Hierarchical Bayesian Methods for Survey Sampling and Other Applications

by

Adrijo Chakraborty

(Under the Direction of Gauri Sankar Datta and Abhyuday Mandal)

Abstract

In model-based survey sampling hierarchical Bayesian (HB) methods have gained immense popularity. One of the major reasons for this popularity remains the convenience in implementation of HB models using MCMC methods even when the models are complex. An inevitable part of this approach is the elicitation of priors for the parameters involved in the model. Authentic expert information can be incorporated by assigning suitable subjective prior distribution to the parameters. In Bayesian analysis nonsubjective or objective priors are assigned to the parameters when reliable subjective information is unavailable. In survey sampling, situations often arise when subjective prior information is unavailable; in such situations noninformative or objective Bayesian methods have great relevance. In this dissertation we study various noninformative HB models which combine information from multiple sources. These models are extensively used in small area estimation. In the first chapter we discuss the outline of the dissertation. In the next two chapters we provide robust small area estimation models which account for possible presence of outliers under two different scenarios. In the fourth chapter we develop robust Bayesian predictors of small area means which account for the possibility when random small area effects are not present for some small areas. In the fifth chapter we develop various HB methods which combine information from different surveys. The methods proposed in this dissertation involve improper priors. We have analytically shown that the posterior distributions based on these priors are proper under mild conditions.

INDEX WORDS: Combining Surveys, Hierarchical Bayes, Mixture Models, Objective Priors, Outliers, Small Area Estimation.

Hierarchical Bayesian Methods for Survey Sampling and Other Applications

by

Adrijo Chakraborty

B.Sc., The University of Calcutta, 2006

M.Sc., The University of Calcutta, 2008

A Dissertation Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2014

Hierarchical Bayesian Methods for Survey Sampling and Other Applications

by

Adrijo Chakraborty

Approved:

Major Professors: Gauri Sankar Datta
Abhyuday Mandal

Committee: Nicole Lazar
T. N. Sriram
John Stufken
Lily Wang

Electronic Version Approved:

Julie Coffield
Interim Dean of the Graduate School
The University of Georgia
August 2014

Dedication

To my parents.

Acknowledgments

I would like to convey my deepest gratitude to my major adviser Dr. Gauri Sankar Datta for his continuous support for the last few years. I have learned the fundamentals of Bayesian statistics from his Bayesian data analysis class and developed a great interest in this field. His detailed explanations of difficult concept helped me to understand my research problems better. Despite his busy schedule he had always made himself available for appointments whenever I requested, I will always remain grateful for that. My co-major professor Dr. Abhyuday Mandal has always inspired me to work on interesting problems. I am thankful to him for his guidance and encouragement. I would like to extend my sincere gratitude to my committee members Dr. Nicole Lazar, Dr. T. N. Sriram, Dr. John Stufken and Dr. Lily Wang for spending time reading my work and for sharing thoughtful comments. I truly appreciate Dr. Wang and Dr. Stufken for their encouragement and support. Thanks to Dr. Lazar for lending me books from her personal library.

My sincere appreciation goes to all my professors, especially Dr. William P. McCormick, Dr. Ishwar Basawa from University of Georgia and Dr. Gaurangadeb Chattopadhyay, Dr. Bhaswati Ganguly, Dr. Kalyan Das and Dr. Sugata Sen Roy from University of Calcutta. I would like to thank Dr. Bhaswati Ganguli and Dr. Kalyan Das for encouraging me to study abroad. I will always be grateful to Dr. Som Subhra Chatterjee for encouraging me to choose Statistics as my major when I got into college. I am greatly thankful to our graduate coordinator Dr. Lynne Seymour for her guidance and for standing by me in hard times.

My parents and my sister have always stood by me in all difficult times. I will always remain indebted to my parents for all their sacrifices for me. I would like to thank my wife Bhargabi for being a source of support and making my life wonderful. Thanks to all my friends in Athens, without their support things would have been much more difficult.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the past few decades, small area estimation has garnered considerable attention from the researchers of various fields. The importance of small area statistics lies in its wide applications, e.g., in agriculture, education, health care, and other government programs. Government agencies need reliable small area statistics for allocation of funds at national, state, county and other sub-national levels. The U.S. Census Bureau created the Small Area Income and Poverty Estimation (SAIPE) program to get accurate information on income and poverty statistics at various small area levels. In 1994 updated SAIPE county level, school district level estimates were used to allocate more than 7 billion dollars of Federal fund for disadvantaged children (Title I of the Elementary and Secondary Education Act, Citro and Kalton, 2000). In 1998, Statistical Methodology Division of Office of National Statistics, U.K. developed SAEP (Small Area Estimation Project), which aims at deriving political-ward level (about 2000 households) estimates for different variables of interest. The data was collected from General Household Survey and Family Resource Survey (Heady and Clarke, 2003).

Since health features differ from county to county, race to race, health planning is always done at state or county level, i.e., at small area level, demand of reliable small area statistics is often high among the health policy makers. To study the health status of the residents and to assess the utilization of the health care resources in the state and regional level, the National Health Interview Survey is conducted every year in U.S. by the National Center for Health Statistics (Malec et al., 1997). Every year in U.S., the National Health Planning Research and Development act requires the health system agencies to conduct surveys in order to collect and analyze data related to the health status of the people in their respective territories (Nandram, 1999 and Rao, 2003). For state level treatment planning, the U.S. Substance Abuse and Mental Health Administration obtains state level and sub-state level small area estimates for about 20 outcomes related to treatment and mental health using data from National Household Survey of Drug Use and Health Administration. The National Immunization program utilizes estimates of immunization rates for different ethnic groups in various small geographical areas (Malec et al., 1997). The National Center for Health Statistics was the first to apply implicit model-based synthetic estimation on National Health Interview Survey (Gonzalez, 1973).

In small area estimation, indirect estimators are developed by borrowing strength from the related areas or other sources through a suitable model. Indirect estimators are deemed more reliable than direct estimators since borrowing information appropriately leads to an increase in effective sample size. Indirect estimators may also use values of the variable of interest from a different time period, especially when the surveys are repeated over a period of time. Model-based approaches are widely used in indirect estimation. Model-based estimators utilize auxiliary information related to the study variable to borrow information from the related areas through an implicit or explicit linking model. The choice of a model plays an important role. If the model does not perform well then the resulting small area estimates could be biased and less effective.

Many small area models are special cases of a linear mixed model. Consider the following general linear mixed model,

$$Y = X\beta + Zv + e,$$

where Y is the vector of response variable, X is the matrix of explanatory variables, v is the vector of random effects and e is the vector of unobserved random error. The matrices X and Z are known. The regression coefficient β is usually assumed to be fixed but there are cases when some or all components of β are considered as random (Jiang and Lahiri, 2006). It is also assumed that v and e are independently normally distributed, which is expressed by $v \sim N(0, G)$, $e \sim N(0, R)$. The covariance matrices R and G usually involve unknown variance components.

Depending on the availability of the values of the response variable and auxiliary information, small area models are classified into two basic models; namely, area level model and unit-level model. When unit specific information on the response and auxiliary data is available, unit level model could be applied. For area-level models, only summary information for the small area is needed. Both of these models could be considered as special cases of the general linear mixed model. Mixed models typically include area specific random effects in order to capture the between area variation, which remain unexplained by the auxiliary information. Popularity of mixed models in small area estimation is due to their effectiveness in incorporating various sources of information.

Several methods have been developed to estimate small area quantities from both frequentist and Bayesian perspective. In this dissertation, we focus on hierarchical Bayesian (HB) approach for small area estimation. In this approach, complex models can be handled relatively easily with the MCMC technique. Moreover, the uncertainty associated with the estimators can be easily computed. Specification of priors for the model parameters in the model is one of the main concerns of Bayesian approach. Subjective priors are recommended when they are available, however, in practice these priors are often unavailable. All the methods developed in this dissertation are based on nonsubjective or objective priors.

In the next three chapters of this dissertation we develop robust Bayesian small area estimation methods. Survey estimates can be highly sensitive to the presence of outliers in the data, particulary in the context of small area estimation, when it is likely to have smaller sample sizes for some areas. In Chapter 2, we propose a robust hierarchical Bayesian small area estimation method for unit-level data. We also develop an outlier detection method which successfully detects the outliers in the data. In their seminal work, Fay and Herriot (1979) model assume that the random area effects are normally distributed although the justification of such assumption is not clear. In Chapters 3 and 4 we relax this normality assumption and propose two different alternatives to the standard Fay-Herriot model with the justification of proposing these models.

Various surveys are conducted every year by different survey agencies to estimate number of occupied households in United States. Although the objective of each survey may be the same but the results differ to a considerable extent. Such a situation may create ambiguity among the policy makers who make decisions based on these estimates. This motivates us to develop a methodology which combines these estimates and provide a single accurate estimate every year. We propose different methods to combine these estimates, some of these methods use auxiliary information and incorporate them in the model. In Chapter 5, we study the performances of these methods and assess them in terms of gain in precision.

Chapter 2

Bayesian small area estimation for unit-level data in presence of outliers

2.1 Introduction

Small area estimation has secured an important place in survey sampling due to its wide range of applicability in both government and private sectors. Sample surveys are conducted to provide adequately accurate estimates of population characteristics of interest, such as population mean, total etc. Situations frequently arise when researchers need estimates of such parameters for various subpopulations or subdomains beyond those for the entire population. A subdomain could be a geographical region, such as state, county, municipality, health service area or a socio-demographic group, for instance, a certain age-sex-race group of a population.

If an estimator uses the domain specific sample data only, then it is referred to as a direct estimator. Precision of a direct estimator depends on the domain specific sample size, and consequently it may not be reliable if the domain sample size is small. A domain is regarded as a small area or small domain if the domain sample size is not large enough to produce a direct estimator with adequate precision. Model-based indirect estimators, based on a reasonable model are considered to be more reliable in small area estimation. Precision of a direct estimator may not be adequate when the sample size for some subpopulations are small. Indirect estimators combine information from related areas or other sources thorough a linking model which increases the effective sample size and helps improving the precision. A linking model can be implicit or explicit. Both implicit and explicit models link the small areas through population level auxiliary information, explicit models incorporate between area variation by considering area specific effects in the model. In the context of explicit small area models linear mixed models are very popular. Application of linear mixed models in small area estimation has been widely discussed in Ghosh and Rao (1994), Rao (2003), Jiang and Lahiri (2006) and Datta (2009). Empirical Bayes (EB), Empirical Best Linear Unbiased Prediction (EBLUP) and Hierarchical Bayes (HB) are the well known methods in small area estimation.

Presence of outliers may affect small area estimates significantly if the estimation method does not account for it. Discarding the outliers may not always be an acceptable solution since it involves the risk of losing valuable information, particularly when the outliers represent some part of the population (Chambers, 1986). Sinha and Rao (2009) have shown that performance of well known methods such as EBLUP is highly affected by the presence of outliers. In their proposed robust method, they use Huber's influence function to reduce the adverse effect of outliers in the estimation process. Chambers et al. (2013) proposed robust small area estimation method based on M-Quantile estimation approach and discuss the success of their method by comparing its performance with other classical frequentist methods.

In this chapter, we propose an HB method for estimating small area means using unit-level data. The basic setup of our model is similar to the nested error regression model (Battese, Harter and Fuller, 1988). We suggest that sampling errors follow a two-component normal mixture model. Previously, Gershunskaya and Lahiri (2009) discussed a two-component normal mixture model for unit-level data and proposed a frequentist robust estimation method. Our method assumes that outliers come from a distribution which has a larger variance compared to the rest of the data. We impose this condition by our choice of prior for the error variance components. We propose noninformative priors for the model parameters since elicitation of subjective priors require authentic historical data, which may not always be available or accessible. Since our choices of priors include improper priors, the propriety of the posterior distribution is needed to be ensured. We provide a set of sufficient conditions under which we analytically show the resulting posterior distribution is proper.

In the following section we briefly discuss unit-level nested error regression model. In Section 2.3 we introduce our proposed method and describe implementation procedures. Results of simulation studies are reported in Section 2.4. We analyze, in Section 2.5, the county level corn data previously introduced and analyzed by Battese et al. (1988). In Section 2.7 a detailed proof of the propriety of the posterior distribution resulting from the proposed model is provided.

2.2 Unit-Level Models

Let y_{ij} be the value of the response variable y and $x_{ij} = (x_{ij1}, \ldots, x_{ijp})^T$ be the value of the auxiliary variable x for the j^{th} unit of the i^{th} small area. A basic unit-level nested error regression model (Battese et al., 1988) is given by

$$y_{ij} \;=\; x_{ij}^T\beta + v_i + e_{ij}, \quad j = 1, \ldots, N_i, \; i = 1, \ldots, m, \tag{2.2.1}$$

where N_i is the number of units in the i^{th} small area and m is the number of small areas. In this model, sampling errors e_{ij}'s and model error v_i's are independently distributed and e_{ij}'s and identically distributed as $N(0, \sigma_e^2)$ random variables. Random small area effects $v_i \overset{\text{iid}}{\sim} N(0, \sigma_v^2)$, $i = 1, \ldots, m$. The vector of regression coefficients is denoted by β $(p \times 1)$.

In this setup, we are assuming a noninformative sampling of the units, so the model given by (2.2.1) is also appropriate for the sampled units. A unit-level model assumes that auxiliary information is available for each unit in the population. However, to estimate the small area totals or means of the response variable, based on (2.2.1), in addition to the values of the auxiliary variables for the sampled units, it is sufficient to know the population means of the auxiliary variables only.

In the finite population sampling scenario, if N_i is large for the i^{th} small area, then the finite population small area mean ($\theta_i = N_i^{-1} \sum_{j=1}^{N_i} y_{ij}$) for that area is approximated by:

$$\theta_i \approx \bar{X}_i^T \beta + v_i, \tag{2.2.2}$$

where $\bar{X}_i = (\bar{X}_{i1}, \ldots, \bar{X}_{ip})$; $\bar{X}_{ik}^T$ represents the population mean of the k^{th} auxiliary variable for the i^{th} small area, $k = 1, \ldots, p$.

There are many applications of the unit-level nested error regression models. Battese et al. (1988) used model (2.2.1) and performed a variance components analysis to estimate mean areas under corn and soybeans for the counties in North Central Iowa. Datta and Ghosh (1991) proposed a hierarchical Bayesian prediction method for general linear mixed model framework, focussing on small area estimation. The model proposed by Datta and Ghosh (1991) is given by:

1. conditional on β, v, λ= $(\lambda_1, \ldots, \lambda_t)^T$ and r , Y $\sim N(X\beta + Zv, \frac{1}{r}\Omega)$,

2. conditional on λ and r, $v \sim N(0, \frac{1}{r}D(\lambda))$,

where Y is an $N \times 1$ vector of response variable, the matrices X $(N \times p)$ and Z $(N \times s)$ are known. The matrix related to sampling variance covariance Ω is a known positive definite matrix. Structure of the positive definite matrix $D(\lambda)$ is known but $\lambda = (\lambda_1, \ldots, \lambda_t)^T$ is unknown. To complete the hierarchical Bayes model specification, Datta and Ghosh (1991) used the following prior distributions: β, r, $\lambda_1 r$,...,$\lambda_t r$ are mutually independent with $\beta \sim Uniform(R^p)$, $r \sim Gamma(\frac{1}{2}g_o, \frac{1}{2}a_o)^1$; $\lambda_i r \sim Gamma(\frac{1}{2}g_i, \frac{1}{2}a_i)$, $i = 1, \ldots, t$ where

$a_i, g_i \geq 0, i = 0, \ldots t$. For $Z = \oplus_{i=1}^{m} 1_{N_i}$ and $D(\lambda) = \lambda I_m$, this model reduces to Bayesian nested error regression model. You and Rao (2003) applied Bayesian nested error regression model with a vague prior for the regression coefficients.

2.3 A Hierarchical Mixture Model for Unit-Level Data

2.3.1 Proposed Model

As before, suppose, there are m small areas. Let n_i denote the number of sampled units for i^{th} small area, y_{ij} be the random variable denoting the response for the j^{th} sampled unit in the i^{th} small area and $x_{ij} = (x_{ij1}, \ldots, x_{ijp})^T$ be the $p \times 1$ vector of auxiliary variables for that unit. In order to account for possible presence of representative outliers in the unit-level data, we propose the following model:

(I) conditional on $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$, v_i, z_{ij}, p_e, σ_1^2, σ_2^2 and σ_v^2,

$$y_{ij} \sim z_{ij} N(x_{ij}^T\boldsymbol{\beta}+v_i, \sigma_1^2)+(1 - z_{ij})\, N(x_{ij}^T\boldsymbol{\beta}+v_i, \sigma_2^2) \text{ for } j = 1, \ldots, N_i, i = 1, \ldots, m.$$

(II) The indicator variables z_{ij}'s are iid with $P(z_{ij} = 1|p_e) = p_e$ and $P(z_{ij} = 0|p_e) = 1 - p_e$, $j = 1, \ldots, n_i$, $i = 1, \ldots, m$. Also, z_{ij}'s are independent of v_i's, $\boldsymbol{\beta}$, σ_1^2, σ_2^2 and σ_v^2.

(III) Conditional on $\boldsymbol{\beta}$, z, p_e, σ_1^2, σ_2^2 and σ_v^2, $v_i \overset{\text{iid}}{\sim} N(0, \sigma_v^2)$ for $i = 1, \ldots, m$.

In this chapter, we carry out an objective Bayesian analysis by assigning non-informative priors to the model parameters. However, subjective priors could also be assigned when such subjective information is available. The following noninformative priors to the parameters are suggested: $\boldsymbol{\beta}$, (σ_1^2, σ_2^2), p_e and σ_v^2 are assumed to be mutually independent, $\boldsymbol{\beta} \sim Uniform(R^p)$, $\sigma_v^2 \sim Uniform(0, \infty)$, $\pi(\sigma_1^2, \sigma_2^2) \propto \frac{1}{(\sigma_2^2)^2} I(\sigma_2^2 > \sigma_1^2)$ and $p_e \sim Uniform(0, 1)$. In the proposed model, we assume that the outlying observations come from a distribution with the larger sampling variance. This assumption prompts us to assign the proposed prior for (σ_1^2, σ_2^2).

Since stage (I) of the model is a mixture model, it is important to use a partially proper prior for the variance parameters σ_1^2 and σ_2^2. By partially proper prior we mean that conditional on σ_1^2 ($\sigma_1^2 > 0$), σ_1^2 has a proper density, and conditional on σ_2^2 ($\sigma_2^2 > 0$), σ_1^2 has a proper density.

Since improper prior distribution has been used in the HB model above, it is important to ensure the propriety of the resulting posterior distribution in order to avoid misleading results based on improper posteriors (cf. Hobert and Casella, 1996). In the following theorems, we provide sufficient conditions for the propriety of the resulting posterior distribution based on the proposed model. Let, $n = \sum_{i=1}^{m} n_i$.

Theorem 2.3.1 *The following conditions are sufficient for the propriety of the posterior distribution under the proposed model:*

(a) $n_i \geq 2$ *for* $i = 1, \ldots, m$,

(b) $n \geq 2m + 2p - 1$,

(c) $m \geq p + 6$.

A detailed proof of Theorem 2.3.1 is provided in Section 2.7. While the Theorem 2.3.1 appears to be too restrictive, the following corollary and the lemma show that it is not the case.

Corollary 2.3.2 *If there exists a set* S *of* μ $(1 \leq \mu \leq m)$ *small areas, such that*

(a) $l_i \geq 2$, l_i *being the number of sampled units from the* i^{th} *small area,* $i \in S$,

(b) $\sum_{i \in S} l_i \geq 2\mu + 2p - 1$;

(c) $\mu \geq p + 6$,

then the posterior distribution under the proposed model will be proper.

Proof of Corollary 2.3.2: Let us state the following Lemma which has also been used in the proof of Theorem 2.3.1.

Lemma 2.3.3 *Let $\theta \sim \pi(\theta)$ and $d|\theta \sim f(d|\theta)$, we partition d as $d = (d^{(1)T}, d^{(2)T})^T$. If the posterior distribution of $\theta|d^{(1)}$ is proper, then the posterior distribution of $\theta|d$ is also proper.*

Proof of Lemma 2.3.3: Let $\pi(\theta|d)$ and $\pi(\theta|d^{(1)})$ be the posterior distributions of θ based on d and $d^{(1)}$ respectively. Let, $f(d^{(1)}|\theta) = \int f(d^{(1)}, d^{(2)}|\theta)\,\mathrm{d}d^{(2)}$. Then,

$$\pi(\theta|d) \propto f(d|\theta)\pi(\theta) \text{ and}$$
$$\pi(\theta|d^{(1)}) \propto f(d^{(1)}|\theta)\pi(\theta),$$

We assume that, $\int \pi(\theta|d^{(1)})\,\mathrm{d}\theta < \infty$. This is equivalent to, $\int f(d^{(1)}|\theta)\pi(\theta)\,\mathrm{d}\theta < \infty$. Now,

$$\begin{aligned}\int\int f(d|\theta)\pi(\theta)\,\mathrm{d}\theta\,\mathrm{d}d^{(2)} &= \int\left[\int f(d|\theta)\,\mathrm{d}d^{(2)}\right]\pi(\theta)\,\mathrm{d}\theta \\ &= \int f(d^{(1)}|\theta)\pi(\theta)\,\mathrm{d}\theta \\ &< \infty.\end{aligned}$$

This necessarily implies that $\int f(d|\theta)\pi(\theta)\,\mathrm{d}\theta < \infty$. This ensures that $\int \pi(\theta|d)\,\mathrm{d}\theta < \infty$.□

Suppose, there exists $\mu\,(\leq m)$ small areas which satisfy the conditions (a), (b) and (c) of Theorem 2.3.1. Let S_μ be the set of those small areas and S_μ^c contain rest of the small areas. Let us partition the responses for the sampled units as follows:

$$Y^{(1)} = \{y_{ij} : i \in S_\mu; j = 1, \ldots, n_i\} \text{ and } Y^{(2)} = \{y_{ij} : i \in S_\mu^c; j = 1, \ldots, n_i\}.$$

Let θ be the set of model parameters. From Theorem 2.3.1 we can say that $f(\theta|Y^{(1)})$ is proper. Now, applying Lemma 2.3.3, we can say $f(\theta|Y) = f(\theta|Y^{(1)}, Y^{(2)})$ is proper. Hence the proof of Corollary 2.3.2.

Once we ensure the propriety of the posterior resulting from the model, we can perform Gibbs sampling to implement the model. For Gibbs sampling, we need the full set of conditional posterior distributions.

According to our proposed model, the joint pdf of $y = \{y_{ij}; j = 1, \ldots, n_i, i = 1\ldots, m\}$, z $= \{z_{ij}; j = 1, \ldots, n_i, i = 1\ldots, m\}$, $\beta, v = (v_1, \ldots, v_m)^T$, $\sigma_1^2, \sigma_2^2, \sigma_v^2, z, p_e$ is

$$f(y, z, \beta, v, p_e, \sigma_1^2, \sigma_2^2, \sigma_v^2)$$

$$\propto \frac{\exp\left(-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(\frac{(y_{ij}-x_{ij}^T\beta-v_i)^2}{\sigma_1^2}z_{ij} + \frac{(y_{ij}-x_{ij}^T\beta-v_i)^2}{\sigma_2^2}(1-z_{ij})\right)\right)}{(\sigma_1^2)^{\sum_{i=1}^{m}\sum_{j=1}^{n_i}\frac{z_{ij}}{2}}(\sigma_2^2)^{\sum_{i=1}^{m}\sum_{j=1}^{n_i}\frac{(1-z_{ij})}{2}}}$$

$$\times\frac{\exp\left(-\frac{1}{2}\sum_{i=1}^{m}\frac{v_i^2}{\sigma_v^2}\right)}{(\sigma_v^2)^{\frac{m}{2}}} \times p_e^{\sum_{i=1}^{m}\sum_{j=1}^{n_i} z_{ij}}(1-p_e)^{\sum_{i=1}^{m}\sum_{j=1}^{n_i}(1-z_{ij})} \times \frac{I(\sigma_1^2 < \sigma_2^2)}{(\sigma_2^2)^2}.$$

In order to implement Gibbs sampling (Gelfand and Smith, 1990), we need the following full set of conditional distributions:

(I) $\beta|y, z, \sigma_1^2, \sigma_2^2, \sigma_v^2, p_e, v \sim N_p\Big(S_\beta \sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-v_i)x_{ij}\Big(\frac{z_{ij}}{\sigma_1^2} + \frac{1-z_{ij}}{\sigma_2^2}\Big), S_\beta\Big)$,
where $S_\beta = \Big[\sum_{i=1}^{m}\sum_{j=1}^{n_i} x_{ij}x_{ij}^T\Big(\frac{z_{ij}}{\sigma_1^2} + \frac{1-z_{ij}}{\sigma_2^2}\Big)\Big]^{-1}$.

(II) $v_i|y, \beta, z, \sigma_1^2, \sigma_2^2, \sigma_v^2, p_e \sim N\Big(\varphi_i \sum_{j=1}^{n_i}(y_{ij} - x_{ij}^T\boldsymbol{\beta})\Big(\frac{z_{ij}}{\sigma_1^2} + \frac{1-z_{ij}}{\sigma_2^2}\Big), \varphi_i\Big)$,
where $\varphi_i = \Big(\frac{1}{\sigma_v^2} + \sum_{j=1}^{n_i}\Big\{\frac{z_{ij}}{\sigma_1^2} + \frac{1-z_{ij}}{\sigma_2^2}\Big\}\Big)^{-1}$, $i = 1, \ldots, m$.

(III) $z_{ij}|y, \beta, \sigma_1^2, \sigma_2^2, \sigma_v^2, p_e, v \sim Bernoulli(p_{ij}^*)$, $j = 1, \ldots, n_i$, $i = 1, \ldots, m$, where

$$p_{ij}^* = \frac{\frac{p_e}{\sigma_1}\exp\left(-\frac{(y_{ij}-x_{ij}^T\boldsymbol{\beta}-v_i)^2}{2\sigma_1^2}\right)}{\frac{p_e}{\sigma_1}\exp\left(-\frac{(y_{ij}-x_{ij}^T\boldsymbol{\beta}-v_i)^2}{2\sigma_1^2}\right) + \frac{(1-p_e)}{\sigma_2}\exp\left(-\frac{(y_{ij}-x_{ij}^T\boldsymbol{\beta}-v_i)^2}{2\sigma_2^2}\right)}.$$

(IV) $p_e|y, z, \beta, \sigma_1^2, \sigma_2^2, \sigma_v^2, v \sim Beta\Big(\sum_{i=1}^{m}\sum_{j=1}^{n_i} z_{ij} + 1, \sum_{i=1}^{m}\sum_{j=1}^{n_i}(1-z_{ij}) + 1\Big)$.

(V) $\frac{1}{\sigma_v^2}|y, z, \beta, \sigma_1^2, \sigma_2^2, p_e, v \sim Gamma(\frac{m}{2} - 1, \frac{1}{2}\sum_{i=1}^{m} v_i^2)$.

(VI) $\sigma_1^2|y,z,\beta,\sigma_2^2,\sigma_v^2,v,p_e \sim \pi(\sigma_1^2|y,z,\beta,\sigma_2^2,\sigma_v^2,v,p_e)$

$$\propto \frac{\exp\left(-\sum_{i=1}^{m}\sum_{j=1}^{n_i}\frac{(y_{ij}-x_{ij}^T\boldsymbol{\beta}-v_i)^2}{2\sigma_1^2}z_{ij}\right)}{(\sigma_1^2)^{\sum_{i=1}^{m}\sum_{j=1}^{n_i}\frac{z_{ij}}{2}}} I(\sigma_1^2 < \sigma_2^2).$$

(VII) $\sigma_2^2|y,z,\beta,\sigma_1^2,\sigma_v^2,v,p_e \sim \pi(\sigma_2^2|y,z,\beta,\sigma_1^2,\sigma_v^2,v,p_e)$

$$\propto \frac{\exp\left(-\sum_{i=1}^{m}\sum_{j=1}^{n_i}\frac{(y_{ij}-x_{ij}^T\boldsymbol{\beta}-v_i)^2}{2\sigma_2^2}(1-z_{ij})\right)}{(\sigma_2^2)^{2+\sum_{i=1}^{m}\sum_{j=1}^{n_i}\frac{(1-z_{ij})}{2}}} I(\sigma_1^2 < \sigma_2^2).$$

Generating samples from (I)$-$(V) is straightforward, based on standard distributions such as normal, beta, gamma and Bernoulli. Steps (VI) and (VII) correspond to truncated gamma distribution. (VI) and (VII) may admit a closed form depending on the values of $\sum_{i=1}^{m}\sum_{j=1}^{n_i} z_{ij}$. Given the values of the other parameters, we generate samples from (VI) as follows:

- When $\sum_{i=1}^{m}\sum_{j=1}^{n_i} z_{ij} \geq 3$, we draw σ_1^2 from a truncated *Inverse-Gamma* (IG) distribution with shape $= (\sum_{i=1}^{m}\sum_{j=1}^{n_i} z_{ij} - 2)/2$, rate $= \frac{1}{2}\Big[\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - x_{ij}^T\boldsymbol{\beta} - v_i)^2 z_{ij}\Big]$ and the upper truncation point σ_2^2.

- When $\sum_{i=1}^{m}\sum_{j=1}^{n_i} z_{ij} = 1$ or 2, we generate samples from the importance density $g(\sigma_1^2) = \frac{1}{2\sqrt{\sigma_1^2}\sqrt{\sigma_2^2}} I(\sigma_1^2 < \sigma_2^2)$ and perform an acceptance-rejection sampling to generate samples for σ_1^2.

- When $\sum_{i=1}^{m}\sum_{j=1}^{n_i} z_{ij} = 0$, we draw σ_1^2 from *Uniform*$(0, \sigma_2^2)$.

Similarly, given the other parameters, we draw samples for σ_2^2 as follows:

- We draw σ_2^2 from a truncated *Inverse-Gamma* (IG) distribution with shape $=$ $(\sum_{i=1}^{m}\sum_{j=1}^{n_i}(1 - z_{ij})/2)+1$ and rate $= \Big[\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - x_{ij}^T\boldsymbol{\beta} - v_i)^2(1 - z_{ij})\Big]$ and the lower truncation point σ_1^2, if at least one $z_{ij} \neq 1$, $j = 1, \ldots, n_i$ and $i = 1, \ldots, m$.

- We draw σ_2^2 from a *Pareto* distribution with shape = 1 and scale = σ_1^2 if $z_{ij} = 1$, for all i, j.

The objective of our method is to estimate the small area means. Recall θ_i is the finite population mean for the i^{th} small area. If the number of population units in the i^{th} small area, N_i is large, $\theta_i \approx \bar{X}_i^T \beta + v_i$, where $\bar{X}_i = (\bar{X}_{i1}, \ldots, \bar{X}_{ip})^T$, $\bar{X}_{ik}$ represents the population mean of the k^{th} auxiliary variable for the i^{th} small area, $k = 1, \ldots, p$.

Gibbs sampling technique can be implemented in order to obtain the Bayes estimates of small area means and other model parameters. Let, $\{\beta^{(d)}, v^{(d)} = (v_1^{(d)}, .., v_m^{(d)}),$ $\sigma_1^{2(d)}, \sigma_2^{2(d)}, \sigma_v^{2(d)}, z^{(d)}, p_e^{(d)}\}$ be the d^{th} draw from the posterior distribution generated from the Gibbs sampler based on our proposed model, $d = b+1, \ldots, b+D$, where b and D are the burn-in sample size and the total size of the Gibbs sampler respectively. Now, the HB estimator of θ_i based on our proposed model is:

$$\hat{\theta}_i^{HB} = \frac{1}{D} \sum_{d=b+1}^{b+D} \theta_i^{(d)},$$

where, $\theta_i^{(d)} = \bar{X}_i^T \beta^{(d)} + v_i^{(d)}$. Similarly, a measure of variability, the posterior variance for θ_i is computed as (Rao, 2003):

$$\hat{V}_i^{HB} = \frac{1}{D-1} \sum_{d=b+1}^{b+D} (\theta_i^{(d)} - \hat{\theta}_i^{HB})^2, \ i = 1, \ldots, m.$$

2.3.2 Outlier Detection

In our proposed model, we assume that the outliers come from the distribution with larger variance. Based on this assumption, we propose to use $P(z_{ij} = 0|y)$, $j = 1, \ldots, n_i, i = 1, \ldots, m$ as a measure of outlier detection. If the posterior probability $P(z_{ij} = 0|y)$ is high for an observation then we suspect that the observation might have come from the distribution with the larger variance.

For φ being the vector of model parameters, we estimate $P(z_{ij} = 0|y)$ by $E[P(z_{ij} = 0|y, \varphi)|y]$,

$$\widehat{P}(z_{ij} = 0|y) = \frac{1}{D} \sum_{d=b+1}^{b+D} \left\{1 - p^*_{ij}(\varphi^d)\right\}, \tag{2.3.1}$$

where $\varphi^{(d)}$ is the d^{th} posterior draw, and b and D are the same as defined in the previous section. We have previously defined p^*_{ij} in Section 2.3.1, while stating the full conditional distribution of z_{ij} in (III).

2.4 Simulation Study

We assess our methodology in two different ways. In Section 2.4.1, we study the performance of our outlier detection method by generating single data sets under different scenario. In Section 2.4.2, we assess our method through a model-based repetitive simulation.

2.4.1 Implementation of the outlier detection method

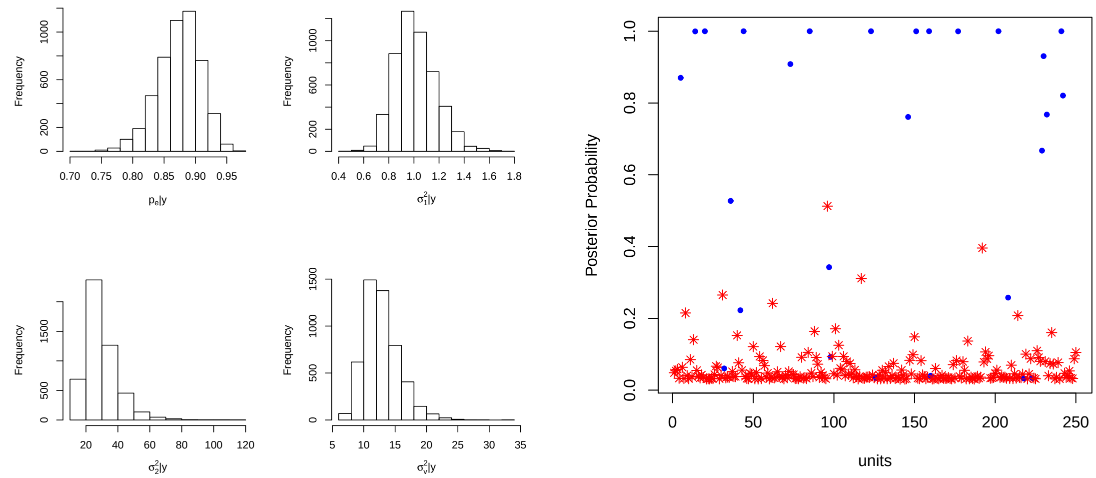
In this Section we study the performance of our proposed model through simulation, under different scenarios. We generate values y_{ij} of the response variable from the following model

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + v_i + e_{ij} \quad j = 1, \ldots, n_i, \ i = 1, \ldots, m, \tag{2.4.1}$$

where m represents the number of small areas and n_i is the number of sampled units from the i^{th} small area. The area specific random effects, v_i's are generated from $N(0, 4^2)$ for all i. The auxiliary variables x_1 and x_2 are generated from $N(12, 3^2)$ and $N(10, 9^2)$, respectively; they are generated only once for the entire study. Sampling errors (e_{ij}) are generated based on the following two simulation setups:

(i) e_{ij}'s are generated from the mixture distribution: $\gamma N(0,1^2) + (1-\gamma)N(0,5^2)$ for all i,j. We generate γ's from a Bernoulli distribution where $P(\gamma = 1) = p_e$. We choose $p_e = 0.9$.

(ii) e_{ij}'s are are generated from $N(0,1^2)$.

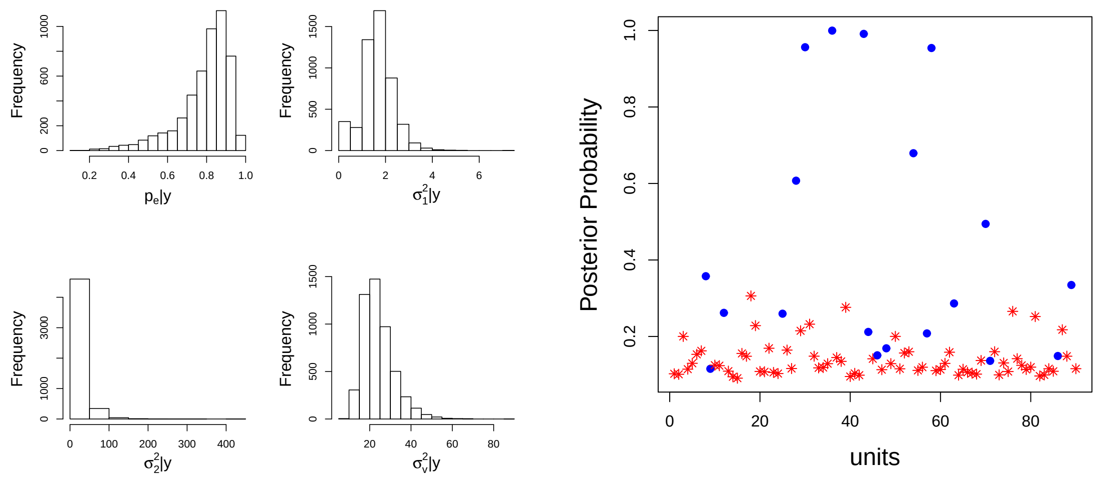


(a) (b)

Figure 2.1: *(a) Histograms of the posterior simulations for different parameters and (b) outlier detection, under simulation setup (i), when* $m = 50, n_i = 5$ *for all* i. *The blue dots represent observations from the outlying distribution* $N(0,5^2)$ *and red stars represent observations from* $N(0,1^2)$.

In the literature, the distribution with larger variance is often referred to as "contaminated distribution" or "outlier distribution". We study the performance of our model for the scenario when outliers are present as well as for the scenario when no outliers are present.

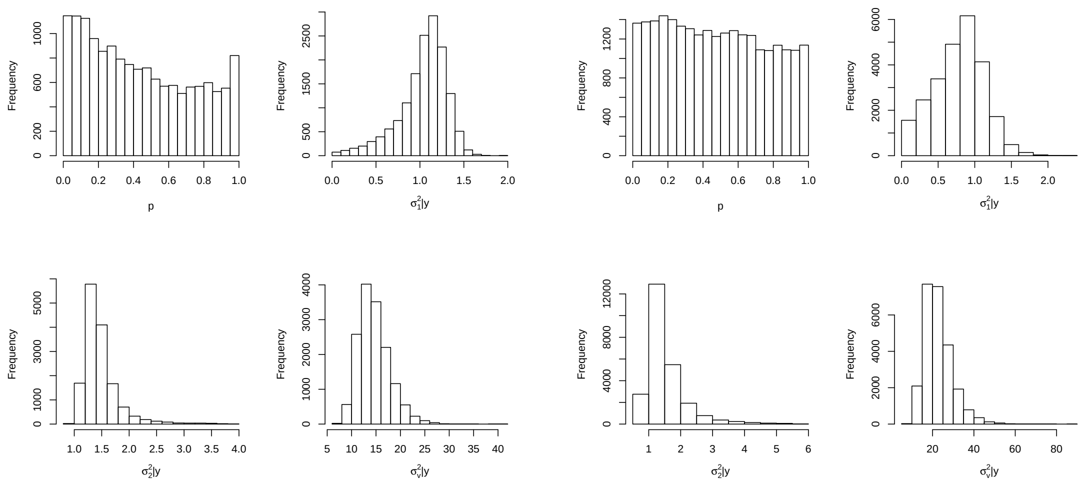
Performance of the proposed method may depend on the number of small areas (m) as well as the number of selected units (n_i) from each small area. So we perform the simulation study for $m = 50$, $n_i = 5$ for all i and also for $m = 30$, $n_i = 3$ for all i. Choice of the regression coefficients $\beta_0 = 10$, $\beta_1 = 2$ and $\beta_2 = 2.9$ remains fixed throughout the simulation study. Sufficient conditions of Theorem 2.3.1 are satisfied, for our choices of m, n_i's and p (number of covariates/ auxiliary variables in the model including intercept).



(a) (b)

Figure 2.2: *(a) Histograms of the posterior simulations for different parameters and (b) outlier detection, under simulation setup (i), when $m = 30, n_i = 3$ for all i. The blue dots represent observations from the outlying distribution $N(0, 5^2)$ and red stars represent observations from $N(0, 1^2)$.*

We illustrate our methodology by analyzing the county crop data of Battese et al. (1988) in Section 2.5. Battese et al. (1988) utilized two auxiliary variables to predict county crops which is our reason to use two auxiliary variables in the simulation study. To carry out our proposed Bayesian method to the simulated data sets for different simulation setups, we run 5 independent chains (each of length 10,000) of Gibbs sampler, we discard the first 50% draws of each chain. Potential scale reduction factors are calculated for β_0, β_1, β_2, σ_1^2, σ_2^2, σ_v^2 and p_e to examine convergence of the Gibbs sampler. Model parameters are estimated by posterior means and estimates of the standard errors of the parameters are obtained by calculating the standard deviations of the posterior draws for different parameters. We calculate and plot outlier probabilities for each observation.



(a) (b)

Figure 2.3: *Histograms of the posterior simulations for different parameters under simulation setup (ii), (a)* $m = 50, n_i = 5$ *for all* i *and (b)* $m = 30, n_i = 3$ *for all* i.

Table 2.1 shows posterior mean, standard deviation and quantiles of the simulated draws for simulation setup (i) when $m = 50$ and $n_i = 5$ for all i. In Figure 2.1 we see the histogram plots of the simulated values from the posterior distribution after discarding the first half of each independent Gibbs sampling chain. The simulation results indicate that the estimates are reasonably close to the true values. We expected the data to contain around 10% of the values from the distribution with larger variance. Hence we do not expect much information on σ_2^2 from the data. That could be a reason for getting a larger posterior standard deviation associated with the estimate of σ_2^2. From Table 2.2, we see that the parameter estimates obtained by our method are also close when sample size is small. From Figure 2.1 (b) and Figure 2.2 (b), we see that the posterior probabilities of being in the outlier distribution are considerably high for some of the observations, which is an indication that our outlier detection method has detected the outliers properly.

Figures 2.3 (a) and (b) show that posterior distribution of p_e is nearly as flat as the assumed prior. By estimating p_e we estimate proportion of observations coming from the

Table 2.1: Summary of posterior simulations for *simulation setup (i)*, when $m = 50$, $n_i = 5$.

Parameter	True Values	Posterior Mean	Posterior sd	Simulated Quantiles 2.5%	Median	97.5%
β_o	10.00	9.30	0.58	8.13	9.31	10.43
β_1	2.00	2.03	0.03	1.98	2.03	2.08
β_2	2.90	2.91	0.01	2.89	2.91	2.93
σ_1^2	1.00	0.85	0.16	0.54	0.84	1.18
σ_2^2	25.00	21.02	6.90	11.5	19.79	37.95
σ_v^2	16.00	11.97	2.63	7.82	11.62	18.07
p_e	0.90	0.83	0.05	0.73	0.84	0.91

Table 2.2: Summary of posterior simulations for *simulation setup (i)*, when $m = 30$, $n_i = 3$.

Parameter	True Values	Posterior Mean	Posterior sd	Simulated Quantiles 2.5%	Median	97.5%
β_o	10.00	10.42	1.14	8.13	10.43	12.596
β_1	2.00	1.98	0.06	1.86	1.98	2.11
β_2	2.90	2.91	0.02	2.88	2.91	2.95
σ_1^2	1.00	1.08	0.40	0.47	1.01	2.04
σ_2^2	25.00	30.86	19.17	11.58	26.05	78.43
σ_v^2	16.00	19.27	6.05	10.75	18.21	34.15
p_e	0.90	0.82	0.08	0.63	0.83	0.93

"true" underlying distribution (distribution with smaller variance). In our simulation setup (ii), we note that the parameter p_e is not a part of the model. Hence, there is little information contained about the outlier distribution in the data. This suggests that the posterior distribution of p_e may look very similar to its prior. For sampling scheme (i), the posterior distribution of p_e changed significantly from its prior because we expect around 10% of the observations to come from the outlier distribution.

Table 2.3: Summary of posterior simulations for *simulation setup (ii)*, when $m = 50$, $n_i = 5$.

Parameter	True Values	Posterior Mean	Posterior sd	Simulated Quantiles 2.5%	Median	97.5%
β_o	10.00	10.14	0.62	8.94	10.13	11.38
β_1	2.00	1.96	0.03	1.91	1.96	2.02
β_2	2.90	2.90	0.01	2.88	2.897	2.91
σ_1^2	1.00	1.04	0.27	0.32	1.09	1.44
σ_2^2	25.00	1.56	3.32	1.10	1.40	2.68
σ_v^2	16.00	14.60	3.22	9.58	14.16	22.04
p_e	0.90	0.43	0.30	0.014	0.39	0.98

Table 2.4: Summary of posterior simulations for *simulation setup (ii)* when $m = 30$, $n_i = 3$.

Parameter	True Values	Posterior Mean	Posterior sd	Simulated Quantiles 2.5%	Median	97.5%
β_o	10.00	10.45	1.13	8.27	10.49	12.71
β_1	2.00	1.99	0.05	1.89	1.99	2.01
β_2	2.90	2.889	0.017	2.86	2.889	2.922
σ_1^2	1.00	0.78	0.34	0.096	0.81	1.41
σ_2^2	25.00	1.67	4.49	0.86	1.35	3.84
σ_v^2	16.00	22.75	6.83	12.93	21.58	39.50
p_e	0.90	0.48	0.29	0.02	0.47	0.97

2.4.2 Model-based simulation

In this section we compare two hierarchical Bayesian (HB) methods through a simulation study. We generate a population of m small areas with N_i units in i^{th} area. In this study the choices of m are $m = 20, 40$ and choices of N_i's are $N_i = 80, 200$. Population values are generated from:

$$Y_{ij} \quad = \quad \beta_0 + \beta_1 X_{ij} + v_i + e_{ij} \quad j = 1, \ldots, N_i, i = 1, \ldots, m, \tag{2.4.2}$$

where Y_{ij} and X_{ij} are population level response value and auxiliary information for the j^{th} population unit in the i^{th} small area. Here X_{ij}'s are drawn from $N(1, 1^2)$, the choices $\beta_0 = \beta_1 = 1$ remain same for the entire study. Random effect v_i's are generated from $N(0, 1)$.

As in Section 2.4.1, e_{ij}'s are drawn from a two component normal mixture distribution $\gamma N(0,1^2)+(1-\gamma)N(0,5^2)$, where $P(\gamma=1)$=p_e. We set values of p_e, as $p_e=0.9$, 0.8 and 0.5. The choices p_e=0.9, 0.8 and 0.5 imply that 10%, 20% and 50% e_{ij}'s are drawn from $N(0,5^2)$ respectively. We also generate another population whose sampling errors are all drawn from $N(0,1^2)$.

Once the population data sets are created, we randomly select n_i observations (without replacement) from each area and construct a sample data set ($n_i=4$, 10). We generate 100 such data sets from the same population. We apply our method and the method by Datta and Ghosh (1991) to estimate small area means for each area based on the sample data sets. Let $\hat{\theta}_i$ be the estimated small area mean for the i^{th} area and θ_i be the true population mean for the i^{th} area. We compute average squared deviation: $\frac{1}{m}\sum_{i=1}^{m}(\hat{\theta}_i-\theta_i)^2$ for each data set. In Table 2.5 we present overall mean of these 100 average squared deviations based on all sample data sets under different simulation configurations.

From the results presented in Table 2.5, we note that when the samples are drawn from a population with contamination, mean average squared deviations are considerably smaller for the proposed method. On the other hand, if the population does not have contamination, the average mean squared deviations for the two competing methods are almost the same.

2.5 Data Analysis

In this section, we illustrate the proposed methodology by analyzing the county crop data previously analyzed by Battese et al. (1988) and Datta and Ghosh (1991), Prasad and Rao (1990). The basic purpose of their analysis was to predict the areas under corn and soybeans in terms of number of hectares, for 12 counties in North Central Iowa. Number of hectares of corn and soybeans in the 37 sample segments of 12 counties were obtained by farm interview survey, conducted by USDA. The auxiliary variables considered by Battese et al. (1988) were the number of pixels classified as corn and soybean. The auxiliary information was

Table 2.5: *Overall mean squared deviations for two methods under different simulation configuration. DG ≡ Datta and Ghosh method and NM ≡ New Method.*

Population size	Sample size	Contamination level							
		None		10%		20 %		50%	
		DG	NM	DG	NM	DG	NM	DG	NM
m=20									
$N_i = 80$	$n_i = 4$	0.18	0.17	0.70	0.37	0.60	0.32	0.97	0.79
	$n_i = 10$	0.08	0.08	0.27	0.15	0.38	0.16	0.73	0.38
$N_i = 200$	$n_i = 4$	0.22	0.23	0.66	0.34	0.44	0.28	1.31	0.91
	$n_i = 10$	0.10	0.10	0.29	0.13	0.31	0.16	0.84	0.42
m=40									
$N_i = 80$	$n_i = 4$	0.19	0.19	0.41	0.24	0.61	0.34	0.58	0.47
	$n_i = 10$	0.08	0.08	0.25	0.11	0.36	0.18	0.47	0.32
$N_i = 200$	$n_i = 4$	0.21	0.21	0.53	0.30	0.64	0.34	0.90	0.61
	$n_i = 10$	0.09	0.09	0.25	0.13	0.36	0.16	0.67	0.34

obtained from the LANDSAT satellite data. Selected sampled segments for some counties are as low as one.

Battese et al. (1988) considered the reported hectares of corn for the second segment (unit) of the Hardin county as unusual and deleted the observation while analyzing the data. Sinha and Rao (2009) did not exclude that observation from the data and studied the performance of their proposed robust method. We apply our proposed model to the corn data based on 37 observations. There are two predictors in the data, namely, the number of pixels per segment classified as corn and soybeans.

In this application, we have $m = 12$, $n = 37$ and $p = 3$ (considering the intercept). Out of 12 counties in the data, 9 counties have sample size at least 2. Total sample size for these 9 counties is 34. Hence, we can apply our model to the data set since the sufficient conditions in Theorem 2.3.1 are satisfied.

Since, N_i's (population unit for i^{th} county) are large in this data, we predict the mean hectares of corn per segment by predicting

$$\theta_i = \beta_0 + \beta_1 \bar{X_{1i}} + \beta_2 \bar{X_{2i}} + v_i, \tag{2.5.1}$$

where population mean number of pixels classified as corn ($\bar{X_{1i}}$) and soybean ($\bar{X_{2i}}$) for each county which are reported in Battese et al. (1988).

Table 2.6: Bayesian inference for the county corn data (Battese et al., 1988).

Parameter	Posterior Mean	Posterior sd	Posterior Quantiles 2.5%	Median	97.5%
β_0	28.6168	32.0841	-37.7516	29.6295	88.7162
β_1	0.3552	0.0651	0.2315	0.3540	0.4880
β_2	-0.0645	0.0749	-0.2061	-0.0667	0.0865
σ_1^2	182.5974	96.5756	26.3237	168.9364	407.1334
σ_2^2	922.9761	6662.8240	217.9412	474.3761	3575.0814
σ_v^2	209.5564	196.9220	18.8630	158.8570	704.2600
p_e	0.5975	0.2808	0.0413	0.6521	0.9756

We simulate samples from the posterior distribution by applying Gibbs Sampling. We run 5 chains and $10,000$ iterations each. The summary of the posterior simulations is presented in Table 2.6. Histograms of the posterior simulation are also presented in Figure 2.4 (a). In order to check the convergence, we calculated potential scale reduction factor for each parameter to confirm the convergence. From Figure 2.4 (b) we see that our method successfully identifies that the second observation of Hardin county is an outlier. We should further note from the posterior mean and the posterior standard deviation of p_e that these quantities are very close to the mean and the standard deviation of a $Uniform(0,1)$ distribution. This essentially indicates that there are not many outliers in the data. This is also confirmed by the histogram of p_e in Figure 2.4 (a). The robust maximum likelihood estimates of the regression parameters obtained and reported by Sinha and Rao (2009) are: $\hat{\beta_0} = 29.14$, $\hat{\beta_1} = 0.3576$ and $\hat{\beta_2} = -0.0694$, which are very close to our HB estimates reported in Table 2.6. We compare our estimates of mean hectares of corn and standard error

Table 2.7: Predicted hectares of corn along with measure of prediction error obtained from two different methods.

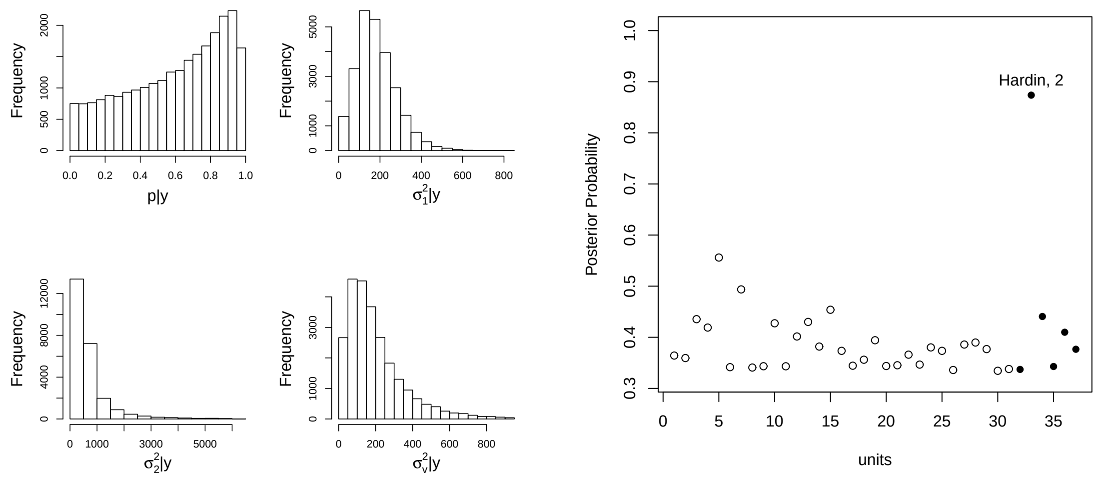
	Proposed Method		Sinha-Rao (2009) Method*	
County	HB estimate	Posterior sd	REBLUP	RMSPE
Cerro Gordo	124.25	10.63	123.7	9.8
Hamilton	125.98	10.48	125.3	9.7
Worth	108.06	11.88	110.3	9.5
Humboldt	112.72	10.46	114.1	8.6
Franklin	142.10	8.40	140.8	7.3
Pocahontas	111.39	7.71	110.8	7.2
Winnebago	114.19	8.06	115.2	7.0
Wright	122.27	7.52	122.7	7.7
Webster	114.28	6.82	113.5	6.3
Hancock	123.77	6.28	124.1	6.3
Kossuth	108.48	6.95	109.5	6.1
Hardin	135.21	7.38	136.9	6.3

*Reported in Table 8 of Sinha and Rao (2009).

of our estimate (measured by posterior standard deviation) with the estimates reported in Sinha and Rao (2009) in Table 2.7. We find that while the point estimates are in remarkably close agreement, majority of the posterior standard deviations are higher than RMSPE of Sinha and Rao (2009), some are almost 20% larger.

2.6 Discussion

Our proposed hierarchical Bayesian mixture model accounts for possible presence of outliers in the data. We show that estimates obtained by our proposed model are reasonably close to the true values and the outlier detection method discussed in this chapter successfully recognize any outliers if they are present in the data. We plan to compare our Bayesian methods with the robust frequestist methods thorough an extensive model-based simulation study. In our proposed model for unit-level data, we assume that random area effects are



(a) (b)

Figure 2.4: *(a) Histograms of the posterior simulations for different parameters and (b) outlier detection for county crop corn data (Battese et al., 1988). The black solid circles in (b) represent observations from Hardin county.*

modeled by a normal distribution. However, outliers may exist among random area effects. In order to address this issue, Datta and Lahiri (1995) proposed a Cauchy prior for the outlying small area random effects. Our goal of future research is to propose and study a hierarchical Bayesian model which would account for possible presence of outliers in the random area effects. We plan to do this by using a two-component mixture of normal distribution for the random effects.

2.7 Proof of the Theorem

Proof of Theorem 2.3.1: We assume that there are at least two sampled units for each small area, i.e. $n_i \geq 2$, $i = 1, \ldots, m$; and $n \geq 2m+2p-1$, where $n = \sum_{i=1}^m n_i$. At first, we consider the case when $n = 2m + 2p - 1$, the argument can be extended to the case $n > 2m + 2p - 1$ by applying Lemma 2.3.3. Under the proposed model, the joint pdf of y_{ij}'s, $j = 1, \ldots, n_i$,

$i = 1, \ldots, m$; $v\,(m \times 1)$, $\beta\,(p \times 1)$, σ_1^2, σ_2^2, σ_v^2 and p_e is given by

$$
\begin{aligned}
& f(y, v, \beta, \sigma_2^2, \sigma_1^2, \sigma_v^2, p_e) \\
& \qquad \propto \sum_{\Omega} \Big[\prod_{i=1}^{m} \Big\{ \prod_{k=1}^{n_{i1}} \frac{p_e}{\sqrt{\sigma_1^2}} \exp\Big(-\frac{1}{2} \frac{(y_{ij_k} - x_{ij_k}^T \beta - v_i)^2}{\sigma_1^2} \Big) \Big\} \\
& \qquad\qquad \times \Big\{ \prod_{k=n_{i1}+1}^{n_i} \frac{(1-p_e)}{\sqrt{\sigma_2^2}} \exp\Big(-\frac{1}{2} \frac{(y_{ij_k} - x_{ij_k}^T \beta - v_i)^2}{\sigma_2^2} \Big) \Big\} \Big] \\
& \qquad\qquad \times \frac{1}{(\sigma_v^2)^{\frac{m}{2}}} \exp\Big(-\frac{1}{2} \sum_{i=1}^{m} \frac{v_i^2}{\sigma_v^2} \Big) \times \frac{1}{(\sigma_2^2)^2} I(\sigma_1^2 < \sigma_2^2) \qquad (2.7.1)
\end{aligned}
$$

The summation $\sum_{\Omega}$ and the quantities n_{i1}, n_{i2}, $i = 1, \ldots, m$ are explained below. Let $z_{ij} = 1$, if the j^{th} sampled unit of the i^{th} small area corresponds to the mixture component σ_1^2 and $z_{ij} = 0$ otherwise. Ω contains all possible choices of $\mathbf{z} = (z_{11}, \ldots, z_{mn_m})$ vector. Hence the cardinality of Ω is 2^n. For a given z, let $n_{i1} = \sum_{j=1}^{n_i} z_{ij}$ and $n_{i2} = n_i - n_{i1}$ for $i = 1, \ldots, m$. Then n_{i1} is the number of units from the i^{th} small area whose sampling variance corresponds to the mixture component σ_1^2. The remaining n_{i2} units from the i^{th} small area corresponds to the mixture component σ_2^2.

Define, $S_1 = \{i : n_{i1} > 0\}$ and $S_2 = \{i : n_{i2} > 0\}$. Clearly, $S_1 \cup S_2 = \{1, \ldots, m\}$ and $S_1 \cap S_2$ may not be an empty set. Let m_i be the cardinality of S_i, $i = 1, 2$, then $m \le m_1 + m_2$. Note that n_{i1} or n_{i2} can be zero for some i. Define, $n_1^* = \sum_{i \in S_1} n_{i1}$ and $n_2^* = \sum_{i \in S_2} n_{i2}$.

From (2.7.1), a typical term under the sum over Ω is,

$$
\begin{aligned}
& \varphi(y, v, \beta, \sigma_1^2, \sigma_2^2, \sigma_v^2, p_e) \\
& \quad = C \times p_e^{n_1^*} (1-p_e)^{n_2^*} \times \frac{1}{(\sigma_1^2)^{\frac{n_1^*}{2}}} \times \exp\Big(-\frac{1}{2\sigma_1^2} \sum_{i \in S_1} \sum_{k=1}^{n_{i1}} (y_{ij_k} - x_{ij_k}^T \beta - v_i)^2 \Big) \\
& \qquad\qquad \times \frac{1}{(\sigma_2^2)^{\frac{n_2^*}{2}}} \times \exp\Big(-\frac{1}{2\sigma_2^2} \sum_{i \in S_2} \sum_{k=n_{i1}+1}^{n_i} (y_{ij_k} - x_{ij_k}^T \beta - v_i)^2 \Big) \\
& \qquad\qquad \times \frac{1}{(\sigma_v^2)^{\frac{m}{2}}} \exp\Big(-\frac{1}{2} \sum_{i=1}^{m} \frac{v_i^2}{\sigma_v^2} \Big) \times \frac{I(\sigma_1^2 < \sigma_2^2)}{(\sigma_2^2)^2}, \qquad (2.7.2)
\end{aligned}
$$

where C is a generic constant. In order to check the integrability of $f(y, v, \beta, \sigma_2^2, \sigma_1^2, \sigma_v^2, p_e)$ with respect to $\beta,\ v,\ \sigma_1^2,\ \sigma_2^2,\ \sigma_v^2,\ p_e$ in (2.7.1), we need to check the integrability of each typical term in (2.7.1) with respect to $\beta,\ v,\ \sigma_1^2,\ \sigma_2^2,\ \sigma_v^2,\ p_e$.

Integrability of $\varphi(y, v, \beta, \sigma_1^2, \sigma_2^2, \sigma_v^2, p_e)$ with respect to p_e is obvious.

$$\begin{aligned}
&\int \varphi(y, v, \beta, \sigma_1^2, \sigma_2^2, \sigma_v^2, p_e)\,\mathrm{d}p_e \\
&\quad = C \times \frac{1}{(\sigma_1^2)^{\frac{n_1^*}{2}}} \times \exp\left(-\frac{1}{2\sigma_1^2}\sum_{i\in S_1}\sum_{k=1}^{n_{i1}}(y_{ij_k} - x_{ij_k}^T\beta - v_i)^2\right) \\
&\qquad \times \frac{1}{(\sigma_2^2)^{\frac{n_2^*}{2}}} \times \exp\left(-\frac{1}{2\sigma_2^2}\sum_{i\in S_2}\sum_{k=n_{i1}+1}^{n_i}(y_{ij_k} - x_{ij_k}^T\beta - v_i)^2\right) \\
&\qquad \times \frac{1}{(\sigma_v^2)^{\frac{m}{2}}}\exp\left(-\frac{1}{2}\sum_{i=1}^{m}\frac{v_i^2}{\sigma_v^2}\right) \times \frac{I(\sigma_1^2 < \sigma_2^2)}{(\sigma_2^2)^2}
\end{aligned} \tag{2.7.3}$$

Let us partition y and X as follows, $y_1 = \text{col}_{i\in S_1}\text{col}_{1\le k\le n_{i1}} y_{ij_k}$; $X_1 = \text{col}_{i\in S_1}\text{col}_{1\le k\le n_{i1}} x_{ij_k}^T$ and $y_2 = \text{col}_{i\in S_2}\text{col}_{n_{i1}+1\le k\le n_i}\, y_{ij_k}$; $X_2 = \text{col}_{i\in S_2}\text{col}_{n_{i1}+1\le k\le n_i}\, x_{ij_k}^T$. Also, $Z_1 = \bigoplus_{i=1}^{m} 1_{n_{i1}}$ and $Z_2 = \bigoplus_{i=1}^{m} 1_{n_{i2}}$. We rewrite (2.7.3) as:

$$\begin{aligned}
&\int \varphi(y, v, \beta, \sigma_2^2, \sigma_1^2, \sigma_v^2, p_e)\,\mathrm{d}p_e \\
&\quad = C \times \frac{1}{(\sigma_1^2)^{\frac{n_1^*}{2}}} \times \exp\left(-\frac{1}{2\sigma_1^2}(y_1 - X_1\beta - Z_1v)^T(y_1 - X_1\beta - Z_1v)\right) \\
&\qquad \times \frac{1}{(\sigma_2^2)^{\frac{n_2^*}{2}}} \times \exp\left(-\frac{1}{2\sigma_2^2}(y_2 - X_2\beta - Z_2v)^T(y_2 - X_2\beta - Z_2v)\right) \\
&\qquad \times \frac{1}{(\sigma_v^2)^{\frac{m}{2}}}\exp\left(-\frac{1}{2}\sum_{i=1}^{m}\frac{v_i^2}{\sigma_v^2}\right) \times \frac{I(\sigma_1^2 < \sigma_2^2)}{(\sigma_2^2)^2}
\end{aligned} \tag{2.7.4}$$

Note that, there are m_1 and m_2 components of v are involved in Z_1v and Z_2v respectively. Let the rank of $X_1\,(n_1^* \times p)$ and $X_2\,(n_2^* \times p)$ be p_1 and p_2 respectively, where $p_1 + p_2 \ge p$. We now state the lemma below

Lemma 2.7.1 *If $n = 2m + 2p - 1$ and $m \ge p + 6$, then one of the following conditions must hold. (a) $n_1^* \ge m_1 + p_1$, $m_1 > 3$ or (b) $n_2^* \ge m_2 + p_2$, $m_2 > 3$.*

The proof of Lemma 2.7.1 is provided in Section 2.7.1. Without loss of generality, for the rest of the proof, we assume that $n_1^* \geq m_1 + p_1$ and $m_1 > 3$. Had we assumed $n_2^* > m_2 + p_2$, $m_2 > 3$, it will lead us to establish the same results. Note that we do not have to make separate assumptions for n_1^* and m_1, they come from the assumptions $n \geq 2m + 2p - 1$ and $m \geq p + 6$.

Without loss of generality, we assume that the rows of X_1 are arranged such that the first p_1 rows are linearly independent. These rows constitute a sub matrix of $X_{11}(p_1 \times p)$, the rest of the rows of X_1 can be expressed as $A_{21}X_{11}$ for some matrix $A_{21}((n_1^* - p_1) \times p_1)$. Similarly, we assume that first p_2 rows of X_2 are so arranged that they are linearly independent. Since, $p_2 \geq p - p_1$, we further assume that the first $(p - p_1)$ of these p_2 rows are linearly independent of the rows of X_{11}, we denote this portion of X_2 as the sub matrix $X_{211}((p - p_1) \times p)$.
Let, X_{212} consists next $p_2 - (p - p_1)$ linearly independent rows of X_2, and X_{22} contains the remaining $(n_2^* - p_2)$ rows. Hence,

$$X_1 = \begin{pmatrix} X_{11} \\ A_{21}X_{11} \end{pmatrix} \qquad \text{and} \qquad X_2 = \begin{pmatrix} X_{211} \\ X_{212} \\ X_{22} \end{pmatrix}$$

where, $\text{rank}(X_{11}) = p_1$.
According to the construction of the matrices, $X_{212} = H \begin{pmatrix} X_{11} \\ X_{211} \end{pmatrix}$ for some $H = \begin{pmatrix} H_{21} & H_{22} \end{pmatrix}$, note that $H_{21} \neq 0$. we can write, $X_{22} = A_{22} \begin{pmatrix} X_{211} \\ X_{212} \end{pmatrix}$ for some $A_{22}((n_2^* - p_2) \times p_2)$.

We consider the transformation: $\rho_1 = X_{11}\beta$ and $\rho_2 = X_{211}\beta$; $\rho = \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix}$.

Now, $X_1\beta = \begin{pmatrix} X_{11} \\ A_{21}X_{11} \end{pmatrix} \beta = \begin{pmatrix} I_{p_1} \\ A_{21} \end{pmatrix} X_{11}\beta = M_1\rho_1$, where $M_1 = \begin{pmatrix} I_{p_1} \\ A_{21} \end{pmatrix}$ and $\text{rank}(M_1) = p_1$.

Similarly,

$$X_{212}\beta = H \begin{pmatrix} X_{11}\beta \\ X_{211}\beta \end{pmatrix} = H \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} = H\rho \text{ and}$$

$$X_{22}\beta = A_{22} \begin{pmatrix} X_{211} \\ X_{212} \end{pmatrix} \beta = A_{22} \begin{pmatrix} \rho_2 \\ H\rho \end{pmatrix} = A_{22} \begin{pmatrix} 0 & I \\ H_{21} & H_{22} \end{pmatrix} \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} = A_{22}^*\rho,$$

hence, $X_2\beta = \begin{pmatrix} X_{211} \\ X_{212} \\ X_{22} \end{pmatrix} \beta = \begin{pmatrix} \rho_2 \\ H\rho \\ A_{22}^*\rho \end{pmatrix} = \begin{pmatrix} \rho_2 \\ G\rho \end{pmatrix}$, where $G = \begin{pmatrix} H \\ A_{22}^* \end{pmatrix}$, we partition y_2 and Z_2 according to the partitioned rows of X_2, i.e., $y_2 = \begin{pmatrix} y_{211} \\ y_{212} \\ y_{22} \end{pmatrix} = \begin{pmatrix} y_{211} \\ y_2^* \end{pmatrix}$, where $y_2^* = \begin{pmatrix} y_{212} \\ y_{22} \end{pmatrix}$ and $Z_2 = \begin{pmatrix} Z_{211} \\ Z_{212} \\ Z_{22} \end{pmatrix} = \begin{pmatrix} Z_{211} \\ Z_{22}^* \end{pmatrix}$, where $Z_{22}^* = \begin{pmatrix} Z_{212} \\ Z_{22} \end{pmatrix}$.

After these transformations, we can rewrite the right hand side of (2.7.4) as

$$\begin{aligned}
&= C \times \frac{1}{(\sigma_1^2)^{\frac{n_1^*}{2}}} \exp\left(-\frac{1}{2\sigma_1^2}(y_1 - M_1\rho_1 - Z_1v)^T(y_1 - M_1\rho_1 - Z_1v)\right) \\
&\quad \times \frac{1}{(\sigma_2^2)^{\frac{p-p_1}{2}}} \exp\left(-\frac{1}{2\sigma_2^2}(y_{211} - \rho_2 - Z_{211}v)^T(y_{211} - \rho_2 - Z_{211}v)\right) \\
&\quad \times \frac{1}{(\sigma_2^2)^{\frac{n_2^*-(p-p_1)}{2}}} \exp\left(-\frac{1}{2\sigma_2^2}(y_2^* - G\rho - Z_2^*v)^T(y_2^* - G\rho - Z_2^*v)\right) \\
&\quad \times \frac{1}{(\sigma_v^2)^{\frac{m}{2}}} \exp\left(-\frac{v^Tv}{2\sigma_v^2}\right) \times \frac{I(\sigma_1^2 < \sigma_2^2)}{(\sigma_2^2)^2} \\
&= \tilde{\varphi}(y_1, y_{211}, y_2^*, v, \rho_1, \rho_2, \sigma_1^2, \sigma_2^2, \sigma_v^2, p_e).
\end{aligned} \tag{2.7.5}$$

We integrate with respect to y_2^*, ρ_2 and σ_2^2 respectively,

$$\begin{aligned}
&\int \tilde{\varphi}(y_1, y_{211}, y_2^*, v, \rho_1, \rho_2, \sigma_1^2, \sigma_2^2, \sigma_v^2, p_e)\, \mathrm{d}y_2^*\, \mathrm{d}\rho_2\, \mathrm{d}\sigma_2^2 \\
&\qquad = C \times \frac{1}{(\sigma_1^2)^{\frac{n_1^*}{2}}} \exp\left(-\frac{1}{2\sigma_1^2}\left(y_1 - M_1\rho_1 - Z_1 v\right)^T \left(y_1 - M_1\rho_1 - Z_1 v\right)\right) \\
&\qquad\quad \times \frac{1}{(\sigma_v^2)^{\frac{m}{2}}} \exp\left(-\frac{v^T v}{2\sigma_v^2}\right) \times \int \frac{I(\sigma_1^2 < \sigma_2^2)}{(\sigma_2^2)^2}\, \mathrm{d}\sigma_2^2 \\
&\qquad = C \times \frac{1}{(\sigma_1^2)^{\frac{n_1^*+2}{2}}} \exp\left(-\frac{1}{2\sigma_1^2}\left(y_1 - M_1\rho_1 - Z_1 v\right)^T \left(y_1 - M_1\rho_1 - Z_1 v\right)\right) \\
&\qquad\quad \times \frac{1}{(\sigma_v^2)^{\frac{m}{2}}} \exp\left(-\frac{v^T v}{2\sigma_v^2}\right). \qquad (2.7.6)
\end{aligned}$$

As we mentioned earlier, there are m_1 components of v involved in $Z_1 v$. We can write those m_1 components as $v^{(1)} = (v_{i_1}, \ldots, v_{i_{m_1}})^T$. Now, $Z_1 v$ can be reduced to $Z_1^{(1)} v^{(1)}$, where $Z_1^{(1)} = \bigoplus_{j=1}^{m_1} 1_{n_{i_j 1}}$. Clearly, $\mathrm{rank}(Z_1^{(1)}) = m_1$ and $n_1^* = \sum_{j=1}^{m_1} n_{i_j 1}$. We integrate out $v^{(2)}=\{v_l : l \in S \setminus S_1\}$,

$$\begin{aligned}
&\int \tilde{\varphi}(y_1, y_{211}, y_2^*, v, \rho_1, \rho_2, \sigma_1^2, \sigma_2^2, \sigma_v^2, p_e)\, \mathrm{d}y_2^*\, \mathrm{d}\rho_2\, \mathrm{d}\sigma_2^2\, \mathrm{d}v^{(2)} \\
&\quad = C \times \frac{1}{(\sigma_1^2)^{\frac{n_1^*+2}{2}}} \exp\left(-\frac{1}{2\sigma_1^2}\left(y_1 - M_1\rho_1 - Z_1^{(1)} v^{(1)}\right)^T \left(y_1 - M_1\rho_1 - Z_1^{(1)} v^{(1)}\right)\right) \\
&\quad\quad \times \frac{1}{(\sigma_v^2)^{\frac{m_1}{2}}} \exp\left(-\frac{v^{(1)T} v^{(1)}}{2\sigma_v^2}\right). \\
&\quad = C \times \frac{1}{(\sigma_1^2)^{\frac{n_1^*+2}{2}}} \exp\left\{-\frac{{y_1^*}^T \left\{I - M_1(M_1^T M_1)^{-1} M_1^T\right\} y_1^*}{2\sigma_1^2}\right\} \\
&\quad\quad \times \exp\left(-\frac{(\rho_1 - \hat{\rho}_1)^T (M_1^T M_1)^{-1} (\rho_1 - \hat{\rho}_1)}{2\sigma_1^2}\right) \times \frac{1}{(\sigma_v^2)^{\frac{m_1}{2}}} \exp\left(-\frac{v^{(1)T} v^{(1)}}{2\sigma_v^2}\right) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (2.7.7)
\end{aligned}$$

where $y_1^* = y_1 - Z_1^{(1)} v^{(1)}$ and $\hat{\rho}_1 = (M_1^T M_1)^{-1} M_1^T y_1^*$.

We integrate with respect to ρ_1;

$$\begin{aligned}
&\int \tilde{\varphi}(y_1, y_{211}, y_2^*, v, \rho_1, \rho_2, \sigma_1^2, \sigma_2^2, \sigma_v^2, p_e)\, \mathrm{d}y_2^*\, \mathrm{d}\rho_2\, \mathrm{d}\sigma_2^2\, \mathrm{d}v^{(2)} \mathrm{d}\rho_1 \\
&= C \times \frac{1}{(\sigma_1^2)^{\frac{n_1^* - p_1 + 2}{2}}} \exp\left(-\frac{(y_1 - Z_1^{(1)} v^{(1)})^T R_1 (y_1 - Z_1^{(1)} v^{(1)})}{2\sigma_1^2} \right) \\
&\quad \times \frac{1}{(\sigma_v^2)^{\frac{m_1}{2}}} \exp\left(-\frac{v^{(1)T} v^{(1)}}{2\sigma_v^2} \right) \qquad (2.7.8)
\end{aligned}$$

where $R_1 = \left(I - M_1 (M_1^T M_1)^{-1} M_1^T\right)$

Before we proceed, let us state the following lemma.

Lemma 2.7.2 *The following results hold: (a)* rank $[R_1 Z_1^{(1)}]$ = rank$\begin{pmatrix} M_1 & Z_1^{(1)} \end{pmatrix}$- rank$(M_1)$, *(b)* rank$(R_2)$ $= n_1^*-$ rank$\begin{pmatrix} M_1 & Z_1^{(1)} \end{pmatrix}$.

Proof of Lemma 2.7.2 is given in Section 2.7.1. From Lemma 2.7.2, we have rank(R_2) = n_1^*- rank$\begin{pmatrix} M_1 & Z_1^{(1)} \end{pmatrix}$ $= n_1^* - (p_1 + m_1 - 1)$. rank$\begin{pmatrix} M_1 & Z_1^{(1)} \end{pmatrix}$ = rank(M_1) + rank$(Z_1^{(1)}) - q_1$ $= p_1 + m_1 - q_1$, where q_1 being the number of independent columns of M_1 which are in the column space of $Z_1^{(1)}$.

Note that, $X_1 = M_1 X_{11} \Rightarrow \mathcal{C}(X_1) \subseteq \mathcal{C}(M_1)$. Again, $M_1 = X_1 X_{11}^T (X_{11} X_{11}^T)^{-1} \Rightarrow \mathcal{C}(M_1) \subseteq \mathcal{C}(X_1)$. Hence, $\mathcal{C}(M_1) = \mathcal{C}(X_1)$.

Since there is an intercept term in the model, the vector $1_{n_1^*} \in \mathcal{C}(X_1)$, which is also obtained by adding the columns of $Z_1^{(1)}$. Therefore, q_1 is at least 1. Now, we assume that,

$$\sum_{l=1}^{m_1} \sum_{j=1}^{n_{i_l 1}} (X_{i_l j_k} - \bar{X}_{i_l \cdot k})^2 > 0; \quad k = 2, \ldots, p, \qquad (2.7.9)$$

$X_{i_l j_k}$ being the value of the k^{th} auxiliary variable for the j^{th} unit of the i_l^{th} small area, where, $i_l \in S_1$ and $\bar{X}_{i_l \cdot k} = \frac{1}{n_{i_l 1}} \sum_{j=1}^{n_{i_l 1}} X_{i_l j_k}$. If the auxiliary variables come from a continuous distribution then the condition (2.7.9) will be satisfied.

This assumption ensures that none of the last $p-1$ columns of X_1 (apart from the intercept term the other columns of X_1) can be spanned by the columns of $Z_1^{(1)}$. Hence, $q_1 = 1$

and $\text{rank}[M_1 Z_1] = m_1 + p_1 - 1$. From Lemma 2.7.2, $\text{rank}(R_2) = n_1^* - \text{rank}\begin{pmatrix} M_1 & Z_1^{(1)} \end{pmatrix} = n_1^* - (p_1 + m_1 - 1) = n_1^* - (p_1 + m_1) + 1 \geq 1$.

Note that we need $\text{rank}(R_2) > 0$, otherwise R_2 will be a null matrix and $y_1^T R_2 y_1$ will be zero with probability 1. Let, $Q_1 = Z_1^{(1)T} R_1 Z_1^{(1)} = (R_1 Z_1^{(1)})^T (R_1 Z_1^{(1)})$, since R_1 is symmetric and idempotent. Therefore, $\text{rank}(Q_1) = \text{rank}[R_1 Z_1^{(1)}] = \text{rank}\begin{pmatrix} M_1 & Z_1^{(1)} \end{pmatrix} - \text{rank}(M_1) = m_1 + p_1 - 1 - p_1 = m_1 - 1 = t_1$ (say).

Let, P_1 be an orthogonal matrix such that, $P_1^T Q_1 P_1 = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{t_1}, 0, \ldots, 0)$, where, $\lambda_1 \geqslant \lambda_2 \cdots \geqslant \lambda_{t_1} > 0$ are the positive eigen values of Q_1. We use the transformation, w=$P_1 v^{(1)}$, Also, we write $P_1 \hat{v}^{(1)} = \hat{w}$.

$$\begin{aligned}\int \tilde{\varphi}(y_1, y_{211}, y_2^*, v, \rho_1, \rho_2, \sigma_1^2, \sigma_2^2, \sigma_v^2, p_e)\, \mathrm{d}y_2^*\, \mathrm{d}\rho_2\, \mathrm{d}\sigma_2^2\, \mathrm{d}v^{(2)}\, \mathrm{d}\rho_1 \\ = C \times \frac{1}{(\sigma_1^2)^{\frac{n_1^* - p_1^* + 2}{2}}} \exp\left\{-\frac{w^T w}{2\sigma_v^2}\right\} \times \frac{1}{(\sigma_v^2)^{\frac{m_1}{2}}} \exp\left\{-\frac{y_1^T R_2 y_1}{2\sigma_1^2}\right\} \\ \times \exp\left\{-\frac{\sum_{j=1}^{t_1} \lambda_j (w_j - \hat{w}_j)^2}{2\sigma_1^2}\right\},\end{aligned} \tag{2.7.10}$$

where $P_1^T \hat{v}^{(1)} = \hat{w}$. We integrate out $(w_{t+1}, \ldots, w_{m_1})$:

$$\begin{aligned}\int \tilde{\varphi}(y_1, y_{211}, y_2^*, v, \rho_1, \rho_2, \sigma_1^2, \sigma_2^2, \sigma_v^2, p_e)\, \mathrm{d}y_2^*\, \mathrm{d}\rho_2\, \mathrm{d}\sigma_2^2\, \mathrm{d}v^{(2)}\, \mathrm{d}\rho_1 \prod_{k=t_1+1}^{m} \mathrm{d}w_k \\ = C \times \frac{1}{(\sigma_1^2)^{\frac{n_1^* - p_1^* + 2}{2}}} \exp\left\{-\frac{y_1^T R_2 y_1}{2\sigma_1^2}\right\} \exp\left\{-\frac{\sum_{j=1}^{t_1} \lambda_j (w_j - \hat{w}_j)^2}{2\sigma_1^2}\right\} \\ \times \frac{1}{(\sigma_v^2)^{\frac{t_1}{2}}} \exp\left\{-\frac{\sum_{j=1}^{t_1} w_j^2}{2\sigma_v^2}\right\}.\end{aligned} \tag{2.7.11}$$

We integrate with respect to σ_1^2 and σ_v^2 using inverse gamma density integration result. For that we need the shape parameters $\frac{n_1^* - p_1}{2}$ and $(\frac{t_1}{2} - 1)$ to be positive, i.e., we need $n_1^* > p_1$ and $t_1 > 2 \Rightarrow m_1 > 3$. Since we already assumed $n_1^* > p_1 + m_1$ and $m_1 > 3$, the shape parameters will be positive. By carrying out integration with respect to σ_1^2 and σ_v^2 we have,

$$\int \tilde{\varphi}(y_1, y_{211}, y_2^*, v, \rho_1, \rho_2, \sigma_1^2, \sigma_2^2, \sigma_v^2, p_e)\, \mathrm{d}y_2^*\, \mathrm{d}\rho_2\, \mathrm{d}\sigma_2^2\, \mathrm{d}v^{(2)}\, \mathrm{d}\rho_1 \prod_{k=t_1+1}^{m} \mathrm{d}w_k\, \mathrm{d}\sigma_1^2\, \mathrm{d}\sigma_v^2 \tag{2.7.12}$$

$$= C \times \frac{1}{\left\{y_1^T R_2 y_1 + \sum_{j=1}^{t_1} \lambda_j (w_j - \hat{w}_j)^2\right\}^{\frac{n_1^* - p_1}{2}}} \times \frac{1}{\left(\sum_{j=1}^{t_1} w_j^2\right)^{\frac{t_1-2}{2}}}$$

$$\leq C \times \frac{1}{\left\{y_1^T R_2 y_1 + \lambda_{t_1} \sum_{j=1}^{t_1} (w_j - \hat{w}_j)^2\right\}^{\frac{n_1^* - p_1}{2}}} \times \frac{1}{\left(\sum_{j=1}^{t_1} w_j^2\right)^{\frac{t_1-2}{2}}} \tag{2.7.13}$$

Note that,

$$\sum_{j=1}^{t_1} w_j^2 \leq 2 \sum_{j=1}^{t_1} \left[(w_j - \hat{w}_j)^2 + \hat{w}_j{}^2\right]$$

$$\Rightarrow \sum_{j=1}^{t_1} (w_j - \hat{w}_j)^2 \geq \frac{1}{2} \sum_{j=1}^{t_1} w_j^2 - \sum_{j=1}^{t_1} \hat{w}_j{}^2.$$

Let us denote $\sum_{j=1}^{t_1} \hat{w}_j{}^2$ by d^2, then for any $\epsilon > 0$,

$$\int \tilde{\varphi}(y_1, y_{211}, y_2^*, v, \rho_1, \rho_2, \sigma_1^2, \sigma_2^2, \sigma_v^2, p_e)\, \mathrm{d}y_2^*\, \mathrm{d}\rho_2\, \mathrm{d}\sigma_2^2\, \mathrm{d}v^{(2)}\, \mathrm{d}\rho_1 \times \prod_{k=t+1}^{m} \mathrm{d}w_k\, \mathrm{d}\sigma_1^2\, \mathrm{d}\sigma_v^2 \tag{2.7.14}$$

$$\leq C_1 \times \frac{1}{\{y_1^T R_2 y_1\}^{\frac{n_1^* - p_1}{2}}} \times \frac{\mathrm{I}\left(\sum_{j=1}^{t_1} w_j^2 \leq 2d^2 + \epsilon\right)}{\left(\sum_{j=1}^{t_1} w_j^2\right)^{\frac{t_1-2}{2}}}$$

$$+ C_2 \times \frac{1}{\lambda_{t_1}^{\frac{(n_1^* - p_1)}{2}} \left[\frac{1}{2} \sum_{j=1}^{t_1} w_j^2 - d^2\right]^{\frac{n_1^* - p_1}{2}}} \times \frac{\mathrm{I}\left(\sum_{j=1}^{t_1} w_j^2 > 2d^2 + \epsilon\right)}{\left(\sum_{j=1}^{t_1} w_j^2\right)^{\frac{t_1-2}{2}}} \tag{2.7.15}$$

C_1 and C_2 are positive constants.

Consider the following transformation,

$$(w_1, \ldots, w_{t_1}) \to (\alpha, \theta_1, \ldots, \theta_{t_1-1}),$$

$$\begin{aligned}
w_1 &= \alpha \cos\theta_1 \cos\theta_2 \ldots \cos\theta_{t_1-2} \cos\theta_{t_1-1} \\
w_2 &= \alpha \cos\theta_1 \cos\theta_2 \ldots \cos\theta_{t_1-2} \sin\theta_{t_1-1} \\
&\vdots \\
w_{t_1-1} &= \alpha \cos\theta_1 \sin\theta_2 \\
w_{t_1} &= \alpha \sin\theta_1,
\end{aligned}$$

$0 < \alpha < \infty$; -$\frac{\pi}{2} < \theta_i < \frac{\pi}{2}$, $i = 1, \ldots, t_1-2$ and $0 < \theta_{t_1-1} < 2\pi$. The jacobian of transformation is given by $J = \alpha^{t_1-1}(\cos\theta_1)^{t_1-2}(\cos\theta_2)^{t_1-3} \ldots \cos\theta_{t_1-2}$.

Now, the right side of (2.7.14) is:

$$\begin{aligned}
&= C_1 \times \frac{1}{(y_1^T R_2 y_1)^{\frac{(n_1^*-p_1)}{2}}} \times \frac{\alpha^{t_1-1}}{(\alpha^2)^{t_1-2}} \times (\cos\theta_1)^{t_1-2} \ldots (\cos\theta_{t_1-2})\, \mathrm{I}(\alpha^2 \leq 2d^2 + \epsilon) \\
&\quad + C_2 \times \frac{1}{\lambda_{t_1}^{\frac{(n_1^*-p_1)}{2}} [\frac{\alpha^2}{2} - d^2]^{\frac{n_1^*-p_1}{2}} (\alpha^2)^{\frac{t_1-2}{2}}} (\cos\theta_1)^{t_1-2} \ldots (\cos\theta_{t_1-2})\, \mathrm{I}(\alpha^2 > 2d^2 + \epsilon)
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\int \tilde{\varphi}(y_1, y_{211}, y_2^*, v, \rho_1, \rho_2, \sigma_1^2, \sigma_2^2, \sigma_v^2, p_e)\, \mathrm{d}y_2^*\, \mathrm{d}\rho_2\, \mathrm{d}\sigma_2^2\, \mathrm{d}v^{(2)}\, \mathrm{d}\rho_1 \\
&\qquad\qquad \times \prod_{k=t+1}^{m} \mathrm{d}w_k \prod_{j=1}^{t_1-1} \mathrm{d}\theta_j\, \mathrm{d}\sigma_1^2\, \mathrm{d}\sigma_v^2 \\
&\quad \leq C_1 \frac{1}{(y_1^T R_2 y_1)^{\frac{n_1^*-p_1}{2}}} \int_0^{\sqrt{(2d^2+\epsilon)}} \frac{\alpha^{t_1-1}}{(\alpha^2)^{\frac{t_1-2}{2}}}\, \mathrm{d}\alpha \\
&\qquad + C_2 \int_{\sqrt{(2d^2+\epsilon)}}^{\infty} \frac{\alpha^{t_1-1}}{(\lambda_{t_1})^{\frac{(n_1^*-p_1)}{2}} \left[\frac{1}{2}\alpha^2 - d^2\right]^{\frac{(n_1^*-p_1)}{2}} (\alpha^2)^{\frac{t_1-2}{2}}}\, \mathrm{d}\alpha \qquad (2.7.16)
\end{aligned}$$

$$
\begin{aligned}
&= C_1 \frac{1}{(y_1^T R_2 y_1)^{\frac{n_1^*-p_1}{2}}} \int_0^{\sqrt{(2d^2+\epsilon)}} \alpha \, \mathrm{d}\alpha \\
&\quad + C_2 \frac{1}{(\lambda_{t_1})^{\frac{(n_1^*-p_1)}{2}}} \int_{\sqrt{(2d^2+\epsilon)}}^{\infty} \frac{\alpha}{\left[\frac{1}{2}\alpha^2 - d^2\right]^{\frac{(n_1^*-p_1)}{2}}} \mathrm{d}\alpha \\
&= C_1 \frac{1}{(y_1^T R_2 y_1)^{\frac{n_1^*-p_1}{2}}} \left(\frac{2d^2+\epsilon}{2}\right) + C_2 \frac{1}{(\lambda_{t_1})^{\frac{(n_1^*-p_1)}{2}}} \times \frac{1}{\epsilon^{\frac{n_1^*-p_1-2}{2}}} < \infty \qquad (2.7.17)
\end{aligned}
$$

Since, we assumed $n_1^* \geq m_1 + p_1$ and $m_1 > 3$, hence $n_1^* > p_1 + 2$.

So far we have proved that any arbitrary typical term in (2.7.1) satisfying conditions (a), (b) and (c) is integrable. Hence, we can conclude, $f(y, v, \beta, \sigma_1^2, \sigma_2^2, \sigma_v^2, p_e)$ (in (2.7.1)) is integrable with respect to $v, \beta, \sigma_1^2, \sigma_2^2, \sigma_v^2, p_e$ if condition (a), (b) and (c) are satisfied.

2.7.1 Proof of the Lemmas

Lemma 2.7.1 *If $n = 2m + 2p - 1$ and $m \geq p + 6$, then one of the following conditions must hold. (a) $n_1^* \geq m_1 + p_1$, $m_1 > 3$ or (b) $n_2^* \geq m_2 + p_2$, $m_2 > 3$.*

Proof At first we note that at least one of these two conditions $n_1^* \geq m_1 + p_1$ and $n_2^* \geq m_2 + p_2$ holds. In order to establish that, let us assume, $n_1^* \leq m_1 + p_1 - 1$ and $n_2^* \leq m_2 + p_2 - 1$, i.e., $n = n_1^* + n_2^* \leq m_1 + p_1 + m_2 + p_2 - 2 < 2m + 2p - 1$, which contradicts to our assumption that $n = 2m + 2p - 1$.

Note that, $n_1^* = \sum_{i \in S_1} n_{i1} \leq 2m_1 + 2p - 1$. If possible, let $n_1^* > 2m_1 + 2p - 1$, that is, m_1 small areas have more than $2m_1 + 2p - 1$ observations. Since we previously assumed that (in Theorem 2.3.1) $n_i \geq 2$, for all i. Hence the remaining $(m - m_1)$ small areas have at least $2(m - m_1)$ observations overall. Therefore, $n = n_1^* + n_2^* > 2m_1 + 2p - 1 + 2(m - m_1) = 2m + 2p - 1$, which is a contradiction to the previous assumption that $n = 2m + 2p - 1$. With similar arguments we can establish $n_2^* \leq 2m_2 + 2p - 1$.

Now we consider various scenarios and prove that either (a) or (b) holds.

Case - I: $n_1^* \geq m_1 + p_1$ and $m_1 \leq 3$.

We know $n_1^* \leq 2m_1 + 2p - 1$. Hence, $n_1^* \leq 6 + 2p - 1 = 2p + 5$. Also, $n_2^* = n - n_1^* \geq n - (2p + 5) = 2m - 6 \geq m + p \geq m_2 + p_2$.

Now, let us assume, $m_2 \leq 3$, $n_2^* \leq 2m_2 + 2p - 1 \Rightarrow n_2^* \leq 2p + 6 \leq p - 1 + m < m + p$, which contradicts to our earlier assertion $n_2^* \geq m + p$. Hence $m_2 > 3$. Therefore, $n_2^* \geq m_2 + p_2$, $m_2 > 3$; i.e. condition (b) holds.

Case - II: $n_2^* \geq m_2 + p_2$ and $m_2 \leq 3$. With the similar arguments, as in Case - I, it can be shown that condition (a) holds in this case.

Case - III: $n_2^* < m_2 + p_2$, $m_2 > 3$ or $m_2 \leq 3$. In this case, $n_1^* = n - n_2^* > 2m + 2p - 1 - (m_2 + p_2) \geq 2m + 2p - m - p - 1 \geq m_1 + p_1 - 1$. Hence, $n_2^* < m_2 + p_2 \Longrightarrow n_1^* \geq m_1 + p_1$. Again, let us assume, $m_1 \leq 3$. Now, $n_1^* \leq 2m_1 + 2p - 1 \leq 2p + 5 \Longrightarrow n = n_1^* + n_2^* \leq (2p+5) + (m_2 + p_2 - 1) \leq 2p + 5 + m + p - 1 = 3p + m + 4 \Rightarrow n = 2m + 2p - 1 \leq 3p + m + 4 \Longleftrightarrow m \leq p + 5$, which contradicts to our earlier assumption that $m \geq p + 6$. Therefore $m_1 \nleq 3$ in this case, i. e. $m_1 > 3$. Hence, $n_2^* < m_2 + p_2 \Rightarrow n_1^* \geq m_1 + p_1$ and $m_1 > 3$, i.e., condition (a) holds.

Case - IV: $n_1^* < m_1 + p_1$, $m_1 > 3$ or $m_1 \leq 3$. It can be proved that condition (2) will hold in this scenario. Hence, under the proposed model at least one of the conditions stated will hold. $\square$

Lemma 2.7.2 *The following results hold: (a)* rank $[R_1 Z_1^{(1)}]$ = rank$\begin{pmatrix} M_1 & Z_1^{(1)} \end{pmatrix}$- rank$(M_1)$, *(b)* rank$(R_2)$ $= n_1^* -$ rank$\begin{pmatrix} M_1 & Z_1^{(1)} \end{pmatrix}$.

Proof (a) Here, $R_1 = I - M_1(M_1^T M_1)^{-1} M_1^T \Rightarrow R_1 Z^{(1)} = Z_1^{(1)} - M_1(M_1^T M_1)^{-1} M_1^T Z_1^{(1)} \Rightarrow M_1^T (R_1 Z_1^{(1)}) = 0 \Rightarrow$ columns of M_1 are orthogonal to the columns of $R_1 Z_1^{(1)}$. Therefore,

$$\text{rank}\begin{pmatrix} R_1 Z_1^{(1)} & M_1 \end{pmatrix} = \text{rank}(R_1 Z_1^{(1)}) + \text{rank}(M_1).$$

Now,

$$\begin{pmatrix} R_1 Z_1^{(1)} & M_1 \end{pmatrix} = \begin{pmatrix} Z_1^{(1)} & M_1 \end{pmatrix} \begin{pmatrix} I & 0 \\ -(M_1^T M_1)^{-1} M_1^T Z_1^{(1)} & I \end{pmatrix}$$

$\Rightarrow$ rank$\begin{pmatrix} R_1 Z_1^{(1)} & M_1 \end{pmatrix}$ = rank$\begin{pmatrix} Z_1^{(1)} & M_1 \end{pmatrix}$ (since $\begin{pmatrix} I & 0 \\ -(M_1^T M_1)^{-1} M_1^T Z_1^{(1)} & I \end{pmatrix}$ is non singular) $\Rightarrow$ rank$(R_1 Z_1^{(1)})$ = rank$\begin{pmatrix} Z_1^{(1)} & M_1 \end{pmatrix}$ $-$ rank(M_1).

(b) $R_2 = R_1 - R_1 Z_1^{(1)} (Z_1^{(1)T} R_1 Z_1^{(1)})^{-} Z_1^{(1)T} R_1$, R_2 is idempotent.

Therefore, rank(R_2) = rank(R_1) $-$ rank$(R_1 Z_1^{(1)})$ = $(n_1^* -$ rank$(M_1))$ $-$ (rank$[M_1 Z_1^{(1)}]$ $-$ rank$(M_1))$ = n_1^*- rank $\begin{pmatrix} Z_1^{(1)} & M_1 \end{pmatrix}$. □

Chapter 3

A two-component normal mixture alternative to the Fay-Herriot model

3.1 Introduction

In order to improve the precision of a direct estimate in small domains or small areas, model-based small area estimation techniques are used by many government agencies. Fay and Herriot (1979) proposed a basic area level model which utilizes auxiliary information to borrow information from related areas. The U.S. Census Bureau and other agencies apply area-level models to estimate per capita income and poverty counts for various small areas. Their proposed model is given by:

$$\begin{aligned} y_i &= \theta_i + e_i, \\ \theta_i &= x_i^T \beta + v_i; \qquad i = 1, \ldots, m, \end{aligned} \tag{3.1.1}$$

where the sampling error e_i's independently follow $N(0, D_i)$, v_i's are iid with $N(0, \sigma_v^2)$. The area specific auxiliary data for the i^{th} area is $x_i = (x_{i1}, \ldots, x_{iq})^T$ and y_i is a regular survey estimate of the i^{th} small area mean. The regression coefficient β is a q-component vector. The sampling variances D_i's are assumed to be known.

Datta and Lahiri (1995), Bell et al. (2006) and Xie et al. (2007) argued that normality assumption on the random effects may not be appropriate when outliers are present. Datta and Lahiri (1995) suggested a t-distribution for the random effects when the degrees of freedom is known. Xie et al. (2007) proposed a hierarchical Bayesian method when degrees of freedom of t-distribution is unknown. They assumed a gamma prior for the degrees of freedom. The hyperparameters involved in this gamma distribution need to specified, which calls for authentic expert opinion. Bell et al. (2006) argued that under practical circumstances limited information is obtained from the data regarding degrees of freedom and recommended avoiding using a such a prior. They analyzed a data for several fixed values the degrees of freedom instead of assuming a prior and compared the results.

To this end, we propose an alternative robust extension of the Fay-Herriot model. We suggest that the area specific random effects follow a two-component normal mixture distribution. Our model assumes that the outlying areas follow a normal distribution with larger variance. We consider hierarchical Bayesian approach by assigning objective priors to the parameters involved in the model. In this context, we provide sufficient conditions for the posterior distribution to be proper since some of the proposed priors are improper. We implement the proposed method using Markov chain Monte Carlo (MCMC) integration technique.

The chapter is organized as follows. In Section 3.2 we describe the proposed model and discuss the implementation procedure. We analyze a real data applying our proposed method and describe the results in Section 3.3. A detailed proof of the propriety of the posterior distribution under the proposed model is provided in Section 3.5.

3.2 Two-component normal mixture model

A two component normal mixture extension of Fay-Herriot model is given by:

$$\begin{aligned} y_i &= \theta_i + e_i \\ \theta_i &= x_i^T \beta + \delta_i v_{1i} + (1-\delta_i) v_{2i}, \; i = 1, \ldots, m, \end{aligned} \tag{3.2.1}$$

where δ_i, v_{1i}, v_{2i} are independently distributed with $P(\delta_i = 1|p) = p$, $v_{1i} \sim N(0, A_1)$ and $v_{2i} \sim N(0, A_2)$. Here β is a $(r \times 1)$ vector of regression parameters. The sampling errors $e_1, \ldots, e_m$ are independently normally distributed with $e_i \sim N(0, D_i)$, $i = 1, \ldots, m$, where the sampling variance D_i's are assumed to be known. We further assume that e_i's, v_{1i}'s, v_{2i}'s and δ_i's are independently distributed. In (3.2.1), $x_i = (x_{i1}, \ldots, x_{ir})^T$ is the auxiliary data corresponding to the i^{th} area. We define, $X = (x_1, \ldots, x_m)^T$.
We consider the following class of improper priors,

$$\pi(\beta, A_1, A_2, p) = A_1^{-\alpha_1} A_2^{-\alpha_2} I(0 < A_1 < A_2 < \infty), \tag{3.2.2}$$

where α_1, α_2 are suitably chosen. We impose the restriction $A_1 < A_2$, so that we do not have a label switching problem leading to a lack of identifiability. The area specific random effects corresponding to the outlying areas in the model are assumed to follow a normal distribution with larger variance, which remains the motivation behind imposing such a restriction.

Since the model involves improper priors, we provide sufficient conditions that ensure the resulting posterior distribution from the proposed model will be proper.

Theorem 3.2.1 *The resulting posterior distribution from model (3.2.1) will be proper if, (a) $m > r + 2(2 - \alpha_1 - \alpha_2)$, (b) $\alpha_2 > 1$ and $2 - \alpha_1 - \alpha_2 > 0$, where rank$(X) = r$.*

A detailed proof of Theorem 3.2.1 is given in Section 3.5.

The joint conditional pdf of $\theta = (\theta_1, \ldots, \theta_m)^T$, $\beta = (\beta_1, \ldots, \beta_r)^T$, $\delta = (\delta_1, \ldots, \delta_m)^T$, A_1, A_2 and p is given by:

$$\pi(\theta, \beta, \delta, p|y) \propto \left\{\prod_{i=1}^{m} \exp\left\{-\frac{(y_i - \theta_i)^2}{2D_i}\right\}\right\} \prod_{i=1}^{m} \Bigg[\left\{ \frac{1}{\sqrt{A_1}} \times \exp\left\{-\frac{(\theta_i - x_i^T\beta)^2}{2A_1}\right\}\right\}^{\delta_i}$$
$$\times \left\{ \frac{1}{\sqrt{A_2}} \times \exp\left\{-\frac{(\theta_i - x_i^T\beta)^2}{2A_2}\right\}\right\}^{1-\delta_i} p^{\delta_i}(1-p)^{1-\delta_i} \Bigg]$$
$$\times A_1^{-\alpha_1} A_2^{-\alpha_2} \times I(0 < A_1 < A_2). \tag{3.2.3}$$

From (3.2.3), we get the following full conditional distributions,

(I) $\theta_i|\beta, A_1, A_2, \delta, p, y \overset{\text{ind}}{\sim}$
$N\Big(\dfrac{D_i^{-1}y_i + (A_1^{-1}\delta_i + A_2^{-1}(1-\delta_i))x_i^T\beta}{D_i^{-1} + (A_1^{-1}\delta_i + A_2^{-1}(1-\delta_i))}\ ,\ \dfrac{1}{D_i^{-1} + (A_1^{-1}\delta_i + A_2^{-1}(1-\delta_i))}\Big)$,
$i = 1, \ldots, m$;

(II) $$\beta|\theta, \delta, A_1, A_2, p, y \sim N\Bigg(\left[\sum_{i=1}^{m} \left\{\tfrac{\delta_i}{A_1} + \tfrac{(1-\delta_i)}{A_2}\right\} x_i x_i^T\right]^{-1} \left[\sum_{i=1}^{m} \left\{\tfrac{\delta_i}{A_1} + \tfrac{(1-\delta_i)}{A_2}\right\} x_i \theta_i\right],$$
$$\left[\sum_{i=1}^{m} \left\{\tfrac{\delta_i}{A_1} + \tfrac{(1-\delta_i)}{A_2}\right\} x_i x_i^T\right]^{-1} \Bigg);$$

(III) $p|\theta, \delta, A_1, A_2, \beta, y \sim Beta\left(\sum_{i=1}^{m} \delta_i + 1, m - \sum_{i=1}^{m} \delta_i + 1\right)$;

(IV) $A_1|A_2, \theta, \beta, \delta, p, y$ has the pdf $f_1(A_1)$, where,

$$f_1(A_1) \propto A_1^{-(\alpha_1 + \sum_{i=1}^{m} \frac{\delta_i}{2})} \exp\left\{-\sum_{i=1}^{m} \frac{\delta_i(\theta_i - x_i^T\beta)^2}{2A_1}\right\} I(A_1 < A_2)$$

(V) $A_2|A_1, \theta, \beta, \delta, p, y$ has the pdf $f_2(A_2)$, where,

$$f_2(A_2) \propto A_2^{-(\alpha_2 + \sum_{i=1}^{m} \frac{(1-\delta_i)}{2})} \exp\left\{-\sum_{i=1}^{m} \frac{(1-\delta_i)(\theta_i - x_i^T\beta)^2}{2A_2}\right\} I(A_1 < A_2)$$

(VI) $$P(\delta_i = 1|\theta, \beta, p, \delta_{(-i)}, y) = \frac{p \times \exp\left\{-\frac{(\theta_i - x_i^T\beta)^2}{2A_1}\right\} A_1^{-\frac{1}{2}}}{p \times \exp\left\{-\frac{(\theta_i - x_i^T\beta)^2}{2A_1}\right\} A_1^{-\frac{1}{2}} + (1-p) \times \exp\left\{-\frac{(\theta_i - x_i^T\beta)^2}{2A_2}\right\} A_2^{-\frac{1}{2}}},$$
$i = 1, \ldots, m$

The distributions (IV) and (V) are truncated inverse gamma distributions. In (VI), $\delta_{(-i)} = (\delta_1, \ldots, \delta_{i-1}, \delta_{i+1}, \ldots, \delta_m)^T$.

Our goal is to estimate θ_i, i.e., small area mean for i^{th} area, $i = 1, \ldots, m$. We implement Gibbs sampling using the conditional distributions (I)$-$(VI) in order to find posterior mean and standard deviations of θ_i's.

3.3 Data Analysis

United States Census Bureau conducts various surveys every year to obtain direct estimates of poverty rates for each county. Government agencies need these estimates to allocate funds. Due to small sample size, the standard errors corresponding to the direct estimates are large for many counties. In order to provide estimates with better accuracy, advanced small area estimation techniques are recommended. These methods connect the direct estimates with other sources of information using suitable models.

We have county-level direct estimates of poverty rates for 3141 counties out of 3143 counties in the United States. These estimates are obtained from the American Community Survey (ACS). These direct estimates are five year combined estimates ($2007 - 2011$) of poverty rate for each county. We also have standard errors associated to these estimates. From previous studies, we have learned about possible association between poverty rates and foodstamp participation rates. For this data, the correlation between foodstamp participation rate and the direct estimates of poverty rates is 0.81. This high correlation motivated us to choose foodstamp participation rate as an auxiliary variable. The data set is publicly available but the identification of the counties are suppressed.

We apply our proposed method to this data set and demonstrate the results in Table 3.1. Our choice of α_1 and α_2 are 0.3 and 1.3 respectively. We have also performed further analysis with other choices of α_1 and α_2 within the feasible range, but results did not change significantly. From Table 3.1, we see that the posterior mean of $\hat{A}_2 = 0.00619$ is almost ten times of $\hat{A}_1 = 0.00054$. In addition to that, the estimate $\hat{p} = 0.93$ indicates that there are about 7% small areas which have different area specific variability compared to the rest of the areas. We computed the ratios of the sampling variances of the direct estimates to the posterior variances of the estimates corresponding to the HB estimates of the county level poverty rates. The average of these ratios is 1.4787, which implies that we have achieved a significant amount of gain in terms of precision (almost 50% on average) using the proposed method, over the direct survey estimates. We also apply Fay-Herriot model to the data and observed that on average there is about 3% gain in precision. From Figure 3.3, we see that the estimates obtained from Fay-Herriot model and the propsed two-component mixture model do not agree for some small areas.

Table 3.1: Hierarchical Bayes estimates of the model parameters (for the ACS county level poverty rates data).

Parameter	Posterior Mean	Posterior sd	Posterior Quantiles 2.5%	Median	97.5%
β_1	0.04649	0.00130	0.04401	0.04646	0.04908
β_2	0.66048	0.00751	0.64587	0.66066	0.67479
A_1	0.00054	0.00003	0.00049	0.00054	0.00059
A_2	0.00619	0.00103	0.00454	0.00609	0.00854
p	0.92748	0.02367	0.89627	0.92958	0.95304

3.4 Discussion

In this chapter, we proposed a robust alternative to Fay-Herriot model. The proposed hierarchical Bayesian estimation procedure is straightforward. Other robust alternative is a t-distribution for the random effects, which requires information regarding the degrees of freedom. Xie et al. (2007) proposed a method to estimate degrees of freedom, however Bell et al. (2006) pointed out some issues associated with specifying this prior.

Our proposed method does not require any subjective information regarding the parameters. We provide sufficient conditions for the propriety of the resulting posterior distribution.

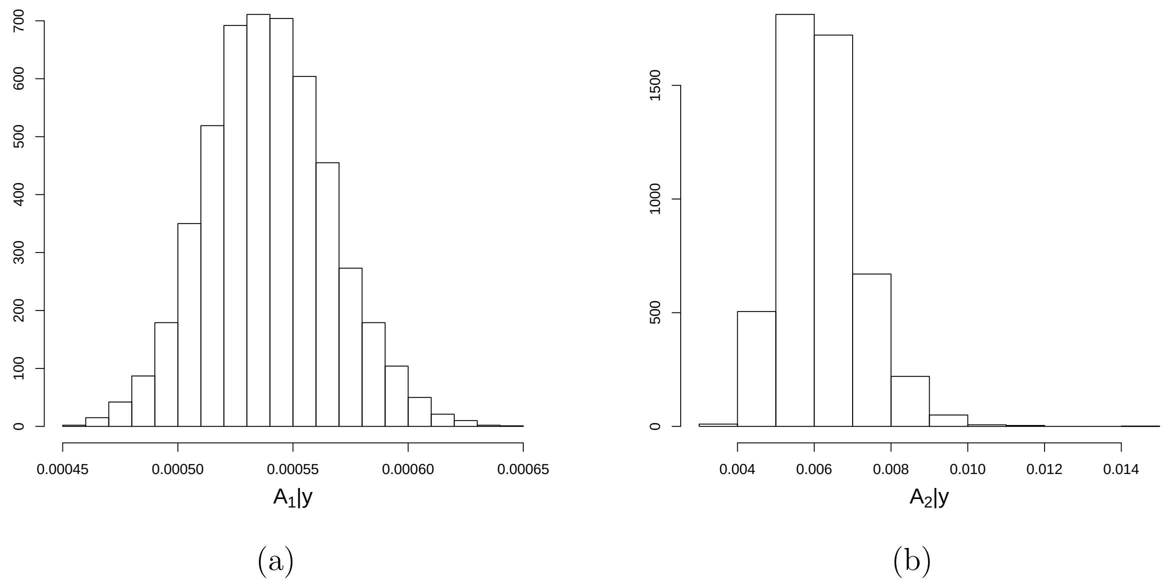


Figure 3.1: *Histograms of the posterior simulations for (a)* A_1 *and (b)* A_2.

We illustrate the method through a data analysis. We have also shown that there is a significant gain in precision using this method over the direct survey estimates.

3.5 Proof of the Theorem

Proof of Theorem 3.2.1: Note that under the proposed mixture model, the likelihood function of the model parameter β, A_1, A_2 and p based on the marginal distribution of $y_1, \ldots, y_m$ is given by,

$$L(\beta, A_1, A_2, p) = C \times \prod_{i=1}^{m} \left[\frac{p}{(A_1 + D_i)^{\frac{1}{2}}} \exp\left\{ -\frac{(y_i - x_i^T \beta)^2}{2(A_1 + D_i)} \right\} + \frac{(1-p)}{(A_2 + D_i)^{\frac{1}{2}}} \exp\left\{ -\frac{(y_i - x_i^T \beta)^2}{2(A_2 + D_i)} \right\} \right], \quad (3.5.1)$$

where C is a generic positive constant not depending on the model parameters. Suppose for

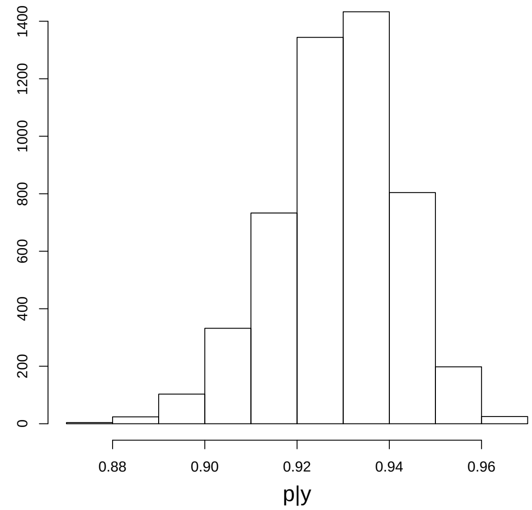


Figure 3.2: *Histogram of the posterior simulations for p.*

$0 < a < b < \infty$ we have $a \le D_i \le b$, $i = 1, \ldots, m$. Since $\dfrac{1}{(A_1 + D_i)^{\frac{1}{2}}}$ is decreasing in D_i and $\exp\left\{-\dfrac{(y_i - x_i^T\beta)^2}{2(A_1 + D_i)}\right\}$ is increasing in D_i, from (3.5.1)

$$L(\beta, A_1, A_2, p) \le C \times \prod_{i=1}^{m} \left[\frac{p}{(A_1 + a)^{\frac{1}{2}}} \exp\left\{-\frac{(y_i - x_i^T\beta)^2}{2(A_1 + b)}\right\} + \frac{(1-p)}{(A_2 + a)^{\frac{1}{2}}} \exp\left\{-\frac{(y_i - x_i^T\beta)^2}{2(A_2 + b)}\right\} \right]. \tag{3.5.2}$$

Let S_1 denote a set of indices $\{i_1, \ldots, i_{k_1}\}$, where $0 \le k_1 \le m$, if $k_1 > 0$, $i_1, \ldots, i_{k_1}$ are distinct indices from $i = 1, \ldots, m$. Similarly, let S_2 denote a set indices $j_1, \ldots, j_{k_2}$, where $0 \le k_2 \le m$, and if $k_2 > 0$, $j_1, \ldots, j_{k_2}$ are distinct indices from $i = 1, \ldots, m$. Furthermore, $S_1 \cap S_2 = \phi$.

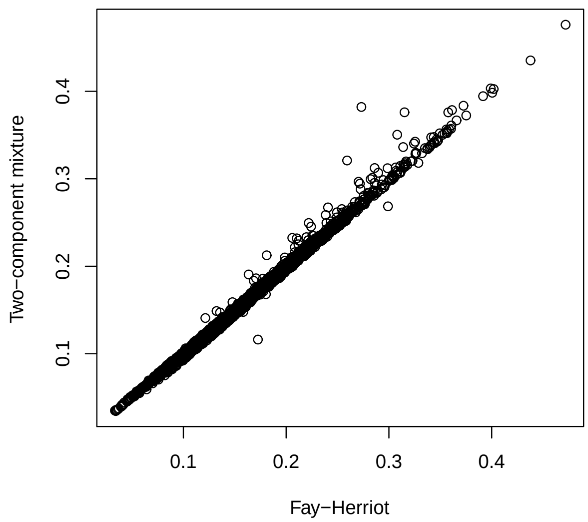


Figure 3.3: *Estimates obtained from the proposed two-component mixture model and the Fay-Herriot model.*

With this notation a typical term from the product of the m factors on the right hand side of (3.5.2) has the following form:

$$t(\beta, A_1, A_2, p) = \frac{p^{k_1}(1-p)^{k_2}}{(A_1+a)^{\frac{k_1}{2}}(A_2+a)^{\frac{k_2}{2}}} \exp\left\{-\frac{1}{2}\left[\sum_{j \in S_1} \frac{(y_j - x_j^T\beta)^2}{A_1+b} + \sum_{j \in S_2} \frac{(y_j - x_j^T\beta)^2}{A_2+b}\right]\right\} \tag{3.5.3}$$

Integrating with respect to p from a typical term of the posterior we get,

$$\begin{aligned} g(\beta, A_1, A_2) = C \times & \frac{A_1^{-\alpha_1} A_2^{-\alpha_2}}{(A_1+a)^{\frac{k_1}{2}}(A_2+a)^{\frac{k_2}{2}}} I(0 < A_1 < A_2 < \infty) \\ & \times \exp\left\{-\frac{1}{2}\left[\sum_{j \in S_1} \frac{(y_j - x_i^T\beta)^2}{A_1+b} + \sum_{j \in S_2} \frac{(y_j - x_i^T\beta)^2}{A_2+b}\right]\right\}. \end{aligned} \tag{3.5.4}$$

To ensure the propriety of the posterior distribution, it is sufficient to verify the integrability of (3.5.4) for each S_1 and S_2. Let us define, $M_1 = \left(x_{i_1}, \ldots, x_{i_{k_1}}\right)^T$. Suppose $r_1 = \text{rank}(M_1)$. Note that, $r_1 = 0$ if $k_1 = 0$. When, $r_1 > 0$, suppose $\{\alpha_1, \ldots, \alpha_{r_1}\} \subset \{i_1, \ldots, i_{k_1}\}$, so that $\left\{x_{\alpha_1}, \ldots, x_{\alpha_{r_1}}\right\}$ is linearly independent. Suppose $\{\gamma_1, \ldots, \gamma_{r-r_1}\} \subset \{j_1, \ldots, j_{k_2}\}$ such that $\left\{x_{\alpha_1}, \ldots, x_{\alpha_{r_1}}, x_{\gamma_1}, \ldots, x_{\gamma_{r-r_1}}\right\}$ is linearly independent. Let us define the $r \times r$ matrix $F = \left(x_{\alpha_1}, \ldots, x_{\alpha_{r_1}}, x_{\gamma_1}, \ldots, x_{\gamma_{r-r_1}}\right)^T$, which is non-singular. Consider the non-singular linear transformation of β by $\phi = F\beta$. From, (3.5.4),

$$
\begin{aligned}
g(\beta, A_1, A_2) \leq C \times & \frac{A_1^{-\alpha_1} A_2^{-\alpha_2}}{(A_1+a)^{\frac{k_1}{2}} (A_2+a)^{\frac{k_2}{2}}} I(0 < A_1 < A_2 < \infty) \\
& \times \exp\left\{-\frac{1}{2}\left[\sum_{u=1}^{r_1} \frac{(y_{\alpha_u} - x_{\alpha_u}^T \beta)^2}{(A_1+b)}\right]\right\} \\
& \times \exp\left\{-\frac{1}{2}\left[\sum_{t=1}^{r-r_1} \frac{(y_{\gamma_t} - x_{\gamma_t}^T \beta)^2}{(A_2+b)}\right]\right\}. \qquad (3.5.5)
\end{aligned}
$$

Note that, for $u = 1, \ldots, r_1$, $x_{\alpha_u}^T \beta = \phi_u$, and for $t = 1, \ldots, r - r_1$, $x_{\alpha_t}^T \beta = \phi_{r_1+t}$. The generic constant C absorbs the jacobian of transformation from β to ϕ.

$$
\begin{aligned}
\int g(\beta, A_1, A_2)\, \mathbf{d}\beta \leq C \times & \frac{A_1^{-\alpha_1 A} A_2^{-\alpha_2}}{(A_1+a)^{\frac{k_1}{2}} (A_2+a)^{\frac{k_2}{2}}} I(0 < A_1 < A_2 < \infty) \\
& \times (A_1+b)^{\frac{r_1}{2}} \times (A_2+b)^{\frac{(r-r_1)}{2}}. \qquad (3.5.6)
\end{aligned}
$$

Note that, for $0 < a < b$, $0 < A_1 < A_2$, $A_1 + a \leq A_1 + b \leq \frac{b}{a}(A_1 + a)$ and $A_2 + a \leq A_2 + b \leq \frac{b}{a}(A_2 + a)$. Then from (3.5.6), we get,

$$
\begin{aligned}
\int g(\beta, A_1, A_2)\, \mathbf{d}\beta \leq C \times & \frac{A_1^{-\alpha_1} A_2^{-\alpha_2}}{(A_1+a)^{\frac{k_1}{2}} (A_2+a)^{\frac{k_2}{2}}} I(0 < A_1 < A_2 < \infty) \\
& \times (A_1+a)^{\frac{-(k_1-r_1)}{2}} \times (A_2+a)^{\frac{-(k_2-r+r_1)}{2}} \\
& = C\, h(A_1, A_2). \qquad (3.5.7)
\end{aligned}
$$

We will now explore the integrability of the function $h(A_1, A_2)$ in (3.5.7). In (3.5.7), it is possible that $k_1 - r_1 = 0$ or $k_2 - r + r_1 = 0$. To ensure the integrability of $h(A_1, A_2)$ with respect to A_1, we need $1 - \alpha_1 > 0$. Actually, this will also be a sufficient condition. Similarly, if $k_2 - r + r_1 = 0$, to ensure the integrability of $h(A_1, A_2)$ with respect to A_2, we need $(1 - \alpha_2) < 0$. Note that, since $0 < A_1 < A_2$,

$$h(A_1, A_2) \leq \frac{A^{-\alpha_1} A^{-\alpha_2}}{(A_1 + a)^{\frac{k_1+k_2-r}{2}}} = \frac{A_1^{-\alpha_1} A_2^{-\alpha_2}}{(A_1 + a)^{\frac{m-r}{2}}}. \tag{3.5.8}$$

$$\begin{aligned}
\int_0^\infty \int_0^\infty h(A_1, A_2)\, \mathrm{d}A_1\, \mathrm{d}A_2 &\leq \int_0^\infty \frac{A_1^{-\alpha_1}}{(A_1 + a)^{\frac{m-r}{2}}} \left(\int_{A_1}^\infty A_2^{-\alpha_2}\, \mathrm{d}A_2 \right) \mathrm{d}A_1 \\
&= (\alpha_2 - 1) \int_0^\infty \frac{A_1^{-\alpha_1-\alpha_2+1}}{(A_1 + a)^{\frac{m-r}{2}}}\, \mathrm{d}A_1 \quad \text{since } \alpha_2 - 1 > 0 \\
&= (\alpha_2 - 1) \int_0^\infty \frac{A_1^{2-\alpha_1-\alpha_2-1}}{(A_1 + a)^{\frac{m-r}{2}-(2-\alpha_1-\alpha_2)+(2-\alpha_1-\alpha_2)}}\, \mathrm{d}A_1 \\
&< \infty,
\end{aligned} \tag{3.5.9}$$

since, $(2 - \alpha_1 - \alpha_2) > 0$ and $\frac{m-r}{2} - (2 - \alpha_1 - \alpha_2) > 0$.

Note that the conditions $\alpha_2 > 1$ and $2 - \alpha_1 - \alpha_2 > 0$ will necessarily imply that $\alpha_1 < 1$. We noted that the condition $\alpha_1 < 1$ would be a necessary and sufficient condition for integrability of $h(A_1, A_2)$ with respect to A_1. Under the conditions $2 - \alpha_1 - \alpha_2 > 0$, $\alpha_2 > 1$, if we take $\alpha_1 = 0$, then we need $1 < \alpha_2 < 2$; and if we take $\alpha_1 = \frac{1}{2}$, then we need $1 < \alpha_2 < \frac{3}{2}$.

Since the integrability conditions for $g(\beta, A_1, A_2)$ do not depend on the indices $\{i_1, \ldots, i_{k_1}\}$ and $\{j_1, \ldots, j_{k_2}\}$ and on the values k_1 and k_2, the conditions $\alpha_2 > 1$, $2 - \alpha_1 - \alpha_2 > 0$ and $m > r + 2(2 - \alpha_1 - \alpha_2)$ will be sufficient to ensure the propriety of the posterior. $\square$

Chapter 4

Robust Bayesian small area estimation for area-level data

4.1 Introduction

In sample surveys, direct estimates based on sample data are computed in order to provide estimates of the population characteristics. When the focus is on a subpopulation, the precision associated to these direct estimates may not be adequate if the sample size corresponding to the subpopulation is not large enough. A small area can be referred to as a small geographic area or a demographic group if the corresponding sample size is small. Since many government and private agencies use small area estimates of required quantities in order to make policies or business strategies, it is necessary to produce reliable small area statistics. To fulfill this increasing demand, a great amount of research has been conducted in the past few decades.

In model-based small area estimation, the objective is to improve the precision of direct estimates by borrowing information from related areas or other surveys in order to improve the precision of direct estimates through appropriately chosen model. These models use relevant auxiliary information as covariates. In this context, linear mixed models are perhaps the most commonly used small area models. Popular small area models such as Fay-Herriot model (Fay and Herriot, 1979) and nested error regression model (Bettese, Harter and Fuller, 1988) are special cases of linear mixed models.

Several approaches have been proposed for estimation of the model parameters and prediction of the small area quantities of interest. Empirical Bayes (EB) and hierarchical Bayes (HB) methods are among the most commonly used estimation methods. It is also important to estimate the variability associated with the estimators. In frequentist approach, approximate expressions of Mean Square Error (MSE) estimators are used (see, for example, Prasad and Rao, 1990; Datta and Lahiri, 2000). In HB approach, posterior standard deviations are usually considered as measure of variability of the estimates.

Model assumptions are inevitable part of small area estimation. Unless an appropriate model is specified results may not be accurate. Fay-Herriot (1979) model incorporates area specific random effect term to account for between area variability. These area specific random effects are assumed to follow normal distribution with unknown variance. Datta and Lahiri (1995) argued against the justification of the assumption of normality for the random effects, particularly in presence of outliers. They proposed a robust estimation method by proposing Cauchy prior for the outlying random effects to account for outliers. Fabrizi and Trivasano (2010) proposed a robust estimation method and suggested exoponential power distribution for random effects.

Datta et al. (2011) discussed the possibility of the random effects not being present and proposed a testing procedure to determine necessity of keeping the random effects in the model for a given data. When the test is not significant a simpler model can be used to produce estimates with better precision. In a different article, Datta et al. (2014) discussed situations when the test is significant but there may still be some areas which does not have any small area effect. To address this problem, Datta et al. (2014) suggested an HB model with a spike-and-slab prior for the random effects. This prior assumes that for each small area random effects may be degenerate at zero with certain probability.

To this end, we suggest a double exponential (Laplace) prior, centered at zero, for the random effects. Double exponential priors are used in Bayesian Lasso regression for variable selection purpose (cf. Park and Casella (2008)). In classical linear regression, penalized least square estimators such as LASSO (Tibshirani, 1996), Ridge and Bridge (Frank and Friedman, 1993) estimators are often used for regression shrinkage and variable selection. Recently, the Bayesian counterpart of some of these frequestist variable selection methods have been developed (Park and Casella 2008; Polson et al., 2014). Methods have also been developed for random effects selection along with variable selection. Bondell et al. (2010), Ibrahim et al. (2011) and Fan and Li (2012) developed efficient techniques for random effects selection under mixed model setup. In the small area context when there are some areas which may not have area specific effect, double exponential prior may be chosen as an alternative to spike-and-slab prior. Double exponential distribution has a spike at its center, which remains as the primary motivation of choosing this distribution.

The choice of priors and hyperprior is an important part of a hierarchical Bayesian model. Elicitation of appropriate subjective priors entails proper expert opinion, which may or may not be available. Even if available, authenticity of the subjective information should also be judged. One can also develop priors based on past data. However, historical data may not also be easily available, for example, in the context of official statistics, historical records sometimes are not accessible due to legal restrictions. In this chapter, we suggest objective priors to the parameters. Specifically, flat priors for both regression parameters and variance of the random effects parameters are assigned. Since both of these priors are improper we analytically show the propriety of the posterior distribution under mild conditions.

As we mentioned before, Fabrizi and Trivasano (2010) proposed a robust Fay-Herriot model where they assumed that the area specific random effects follow an exponential power distribution. They also considered double exponential distribution which is special case of exponential power distribution. They used diffuse uniform priors for the parameters. However, the justification of choosing diffuse prior and the choices of the bounds was not elaborately explained. Berger (2006) pointed out that the results based on uniform priors of bounded sets may highly depend on the choice of bound(s) if the posterior distribution corresponding to the model is improper. This entails the requirement of checking the propriety of posterior distribution based on improper uniform prior distribution.

We extend the Fay-Herriot model by assigning a double exponential prior for the random effects. The tail of a double exponential distribution is thicker than the tail of a normal distribution and it has a spike at the center. We exploit these two properties to produce a robust estimator. Our approach does not require any subjective information and it becomes computationally simple and efficient when we utilize the normal-mixing representation of a double exponential distribution (Andrews and Mallows, 1974) previously used in Park and Casella (2008). In Section 4.2 we discuss the Fay-Herriot model and subsequently describe the proposed extension. In Section 4.3, we perform a simulation study to evaluate the performance of our method. In Section 4.4, we discuss a possible extension of our model.

4.2 Fay-Herriot model and an extension

4.2.1 Fay-Herrot model

Fay and Herriot (1979) proposed a two-level model to improve the precision of direct estimates for small areas using area specific auxiliary data. Suppose there are m small areas and Y_i is the direct estimate of population characteristics for the i^{th} area, $i = 1, \ldots, m$. Let θ_i be the true value of that population characteristics. Their proposed model is given by,

$$
\begin{aligned}
Y_i &= \theta_i + e_i \\
\theta_i &= x_i^T \beta + v_i; \qquad i = 1, \ldots, m,
\end{aligned}
\tag{4.2.1}
$$

where $x_i = (x_{i1}, \ldots, x_{ip})^T$ represents area specific auxiliary data for the i^{th} area. In model (4.2.1), the area specific random effects are denoted by v_i's, $v_i \sim N(0, \sigma_v^2)$ and sampling error for i^{th} small area $e_i \sim N(0, D_i)$, $i = 1, \ldots, m$. The regression parameter β $(p \times 1)$ and random effect variance σ_v^2 are unknown. Sampling variance D_i's are assumed to be known. The Fay-Herriot model considers the area specific random effects term v_i to explain between area variability, it is particularly necessary when the auxiliary data fails to explain this variability. Rao (2003) and Jiang and Lahiri (2006) elaborately discuss various frequentist methods to implement Fay-Herriot model. In order to get hierarchical Bayes estimates of the small area quantities Bayesian formulation of Fay-Herriot model has been suggested by Ghosh (1992), Datta et al. (2005), Datta and Ghosh (2012). The following formulation assumes a noninformative uniform prior for β and A.

$$\begin{aligned} \text{Model } M_{FH}: \quad & y_i|\theta_i, \beta, A \sim N(\theta_i, D_i), \text{ for } i = 1, \ldots, m, \text{ independently;} \\ & \theta_i|\beta, A \sim N(x_i^T\beta, A), \text{ for } i = 1, \ldots, m, \text{ independently;} \\ & \pi(\beta, A) \propto 1, \beta \in \mathbf{R}^p, \ A \in (0, \infty). \end{aligned} \tag{4.2.2}$$

The resulting posterior distribution based on this model will be proper if $m > p+2$. More generally, a sufficient condition for propriety of the posterior distribution for a Bayesian Fay-Herriot model with prior $\pi(\beta, \sigma_v^2) \propto A^{-\alpha}$ is $m > p - 2\alpha + 2$, where $\alpha \in [0, 1)$ (Datta and Ghosh, 2012).

4.2.2 A Normal-double exponential model

The Fay-Herriot model assumes normal distribution for the random effect. In this chapter we suggest a double exponential distribution for the random small area effects. When there are some areas that have no area specific effect, a predictor using this model will function as a sparse predictor. This can be considered to be an application of Bayesian Lasso (Park and Casella, 2008) for random effects selection. On the other hand, prediction based on such a model is expected to be more robust in presence of outliers due to the fact that double exponential distribution has a thicker tail than a normal distribution.

In the following model we suggest a double exponential prior for the area specific random effects. We utilize the normal scale mixture representation of double exponential distribution (Andrews and Mallows, 1974) as discussed in Park and Casella (2008). Suppose there are m small areas, let y_i be the direct estimate and D_i be the sampling variance associated with the estimate for the i^{th} area. Now, our suggested model is,

$$\text{Model } M_{DE}: \qquad y_i|\beta, v_i, A \sim N(x_i^T\beta + v_i, D_i),$$
$$v_i|\beta, A \sim DE(0, A),\ i = 1, \ldots, m, \tag{4.2.3}$$

where $x_i = (x_{i1}, \ldots, x_{ip})^T$ and v_i's are area specific auxiliary variable and the random effect for the i^{th} area, respectively. The model parameters β and A are unknown. In order to produce Bayes predictor for the small area quantities $\theta_i = x_i^T\beta + v_i$, $i = 1, \ldots, m$, we assign the following noninformative prior for β and A,

$$\pi(\beta, A) \propto 1, \beta \in \mathbf{R}^p \text{ and } A \in (0, \infty). \tag{4.2.4}$$

The proposed prior for β and A are improper. Improper priors do not guarantee a proper posteriors. In order to avoid improper posteriors, it is necessary to verify the propriety of the posterior distribution. To this end, we state the following theorem which provides a sufficient condition for the propriety of the posterior distribution.

Theorem 4.2.1 *The posterior distribution resulting from the proposed model (4.2.3) with prior $\pi(\beta, A) \propto 1$, will be proper if $m > p + 2$.*

A proof of Theorem 4.2.1 is provided in Section 4.6. Earlier in Section 4.2, we mention that $m > p + 2$ is also a sufficient condition for the propriety of the posterior resulting from Fay-Herriot model with flat prior for the unknown model parameters.

Theorem 4.2.1 shows that the number of small areas m needed to yield a proper posterior for the new model is as many as it is required to yield a proper posterior for the Bayesian Fay-Herriot model with flat priors.

4.2.3 Computation

Implementation of the model proposed in Section 4.2.2 is straightforward. The implementation procedure is very much along the line of Park and Casella (2008). In order to apply the Bayesian Lasso for regression, Park and Casella (2008) utilized the following representation of double exponential distribution:

$$\frac{\lambda}{2\sqrt{A}} \exp\left\{-\frac{\lambda|v_i|}{\sqrt{A}}\right\} = \frac{1}{\sqrt{2\pi}} \int_0^\infty \frac{1}{\sqrt{(A\tau_i^2)}} \exp\left\{-\frac{v_i^2}{2(A\tau_i^2)}\right\} \times \frac{\lambda^2}{2} \exp\left(-\frac{\lambda^2\tau_i^2}{2}\right) d\tau_i^2. \quad (4.2.5)$$

Motivated by this representation we introduce m latent variables $\tau^2 = (\tau_1^2, \ldots, \tau_m^2)$ and rewrite model (4.2.3) as follows.

(I) Conditional on β and v_i, $y_i \sim N(x_i^T\beta + v_i, D_i)$, $i = 1, \ldots, m$, D_i's are known.

(II) Conditional on A and τ_i^2, $v_i|A, \tau_i^2 \sim N(0, A\tau_i^2)$, i=1,...,m.

We assign the following priors to τ_i^2 given β and A:

$$\pi(\tau_i^2) = \frac{\lambda^2}{2} \exp\left(-\frac{\lambda^2\tau_i^2}{2}\right)$$

The prior distribution $\pi(\beta, A) \propto 1$, remains the same. The value of λ is known. We choose $\lambda^2 = 2$, which makes $var(v_i|A) = A$.

The joint pdf based on the model is,

$$f(y,v,\beta,A,\tau^2) \propto \prod_{i=1}^{m}\left\{\exp\left\{-\frac{(y_i - x_i^T\beta - v_i)^2}{D_i}\right\} \times \frac{1}{(A\tau_i^2)^{\frac{1}{2}}} \times \exp\left\{-\frac{v_i^2}{2(A\tau_i^2)}\right\}\right\}$$
$$\times \prod_{i=1}^{m}\left\{\frac{\lambda^2}{2}\exp\left\{-\frac{1}{2}\lambda^2\tau_i^2\right\}\right\}. \tag{4.2.6}$$

Let $\pi(\tau_i^2|\beta, v, A)$ be the conditional posterior distribution of τ_i^2 given β, v, A, y, from (4.2.6),

$$\begin{aligned}\pi(\tau_i^2|\beta,v,A,y) &\propto \frac{1}{(A\tau_i^2)^{\frac{1}{2}}} \times \exp\left\{-\frac{v_i^2}{2(A\tau_i^2)}\right\} \\ &\quad \times \frac{\lambda^2}{2}\exp\left\{-\frac{1}{2}\lambda^2\tau_i^2\right\} \\ &\propto \frac{1}{(A\tau_i^2)^{\frac{1}{2}}} \times \exp\left\{-\frac{1}{2}\left[\frac{v_i^2}{(A\tau_i^2)} + \lambda^2\tau_i^2\right]\right\}.\end{aligned} \tag{4.2.7}$$

Let us consider the transformation $z = \frac{1}{\tau_i^2}$, hence,

$$\begin{aligned}\pi(z^*|\beta,v,A,y) &\propto \frac{1}{z^{\frac{3}{2}}} \times \exp\left\{-\frac{1}{2}\left[\frac{v_i^2 z^2 + \lambda^2 A}{Az}\right]\right\} \\ &= C \times \frac{1}{z^{\frac{3}{2}}} \times \exp\left\{-\frac{v_i^2}{2}\left[\frac{\left(z - \frac{\lambda\sqrt{A}}{v_i}\right)^2}{Az}\right]\right\} \\ &= C \times \frac{1}{z^{\frac{3}{2}}} \times \exp\left\{-\frac{v_i^2}{2}\left[\frac{\left(z - \frac{\lambda\sqrt{A}}{v_i}\right)^2}{Az}\right]\right\} \\ &= C \times \frac{1}{z^{\frac{3}{2}}} \times \exp\left\{-\frac{\lambda^2}{2}\left[\frac{\left(z - \frac{\lambda\sqrt{A}}{v_i}\right)^2}{z(\frac{\lambda\sqrt{A}}{v_i})^2}\right]\right\},\end{aligned}$$

i.e., $z = \frac{1}{\tau_i^2}|\beta, v, A, y \sim$ Inverse-Gaussian$(\frac{\lambda\sqrt{A}}{v_i}, \lambda^2)$, for $i = 1, \ldots, m$. The conditional distributions for the other parameters in the model are:

1. $\beta|y, v, A, \tau_1^2, \ldots, \tau_m^2 \sim N_p\left((X^TD^{-1}X)X^TD^{-1}(y - v), (X^TD^{-1}X)^{-1}\right)$,

2. $v_i|\beta, A, \tau_i, y \sim N\left(\frac{D_i(y_i - x_i^T\beta)}{(A\tau_i^2 + D_i)}, \left(\frac{1}{D_i} + \frac{1}{A\tau_i^2}\right)^{-1}\right)$, $i = 1, \ldots, m$,

3. $A|\beta, v, \tau_1^2, \ldots, \tau_m^2, y \sim$ Inverse-Gamma$(\frac{m}{2} - 1, \frac{1}{2}\sum_{j=1}^{m}\frac{v_j^2}{\tau_j^2})$.

These full conditional distributions can be used to implement Gibbs sampling in order to get the estimates of the model parameters and predict the small area characteristics.

4.3 Simulation from a sparse random effect model

The goal of developing a small area estimation technique is to estimate the small area quantities accurately and precisely. In this section, we compare the accuracy and precision of two methods based on the estimates obtained from several simulated data sets. We simulate sample data sets from different models with appropriately chosen values of the parameters involved in the model. The two competing methods are implemented to estimate the small area means and then the average square deviation from the true small area mean is measured by computing average squared deviations:

$$\frac{1}{m}\sum_{i=1}^{m}\left(\hat{\theta}_i - \theta_i^*\right)^2, \tag{4.3.1}$$

where θ_i^* is the true small area mean and $\hat{\theta}_i$ be the estimated small area mean for i^{th} small area, $i = 1, \ldots, m$. In order to compare the variability associated with estimates obtained from the two methods we compute the ratio of the posterior standard deviations.

Let there are m small areas, y_i's are the direct estimates and D_i's are the known sampling variances. The first simulation model S_1 is defined below,

$$\text{Model } S_1: \qquad Y_i = \theta_i + e_i,$$
$$\theta_i = x_i^T\beta + \delta_i v_i; \qquad i = 1, \ldots, m,$$

sampling error e_i's are generated from $N(0, D_i)$, with D_i's chosen between 5 to 15; δ_i's are iid Bernoulli random variables with $P(\delta_i = 1) = p^*$. The auxiliary variables $x = (x_1, x_2, x_3)$, x_1^T, x_1 and x_2 are generated from $N(3, 0.9^2)$, χ_2^2 and $Gamma(2, 2)$. The choice of β in the entire simulation study remains $\beta = (1, 2, 3.3, 4.1)^T$. In model S_1, v_i's are generated from $\delta_i I_0 + (1-\delta_i)N(0, \sigma_v^2)$, where I_0 represents a distribution degenerate at zero. We generate 100 data sets from model S_1 for $m = 50$, 100, $p^* = 0.2$, 0.5 and 0.8 and $\sigma_v^2 = 2^2$, 4^2, 5^2, 9^2 and 10^2. By choosing smaller values of p^*, we intend to ensure that many small areas to have $v_i = 0$. The results are demonstrated in Table 4.1 and Table 4.2. In Table 4.2 we see that for larger values of σ_v^2 and smaller values of p^* the new method performs considerably better.

The model S_1 is a two-level model. The first level of the hierarchical model coincides with the first level of both Fay-Herriot model and its new proposed extension. However, the second level of each competing model differs from the second level of model S_1. Hence the generating model does not favor any of the competing models. From Table 4.3 the variability associated with the estimates obtained from the new method is at least as good as that of Fay-Herriot model. It is clear from some chosen values of p^* and σ_v^2, the new methods provides estimates with more precision.

4.4 Exponential power prior

Polson et al. (2014) studied the performance of exponential power prior for regression coefficients and named the estimators based on such priors as Bayesian Bridge estimators. They

Table 4.1: *Average Squared Prediction Error for two different methods. Data sets are simulated from model with $m = 50$ and different choices of p^* and σ_v^2. FH $\equiv$ Fay-Herriot and DE $\equiv$ Double-Exponential*

p^*	Method	$\sigma_v^2 = 2^2$	$\sigma_v^2 = 4^2$	$\sigma_v^2 = 5^2$	$\sigma_v^2 = 9^2$	$\sigma_v^2 = 10^2$
0.2	DE	1.6807	3.0912	3.6623	4.9106	5.1446
	FH	1.7015	3.2035	3.9435	6.0245	6.3512
0.5	DE	2.6914	5.0504	5.8938	7.6133	7.8183
	FH	2.6287	4.9992	5.8978	7.9431	8.2118
0.8	DE	3.3661	5.8708	6.5853	8.0521	8.2425
	FH	3.2438	5.7336	6.5414	8.1683	8.3608

Table 4.2: *Average Squared Prediction Error for two different methods. Data sets are simulated from model with $m = 100$ and different choices of p^* and σ_v^2. FH $\equiv$ Fay-Herriot model and DE $\equiv$ Double-Exponential model*

p^*	Method	$\sigma_v^2 = 2^2$	$\sigma_v^2 = 4^2$	$\sigma_v^2 = 5^2$	$\sigma_v^2 = 9^2$	$\sigma_v^2 = 10^2$
0.2	DE	1.1038	2.1744	2.7162	4.2643	4.5086
	FH	1.1074	2.3679	3.0492	5.3865	5.8170
0.5	DE	1.7899	4.4283	4.8250	5.3400	6.4901
	FH	1.8072	4.6072	5.1375	5.9904	7.3310
0.8	DE	2.6090	5.2354	6.0462	7.9422	7.7978
	FH	2.5913	5.1939	6.0385	8.08	7.9156

suggested an efficient Gibbs sampling procedure to implement this method. In the context of robust small area estimation exponential power prior for the area specific random effect has previously been suggested by Fabrizi and Trivasano (2010).

Table 4.3: *Ratio of posterior standard deviations (FH/DE) for two different methods based on the data sets are simulated from model with* $m = 50, 100$ *and different choices of* p^* *and* σ_v^2.

p^*	m	$\sigma_v^2 = 2^2$	$\sigma_v^2 = 4^2$	$\sigma_v^2 = 5^2$	$\sigma_v^2 = 9^2$	$\sigma_v^2 = 10^2$
0.2	$m = 50$	1.02	1.02	1.03	1.05	1.06
	$m = 100$	1.04	1.04	1.04	1.08	1.08
0.5	$m = 50$	1.01	1.03	1.03	1.03	1.02
	$m = 100$	1.02	1.05	1.04	1.03	1.05
0.8	$m = 50$	1.00	1.01	1.02	1.02	1.02
	$m = 100$	1.01	1.03	1.02	1.02	1.02

An exponential power (EP) distribution center at 0 and scale σ and power φ is denoted by $\text{EP}(0, \sigma, \varphi)$ has the following form,

$$f(x|\sigma, \varphi) = c\frac{1}{\sigma} \exp\left\{-\left|\frac{x}{\sigma}\right|^{\varphi}\right\} \quad \text{where } x \in \mathbf{R},\ \sigma \in \mathbf{R}^{+} \text{ and } \varphi > 0\ . \tag{4.4.1}$$

Fabrizi and Trivasano (2010) used a slightly different formulation of exponential power. For, $\varphi = 1$, the distribution (4.4.1) is equivalent to a double exponential distribution with mean 0 and variance σ^2. The, choice $\varphi = 2$ leads to a normal distribution with mean 0 and variance σ^2.

An exponential power distribution (4.4.1) with suitable choice of φ has heavier tail than double exponential distribution. That motivates us to study the following extension of Fay-Herriot model.

$$\begin{aligned} \text{Model } M_{EP}: \qquad & y_i|\beta, v_i, A \sim N(x_i^T\beta + v_i, D_i), \\ & v_i|\beta, A \sim EP(0, \sqrt{A}, \varphi),\ \ i = 1, \ldots, m, \\ & \pi(\beta, A) \propto 1,\ \ \beta \in \mathbf{R}^p \text{ and } \ A \in (0, \infty). \end{aligned} \tag{4.4.2}$$

This is, of course a generalization of (4.2.3). Since, (4.4.2) involves improper priors we state the following theorem which provides a sufficient condition for the propriety of the posterior distribution resulting from the model.

Theorem 4.4.1 *The posterior distribution resulting from the model (4.4.2) with* $0 < \varphi < 2$, *will be proper if* $m > p + 2$.

Proof: The joint pdf obtained from the model (4.4.2) is:

$$f(y, v, \beta, A) \propto C \times \exp\left\{-\frac{1}{2}(y^* - X\beta)^T D^{-1}(y^* - X\beta)\right\} \times \frac{1}{\left(\sqrt{A}\right)^m} \exp\left\{-\sum_{i=1}^{m}\left|\frac{v_i}{\sqrt{A}}\right|^{\varphi}\right\}, \tag{4.4.3}$$

where $y^* = y - v$, $y = (y_1, \ldots, y_m)^T$, $v = (v_1, \ldots, v_m)^T$ and X is an $(m \times p)$ matrix. Let us consider the transformation, $u^{-\frac{1}{\varphi}} = \sqrt{A}$, the Jacobian of transformation is, $\dfrac{u^{-\left(\frac{2}{\varphi}+1\right)}}{\varphi}$. Hence,

$$\begin{aligned} f(y, v, \beta, u) = C \times \exp&\left\{-\frac{1}{2}(y^* - X\beta)^T D^{-1}(y^* - X\beta)\right\} \\ \times \exp&\left\{-u\sum_{i=1}^{m}|v_i|^{\varphi}\right\} \times u^{\frac{m-2}{\varphi}-1}. \end{aligned} \tag{4.4.4}$$

Now,

$$\begin{aligned} \int f(y, \beta, v, u)\,\mathrm{d}u &= C \times \exp\left\{-\frac{1}{2}(y^* - X\beta)^T D^{-1}(y^* - X\beta)\right\} \\ &\quad \times \int \exp\left\{-u\sum_{i=1}^{m}|v_i|^{\varphi}\right\} \times u^{\frac{m-2}{\varphi}-1}\,\mathrm{d}u \\ &= C \times \exp\left\{-\frac{1}{2}(y^* - X\beta)^T D^{-1}(y^* - X\beta)\right\} \\ &\quad \times \left[\left(\sum_{i=1}^{m}|v_i|^{\varphi}\right)^{\frac{m-2}{\varphi}}\right]^{-1} \end{aligned} \tag{4.4.5}$$

Before proceeding further we state the following inequality,

$$\sum_{i=1}^{m} |a_i|^{\varphi} \geq \left(\sum_{i=1}^{m} a_i^2 \right)^{\frac{\varphi}{2}}, \quad \text{where } 0 < \varphi < 2, \; a_i \in \mathbf{R}, i = 1, \dots, m. \tag{4.4.6}$$

Using (4.4.6), we have,

$$\int f(y, \beta, v, u) \, \mathrm{d}u \leq C \times \exp\left\{ -\frac{1}{2}(y^* - X\beta)^T D^{-1}(y^* - X\beta) \right\} \times \left[\left(\sum_{i=1}^{m} v_i^2 \right)^{\frac{m-2}{2}} \right]^{-1} \tag{4.4.7}$$

The right side of (4.4.7) is equivalent to (4.6.2) in Section (4.6). It has been shown that the right side is integrable with respect to β and v. If we follow the steps similarly as in the proof of Theorem 4.2.1 it can be shown that the right side of (4.4.7) is integrable when $m > p + 2$. □

4.5 Discussion

In this chapter, we propose an objective Bayesian small area model which accounts for the presence of sparsity in the random effects. We perform a comparison study through simulation and conclude that our method performs well when area specific random effects are absent for some small areas. We provide sufficient condition for propriety of the posterior resulting from the model. The sufficient condition is same as the sufficient condition for propriety of a posterior for a Bayesian Fay-Herriot model (Datta and Ghosh, 2012). We discuss possibility of further extension of the model using exponential power prior for the random effects, we also provide a sufficient condition the propriety of the posterior distribution resulting from the model.

4.6 Proof of the Theorem

Proof of Theorem 4.2.1: The joint pdf obtained from model M_{DE} is given by:

$$f(y,\beta,v,A) = C \times \exp\left\{-\frac{1}{2}(y^*-X\beta)^T D^{-1}(y^*-X\beta)\right\} \times \frac{1}{A^{\frac{m}{2}}} \times \exp\left\{-\sum_{i=1}^{m}\frac{|v_i|}{\sqrt{A}}\right\}, \tag{4.6.1}$$

where $y^* = y - v$, $y = (y_1, \ldots, y_m)^T$, $v = (v_1, \ldots, v_m)^T$ and $X = (x_1, \ldots, x_p)$ is of order $m \times p$ matrix, we assume $\text{rank}(X) = p$. Consider the transformation $w = \frac{\sum_{i=1}^m |v_i|}{\sqrt{A}}$, the jacobian of transformation is $\frac{(\sum_{i=1}^m |v_i|)^2}{w^3}$. Therefore,

$$f(y,\beta,v,w) = C \times \exp\left\{-\frac{1}{2}(y^*-X\beta)^T D^{-1}(y^*-X\beta)\right\} \times \frac{w^{m-3}}{(\sum\limits_{i=1}^{m}|v_i|)^{m-2}} \times \exp\{-w\}.$$

Now,

$$\begin{aligned}\int f(y,\beta,v,w)\,\mathrm{d}w &= C \times \frac{1}{(\sum\limits_{i=1}^{m}|v_i|)^{m-2}} \times \exp\left\{-\frac{1}{2}(y^*-X\beta)^T D^{-1}(y^*-X\beta)\right\} \\ &\quad \times \int w^{m-3} \times \exp\{-w\}\,\mathrm{d}w \\ &= C \times \frac{1}{(\sum\limits_{i=1}^{m}|v_i|)^{m-2}} \times \exp\left\{-\frac{1}{2}(y^*-X\beta)^T D^{-1}(y^*-X\beta)\right\},\end{aligned}$$

since $m-2>0$.

$$\le C \times \frac{1}{(\sum\limits_{i=1}^{m}v_i^2)^{\frac{m-2}{2}}} \times \exp\left\{-\frac{1}{2}(y^*-X\beta)^T D^{-1}(y^*-X\beta)\right\} \tag{4.6.2}$$

Since the inequality $(\sum\limits_{i=1}^{m}|v_i|)^2 \ge (\sum\limits_{i=1}^{m}v_i^2) \iff \frac{1}{(\sum\limits_{i=1}^{m}|v_i|)^{m-2}} \le \frac{1}{(\sum\limits_{i=1}^{m}v_i^2)^{\frac{(m-2)}{2}}}$ holds for $v_i \in \mathbb{R}$, for all i and $m-2>0$.

The exponent in (4.6.2) can be expressed as

$$(y^* - X\beta)^T D^{-1}(y^* - X\beta) = y^* \left\{D^{-1} - D^{-1}X^T(X^TD^{-1}X)^{-1}X^TD^{-1}\right\} y^* + (\beta - \hat{\beta})^T(X^TD^{-1}X)(\beta - \hat{\beta}),$$

where $\hat{\beta}$=$(X^TD^{-1}X)^{-1}X^TD^{-1}y^*$, Therefore,

$$\begin{aligned}\int f(y,\beta,v,w)\,\mathrm{d}w\,\mathrm{d}\beta &\leq C \times \frac{1}{(\sum_{i=1}^{m} v_i^2)^{\frac{m-2}{2}}} \times \exp\left\{-\frac{1}{2}y^{*T}Qy^*\right\} \\ &\quad \times \int \exp\left\{-\frac{1}{2}(\beta - \hat{\beta})^T(X^TD^{-1}X)(\beta - \hat{\beta})\right\} \mathrm{d}\beta \\ &= C \times \frac{1}{(\sum_{i=1}^{m} v_i^2)^{\frac{m-2}{2}}} \times \exp\left\{-\frac{1}{2}y^{*T}Qy^*\right\}, \end{aligned} \tag{4.6.3}$$

where $Q = \left\{D^{-1} - D^{-1}X^T(X^TD^{-1}X)^{-1}X^TD^{-1}\right\}$. Since D is nonsingular, rank(Q) =rank(QD) = rankQ^*, where $Q^* = (I - D^{-1}X(X^TD^{-1}X)^{-1}X^T)$, which is symmetric and idempotent. Hence, rank$(Q^*) = \mathrm{tr}(I) - \mathrm{tr}(D^{-1}X(X^TD^{-1}X)^{-1}X^T) = m - p = t$ (say).

Let, $P^{m\times m}$ be an orthogonal matrix such that, $P^TQP = diag(\lambda_1, \ldots, \lambda_t, \ldots, 0)$, where $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_t > 0$ and $t = m - p$. Consider, $u = P^Tv$, therefore,

$$\begin{aligned} y^{*T}Qy^* &= (y - v)^TQ(y - v) \\ &= (y - Pu)^TQ(y - Pu) = (P^Ty - u)^TP^TQP(P^Ty - u) \\ &= (s - u)^TP^TQP(s - u) \quad \text{where } s = P^Ty \\ &= \sum_{j=1}^{t} \lambda_j(s_j - u_j)^2. \end{aligned}$$

Hence from (4.6.3),

$$\int f(y,\beta,u,w)\,\mathrm{d}w\,\mathrm{d}\beta \leq C \times \frac{1}{(\sum\limits_{i=1}^{m} u_i^2)^{\frac{m-2}{2}}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{t}\lambda_j(s_j-u_j)^2\right\} \tag{4.6.4}$$

Now, at first we integrate (4.6.5) with respect to u_m

$$\begin{aligned}
\int f(y,\beta,u,w)\,\mathrm{d}w\,\mathrm{d}\beta\,\mathrm{d}u_m &\leq C \times \exp\left\{-\frac{1}{2}\sum_{j=1}^{t}\lambda_j(s_j-u_j)^2\right\} \\
&\quad \times \int_{-\infty}^{\infty} \frac{1}{(\sum\limits_{i=1}^{m} u_i^2)^{\frac{m-2}{2}}}\,\mathrm{d}u_m \\
&= C \times \exp\left\{-\frac{1}{2}\sum_{j=1}^{t}\lambda_j(s_j-u_j)^2\right\} \\
&\quad \times \int_{-\infty}^{\infty} \frac{1}{(\sum\limits_{i=1}^{m-1} u_i^2 + u_m^2)^{\frac{m-2}{2}}}\,\mathrm{d}u_m \\
&= C \times \exp\left\{-\frac{1}{2}\sum_{j=1}^{t}\lambda_j(s_j-u_j)^2\right\} \\
&\quad \times \int_{-\infty}^{\infty} \frac{1}{\left(\left\{\sqrt{\sum_{i=1}^{m-1} u_i^2}\right\}^2 + u_m^2\right)^{\frac{m-3+1}{2}}}\,\mathrm{d}u_m \\
&= C \times \exp\left\{-\frac{1}{2}\sum_{j=1}^{t}\lambda_j(s_j-u_j)^2\right\} \\
&\quad \times \frac{1}{\left(\sum_{i=1}^{m-1} u_i^2\right)^{\frac{m-2-1}{2}}} \qquad (4.6.5)\\
&\quad \text{(since } m-3>0\text{)}.
\end{aligned}$$

Similarly, after successive integration with respect to $u_{m-1}, \ldots u_{t+1}$,

$$\begin{aligned}\int f(y,\beta,u,w)\,\mathrm{d}w\,\mathrm{d}\beta \prod_{i=t+1}^{m} \mathrm{d}u_i &\leq C \times \exp\left\{-\frac{1}{2}\sum_{j=1}^{t} \lambda_j (s_j - u_j)^2\right\} \\ &\quad \times \frac{1}{\left(\sum_{i=1}^{t} u_i^2\right)^{\frac{m-2-(m-t)}{2}}} \\ &= C \times \exp\left\{-\frac{1}{2}\sum_{i=1}^{t} \lambda_j (s_j - u_j)^2\right\} \\ &\quad \times \frac{1}{\left(\sum_{i=1}^{t} u_i^2\right)^{\frac{t-2}{2}}} \end{aligned} \tag{4.6.6}$$

since $t-2>0$, i.e., $m-p>2$. Let, $\min\limits_{1\leq j\leq t} \lambda_j = \frac{1}{\delta^2} > 0$,

$$\begin{aligned}\int f(y,\beta,u,w)\,\mathrm{d}w\,\mathrm{d}\beta \prod_{i=t+1}^{m} \mathrm{d}u_i \prod_{i=1}^{t} \mathrm{d}u_i &\leq C \times \int \exp\left\{-\frac{1}{2}\sum_{j=1}^{t} \lambda_j (s_j - u_j)^2\right\} \\ &\quad \times \frac{1}{\left(\sum_{i=1}^{t} u_i^2\right)^{\frac{t-2}{2}}} \prod_{i=1}^{t} \mathrm{d}u_i \\ &\leq C \times \int \exp\left\{-\frac{1}{2\delta^2}\sum_{i=1}^{t} (s_j - u_j)^2\right\} \\ &\quad \times \frac{1}{\left(\sum_{i=1}^{t} u_i^2\right)^{\frac{t-2}{2}}} \prod_{i=1}^{t} \mathrm{d}u_i. \end{aligned} \tag{4.6.7}$$

If $X_i \sim N(\mu_i, \sigma^2)$, independently, $i = 1, \ldots, t$ then $\sum\limits_{i=1}^{t} \frac{X_i^2}{\sigma^2} \sim \chi_t^2(\alpha)$, where $\alpha = \sum\limits_{i=1}^{t} \frac{\mu_i^2}{\sigma^2}$ is the non-centrality parameter and t is the degrees of freedom. Using this fact in (4.6.7),

$$
\begin{aligned}
\int f(y,\beta,u,w)\,\mathrm{d}w\,\mathrm{d}\beta \prod_{i=t+1}^{m}\mathrm{d}u_i \prod_{i=1}^{t}\mathrm{d}u_i \le{}& C\times \int \exp\left\{-\frac{1}{2\delta^2}\sum_{i=1}^{t}(s_j-u_j)^2\right\} \\
&\times \frac{1}{\left(\sum_{i=1}^{t}u_i^2\right)^{\frac{t-2}{2}}}\prod_{i=1}^{t}\mathrm{d}u_i \\
={}& C\times \mathrm{E}\left[\frac{1}{(\chi_t^2(\alpha))^{\frac{t-2}{2}}}\right] \\
={}& C\times \sum_{i=0}^{\infty}\frac{\exp\{-\frac{\lambda}{2}\}(\frac{\lambda}{2})^i}{i!}\times \mathrm{E}\left[\frac{1}{(\chi_{t+2i}^2)^{\frac{t-2}{2}}}\right] \\
\le{}& C\times \sum_{i=0}^{\infty}\frac{\exp\{-\frac{\lambda}{2}\}(\frac{\lambda}{2})^i}{i!}\times \mathrm{E}\left[\frac{1}{(\chi_{t}^2)^{\frac{t-2}{2}}}\right] \qquad (4.6.8)
\end{aligned}
$$

In (4.6.8), we use the fact that for $\lambda_2 \ge \lambda_1 > 0$,

$$
\mathrm{E}\left[g(\chi_{\lambda_2}^2)\right] \le \mathrm{E}\left[g(\chi_{\lambda_1}^2)\right], \qquad (4.6.9)
$$

where $g(x)$ is a nonnegative decreasing function in x. The result (4.6.9) holds for χ^2 random variables since $\chi_{\lambda_2}^2 \overset{\mathrm{st}}{\ge} \chi_{\lambda_1}^2$ for $\lambda_2 \ge \lambda_1$. In our case, $g(x) = \frac{1}{x^{\frac{t}{2}-1}}$ is decreasing in x since $t > 2$, i.e., $m > p+2$. From the right hand side of (4.6.8),

$$
\begin{aligned}
&C\times \sum_{i=0}^{\infty}\frac{\exp\{-\frac{\lambda}{2}\}(\frac{\lambda}{2})^i}{i!}\times \mathrm{E}\left[\frac{1}{(\chi_{t}^2)^{\frac{t-2}{2}}}\right] \\
&= C\times \left(\sum_{i=0}^{\infty}\frac{\exp\{-\frac{\lambda}{2}\}(\frac{\lambda}{2})^i}{i!}\right)\left(\int_0^{\infty}\frac{\exp\left\{-\frac{w}{2}\right\}w^{\frac{t}{2}-1}}{w^{\frac{t-2}{2}}}\mathrm{d}w\right) \\
&= C\times 2 < \infty. \qquad (4.6.10)
\end{aligned}
$$

Hence,

$$\int f(y,\beta,u,w)\,\mathrm{d}w\,\mathrm{d}\beta\prod_{i}^{m}\mathrm{d}u_i<\infty.$$
$$\Longrightarrow\int f(y,\beta,v,A)\,\mathrm{d}v\,\mathrm{d}\beta\mathrm{d}A<\infty, \tag{4.6.11}$$

which implies the posterior distribution $f(\beta,v,A|y)$ is proper. □

Chapter 5

Hierarchical Bayesian Methods for Combining Surveys

5.1 Introduction

In order to estimate the number of occupied housing units (households), many surveys are conducted by the United States Census Bureau. In Table 5.1, we report the household estimates from 2002 to 2011, obtained by the Current Population Survey (CPS), the Housing Vacancy Survey (HVS), the American Community Survey (ACS) and the American Housing Survey (AHS). Differences among the survey estimates are noticeable in Table 5.1. Estimates obtained by the CPS are consistently high over the years and the estimates from the HVS and the AHS are typically low. In order to combine the estimates obtained by these surveys, we propose and discuss various hierarchical Bayesian (HB) models. In this chapter, we study and compare the combined estimates obtained from these HB methods.

Table 5.1: *Estimates of households, obtained in different surveys (numbers in 1000s).*

Survey	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
CPS/ASEC	111278	112000	113343	114384	116011	116783	117181	117538	119927	121084
HVS	104994	105636	106971	108667	109736	110173	110475	112295	112899	113533
ACS	107367	108420	109902	111091	111617	112378	113101	113616	114567	114992
AHS	.	105842	.	108871	.	110692	.	111806	.	114907

Table 5.2: *Standard errors of the estimates obtained in different surveys (numbers in 1000s).*

Survey	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
CPS/ASEC	260	260	235	234	261	261	261	262	262	262
HVS	185	182	179	204	194	187	181	174	173	171
ACS	.	.	.	144	146	144	147	161	163	180
AHS	.	165	.	218	.	231	.	238	.	396

5.2 A hierarchical Bayesian model to combining several unbiased survey estimates

Let θ be a population characteristic of interest and suppose estimates of θ are available from m different surveys. Moreover, suppose that these surveys are repeated over time, annually or biennially. Thus some surveys may not have been conducted over every time point of interest. While one or more surveys are conducted at every time point 1 to T, not all surveys are done at every time point. Suppose the i^{th} survey is conducted at time points belonging to a set $S_i \subset \{1, 2, \ldots, T\}$. We assume that, $\bigcup_{i=1}^{m} S_i = \{1, 2, \ldots, T\}$, i.e., at least one of the surveys is conducted every year. To estimate θ_t, the population characteristic of interest at time t, we consider the following model:

$$y_{it} = \theta_t + e_{it}, \qquad t \in S_i \subset \{1, 2, \ldots, T\}, \tag{5.2.1}$$

where, y_{it} is the estimate of θ_t from the i^{th} survey at the t^{th} time point. We assume that sampling errors $e_{it} \sim N(0, \sigma^2_{it})$, $t \in S_i$, $i = 1 \ldots, m$, are independently distributed. We assume that σ^2_{it}'s are known. Now, we propose the following random walk model for θ_t:

$$\theta_t = \theta_{t-1} + e^*_t, \qquad t = 1, 2, \ldots, T, \tag{5.2.2}$$

where, e^*_t's are independently distributed with a truncated normal distribution truncated above 0, with variance $\sigma^2_{e^*}$. We assume that $\sigma^2_{e^*}$ and θ_o are unknown. Our proposed hierarchical Bayesian model for estimating the number of households is:

$$\text{Model } M_1 : \qquad \begin{aligned} y_{it}|\theta_0, \theta_t, \sigma^2_{e^*} &\overset{\text{ind}}{\sim} N(\theta_t, \sigma^2_{it}),\ t \in S_i,\ i = 1, \ldots, m, \\ \theta_t &= \theta_{t-1} + e^*_t,\ t = 1, \ldots, T, \\ e^*_t|\sigma^2_{e^*} &\overset{\text{iid}}{\sim} \text{truncated } N(0, \sigma^2_{e^*}), \end{aligned} \tag{5.2.3}$$

with lower truncation point 0. In model M_1, values of σ^2_{it}'s are known. The values of θ_0 and $\sigma^2_{e^*}$ are not available, so we assign the following noninformative priors to those parameters: θ_o and $\sigma^2_{e^*}$ are independently distributed with Uniform$(0, \infty)$.

Since we assume improper prior to some parameters in the model, the propriety of the posterior distribution resulting from the model need to be ensured. Theorem 5.2.1 provides sufficient conditions for the propriety of the posterior density for the model stated above.

From Table 5.2 we see that standard errors are not available from the American Community Survey from 2002−2004. Also, the American Housing Survey estimates along with the standard errors are missing at every alternative year from 2002−2011 (Tables 5.1 and 5.2). Let us introduce indicator variables δ_{it}'s, such that, $\delta_{it} = 1$ if data from the i^{th} survey is available at time t and $\delta_{it} = 0$ otherwise, $i = 1 \ldots, m$ and $t = 1, \ldots, T$. We also define, $n_t = \sum_{i=1}^{m} \delta_{it}$, $t = 1, \ldots, T$.

Theorem 5.2.1 *The posterior distribution resulting from the proposed model will be proper if (a) $n_t > 0$ for all t, and (b) the number of time points $T > 3$.*

Proof of the theorem is provided in Section 5.6.1. Since there are some missing y_{it}'s along with σ_{it}^2's, we define the variable r such that, $r_{it} = 0$ if σ_{it}^2 is missing and $r_{it} = \frac{1}{\sigma_{it}^2}$ otherwise. The following full conditional distributions obtained below will be essential to perform a Gibbs Sampling.

(a) $\theta_T|\theta_0, \theta_1, \ldots, \theta_{T-1}, \sigma_{e^*}^2, y \sim$ truncated Normal with mean $= \dfrac{\sum_{i=1}^{m} r_{iT} y_{iT} + \sigma_{e^*}^{-2}\theta_{T-1}}{\sum_{i=1}^{m} r_{iT} + \sigma_{e^*}^{-2}}$ and variance $= (\sum_{i=1}^{m} r_{iT} + \sigma_{e^*}^{-2})^{-1}$ with lower truncation point θ_{T-1}.

(b) $\theta_t|\theta_0, \theta_1, ., \theta_{t-1}, \theta_{t+1}, ., \theta_T, \sigma_{e^*}^2, y \sim$ truncated Normal with mean $= \dfrac{\sum_{i=1}^{m} r_{it} y_{it} + \sigma_{e^*}^{-2}(\theta_{t-1} + \theta_{t+1})}{\sum_{i=1}^{m} r_{it} + 2\sigma_{e^*}^{-2}}$ and variance $= (\sum_{i=1}^{m} r_{it} + 2\sigma_{e^*}^{-2})^{-1}$ truncated in $(\theta_{t-1}, \theta_{t+1})$; $t = 1, \ldots, T-1$.

(c) $\theta_0|\theta_1, \ldots, \theta_T, \sigma_{e^*}^2, y \sim$ truncated Normal with mean $= \theta_1$ and variance $= \sigma_{e^*}^2$ truncated in $(0, \theta_1)$.

(d) $\sigma_{e^*}^2|\theta_0, \ldots, \theta_T, y \sim$ *Inverse-Gamma* (IG) with shape $= \frac{T}{2} - 1$, rate $= \sum_{t=1}^{T} \dfrac{(\theta_t - \theta_{t-1})^2}{2}$.

(If X $\sim$ *Inverse-Gamma*(α, β), then the pdf of X is $f(x) \propto x^{-\alpha-1} \exp\{-\frac{\beta}{x}\}$, where α is the shape and β is the rate parameter.)

We perform a Gibbs sampling to get the HB estimates of θ_t's. Table 5.4 presents the proposed HB estimates and posterior standard deviations of θ_t's. Figure 5.1 shows the proposed combined estimates and the survey estimates. From Table 5.3 we get the estimates of θ_0 and $\sigma_{e^*}^2$ obtained by our method. In Figure 5.3, we plot histograms based on the simulated values from the posterior distribution of $\sigma_{e^*}^2$. In Table 5.6 we provide some details about the simulated values from the posterior distribution of $\sigma_{e^*}^2$. Table 5.5 and Figure 5.2 (b) show that the posterior standard deviations associated with the Bayes estimates of

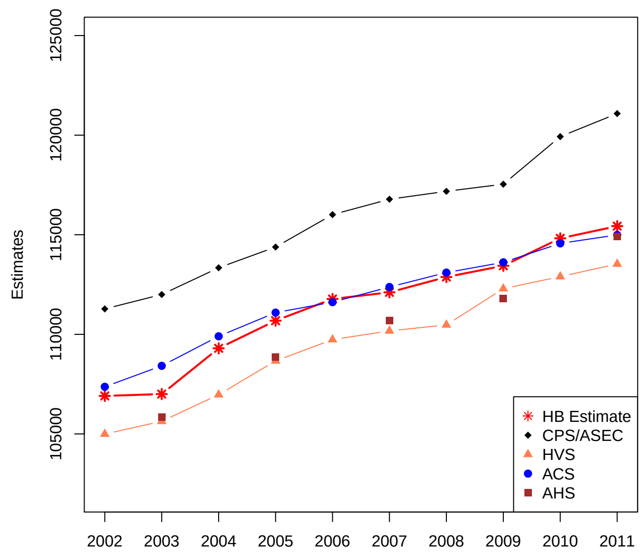
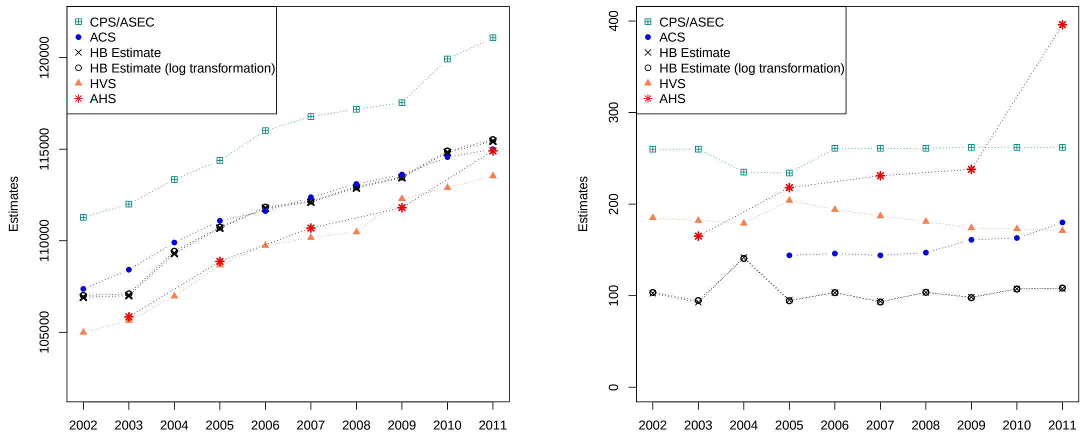


Figure 5.1: *Proposed HB estimates based on model M_1 and other survey estimates.*

θ_t obtained by our method are considerably lower than the standard errors of the survey estimates. This implies we achieve significant gain in precision by applying model M_1.

Table 5.3: *Summary of the posterior simulations (numbers in 1000s).*

	Posterior	Posterior	Simulated Quantiles		
Parameter	Mean	sd	2.5%	Median	97.5%
θ_0	105756.25	995.92	105296.47	105993.36	106489.93
$\sigma^2_{e^*}$	2369485.65	1921155.29	1300676.22	1854130.76	2774842.83



(a) (b)

Figure 5.2: *(a) Proposed HB estimates based on models M_1 and M_2 (with and without log transformation) and other survey estimates. (b) Posterior standard deviations of the proposed HB estimates based on models M_1 and M_2, and the standard errors corresponding to the other survey estimates.*

5.2.1 Log Transformation

Since the values of household estimates are large and positive, we consider the following transformation: let, $y_{it}^* = log(y_{it})$ and $\theta_t^* = log(\theta_t)$. Now, we can rewrite equation (5.2.1) as,

$$y_{it}^* = \theta_t^* + \epsilon_{it}, \qquad t \in S_i \subset \{1, 2, \ldots, T\}.$$

We assume that sampling errors $\epsilon_{it} \sim N(0, \tau_{it}^2)$. Previously, we assumed that $\text{Var}(y_{it|\theta_t}) = \sigma_{it}^2$, where σ_{it}^2's are known. Now, $\tau_{it}^2 = \text{Var}(y_{it}^*|\theta_t^*) = \text{Var}(log(y_{it}|\theta_t)) \approx \dfrac{\sigma_{it}^2}{y_{it}^2}$, using Taylor series expansion. We obtain, the values of τ_{it}^2's using this approximation. Similarly, as in equation (5.2.2), we assume,

$$\theta_t^* = \theta_{t-1}^* + \varepsilon_t, \qquad t = 1, 2, \ldots, T,$$

Table 5.4: *HB estimates and posterior standard deviations based on model M_1 (numbers in 1000s).*

Year	$\hat{\theta}_t$	Posterior sd	Year	$\hat{\theta}_t$	Posterior sd
2002	106909.21	103.48	2007	112110.28	92.93
2003	107002.75	93.73	2008	112877.75	103.76
2004	109300.26	141.15	2009	113443.00	97.42
2005	110688.17	94.65	2010	114823.40	107.55
2006	111775.22	103.65	2011	115433.41	107.08

Table 5.5: *Posterior standard deviations and the standard errors (numbers in 1000s).*

Year	Proposed method (M_1) Posterior sd	CPS/ASEC s.e	HVS s.e	ACS s.e	AHS s.e
2002	103.48	260	185	.	.
2003	93.73	260	182	.	165
2004	141.15	235	179	.	.
2005	94.65	234	204	144	218
2006	103.65	261	194	146	.
2007	92.93	261	187	144	231
2008	103.76	261	181	147	.
2009	97.42	262	174	161	238
2010	107.55	262	173	163	.
2011	107.08	262	171	180	396

where, ε_t's are independently distributed with a truncated normal distribution truncated above 0, with variance σ_ε^2. We assume that σ_ε^2 and θ_o^* are unknown. Now, with this re-parametrization, the proposed hierarchical Bayesian model in Section 5.2 could be rewritten as,

$$\text{Model } M_2: \quad y_{it}^*|\theta_0^*, \theta_t^*, \sigma_\varepsilon^2 \overset{\text{ind}}{\sim} N(\theta_t^*, \tau_{it}^2),\ t \in S_i,\ i = 1, \ldots, m,$$

$$\theta_t^* = \theta_{t-1}^* + \varepsilon_t,\ t = 1, \ldots, T,$$

$$\varepsilon_t|\sigma_\varepsilon^2 \overset{\text{iid}}{\sim} \text{truncated } N(0, \sigma_\varepsilon^2), \tag{5.2.4}$$

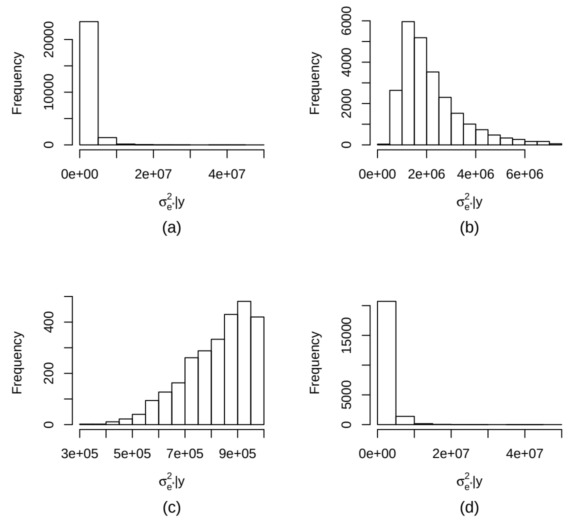


Figure 5.3: *Histograms of the posterior simulations for* $\sigma^2_{e^*}$ *(a) based an all simulated values (b) after dropping upper* 2.5% *observations (c) after dropping observations larger than* 10^6 *(d) after dropping observations smaller than* 10^6*.*

with lower truncation point 0. We assume that, σ^2_ε and θ^*_o are independently distributed with $\sigma^2_\varepsilon \sim$ Uniform$(0, \infty)$ and $\theta^*_o \sim$ Uniform$(-\infty, \infty)$.

The resulting posterior distribution from this model will be proper if the sufficient conditions stated in Theorem 5.2.1 are satisfied. In order to estimate the parameters in this model, we use Gibbs sampling technique. Full conditional posterior distributions could be obtained by simple modifications of the full conditional distributions mentioned in Section 5.2. We run 5 chains and 10,000 iterations for each chain. We discard first 50% observations of each chain and compute our estimates based on the remaining observations. Table 5.7 shows the summary of the posterior inference for σ^2_ε . Histograms based on the posterior simulations for σ^2_ε are shown in Figure 5.4.

Table 5.6: *Details about the posterior simulations of* $\sigma^2_{e^*}$.

Simulated values of $\sigma^2_{e^*}\|y$	Proportion
$< 10^6$	0.10696
$10^6 - 5 \times 10^6$	0.82868
$5 \times 10^6 - 9 \times 10^6$	0.05176
$> 9 \times 10^6$	0.0126

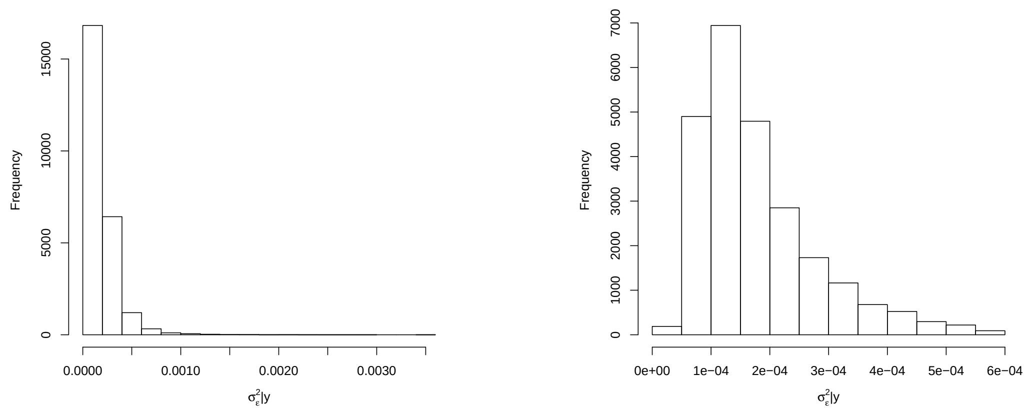
Using the estimates of θ_t^* (say, $\hat{\theta}_t^*$), we can get the estimates of θ_t (say, $\hat{\theta}_t$) by the transformation $\hat{\theta}_t = \mathrm{E}\left[\exp(\hat{\theta}_t^*)|y\right]$. In Table 5.8 we present the estimates of θ_t and the posterior standard deviations corresponding to the estimates. From Figure 5.2(a) we see that the estimates obtained by considering a log transformation almost coincide with the estimates obtained without considering a transformation. This applies to the posterior standard deviations as well (Figure 5.2(b)).

Table 5.7: Summary of the posterior simulation for σ^2_ε (numbers in 1000s)

	Posterior	Posterior	Simulated Quantiles		
Parameter	Mean	sd	2.5%	Median	97.5%
σ^2_ε	0.00019	0.00015	0.00006	0.00015	0.00058

Table 5.8: *HB estimates and posterior standard deviations (numbers in 1000s).*

Year	$\hat{\theta}_t$	Posterior sd	Year	$\hat{\theta}_t$	Posterior sd
2002	107012.90	103.46	2007	112166.94	93.08
2003	107100.13	94.67	2008	112943.28	103.66
2004	109426.22	140.62	2009	113486.02	97.81
2005	110739.88	94.46	2010	114901.45	107.21
2006	111831.94	103.30	2011	115523.07	108.31



(a) (b)

Figure 5.4: *Histograms of the posterior simulations for* σ^2_ε *(a) based an all simulated values (b) after dropping upper* 2.5% *observations.*

5.3 A Hierarchical Bayesian method using auxiliary data

Population size and the number of occupied households have a natural relationship. That motivates us to use total population size as an auxiliary variable in order to improve the survey estimates. We assume that number of households and total population size is linearly related. Before we describe the model we define the following quantities:

$$Z_t = \frac{\sum_{i=1}^{m} r_{it} D_{it}^{-1} y_{it}}{\sum_{i=1}^{m} r_{it} D_{it}^{-1}}, \qquad D_t = \left(\sum_{i=1}^{m} \frac{r_{it}}{D_{it}} \right)^{-1} \quad i = 1, \ldots, m, \ t = 1, \ldots, T, \tag{5.3.1}$$

where, $r_{it} = 1$ if the i^{th} survey produces estimator y_{it} and variance of the estimator in year t, and it is 0 otherwise. The sampling variance corresponding to y_{it} is denoted by D_{it}.

Let θ_t represents the true number of households at time t. Let us consider the following model,

$$\text{Model } M_3 : \qquad Z_t|\theta_t \sim N(\theta_t, D_t),\ \theta_t = x_t^T\beta + \delta_t$$
$$\delta_t|\sigma_\delta^2 \sim N(0, \sigma_\delta^2),\ t = 1, \ldots, T, \tag{5.3.2}$$

where $x_t = (1, P_t)^{'}$, P_t is the total population in the United States at time t. We assign a flat prior for the unknown quantities $\beta\,(2 \times 1)$ and σ_δ^2 in model M_3,

$$\pi(\beta, \sigma_\delta^2) \propto 1,\ \beta \in \mathbf{R^2},\ \sigma_\delta^2 \in \mathbf{R}^+ \tag{5.3.3}$$

The posterior distribution resulting from this model will be proper if $T \geq p + 3$, p being the dimension of β. Here, $p = 2$ and $T = 10$ hence the sufficient condition for propriety of the posterior is satisfied for this data.

The model (5.3.2) can be implemented by applying a Gibbs sampler using the following full conditional distributions. Let $\delta = (\delta_1, \ldots, \delta_T)^T$, $Z = (Z_1, \ldots, Z_T)^T$, $D = diag(D_1, \ldots, D_T)$

(I) $\beta|Z, \delta, \sigma_\delta^2 \sim N_2((X^T D^{-1} X)^{-1} X^T D^{-1}(Z - \delta), (X^T D^{-1} X)^{-1})$

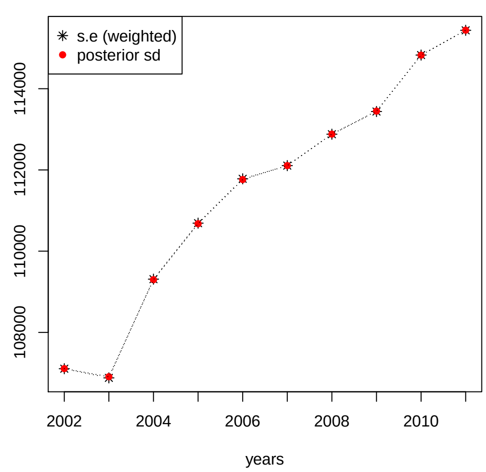
(II) $\delta_t|Z, \beta, \sigma_\delta^2 \sim N\left(\dfrac{D_t^{-1}(Z_t - x_t^T\beta)}{D_t^{-1} + \sigma_\delta^{-2}}, \dfrac{1}{D_t^{-1} + \sigma_\delta^{-2}}\right),\ t = 1, \ldots, T.$

(III) $\dfrac{1}{\sigma_\delta^2}|Z, \beta, \delta \sim Gamma\left(\dfrac{T}{2} - 1, \dfrac{1}{2}\sum_{t=1}^{T}\delta_t^2\right).$

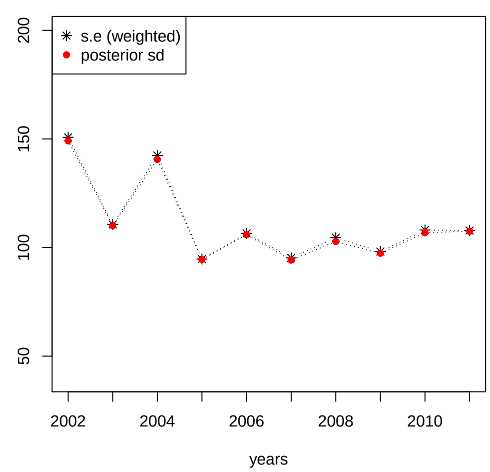
Results obtained by implementing model M_3 are described in Table 5.9 and in Figure 5.5.

Table 5.9: *Estimates obtained by weighted average of the survey estimates and the model estimates obtained by using model M_3 (numbers in 1000s).*

Year (t)	Z_t	$\sqrt{D}_t$	$\hat{\theta}_t$	Posterior *sd*
2002	106909.21	150.74	107213.4	149.13
2003	107002.75	110.62	108212.1	110.08
2004	109300.26	142.40	110012.7	140.55
2005	110688.17	94.68	109145.2	94.51
2006	111775.22	106.50	110012.7	105.92
2007	112110.28	95.24	111876.6	94.19
2008	112877.75	104.55	112852.3	102.78
2009	113443.00	98.14	113878.9	97.38
2010	114823.40	108.07	114819.6	106.81
2011	115433.41	107.82	115740.7	107.50



(a)



(b)

Figure 5.5: *Estimates obtained by weighted average of the survey estimates and the model estimates by using model M_3*

5.4 Some new models accounting for sampling bias

From Figure 5.1 we see that there is a considerable difference in the survey estimates of households ever year. To verify whether the surveys estimate the same quantity, we conduct hypothesis test to test the equality of $\mu_{it} = E(y_{it})$ among the surveys for each t, where y_{it} is the estimate of number of occupied household obtained from the i^{th} survey at the t^{th} year, for $i = 1\ldots, m$ and $t = 1, \ldots, T$. For each year the null hypothesis that the surveys are estimating the same quantity was rejected convincingly. Motivated by this result we introduce a bias term for each survey in the new model described below.

$$\begin{aligned} y_{it} &= \theta_{it} + e_{it}, \\ \theta_{it} &= h_t + \alpha_i + b_{it}, \\ h_t &= \beta_0 + \beta x_t + \eta_t,\ t \in S_i,\ i = 1, \ldots, m, \end{aligned} \tag{5.4.1}$$

where α_i is the bias associated with the i^{th} survey, h_t is the true number of households at the year t. We impose an additive constraint $\sum_{i=1}^{m} \alpha_i = 0$ among the biases in the model. In the model, $e_{it} \sim N(0, D_{it})$, $b_{it} \sim N(0, \sigma_b^2)$, $\eta_t \sim N(0, \sigma_\eta^2)$, independently. The sampling variances D_{it}'s are known but the model variances σ_b^2 and σ_η^2 are unknown. Let, $\mu_i = \beta_0 + \alpha_i$, $i = 1, \ldots, m$. Since, $\sum_{i=1}^{m} \alpha_i = 0$, $\frac{1}{m} \sum_{i=1}^{m} \mu_i = \beta_0$.

We rewrite model (5.4.1) in the following form.

$$\begin{aligned}
\text{Model } M_4: \ y_{it}|\alpha_i, h_t, b_{it} &\sim N(h_t + \alpha_i + b_{it}, D_{it}),\ t \in S_i,\ i = 1, \ldots, m,\\
h_t &= \frac{1}{m}\sum_{i=1}^{m} \mu_i + \beta x_t + \eta_t,\\
\eta_t|\sigma_\eta^2 &\sim N(0, \sigma_\eta^2),\\
b_{it}|\sigma_b^2 &\sim N(0, \sigma_b^2),\ t = 1, \ldots, T,\ i = 1, \ldots, m,\\
\pi(\mu, \beta, \sigma_b^2, \sigma_\eta^2) &\propto 1,\\
\frac{1}{m}\sum_{i=1}^{m} \mu_i &= \beta_0,\ \text{where } \mu_i = \beta_0 + \alpha_i. \qquad (5.4.2)
\end{aligned}$$

where, $\mu = (\mu_1, \ldots, \mu_m)^T$, $\beta \in \mathbf{R}$ and σ_b^2, $\sigma_\eta^2 \in \mathbf{R}^+$.

Theorem 5.4.1 *The posterior distribution resulting from Model M_4 will be proper if $T > 4$ and $m(T-1) > 5$.*

The proof of the theorem is provided in Section 5.6.2. The proof is provided for a balanced case assuming data are available for all m surveys over T time points. This proof can be modified for an unbalanced case as well.

The joint pdf of $y, \mu, \beta, \eta, b, \sigma_b^2, \sigma_\eta^2$ from model M_4 is given by,

$$\begin{aligned}
\pi(y, \mu, \beta, \eta, b, \sigma_b^2, \sigma_\eta^2) &= C \times \exp\left\{-\frac{1}{2}(y - Xw - Z_1\eta - b)^T D^{-1}(y - Xw - Z_1\eta - b)\right\}\\
&\times \frac{1}{(\sigma_\eta^2)^{\frac{T}{2}}} \times \exp\left\{-\frac{1}{2}\frac{\eta^T\eta}{2\sigma_\eta^2}\right\} \times \frac{1}{(\sigma_b^2)^{\frac{n}{2}}} \times \exp\left\{-\frac{1}{2} \times \frac{b^T b}{2\sigma_b^2}\right\}, \qquad (5.4.3)
\end{aligned}$$

where, $Z_1 = \bigoplus_{t=1}^{T} 1_{n_t}$, $n_t = \sum_{i=1}^{m} = \delta_{it}$, where $\delta_{it} = 1$ if data from i^{th} survey is available at time t and $\delta_{it} = 0$ otherwise; $n = \sum_{t=1}^{T} n_t$. Here, $w = (\mu_1, \ldots, \mu_m, \beta)^T$ and,

$$X = \begin{pmatrix} I_{n_1} & x_1 1_{n_1} \\ I_{n_2} & x_2 1_{n_2} \\ \vdots & \\ I_{n_T} & x_T 1_{n_T} \end{pmatrix}.$$

We denote the identity matrix of order $n_t \times n_t$ by I_{n_t}.

In (5.4.3), $y = (y_{11}, \ldots, y_{n_1 1}, y_{12}, \ldots, y_{n_2 2}, \ldots, y_{n_T T})^T$, $\eta = (\eta_1, \ldots, \eta_T)^T$

$D = diag(D_{11}, \ldots, D_{n_1 1}, D_{12}, \ldots, D_{n_2 2}, \ldots, D_{n_T T})$, $b = (b_{11}, \ldots, b_{n_1 1}, b_{12}, \ldots, b_{n_2 2}, \ldots, b_{n_T T})^T$.

Let us define, $f = y - Z_1\eta - b$, $g = y - Xw - b$ and $h = y - Xw - Z_1\eta$. The full conditional distributions obtained from (5.4.3) are given below,

(I) $w|y, \eta, b, \sigma_\eta^2, \sigma_b^2 \sim N((X^T D^{-1} X)^{-1} X^T D^{-1} f, (X^T D^{-1} X)^{-1})$,

(II) $\eta|y, w, b, \sigma_\eta^2, \sigma_b^2 \sim N\left((\sigma_\eta^{-2} I_T + Z_1^T D^{-1} Z_1)^{-1} Z_1^T D^{-1} g, (\sigma_\eta^{-2} I_T + Z_1^T D^{-1} Z_1)^{-1}\right)$,

(III) $b|y, w, \eta, \sigma_\eta^2, \sigma_b^2 \sim N\left((\sigma_b^{-2} I_n + D^{-1})^{-1} D^{-1} h, (\sigma_b^{-2} I_n + D^{-1})^{-1}\right)$,
where $n = (\sum_{t=1}^{T} n_t)$,

(IV) $\frac{1}{\sigma_\eta^2}|y, w, b, \eta, \sigma_b^2 \sim Gamma\left(\frac{T}{2} - 1, \frac{\eta^T \eta}{2}\right)$,

(V) $\frac{1}{\sigma_b^2}|y, w, b, \eta, \sigma_\eta^2 \sim Gamma\left(\frac{n}{2} - 1, \frac{b^T b}{2}\right)$.

We implement a Gibbs sampler using these conditional distributions. Estimates of h_t obtained from model M_4 and the standard deviation associated with the estimates from 2002 to 2011 are given in the first and third column of Table 5.10. From the fourth column of Table 5.10, we see that the posterior standard deviations associated with the estimates are on average larger than the sampling standard errors. This may be caused by using too many parameters in the model.

We consider another model which is almost same as Model M_4 but involves less number of parameters.

$$\begin{aligned}
\text{Model } M_5: \quad & y_{it}|h_t, \alpha_i \sim N(h_t + \alpha_i, D_{it}), \\
& h_t = \frac{1}{m}\sum_{i=1}^{m} \mu_i + \beta x_t + \eta_t, \\
& \eta_t|\sigma_\eta^2 \sim N(0, \sigma_\eta^2),\ t \in S_i,\ i = 1, \dots, m, \\
& \pi(\mu, \beta, \sigma_\eta^2) \propto 1, \\
& \frac{1}{m}\sum_{i=1}^{m} \mu_i = \beta_0,\ \text{where } \mu_i = \beta_0 + \alpha_i. \qquad (5.4.4)
\end{aligned}$$

Notation used in model M_5 has the same meaning as that defined before. The required full conditional distributions for model M_5 can be obtained with a little modification to the full conditional distributions corresponding to model M_4. We implement model M_5 and compute the estimates of number of households and the posterior standard deviations associated with the estimates given in Table 5.10. From Table 5.10 we see that while the posterior standard deviations are considerably small for model M_5, the point estimates obtained using model M_5 are similar to the estimates obtained from model M_4 to a large extent. In Table 5.11 we estimate the bias for the surveys, where α_1 represents bias for CPS/ASEC, α_2 represents bias for HVS, α_3 is the bias for ACS and α_4 is the bias for AHS. In Table 5.12 we show the bias adjusted survey estimates.

5.5 Discussion

In this chapter we study various estimation methods which combine estimates from different surveys. We have shown that considerable gain in terms of precision can be achieved using some of these methods. In the last section, we considered bias in the model and performed an exploratory analysis. Number of households estimated by different surveys differ considerably, which may create ambiguity among the researchers and impact decisions of the

Table 5.10: *HB estimates based on model M_4 and M_5 (numbers in 1000s)*

	Estimate		Posterior SD	
Year	M_4	M_5	M_4	M_5
2002	107156.26	107136.87	251.31	151.33
2003	107840.94	107786.69	211.06	112.93
2004	109140.83	109129.91	241.40	142.17
2005	110703.93	110869.47	183.58	94.77
2006	111730.94	111796.26	206.31	109.90
2007	112477.30	112495.83	172.28	95.77
2008	113046.61	113024.20	197.38	107.80
2009	113923.89	113934.36	180.77	97.56
2010	115131.95	115032.04	204.64	110.53
2011	115974.39	115788.58	188.92	108.81

Table 5.11: *HB estimates of bias for each survey using model M_5*

i	$\hat{\alpha}_i$	Posterior sd
1	4233.42	7441.85
2	-2156.89	7441.54
3	-182.91	7446.68
4	-1893.62	7440.90

government organizations. Our proposed methods successfully combine the survey estimates which could be helpful to the researchers.

Table 5.12: *Bias corrected estimates of households from* 2002−2011 *for three different surveys based on model* M_5 *(numbers in 1000s).*

Survey	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
CPS/ASEC	107044.58	107766.58	109109.58	110150.58	111777.58	112549.58	112947.58	113304.58	115693.58	116850.58
HVS	107150.89	107792.89	109127.89	110823.89	111892.89	112329.89	112631.89	114451.89	115055.89	115689.89
ACS	107549.91	108602.91	110084.91	111273.91	111799.91	112560.91	113283.91	113798.91	114749.91	115174.91
AHS	.	107735.62	.	110764.62	.	112585.62	.	113699.62	.	116800.62

5.6 Prof of the theorems

5.6.1 Proof of Theorem 5.2.1

From model M_1, the joint posterior distribution of θ_0, θ_1, θ_2,…,θ_T is given by,

$$\begin{aligned}
&\pi^*(\theta_0, \theta_1, \ldots, \theta_T, \sigma^2_{e^*}|y) \\
&\qquad = C \times \exp\left\{-\frac{1}{2}\sum_{i=1}^{m}\sum_{t\in S_i} r_{it}(y_{it}-\theta_t)^2\right\} \\
&\qquad\qquad \times \frac{1}{(\sigma^2_{e^*})^{\frac{T}{2}}} \times \exp\left\{-\frac{1}{2\sigma^2_{e^*}}\sum_{t=1}^{T}(\theta_t-\theta_{t-1})^2\right\}\prod_{t=1}^{T}\mathrm{I}(\theta_t>\theta_{t-1}), \qquad (5.6.1)
\end{aligned}$$

where C is a generic positive constant, $r_{it}=0$ if σ^2_{it} is missing and $r_{it}=\dfrac{1}{\sigma^2_{it}}$ if σ^2_{it} is available, $t\in S_i$, $i=1,\ldots,m$.

Now, from (5.6.1),

$$\begin{aligned}
&\int \pi^*(\theta_0, \theta_1, \ldots, \theta_T, \sigma^2_{e^*}|y)\mathrm{d}\theta_0\prod_{t=1}^{T}\mathrm{d}\theta_t\mathrm{d}\sigma^2_{e^*} \\
&\qquad \le \int \pi(\theta_0, \theta_1, \ldots, \theta_T, \sigma^2_{e^*}|y)\mathrm{d}\theta_0\prod_{t=1}^{T}\mathrm{d}\theta_t\mathrm{d}\sigma^2_{e^*}, \qquad (5.6.2)
\end{aligned}$$

where,

$$\pi(\theta_0, \theta_1, \ldots, \theta_T, \sigma^2_{e^*}|y)$$
$$= C \times \exp\left\{-\frac{1}{2}\sum_{i=1}^{m}\sum_{t\in S_i} r_{it}(y_{it}-\theta_t)^2\right\}$$
$$\times \frac{1}{(\sigma^2_{e^*})^{\frac{T}{2}}} \times \exp\left\{-\frac{1}{2\sigma^2_{e^*}}\sum_{t=1}^{T}(\theta_t-\theta_{t-1})^2\right\}. \quad (5.6.3)$$

From the right hand side of (5.6.3),

$$\sum_{i=1}^{m}\sum_{t\in S_i} r_{it}(y_{it}-\theta_t)^2$$
$$= (\theta^*-\mu)^T\Sigma^{-1}(\theta^*-\mu) + \sum_{i=1}^{m}\sum_{t\in S_i} r_{it}y_{it}^2 - \sum_{t=1}^{T} r_{.t}\mu_t^2, \quad (5.6.4)$$

where, $\theta^* = (\theta_1, \ldots, \theta_T)^T$, $\mu = (\mu_1, \ldots, \mu_T)^T$, $\mu_t = (\sum\limits_{i=1}^{m} r_{it})^{-1}(\sum\limits_{i=1}^{m} r_{it}y_{it})$, $t = 1, \ldots, T$ and $\Sigma = diag(r_{.1}, r_{.2}, \ldots, r_{.T})$, where, $r_{.t} = \sum\limits_{i=1}^{m} r_{it}$, $t = 1, \ldots, T$. Note that, we have assumed $n_t > 0$, $n_t = \sum\limits_{i=1}^{m} \delta_{it}$, where $\delta_{it} = 1$ if data from i^{th} survey is available at time t and $\delta_{it} = 0$ otherwise. This implies, $r_{.t} > 0$, for all t. Also,

$$\sum_{t=1}^{T}(\theta_t-\theta_{t-1})^2 = (\theta_0-\theta_1)^2 + \sum_{t=2}^{T}(\theta_t-\theta_{t-1})^2$$
$$= (\theta_0-\theta_1)^2 + \theta^{*T}(B^TB)\theta^*, \quad (5.6.5)$$

where, B is a $(T-1\times T)$ matrix and rank$(B^TB) = T-1$.

Hence,

$$\begin{aligned}
&\pi(\theta_0, \theta_1, \ldots, \theta_T, \sigma^2_{e^*}|y) \\
&\quad \leq C \times \exp\left\{-\frac{1}{2}(\theta^* - \mu)^T \Sigma^{-1}(\theta^* - \mu)\right\} \\
&\qquad \times \frac{1}{(\sigma^2_{e^*})^{\frac{T}{2}}} \exp\left\{-\frac{1}{2}\frac{\left[(\theta_0 - \theta_1)^2 + \theta^{*T}(B^T B)\theta^*\right]}{\sigma^2_{e^*}}\right\}
\end{aligned} \tag{5.6.6}$$

$$\begin{aligned}
&\int \pi(\theta_0, \theta_1, \ldots, \theta_T, \sigma^2_{e^*}|y)\mathrm{d}\theta_0 \\
&\quad \leq C \times \exp\left\{-\frac{1}{2}(\theta^* - \mu)^T \Sigma^{-1}(\theta^* - \mu)\right\} \\
&\qquad \times \frac{1}{(\sigma^2_{e^*})^{\frac{(T-1)}{2}}} \times \exp\left\{-\frac{1}{2\sigma^2_{e^*}}\theta^{*T}(B^T B)\theta^*\right\}.
\end{aligned} \tag{5.6.7}$$

Let $\xi_1, \ldots, \xi_T$ be the T positive eigenvalues of Σ such that $\xi_1 < \xi_2 < \cdots < \xi_T$. Let R be an orthogonal matrix such that,

$$\begin{aligned}
R^T \Sigma R = diag(\xi_1, \ldots, \xi_T) < 2\xi_T I \\
\Rightarrow \Sigma < 2\xi_T I \\
\Rightarrow -(\theta^* - \mu)\Sigma^{-1}(\theta^* - \mu) < -\frac{(\theta^* - \mu)^T(\theta^* - \mu)}{2\xi_T},
\end{aligned} \tag{5.6.8}$$

Hence,

$$\begin{aligned}
&\int \pi(\theta_0, \theta_1, \ldots, \theta_T, \sigma^2_{e^*}|y)\mathrm{d}\theta_0 \\
&\quad \leq C \times \exp\left\{-\frac{1}{2}\frac{(\theta^* - \mu)^T(\theta^* - \mu)}{2\xi_T}\right\} \\
&\qquad \times \frac{1}{(\sigma^2_{e^*})^{\frac{(T-1)}{2}}} \times \exp\left\{-\frac{1}{2\sigma^2_{e^*}}\theta^{*T}(B^T B)\theta^*\right\},
\end{aligned} \tag{5.6.9}$$

Let P be an orthogonal matrix such that, $P^T(B^T B)P = diag(\lambda_1, \ldots, \lambda_{T-1}, 0)$, where λ_i's are positive eigenvalues of $B^T B$. Consider the transformation, $\theta^* = P\alpha$, where $\alpha =$

$(\alpha_1, \ldots, \alpha_T)^T$. Also, define $\nu = P\mu$. Now, from (5.6.9),

$$\begin{aligned}
&\int \tilde{\pi}(\theta_0, \alpha_1, \ldots, \alpha_T, \sigma^2_{e^*}|y)\mathrm{d}\theta_0 \\
&\quad \leq C \times \exp\left\{-\frac{1}{2}\frac{(\alpha-\nu)^T(\alpha-\nu)}{2\xi_T}\right\} \\
&\qquad \times \frac{1}{(\sigma^2_{e^*})^{\frac{(T-1)}{2}}} \times \exp\left\{-\frac{1}{2\sigma^2_{e^*}}\theta^{*T}(B^TB)\theta^*\right\} \\
&\quad = C \times \exp\left\{-\frac{1}{2}\sum_{i=1}^{T}\frac{(\alpha_i-\nu_i)^2}{2\xi_T}\right\} \\
&\qquad \times \frac{1}{(\sigma^2_{e^*})^{\frac{(T-1)}{2}}} \times \exp\left\{-\frac{1}{2\sigma^2_{e^*}}\sum_{i=1}^{T-1}\lambda_i\alpha_i^2\right\} \\
&\quad = C \times \exp\left\{-\frac{1}{2}\sum_{i=1}^{T-1}\alpha_i^2\left(\frac{1}{2\xi_T}+\frac{\lambda_i}{\sigma^2_{e^*}}\right) - 2\alpha_i\frac{\nu_i}{2\xi_T}\right\} \\
&\qquad \times \frac{1}{(\sigma^2_{e^*})^{\frac{(T-1)}{2}}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{T-1}\frac{\nu_i^2}{2\xi_T}\right\} \times \exp\left\{-\frac{1}{2}\frac{(\alpha_T-\nu_T)^2}{2\xi_T}\right\} \\
&\quad = C \times \exp\left\{-\frac{1}{2}\sum_{i=1}^{T-1}\left(\frac{1}{2\xi_T}+\frac{\lambda_i}{\sigma^2_{e^*}}\right)\left(\alpha_i - \frac{\nu_i}{2\xi_T}\left(\frac{1}{2\xi_T}+\frac{\lambda_i}{\sigma^2_{e^*}}\right)^{-1}\right)^2\right\} \\
&\qquad \times \frac{1}{(\sigma^2_{e^*})^{\frac{(T-1)}{2}}} \times \exp\left\{\frac{1}{2}\sum_{i=1}^{T-1}\frac{\nu_i^2}{(2\xi_T)^2}\left(\frac{1}{2\xi_T}+\frac{\lambda_i}{\sigma^2_{e^*}}\right)^{-1} - \frac{1}{2}\sum_{i=1}^{T-1}\frac{\nu_i^2}{2\xi_T}\right\} \\
&\qquad \times \exp\left\{-\frac{1}{2}\frac{(\alpha_T-\nu_T)^2}{2\xi_T}\right\} \qquad (5.6.10)
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\int \tilde{\pi}(\theta_0, \alpha_1, \ldots, \alpha_T, \sigma^2_{e^*}|y)\,\mathrm{d}\theta_0 \prod_{i=1}^{T}\mathrm{d}\alpha_i \\
&\quad \leq C \times \frac{1}{(\sigma^2_{e^*})^{\frac{(T-1)}{2}}}\left(\frac{1}{2\xi_T}+\frac{\lambda_{\min}}{\sigma^2_{e^*}}\right)^{-\frac{(T-1)}{2}} \times \exp\left\{-\frac{1}{2}\sum_{i=1}^{T-1}\frac{\nu_i^2\lambda_i}{(\sigma^2_{e^*}+2\xi_T\lambda_i)}\right\} \\
&\quad \leq C \times \frac{1}{(\sigma^2_{e^*}+2\xi_T\lambda_{\min})^{\frac{(T-1)}{2}}}, \qquad (5.6.11)
\end{aligned}$$

where, $\lambda_{\min} = \min(\lambda_1, \ldots, \lambda_{T-1})$. Finally,

$$\begin{aligned} &\int \tilde{\pi}(\theta_0, \alpha_1, \ldots, \alpha_T, \sigma^2_{e^*}|y)\, \mathrm{d}\theta_0 \prod_{i=1}^{T} \mathrm{d}\alpha_i \mathrm{d}\sigma^2_{e^*} \\ &\qquad \le C \int \frac{1}{(\sigma^2_{e^*} + 2\xi_T \lambda_{\min})^{\frac{(T-1)}{2}}} \mathrm{d}\sigma^2_{e^*} \\ &\qquad < \infty, \end{aligned} \tag{5.6.12}$$

since $\dfrac{T-1}{2} > 1$, i.e., $T > 3$ provided by the sufficient condition. Hence the proof. □

5.6.2 Proof of Theorem 5.4.1

The proof is provided for a balanced case assuming data are available for all m surveys over T time points. From (5.4.1) and the formulation of model M_4,

$$y_{it} = \mu_i + \beta x_t + b_{it} + \eta_t + e_{it}, \tag{5.6.13}$$

where, $b_{it} \overset{\text{iid}}{\sim} N(0, \sigma_b^2)$, $\eta_t \overset{\text{iid}}{\sim} N(0, \sigma_\eta^2)$ and $e_{it} \overset{\text{ind}}{\sim} N(0, D_{it})$, D_{it}'s are known, $i = 1\ldots, m$, $t = 1, \ldots, T$.

Let, $y_t = (y_{1t}, \ldots, y_{mt})^T$, $b_t = (b_{1t}, \ldots, b_{mt})^T$, $\mu = (\mu_1, \ldots, \mu_m)^T$, $e_t = (e_{1t}, \ldots, e_{mt})^T$. Let, $1_m = (1, \ldots, 1)^T$ and let $J_m = 1_m 1_m^T$ be an $(m \times m)$ matrix.

Then,

$$\begin{aligned} & y_t = \mu + (x_t 1_m)\beta + \eta_t 1_m + b_t + e_t, \\ \Rightarrow\quad & y = X\beta^* + \epsilon, \end{aligned} \tag{5.6.14}$$

where, $\beta^* = (\mu_1, \ldots, \mu_m, \beta)^T$, $\epsilon = R\eta + b + e$, $\eta = (\eta_1, \ldots, \eta_T)^T$, and $R = I_T \otimes 1_m$, I_T is $(T \times T)$ identity matrix. From (5.6.14),

$$Var(\epsilon) = \Sigma = D + \sigma_b^2 I_{mT} + \sigma_\eta^2 I_T \otimes J_m, \tag{5.6.15}$$

where $D = diag(D_{11}, \dots, D_{mT})$ and I_{mT} is an $(mT \times mT)$ diagonal matrix. Based on this re-parametrization, we rewrite model M_4 as,

$$\begin{aligned} y|\sigma_b^2, \sigma_\eta^2, \beta^* &\sim N(X\beta^*, \Sigma), \\ \pi(\sigma_b^2, \sigma_\eta^2, \beta^*) &= 1. \end{aligned} \tag{5.6.16}$$

Hence, the joint posterior distribution based on (5.6.16) is given by,

$$\pi(\sigma_\eta^2, \sigma_b^2, \beta^*|y) = C \times \frac{1}{|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(y - X\beta^*)^T \Sigma^{-1} (y - X\beta^*)\right\}, \tag{5.6.17}$$

where C is a positive generic constant.

$$\begin{aligned} \int \pi(\sigma_\eta^2, \sigma_b^2, \beta^*|y)\, \mathrm{d}\beta^* &= C \times \frac{|X^T\Sigma^{-1}X|^{-\frac{1}{2}}}{|\Sigma|^{\frac{1}{2}}} \\ &\quad \times \exp\left\{-\frac{1}{2} y^T (\Sigma^{-1} - \Sigma^{-1}X(X^T\Sigma^{-1}X)X^T\Sigma^{-1})^T y\right\} \qquad (5.6.18) \\ \Rightarrow \int \pi(\sigma_\eta^2, \sigma_b^2, \beta^*|y)\, \mathrm{d}\beta^* &\le C \times \frac{|X^T\Sigma^{-1}X|^{-\frac{1}{2}}}{|\Sigma|^{\frac{1}{2}}}. \qquad (5.6.19) \end{aligned}$$

Using the matrix version of Schwartz inequality (Harville, 2008) we have,

$$\begin{aligned} |X^TX|^2 &= |X^T\Sigma^{-\frac{1}{2}}\Sigma^{\frac{1}{2}}X|^2 \\ &\le |X^T\Sigma^{-1}X||X^T\Sigma X| \\ \text{i.e.,} \quad |X^T\Sigma^{-1}X|^{-\frac{1}{2}} &\le \frac{|X^T\Sigma X|^{\frac{1}{2}}}{|X^TX|}. \qquad (5.6.20) \end{aligned}$$

From (5.6.19) and (5.6.20),

$$\int \pi(\sigma_\eta^2, \sigma_b^2, \beta^*|y)\mathrm{d}\,\beta^* \le C \times \frac{|X^T\Sigma X|^{\frac{1}{2}}}{|X^TX||\Sigma|^{\frac{1}{2}}}. \tag{5.6.21}$$

Let, $\omega_1 = \min\limits_{\substack{1\le i\le m\\ 1\le t\le T}} (D_{it})$ and $\omega_2 = \max\limits_{\substack{1\le i\le m\\ 1\le t\le T}} (D_{it})$. From (5.6.15),

$$\begin{aligned}\Sigma &= D + \sigma_b^2 I_{mT} + \sigma_\eta^2 I_T \otimes J_m,\\ &\ge \omega_1 I_{mT} + \sigma_b^2 I_{mT} + \sigma_\eta^2 I_T \otimes J_m,\\ &= I_T \otimes \left[\omega_1 I_m + \sigma_b^2 I_m + \sigma_\eta^2 J_m\right],\end{aligned} \tag{5.6.22}$$

where I_m is an $(m \times m)$ diagonal matrix. Therefore,

$$|\Sigma|^{-\frac{1}{2}} \le \left[(\omega_1 + \sigma_b^2 + m\sigma_\eta^2)(\omega_1 + \sigma_b^2)^{(m-1)}\right]^{-\frac{T}{2}} \tag{5.6.23}$$

Also,

$$\begin{aligned}X^T \Sigma X &\le \sum_{t=1}^{T} X_t^T \left[\omega_2 I_{mT} + \sigma_b^2 I_{mT} + \sigma_\eta^2 I_T \otimes J_m\right] X_t\\ &= (\omega_2 + \sigma_b^2)\Psi + \sigma_\eta^2\ \Gamma,\end{aligned} \tag{5.6.24}$$

where $\Psi = \sum\limits_{t=1}^{T} X_t^T X_t$, $\Gamma = \sum\limits_{t=1}^{T} X_t^T J_m X_t$ and $X_t = \begin{pmatrix} I_m & x_t 1_m \end{pmatrix}$. Here, rank$(\Gamma) = 2$. The matrix Ψ $(m+1 \times m+1)$ is of full rank, i.e,, rank$(\Psi) = m+1$. Hence, rank$(\Psi^{-\frac{1}{2}}\Gamma\Psi^{\frac{1}{2}}) = 2$. Let, L be an orthogonal matrix such that, $L(\Psi^{-\frac{1}{2}}\Gamma\Psi^{\frac{1}{2}})L^T = diag(\lambda_1, \lambda_2)$, where λ_1 and λ_2 are two positive eigen values of $\Psi^{-\frac{1}{2}}\Gamma\Psi^{\frac{1}{2}}$.

$$\begin{aligned}|X^T \Sigma X| &\le |\Psi||(\omega_2 + \sigma_b^2) I_m + \sigma_\eta^2(\Psi^{-\frac{1}{2}}\Gamma\Psi^{\frac{1}{2}})|\\ &= |\Psi|||(\omega_2 + \sigma_b^2) L\, L^T + \sigma_\eta^2 L(\Psi^{-\frac{1}{2}}\Gamma\Psi^{\frac{1}{2}})L^T|\\ &= |\Psi|||(\omega_2 + \sigma_b^2) L\, L^T + \sigma_\eta^2 diag(\lambda_1, \lambda_2)|\\ &\le |\Psi|(\omega_2 + \sigma_b^2 + \lambda^* \sigma_\eta^2)^2 (\omega_2 + \sigma_b^2)^{m-1},\end{aligned} \tag{5.6.25}$$

where, $\lambda^* = \max(\lambda_1, \lambda_2)$. Now, from (5.6.21), (5.6.23) and (5.6.25),

$$\begin{aligned}
\int \pi(\sigma_\eta^2, \sigma_b^2, \beta^*|y)\mathrm{d}\,\beta^* &\leq C \times \frac{|X^T \Sigma X|^{\frac{1}{2}}}{|\Sigma|^{\frac{1}{2}}} \\
&\leq C \times \frac{(\omega_2 + \sigma_b^2 + \lambda_* \sigma_\eta^2)(\omega_2 + \sigma_b^2)^{\frac{(m-1)}{2}}}{(\omega_1 + \sigma_b^2 + m\sigma_\eta^2)^{\frac{T}{2}}(\omega_1 + \sigma_b^2)^{\frac{(m-1)T}{2}}} \\
&\leq C \times \frac{(\omega^* + \omega^* \sigma_b^2 + \omega^* \lambda_* \sigma_\eta^2)(\omega^* + \omega^* \sigma_b^2)^{\frac{(m-1)}{2}}}{(\omega^{**} + \omega^{**}\sigma_b^2 + \omega^{**}\sigma_\eta^2)^{\frac{T}{2}}(\omega^{**} + \omega^{**}\sigma_b^2)^{\frac{(m-1)T}{2}}}, \qquad (5.6.26)
\end{aligned}$$

where, $\omega^* = \max(\omega_2, \lambda^*, 1)$ and $\omega^{**} = \min(\omega_1, m, 1)$. Hence,

$$\int \pi(\sigma_\eta^2, \sigma_b^2, \beta^*|y)\mathrm{d}\,\beta^* \leq C \times \frac{(1 + \sigma_b^2 + \sigma_\eta^2)(1 + \sigma_b^2)^{\frac{(m-1)}{2}}}{(1 + \sigma_b^2 + \sigma_\eta^2)^{\frac{T}{2}}(1 + \sigma_b^2)^{\frac{(m-1)T}{2}}}, \qquad (5.6.27)$$

where C is generic constant. Therefore,

$$\int \pi(\sigma_\eta^2, \sigma_b^2, \beta^*|y)\,\mathrm{d}\beta^* \leq C \times ((1 + \sigma_b^2 + \sigma_\eta^2)^{-\frac{(T-2)}{2}}(1 + \sigma_b^2)^{-\frac{(m-1)(T-1)}{2}}. \qquad (5.6.28)$$

Now,

$$\begin{aligned}
&\int \pi(\sigma_\eta^2, \sigma_b^2, \beta^*|y)\,\mathrm{d}\beta^*\,\mathrm{d}\sigma_\eta^2\,\mathrm{d}\sigma_b^2 \\
&\quad \leq C \times \int (1 + \sigma_b^2)^{-\frac{(m-1)(T-1)}{2}} \left[\int_0^\infty (1 + \sigma_b^2 + \sigma_\eta^2)^{-\frac{(T-2)}{2}} \mathrm{d}\sigma_\eta^2\right] \mathrm{d}\sigma_b^2 \\
&\quad = C \times \int (1 + \sigma_b^2)^{-\frac{(m-1)(T-1)}{2}} \times \left[(1 + \sigma_b^2)^{\frac{4-T}{2}}\right] \mathrm{d}\sigma_b^2, \qquad (5.6.29)
\end{aligned}$$

provided the sufficient condition $T > 4$ holds. Therefore,

$$\begin{aligned}
&\int \pi(\sigma_\eta^2, \sigma_b^2, \beta^*|y)\,\mathrm{d}\beta^*\,\mathrm{d}\sigma_\eta^2\,\mathrm{d}\sigma_b^2 \\
&\quad \leq C \times \int (1 + \sigma_b^2)^{\frac{(3+m-mT)}{2}}\,\mathrm{d}\sigma_b^2 \\
&\quad \leq C \times \left[(1 + \sigma_b^2)\right]^{\frac{3+m-mT}{2}+1}\Big|_0^\infty < \infty, \qquad (5.6.30)
\end{aligned}$$

since, $\dfrac{3+m-mT}{2}+1<0$ which is ensured by the sufficient condition $m(T-1)>5$. Hence the posterior distribution resulting from model M_4 is proper. This proof can be modified for the unbalanced case. $\square$

Chapter 6

Conclusion

In this dissertation, we propose robust Bayesian small area estimation techniques for both unit-level and area-level data sets. In Chapter 2, we performed an extensive simulation study and showed that the proposed method for unit-level data performed considerably well in presence of outliers. Also, when there were no outliers in the data, the proposed method performed almost as good as the method proposed by Datta and Ghosh (1991). In Chapter 3 we proposed a two-component normal mixture distribution for the random area specific effects in a basic area-level model. The purpose of proposing such a model was to provide a robust alternative to the Fay-Herriot model. In Chapter 4, we proposed double exponential distribution for the random small area effects for area-level data and discussed how this new model may perform well when area specific random effects are absent for some areas. In Chapter 5, we proposed various HB methods to combine survey estimates obtained from different sources. We discussed various noninformative HB models which combine information from multiple sources. We have shown that these models can be used to improve the direct survey based estimates. In future, we would like to develop efficient model selection techniques which select the appropriate model for a given data set. These techniques will be helpful for survey researchers particularly when the scope of external evaluations are restrictive.

Bibliography

[1] Andrews, D. F. and Mallows, C. L. (1974). Scale Mixtures of Normal Distri-butions, *Journal of the Royal Statistical Society, Ser. B*, **36**, 99102.

[2] Albert, J. H. (1998). Computational methods using a Bayesian hierarchical generalised linear model, *J. Amer. Statist. Assoc.* **83** 1037-1044.

[3] Battese, G. E., Harter, R. M. and Fuller, W. A. (1988). An error component model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc.* **83** 28-36.

[4] Berger, J. (2006). The Case for Objective Bayesian Analysis. *Bayesian Analysis* **3** 385-402.

[5] Bell, W. R., Huang, T. H., (2006). Using t-distribution to deal with outliers in small area estimation. *Proceedings of Statistics Canada Synposium 2006 Methodological issues in measuring population health.*

[6] Bondell, H. D., Krishna, A. and Ghosh, S. K. (2010). Joint variable selection for fixed and random effects in linear mixed-effects models. *Biometrics*, **66,** 1069–1077.

[7] Chambers, R., Chandra, H. and Salvati, N. and Tzavidis, N. (2014). Outlier robust small area estimation. *Journal of the Royal Statistical Society B,* **76** 47-69.

[8] Citro, C. and Kalton, G., eds. (2000). Small-Area Income and Poverty Esti- mates: Priorities for 2000 and Beyond. Panel on Estimates of Poverty for Small Geographic Area, Committee on National Statistics, Washington, DC: National Academy Press.

[9] Cochran, W. G. (1977). *Sampling Techniques*, 3rd. ed. Wiley, New York.

[10] Datta, G.S. (2009). Model-based approach to small area estimation", in: Handbook of Statis- tics: Sample Surveys: Inference and Analysis, Volume 29B, Eds.: D. Pfefermann and C.R. Rao, The Netherlands: North-Holland. 251-288.

[11] Datta, G. and Ghosh, M. (1991). Bayesian prediction in linear models: Applications to small area estimation. *Ann. Statist.*, **19,** 1748-1770.

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

[13] Datta, G. S., Lahiri, P., (1995). Robust hierarchical Bayesian estimation of small area characteristics in presence of covariates and outliers. *Journal of Multivariate Anslysis*, **54,** 2, 310–328.

[14] Datta, G. and Ghosh, M. (2012). Small area Shrinkage Estimation. *Statistical Science*, **27,** 95-114.

[15] Datta, G. S. and Mandal, A. (2014). Small area estimation with uncertain random effect, *submitted*.

[16] Datta, G. S., Hall, P. and Mandal, A. (2011), Model selection by testing for the presence of Small-area effects in area-level data, *Journal of American Statistical Association*, 106, 362-374.

[17] Demster, A. P., Rubin, D. B. and Tsutakawa, R. K. (1981). Estimation in covariance components models. *J. Amer. Statist. Assoc.* **76** 341-353.

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

[19] Fan, Y. & Li, R. (2012). Variable selection in linear mixed effect models. *The Annals of Statistics*, **40,** 4, 2043–2068.

[20] Febrizi, E., and Trivisano, C. (2010). Robust linear mixed models for Small Area Estimation. *Journal of Statistical Planning and Inference*, **140,** 433-443.

[21] Gershunskaya, J. (2010). Robust Small Area Estimation Using a Mixture Model. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

[22] Gelfand, A. and Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85** 398-409.

[23] Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). Bayesian Data Analysis, second edition. Chapman and Hall, London.

[24] Geman, s. and Geman, D. (1984). Stochastic realization, Gibbs distributions and Bayesian restoration of images. *IEEE Trans. Pattern Analysis and Machine Intelligence* **6** 721-741.

[25] Ghosh, M. and Lahiri, P. (1987). Robust empirical Bayes estimation of means from stratified samples. *J. Amer. Statist. Assoc.* **82** 1153-1162.

[26] Ghosh, M. and Lahiri, P. (1988). Bayes and empirical Bayes analysis in multistage sampling. In *Statistical Decision Theory and Related Topics IV* (S. S. Gupta and J. O. Berger, eds.)**1** 195-212. Springer, New York.

[27] Ghosh, M. and Lahiri, P. (1989). A hierarchical Bayes approach to small area estimation with auxiliary information. In *Proceedings of the Joint Indo-U. S. Workshop on Bayesian Inference in Statistics and Econometrics.*

[28] Ghosh, M. and Meeden, G. (1986). Empirical Bayes estimation in finite population sampling. *J. Amer. Statist. Assoc.* **81** 1058-1062.

[29] Ghosh, M. and Rao, J.N.K. (1994). Small are estimation: an appraisal. *Statistical Science*, **9,** 55-93.

[30] Gonzalez, J.F, Placek, P.J., and Scott, C. (1966). Synthetic estimation of followback surveys at the National Center for Health Statistics. *Indirect Estimators in U.S. Federal Programs*, Springer-Verlag, New York, 16-27.

[31] Gonzalez, M. E. (1973). Use and Evaluation of Synthetic Estimates. *Proceedings of the Social Statistics Section*, American Statistical Association, 33-36.

[32] Harville, D. A. (1985). Decomposition of prediction error. *J. Amer. Statist. Assoc.* **80** 132-138.

[33] Harville, D. A. (1988). Mixed-model methodology: Theoretical justifications and and future directions. In *Proceedings of the Statistical COmputing Section* 41-49. Amer. Statist.Assoc., Alexandria, Va.

[34] Harville, D. A. (1990). BLUP and beyond. In *Advances in Statistical Methods for Genetic Improvement of Livestock.* (D. Gianola and K. L. Hammond, eds.) 239-276. Springer, New York.

[35] Harville, D. A. and Jeske, D. R. (1989). Mean squared error of estimation and prediction under a general linear model. Preprint 89-37, Statistics Laboratory, Iowa State.

[36] Harville, D. A. (2008). Matrix Algebra from a Statisticians Perspective, Springer-Verlag, New York.

[37] Henderon, C. R. (1963). Selection index and expected genetic advance. In *Statistical Genetics and Plant Breeding* (W. D. Hanson and M. F. Robinson, eds.) 141-163. Publication, 982, NAS-NRC, Wasington, D.C.

[38] Hobert, J.P., Casella, G (1996). The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *J. Amer. Statist. Assoc.*, **91,** 1461-1473.

[39] Ibrahim, J. G., Zhu, H., Garcia, R. I. & Guo, R. (2011). Fixed and random effects selection in mixed effects models. *Biometrics*, **67**, 495–503.

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

[41] Lindley, D. V. and Smith, A. F. M. (1972). Bayes estimates for linear model (with discussion). *J. Roy. Statist. Soc. Ser B* **34** 1-41.

[42] Lui, K. J. and CumberLand, W. G. (1989). A Bayesian approach to small domain estimation. *Journal of Official Statistics* **5** 143-156.

[43] Malec, D. and Sedransk, J. (1985). Bayesian inference for finite population parameters in multistage cluster sampling. *J. Amer. Statist. Assoc.* **80** 897-902.

[44] Malec, D., Sedransk, J., Moriarty, C.L., and LeClere, F.B. (1997). Small area inference for binary variables in National Health Interview Survey. *Journal of the American Statistical Association*, **92,** 815-826.

[45] Morris, C. N. (1988). Determining the accuracy of Bayesian empirical Bayes estimates in the familiar exponential families. In *Statistical Decision Theory and Related Topics IV* (S. S. Gupta and J. O. Berger, eds.)**1** 251-263. Springer, New York.

[46] Nandram, B. and Sedransk, J. (1993). Emperical Bayes estimation of the finite population mean on the current occasion. *Journal of the American Statistical Association*, **88** 994-1000.

[47] Park and G. Casella. The Bayesian lasso (2008). *Journal of the American Statistical Association*, 103(482) 681-686.

[48] Polson, N.G., Scott, J. G. and Windle, J. (2014). The Bayesian bridge. To be published in *Journal of the Royal Statistical Society B.*

[49] Prasad, N. G. N. and Rao, J. N. K. (1990). On the estimation of mean square error of small area predictors. *J. Amer. Statist. Assoc.* **85** 163-171.

[50] Rao, C. R. (1973). *Linear Statistical Inference and Its Applications,* 2nd ed. Wiley, New York.

[51] Rao, J. N. K.(2003). Small area estimation. Wiley-Interscience,Hoboken, NJ.

[52] Royall, R. M. (1978). Prediction models in small area estimation. In *Synthetic Estimates for Small Area* (J. Steinberg, ed.) 63-67. NIDA Monograph Sreies 24. Dept. Health Education and Welfare, Washington, D. C.

[53] Frank, I. and Friedman, J. H. (1993). A statistical view of some chemometrics regression tools (with discussion). *Technometrics*, 35(2) 109-135.

[54] Scott, A. and Smith, T. F. M. (1969). Estimation of multistage surveys. *J. Amer. Statist. Assoc.* **64** 830-840.

[55] Searle, S. R. (1971). *Linear Models.* Wiley, New York.

[56] Sinha, S. K., Rao, J. N. K. (2009). Robust small area estimation. *The Canadian Journal of Statistics*, **37,** 381-399.

[57] Stroud, T. W. F. (1987). Bayes and empirical Bayes approaches to small area estimation. In *Small Area Statistics: An International Symposium.* (R.Platek, J. N. K. Rao, C. E. Sarndal and M. P. Singh, eds.) 124-137. Wiley, New York.

[58] Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso, *Journal of the Royal Statistical Society, Ser. B*, 58 267288.

[59] Xie, D., Raghunathat, T. E. and Lepkowski, J. M., (2007). Estimation of the proportion of overweight individuals in small areas–a robust extenstion of the Fay-Herriot model. *Statistics in Medicine*, **26,** 2699-2715.

[60] You, Y., and Rao, J. N. K. (2003). Pseudo hierarchical Bayes small area estimation combining unit level models and survey weights. *Journal of Statistical Planning and Inference*, **111** 197-208.

[61] Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics.* Wiley, New York.