1}-\Omega^{-1}\left[X_s'(X_s\Omega^{-1}X_s')^{-1}X_s\right]\Omega^{-1}\right)\underset{\sim}{y}_s\right\}.$$

To be able to conclude that $\pi\left(\underset{\sim}{\nu},\underset{\sim}{\beta},\Sigma,\Gamma|\underset{\sim}{y}_s\right)$ is proper, we need to show that $\pi_3\left(\Sigma,\Gamma|\underset{\sim}{y}_s\right)$ is proper. Once we show that $\pi_3\left(\Sigma,\Gamma|\underset{\sim}{y}_s\right)$ is proper, we can conclude that the joint posterior density $\pi\left(\underset{\sim}{\nu},\underset{\sim}{\beta},\sigma^2,\Gamma|\underset{\sim}{y}_s\right)$ is proper. We have a form similar to $\pi_3\left(\Sigma,\Gamma|\underset{\sim}{y}_s\right)$ for the benchmarking multivariate model.

Bibliography

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*, 83, 28 – 36.

Census (2000), "United States Decennial Census 2000,Summary File 1 (using American Fact Finder)," `http://www.factfinder.census.gov/`.

Datta, G., Ghosh, M., Steorts, R., and Maples, J. (2009), "Bayesian Benchmarking with Applications to Small Area Estimation," in *Statistics #2009-01*, U. S. Census Bureau, Statistical Research Division Research Report Series, pp. 1–17, available online at http://www.census.gov/srd/papers/pdf/rss2009-01.pdf.

Datta, G. S., Day, B., and Maiti, T. (1998), "Multivariate Bayesian small area estimation: an application to survey and satellite data," in *Bayesian analysis*, Indian Statistical Institute, vol. 60 of *Sankhyā : The Indian Journal of Statistics. Series A*, pp. 344–362.

Deville, J. C. and Sarndal, C. E. (1992), "Calibration Estimators in Survey Sampling," *Journal of the American Statistical Association*, 87, 376 – 382.

Diciccio, T. J. and Monti, A. C. (2004), "On robust small area estimation using a simple random effects model," *Journal of the American Statistical Association*, 99, 439 – 450.

Fay, R. E. (1987), "Application of multivariate regression to small domain estimation," in *Small Area Statistics*, eds. Platek, R., Rao, J. N. K., Sarndal, C. E., and Singh, M. P., New York: Wiley, pp. 91–102.

Fuller, W. A. and Harter, R. M. (1987), "The Multivariate Components of Variance Model for Small Area Estimation," in *Small Area Statistics*, eds. Platek, R., Rao, J. N. K., Sarndal, C. E., and Singh, M. P., New York: Wiley, pp. 103–123.

Gelfand, A. and Smith, A. (1990), "Sampling based approaches to calculating marginal densities," *Journal of the American Statistical Association*, 85, 398 – 409.

Gelman, A. (2006), "Prior distributions for variance parameters in hierarchical models," *Bayesian Analysis*, 1, 515–533.

Ghosh, M. and Rao, J. N. K. (1994), "Small Area Estimation: An Appraisal," *Statistical Science*, 9, 55–76.

Hillmer, S. C. and Trabelsi, A. (1987), "Benchmarking of Economic Time Series," *Journal of the American Statistical Association*, 82, 1064 – 1071.

Isaki, C., Tsay, J., and Fuller, W. (2000), "Estimation of Census Adjustment Factors," *Survey Methodology*, 26, 31–42.

Nandram, B. and Choi, J. W. (2005), "Hierarchical Bayesian Nonignorable Nonresponse regression Models for Small Areas: An Application to the NHANES Data," *Survey Methodology*, 31, 73 – 84.

— (2010), "A Bayesian Analysis of Body Mass Index From Small Domains Under Nonignorable Nonresponse and Selction," *Journal of the American Statistical Association*, 105, 120 – 135.

Nandram, B., Choi, J. W., Shen, G., and Burgos, C. B. (2006), "Bayesian Predictive Inference Under Informative Sampling and Transformation," *Applied Stochastic Models in Business and Industry*, 22, 559 – 572.

Nandram, B. and Toto, M. C. S. (2010), "Bayesian Predictive Inference for Benchmarking Crop Production for Iowa Counties," *Journal of the Indian Society of Agricultural Statistics*, (Accepted).

Nandram, B., Toto, M. C. S., and Choi, J. W. (2010), "A Bayesian Benchmarking of the Scott-Smith Model for Small Areas," *Journal of Statistical Computation and Simulation*, (Accepted).

Nandram, B., Toto, M. C. S., and Katzoff, M. (2009), "Bayesian Inference for a Stratified Categorical Variable Allowing All Possible Category Choices," *Journal of Statistical Computation and Simulation*, 79, 161 – 179.

NHANES III (1996), "Third National Health and Nutrition and Examination Survey 1988 - 1994," Examination Data File Public Use Data File Documentation Number 76200.

Pfeffermann, D. and Barnard, C. H. (1991), "Some New Estimators for Small-Area Means with Application to the Assessment of Farmland Values," *Journal of Business & Economic Statistics*, 9, 73–84.

Pfeffermann, D., Krieger, A., and Rinott, Y. (1998), "Parametric Distributions of Complex Survey Data under Informative Probability Sampling," *Statistica Sinica*, 8, 1087 – 1114.

Pfeffermann, D. and Sverchkov, M. (2007), "Small-area Estimation under Informative Probability Sampling of Areas and Within Areas," *Journal of the American Statistical Association*, 102, 1427 – 1439.

Pfeffermann, D. and Tiller, R. (2006), "Small-Area Estimation With State-Space Models Subject to Benchmark Constraints," *Journal of the American Statistical Association*, 101, 1387 – 1397.

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

Scott, A. and Smith, T, M. F. (1969), "Estimation in multi-stage surveys," *Journal of the American Statistical Association*, 101, 1387 – 1397.

Sverchkov, M. and Pfeffermann, D. (2004), "Prediction of Finite Population Totals Based on the Sample Distribution," *Survey Methodology*, 30, 79 – 92.

Toto, M. and Nandram, B. (2010a), "A Bayesian Predictive Inference for Small Area Means Incorporating Covariates and Sampling Weights," *Journal of Statistical Planning and Inference*, (in Press).

Toto, M. C. S. and Nandram, B. (2010b), "Bayesian Benchmarking of the Fuller-Harter Multivariate Regression Model for Small Areas," (Working Paper).

Wang, J., Fuller, W. A., and Qu, Y. (2008), "Small area estimation under a restriction," *Survey Methodology*, 34, 29 – 36.

You, Y. and Rao, J. N. K. (2002a), "A Pseudo-Empirical Best Linear Unbiased Prediction Approach to Small Area Estimation Using Survey Weights," *The Canadian Journal of Statistics*, 30, 431 – 439.

— (2002b), "Small area estimation using unmatched sampling and linking models," *The Canadian Journal of Statistics*, 30, 3–15.

— (2003), "Pseudo hierarchical Bayes small area estimation combining unit level models and survey weights," *Journal of Statistical Planning and Inference*, 111, 197 – 208.

You, Y., Rao, J. N. K., and Dick, J. P. (2004), "Benchmarking hierarchical Bayes small area estimators in the Canadian Census undercoverage estimation," *Statistics in transition*, 6, 631 – 640.