Bayesian Analysis of Crime Survey Data with Nonresponse

by

Shiao Liu

A Thesis

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Master of Science

in

Applied Statistics

by

May 2018

APPROVED:

Professor Balgobin Nandram, Major Thesis Advisor

Abstract

Bayesian hierarchical models are effective tools for small area estimation by pooling small datasets together. The pooling procedures allow individual areas to “borrow strength” from each other to desirably improve the estimation. This work is an extension of Nandram and Choi (2002), NC, to perform inference on finite population proportions when there exists non-identifiability of the missing pattern for nonresponse in binary survey data. We review the small-area selection model (SSM) in NC which is able to incorporate the non-identifiability. Moreover, the proposed SSM, together with the individual-area selection model (ISM), and the small-area pattern-mixture model (SPM) are evaluated by real crime data in Stasny (1991). Furthermore, the methodology is compared to ISM and SPM using simulated small area datasets. Computational issues related to the MCMC are also discussed.

Acknowledgements

I would like to express my deepest gratitude to my advisor, Dr. Balgobin Nandram, for his consistent support and warm encouragement through my two years' study at WPI. His patient guide and insightful comments have been highly instrumental in my academic development.

I am also extremely grateful to my parents, Guanghua and Ping, for their unconditional love and support throughout all the peaks and valleys in my life.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The Bureau of Justice Statistics' National Crime Survey (NCS) is the nation's primary source of information on criminal victimization. It is collecting data every year on nonfatal personal crimes (rape or sexual assault, robbery, aggravated and simple assault, and personal larceny) and household property crimes (burglary, motor vehicle theft) both reported and not reported to police. One common problem about the data is the increasing nonresponse rate in recent years, which makes the inference on finite population extremely hard, especially when we don't know how different the observed data and the nonresponse missing data are.

1.1 Nonresponse Missing Mechanisms

We consider an analysis of binary survey data from NCS with uncertainty about ignorability. After the survey, the data are poststratified into various small areas. Due to the limited size of datasets from some of the small areas, individual-area (direct) estimation is regularly misleading and untrustworthy. However, some small-area models, such as hierarchical Bayesian models, are excellent choices for small area estimation of this sort. Since they help to pool small datasets from different areas together, hierarchical Bayesian models can supply each small dataset with additional information by sharing same priors and exchangeability stuctures. Stasny (1991) proposed a empirical Bayes method for a two stage model to solve this problem. Nandram and Choi (2000) refined the model by introducing one more stage and hyperparameters. Also, they explained the corresponding Markov chain Monte Carlo approach to dealing with the computational issues. Here, we are using a multistage hierarchical Bayesian nonresponse model to tackle the problem of small datasets as NC did and making inference on the finite population proportions.

It is often necessary to identify the data missing mechanism before implementing inference. The crucial role of the mechanism of missing data in the analysis was largely ignored until the concept was formalized into three types of missing mechanisms (Rubin 1976a), MCAR, MAR, and MNAR. Missing completely at random (MCAR) happens when the missingness does not depend on the values of the data, missing or observed. Missing at random (MAR) occurs when the missingness depends only on the observed data, and not on the unobserved missing components. Both MCAR and MAR are defined as ignorable missing mechanisms, where the missing data are not different from the observed data. When the missing mechanism depends on the unobserved missing data, it is missing not at random (MNAR), which is classified into nonignorable missing mechanism.

Without prespecifying ignorable or nonignorable missing mechanism, NC's multistage Bayesian hierarchical selection nonresponse model can incorporate the uncertainty about ignorability of missing mechanism. By introducing uncertainty parameters γ_i for i^{th} area, it is able to monitor ignorability or nonignorability of the nonresponse and make more general and accurate inference regardless of the missing mechanism. Analysis of ignorability can also be carried out throughout all the areas, after a sample of γ_i has been drawn from the posterior distribution.

1.2 Empirical Bayes versus Full Bayes

There are two distinct approaches to specifying priors for Bayesian hierarchical models, empirical Bayes (EB) and full Bayes (FB), which have been leading hot debate about Bayesianism and Frequentism for a long time (Berger 2006). EB is to specify the parameters of priors by MLEs based on the data, which unfortunately, cannot avoid the procedure of double using the data. FB is to update all the parameters without specifying anyone of them. FB is essentially subjective to prior beliefs, though not so much by selecting objective-type priors like Jeffrey's priors (Efron 2013) but it will introduce more hyperparameters, which will certainly give rise to higher model complexitiy. In NC, they applied EB to specify the hyperparameters in priors by MLEs. Though, in practice, EB can help shift the workload to computers, it inevitably results in overfitting problems. Also, under Markov Chain Monte Carlo, finding MLEs for the hyperparameters can be computationally troublesome, especially when dealing with high dimensional problems. In my point of view, full Bayesian approach is more neat and philosophically beautiful. The simplicity of the full Bayesian approach makes it smooth both in form and computation. As a consequence, we are using full Bayesian approach through the entire work.

1.3 Pattern-Mixture versus Selection

One discussion about the selection of Bayesian hierarchical nonresponse models is whether to use pattern-mixture model or selection model. Rubin (1987) and Little (1993) figured out these two distinct models for incomplete data analysis. Pattern-mixture models specify the conditional distribution of random variables given that it is observed or missing respectively and the marginal distribution of the binary indicator for whether or not a random variable is missing. Selection models specify the distribution of random variables over respondents and nonrespondents, and also the conditional distribution that a random variable is missing given the true answer of it. There are some papers based on the application and comparison of selection models and pattern-mixture models (Ekholm and Skinner 1998, Baker 1995). As for univariate data with a routine pattern of incomplete values, the literature prefered the selection models. NC also mentioned the selection models are more natural and more convenient for our inference. In this work, we will compare the performance of pattern-mixture model and selection models.

1.4 Small-Area versus Individual-Area

As we have stated previously, when the datasets are too small, individual-area (direct) estimation is regularly misleading and untrustworthy. However, a small-area model can help to borrow information among different datasets to reach an overall estimation. Later, we will show the comparison between small-area model and individual-area model.

In the thesis, we wish to go further to mainly address two issues. First, in order to understand the characteristics of the finite population, we hope to make inference on the finite population proportions based on the data from NCS using the proposed small-area selection model (SSM). Second, the key idea of this work is to show that SSM out-performs the individual-area selection model (ISM) and the small-area pattern-mixture model (SPM) on the inference of finite population proportions.

This thesis is organized as follows. In Chapter 2, we give a brief review of the data from NCS in Stasny (1991) and NC. In Chapter 3, we present a finite population inference using SSM, ISM and SPM to show the advantages of the proposed SSM over other two models in small area estimation. We also discuss the prior distributions that may be used in this case and the assessment of the convergence of the Markov Chain Monte Carlo methods by applying formal and informal convergence criteria. In Chapter 4, we perform a multiple comparison of three models by simulation studies based on various criteria. In Chapter 5, we present conclusions and the future outlook.

Chapter 2

National Crime Survey Data and Small-Area Selection Model

We give a brief review of National Crime Survey data in Stasny (1991) to conduct a real data experiment, and small-area selection model (SSM) in Nandram and Choi (2002) which is the proposed model to compare with the other two models in the following sections.

2.1 National Crime Survey Data

The Bureau of Justice Statistics' National Crime Survey (NCS) is the nation's main source of information on criminal victimization. The data from NCS are used in order to calculate quarterly and yearly estimates of the prevalence of crime. Individuals interviewed are questionaired for the crimes (e.g., rape or sexual assault, robbery, aggravated and simple assault, and personal larceny) committed against them in the previous six months. Stasny (1991) summarized a subset of the regular NCS interview data collected from January 1975 to June 1979 in a table and discussed the design of the NCS. Sampling design features are not showing up in the proposed model, except that we are assuming the data are collected by probability sampling.

We used the data summarized by Stasny (1991), who took a random start at the record for the eighth household in the full NCS data from January 1975 to June 1979 and then systematically every fifteenth record after the start. The selected data are poststratified into 10 areas based on three neighborhood characteristics: (a) urban (U) and rural (R); (b) central city (C), other incorporated place (I), and unincorporated or not a place (N); and (c) low poverty level (L) (9% or fewer of families below the poverty level) and high poverty level (H) (10% or more of families below the poverty level). Because some combination of neighborhood characteristics are impossible, for example a rural area cannot be a central city as observed in Stasny (1991), this poststratification classifies the full data into 10 different areas.

We reproduce the Table 1 in NC as Table 2.1 below to present the selected data conveniently. Throughout, y_i is the number of households committed against crimes in the i^{th} area, r_i is the number of respondents in the i^{th} area, n_i is the total number of households sampled in the i^{th} area, $i = 1, \ldots, \ell$, where $\ell = 10$ in this problem. The nonresponse rate in these areas ranges from 9.4% to 16.9%; There are a lot of reasons for nonresponse, for example a woman may be shy to report a sexual assault committed against her. The areas UNH, RIL, and RIH have relatively few households, thus can be considered as small areas. To be noticed, areas UNH and RIL have relatively high observed nonresponse rates.

Table 2.1: National Crime Survey Data

Area	y	$r-y$	$n-r$	$\hat{p}$	$\hat{\delta}$
UCL	156	555	104	.219	.872
UCH	95	364	73	.207	.862
UIL	162	557	101	.225	.877
UIH	72	262	36	.216	.903
UNL	92	297	79	.237	.831
UNH	15	40	9	.273	.859
RIL	11	36	7	.234	.870
RIH	10	105	20	.087	.852
RNL	35	274	32	.113	.906
RNH	79	413	64	.161	.885

Note: The response is binary: 0 for no crimes and 1 for at least one crime; $\hat{p}$ is the observed proportion of house holds with at least one crime; and $\hat{\delta}$ is the observed proportion of respondents.

2.2 Small-Area Selection Model

If one hopes to incorporate uncertainty about missing mechanism of nonresponse, intuitively, the first choice would be mixing the ignorable model and the nonignorable model together probabilistically. However, the mixture model of this sort has been proved useless in NC. They showed that if we assume the weights of both models to be 0.5, the posterior probability that the ignorable model holds is only 0.03. Therefore, it is pertinent to review the different uncertainty model in NC.

Let y_{ij} and r_{ij} be the characteristic and response variables for the j^{th} individual in the i^{th} area. To be more specific, $y_{ij} = 1$ if the j^{th} individual in the ith area has the characteristic and $y_{ij} = 0$ otherwise; and $r_{ij} = 1$ if j^{th} individual in the i^{th} area is a respondent and $r_{ij} = 0$ otherwise, for $i = 1, \ldots, \ell$, $j = 1, \ldots, n_i$. The nonignorable nonresponse model in NC is given as follows:

$$
\begin{gathered}
y_{ij}|p_i \overset{ind}{\sim} \text{Bernoulli}(p_i), \\
r_{ij}|y_{ij} = s-2, \pi_{is-1} \overset{ind}{\sim} \text{Bernoulli}(\pi_{is-1}) \ , \ s = 2, 3, \\
p_i|\mu_{21}, \tau_{21} \overset{iid}{\sim} \text{Beta}(\mu_{21}\tau_{21}, (1-\mu_{21})\tau_{21}), \\
\pi_{is-1}|\mu_{2s}\tau_{2s} \overset{ind}{\sim} \text{Beta}(\mu_{2s}\tau_{2s}, (1-\mu_{2s})\tau_{2s}).
\end{gathered}
$$

Different from the pattern-mixture model, the nonignorable selection model identifies the parameters π_{i1} and π_{i2}, which are the responding mechanisms of samples without and with the characteristic respectively. We can get the ignorable model by simply letting $\pi_{i1} = \pi_{i2}$. By the specification of p_i, we can understand the proportion of the characteristic of samples in the i^{th} area.

Here, one problem is that we have to make the same assumption for the missing mechanism of all the areas before we fit the model. However, while the assumption works fine for some of the areas, it may not be appropriate for the rest, especially when the number of areas increase as we introduce more neighborhood characteristics. The assumption of this kind would possibly make the estimation of finite population far from accurate, as it sometimes goes to the wrong assumption, making the ignorable area nonignorable or the opposite. Therefore, the small-area selection model (SSM) is necessary to be introduced to incorporate uncertainty about ignorability by using a centering parameter γ_i for the i^{th} area. The key point is to let $\pi_{i2} = \gamma_i \pi_{i1}$ for π_{i2} in the nonignorable model. Similar to the nonignorable model, SSM for $i = 1, \ldots, \ell$, $j = 1, \ldots, n_i$ is given as follows:

$$
\begin{gathered}
y_{ij} | p_i \overset{iid}{\sim} \text{Bernoulli}(p_i), \\
r_{ij} | \pi_i, y_{ij} = 0 \overset{iid}{\sim} \text{Bernoulli}(\pi_i), \\
r_{ij} | \pi_i, y_{ij} = 1, \gamma_i \overset{iid}{\sim} \text{Bernoulli}(\gamma_i \pi_i), \ 0 < \gamma_i \pi_i < 1, \\
p_i | \mu_i, \tau_i \overset{iid}{\sim} \text{Beta}(\mu_1 \tau_1, (1 - \mu_1)\tau_1), \\
\pi_i | \mu_2 \tau_2 \overset{iid}{\sim} \text{Beta}(\mu_2 \tau_2, (1 - \mu_2)\tau_2), \\
\gamma_i \overset{iid}{\sim} \text{Truncated Gamma}(\nu, \nu), \ 0 < \gamma_i < \pi_i^{-1}.
\end{gathered}
$$

Here γ_i is the ratio of the probability of respondents among success to the probability of respondents among non-success for the i^{th} area. If $\gamma_i = 1$, then the i^{th} area has ignorable missing mechanism, otherwise nonignorable.

The hyperparameters are in turn assigned to be independent with proper priors:

$$\nu \sim \text{Gamma}(\eta_3^{(0)}, \nu_3^{(0)}), \quad \mu_1, \mu_2 \overset{iid}{\sim} \text{Uniform}(0,1),$$

$$\tau_1 \sim \text{Gamma}(\eta_1^{(0)}, \nu_1^{(0)}), \;\; \tau_2 \sim \text{Gamma}(\eta_2^{(0)}, \nu_2^{(0)}),$$

$(\eta_s^{(0)}, \nu_s^{(0)})$ are to be specified by maximum likelihood estimates.

In small area estimation, it is a standard practice to assume that the area effects are exchangeable. This assumption is incorporated by setting the area effects to have a common distribution, which is one of the key ideas of NC. By this step, the "borrowing strength" of the small dataset from larger areas can be realized. However, the procedure to specify priors cannot avoid the charge of double using the data. Moreover, it requires unknown additional computational effort to find the maximum likelihood estimates for these hyperparameters when the dimension of data becomes higher.

Chapter 3

Inference on Finite Population Proportions

By data augmentation approach, we introduce a latent variable $\mathbf{z}= (z_1, \ldots, z_l)$, where $z_i = \sum_{j=1}^{n_i} y_{ij}$, which represents the total number of people who have the characteristic in the sample within i^{th} area. Similar to the multivariate form of $\mathbf{z}$, we get $\mathbf{p}, \boldsymbol{\pi}, \boldsymbol{\gamma}, \mathbf{z}, \mathbf{n}, \mathbf{y}, \mathbf{r}$, for example $\mathbf{p}= (p_1, \ldots, p_l)$. Unlike the specification of priors in NC, we choose non-data dependent non-informative priors for μ_s and τ_s:

$$\pi(\mu_s) \overset{ind}{=} \frac{1}{\sqrt{\mu_s(1-\mu_s)}}, \;\; \pi(\tau_s) \overset{ind}{=} \frac{1}{(1+\tau_s)^2}, \text{ where } s = 1, 2 \text{ and } 0 < \mu_s < 1, \;\; \tau_s > 0.$$

3.1 Model Fitting

By incorporating independent priors, we can obtain the joint posterior density of $[\mathbf{p}, \boldsymbol{\pi}, \boldsymbol{\gamma}, \mathbf{z}, \mathbf{n}, \mathbf{y}, \mathbf{r}]$, containing all the quantities in the model:

$$\pi(\mu_1)\pi(\mu_2)\pi(\tau_1)\pi(\tau_2) \times \prod_{i=1}^{l}\{\frac{n_i!}{y_i!(z_i-y_i)!(r_i-y_i)!(n_i-z_i-r_i+y_i)!}\times$$

$$(p_i\gamma_i\pi_i)^{y_i}(p_i(1-\gamma_i\pi_i))^{z_i-y_i}\times((1-p_i)\pi_i)^{r_i-y_i}((1-p_i)(1-\pi_i))^{n_i-z_i-r_i+z_i}\times$$

$$\frac{p_i^{\mu_1\tau_1-1}(1-p_i)^{(1-\mu_1)\tau_1-1}}{B(\mu_1\tau_1,(1-\mu_1)\tau_1)}\frac{\pi_i^{\mu_2\tau_2-1}(1-\pi_i)^{(1-\mu_2)\tau_2-1}}{B(\mu_2\tau_2,(1-\mu_2)\tau_2)}\}.$$

However, the joint posterior density does not have a closed form, so the analysis of posterior cannot be conducted thoroughly. Typical way to do it, we need apply MCMC to draw samples from posterior. Here, we implement the Metropolis-within-Gibbs algorithm to generate samples from the marginal distributions of parameters. Note that $[\mathbf{n},\mathbf{r},\mathbf{y}]$ are already known, it is not necessary to keep them in the posterior density. The steps in the Metropolis-within-Gibbs sampler which we implement to fit the model are:

0) Integrate out $[\mathbf{p},\boldsymbol{\pi},\boldsymbol{\gamma}]$ to get the marginal density of $[\mathbf{z},\boldsymbol{\mu_1},\boldsymbol{\mu_2},\boldsymbol{\tau_1},\boldsymbol{\tau_2}]$.

1) Specify the initial values for $\mu_1,\mu_2,\tau_1,\tau_2$ and latent variable $\mathbf{z}$.

2) Generate $\mu_1,\mu_2,\tau_1,\tau_2$ from the conditional distribution of $[\mu_1,\tau_1,\mu_2,\tau_2 \mid \mathbf{z}]$.

3) Generate $\mathbf{z}$ from the conditional distribution of $[\mathbf{z} \mid \boldsymbol{\mu_1},\boldsymbol{\mu_2},\boldsymbol{\tau_1},\boldsymbol{\tau_2}]$.

4) Set $[\mathbf{p},\boldsymbol{\pi},\boldsymbol{\gamma}]$ as a block, and generate them simultaneous from three independent conditional distribution of $[\mathbf{p},\boldsymbol{\pi},\boldsymbol{\gamma} \mid \boldsymbol{\mu_1},\boldsymbol{\mu_2},\boldsymbol{\tau_1},\boldsymbol{\tau_2},\mathbf{z}] = [\mathbf{p} \mid \boldsymbol{\mu_1},\boldsymbol{\mu_2},\boldsymbol{\tau_1},\boldsymbol{\tau_2},\mathbf{z}] \times [\boldsymbol{\pi} \mid \boldsymbol{\mu_1},\boldsymbol{\mu_2},\boldsymbol{\tau_1},\boldsymbol{\tau_2},\mathbf{z}] \times [\boldsymbol{\gamma} \mid \boldsymbol{\mu_1},\boldsymbol{\mu_2},\boldsymbol{\tau_1},\boldsymbol{\tau_2},\mathbf{z}]$.

5) Iterate from 2).

3.1.1 Generating $\mu_1, \mu_2, \tau_1, \tau_2$

The marginally conditional distribution of $\mu_1, \mu_2, \tau_1, \tau_2$ given $\mathbf{z}, \mathbf{n}, \mathbf{r}, \mathbf{y}$ is formed from all of the terms in the marginal distribution with $\mu_1, \mu_2, \tau_1, \tau_2$. Therefore, we can update $\mu_1, \mu_2, \tau_1, \tau_2$ by the following conditional densities:

$$[\mu_1 \mid \tau_1, \mathbf{z}] \propto \frac{1}{\sqrt{\mu_1(1-\mu_1)}} \times \prod_{i=1}^{l} \{\frac{\Gamma(z_i + \mu_1\tau_1)\Gamma(n_i - z_i + (1-\mu_1)\tau_1)}{\Gamma(\mu_1\tau_1)\Gamma((1-\mu_1)\tau_1)}\},$$

$$[\tau_1 \mid \mu_1, \mathbf{z}] \propto \frac{1}{(1+\tau_1)^2} \times \prod_{i=1}^{l} \{\frac{\Gamma(z_i + \mu_1\tau_1)\Gamma(n_i - z_i + (1-\mu_1)\tau_1)\Gamma(\tau_1)}{\Gamma(n_i + \tau_i)\Gamma(\mu_1\tau_1)\Gamma((1-\mu_1)\tau_1)}\},$$

$$[\mu_2 \mid \tau_2, \mathbf{z}] \propto \frac{1}{\sqrt{\mu_2(1-\mu_2)}} \times \prod_{i=1}^{l} \{\frac{\Gamma(r_i - y_i + \mu_2\tau_2 - 1)\Gamma(n_i - z_i - r_i + y_i + (1-\mu_2)\tau_2)}{\Gamma(\mu_2\tau_2)\Gamma((1-\mu_2)\tau_2)}\},$$

$$[\tau_2 \mid \mu_2, \mathbf{z}] \propto \frac{1}{(1+\tau_2)^2} \times \prod_{i=1}^{l} \{\frac{\Gamma(r_i - y_i + \mu_2\tau_2 - 1)\Gamma(n_i - z_i - r_i + y_i + (1-\mu_2)\tau_2)\Gamma(\tau_2)}{\Gamma(n_i - z_i + \tau_2 - 1)\Gamma(\mu_2\tau_2)\Gamma((1-\mu_2)\tau_2)}\}.$$

Since these conditional densities are not known distributions which we can easily draw samples from, a Metropolis-Hastings accept-reject step is needed. Respectively, we take Beta(10,30), Beta(27,3), Uniform(10,200) and Uniform(10,200) as the proposal distributions for generating μ_1, μ_2, τ_1 and τ_2, and the jumping rates are about 0.375, 0.312, 0.501 and 0.399. Note that increasing the width of the bounds of τ_1 and τ_2 will not change the inference results much. However, it will increase the sample autocorrelation by some degree. The upper bound of τ should be around $\sum_{i=1}^{\ell} \sqrt{r_i} = 177.7$ (Nandram and Sedransk 1993).

3.1.2 Generating $\mathbf{z}$

The marginally conditional distribution of z_i given μ_1, μ_2, τ_1 and τ_2 can be obtained in the similar way:

$$[z_i \mid \mu_1, \mu_2, \tau_1, \tau_2] \propto \frac{\Gamma(z_i + \mu_1\tau_1)\Gamma(n_i - z_i + (1-\mu_1)\tau_1)\Gamma(n_i - z_i - r_i + y_i + (1-\mu_2)\tau_2)}{\Gamma(z_i + 2)\Gamma(n_i - r_i - z_i + y_i + 1)\Gamma(n_i - z_i + \tau_2 - 1)}.$$

As the domain of z_i is $\{y_i,\ y_{i+1},\ \ldots,\ n_i - r_i + y_i\}$, which contains finite number of values, we can calculate the probability of z_i at each possible value, and draw samples of $z_i^{\prime}s$ based on these probabilities. Note that we update all the $z_i^{\prime}s$ simultaneously.

3.1.3 Generating $\mathbf{p}$, $\boldsymbol{\pi}$, $\boldsymbol{\gamma}$

By reparameterization $\phi_i = \gamma_i \pi_i$ or $\boldsymbol{\phi} = \boldsymbol{\gamma} \circ \boldsymbol{\pi}^1$, we can easily see that $\mathbf{p}$, $\boldsymbol{\pi}$ and $\boldsymbol{\phi}$ are independent conditional on $\mathbf{z}$, μ_1, μ_2, τ_1 and τ_2. Therefore, we can obtain the corresponding densities:

$$\phi_i \overset{ind}{\sim} \text{Beta}(y_i + 1,\ z_i - y_i + 1)$$
$$p_i \overset{ind}{\sim} \text{Beta}(z_i + \mu_1\tau_1,\ n_i - z_i + (1 - \mu_1)\tau_1)$$
$$\pi_i \overset{ind}{\sim} \text{Beta}(r_i - y_i + \mu_2\tau_2 - 1,\ n_i - z_i - r_i + y_i + (1 - \mu_2)\tau_2)$$

Since $\mathbf{p}$, $\boldsymbol{\pi}$ and $\boldsymbol{\phi}$ are independent, we can draw them simultaneously in one block. Note that γ_i and π_i are not independent. If we implement Gibbs sampler here for γ_i and π_i, we will get highly autocorrelated samples for both γ_i and π_i due to the shared bound of γ_i and π_i, $0 < \gamma_i\pi_i < 1$. However, ϕ_i and π_i are independent random variables after the reparameterization. In each run, $\gamma_i^{\prime}s$ can be obtained by $\gamma_i = \frac{\phi_i}{\pi_i}$. Therefore, we would like to apply reparameterization to reduce tremendous autocorrelation among the posterior samples of $\boldsymbol{\gamma}$ and $\boldsymbol{\pi}$.

3.2 Inference on Finite Population

Quantity of interest is the finite population proportion of crime rates in these ten areas. We hope to estimate the finite population proportions in spite of the uncertainty of nonresponse missing mechanism. Let $0 < \omega < 1$ be the sampled proportion of the finite population, so $1 - \omega$ is the nonsampled proportion of the finite population. The finite population can be a census of the population in the area. Thus, the sampled proportion is generally known in real studies. The SSM

---

[1] $\circ$ is the Hadamard product.

estimator of i^{th} area is defined as:

$$\hat{p}_{S,i} = \omega \frac{z_i}{n_i} + (1-\omega)\frac{z_i^{'}}{(1/\omega - 1)n_i}, \text{ where } z_i^{'} \overset{ind}{\sim} \text{Binomial}((1/\omega - 1)n_i, p_i).$$

To examine the performance of the SSM estimator, we perform a real-data experiment based on previously reviewed NCS data in comparison to the performance of the ISM estimator, which is same with SSM in almost all the stages except that it does not have the shared priors $\mu_1, \mu_2, \tau_1, \tau_2$. The prior distributions of $\mathbf{p}$, $\boldsymbol{\pi}$, $\boldsymbol{\gamma}$ are uniforms with the same bounds respectively as they are in SSM. Similarly, the posterior distribution of ISM is

$$[p_i, \pi_i, \gamma_i, z_i \mid y_i, r_i] \propto \frac{p_i^{z_i}(1-p_i)^{n_i - z_i}(\gamma_i \pi_i)^{y_i}(1-\gamma_i\pi_i)^{z_i - y_i}\pi_i^{r_i - y_i}(1-\pi_i)^{n_i - z_i - r_i + y_i}}{(z_i - y_i)!(n_i - z_i - r_i + y_i)!}.$$

Based on the similar MCMC steps, we can obtain the compared results of these two models in the inference on finite population crime proportions. Besides, we introduce the SPM and its estimator:

$$r_{ij} \mid \pi_i \overset{ind}{\sim} \text{Bernoulli } (\pi_i),$$
$$y_{ij} \mid r_{ij} = 1,\ p_i \overset{ind}{\sim} \text{Bernoulli } (p_i),$$
$$y_{ij} \mid r_{ij} = 0,\ p_i,\ \gamma_i \overset{ind}{\sim} \text{Bernoulli } (\gamma_i p_i),$$
$$p_i | \mu_1, \tau_1 \overset{iid}{\sim} \text{Beta}(\mu_1\tau_1, (1-\mu_1)\tau_1),$$
$$\pi_i | \mu_2, \tau_2 \overset{iid}{\sim} \text{Beta}(\mu_2\tau_2, (1-\mu_2)\tau_2).$$

Let $y_i = \sum\limits_{r_{ij}=1} y_{ij}$ and $z_i = \sum\limits_{r_{ij}=0} y_{ij}$. Pattern-mixture estimator is $\hat{p}_{P,i} = \omega \dfrac{y_i + z_i}{n_i} + (1-\omega)\dfrac{z_i^{''}}{(1/\omega - 1)n_i}$, where $z_i^{''} \overset{ind}{\sim} \text{Binomial}((1/\omega - 1)n_i, \pi_i p_i + (1-\pi_i)\gamma_i p_i)$.

Table 3.1: Comparison of the Inference on Finite Population Proportions under Three Models

Area	Small-Area Selection Model			Individual-Area Selection Model			Small-Area Pattern-Mixture Model		
	ω=1/5	ω=1/10	ω=1/50	ω=1/5	ω=1/10	ω=1/50	ω=1/5	ω=1/10	ω=1/50
UCL	.251(.032)	.250(.034)	.249(.034)	.248(.040)	.252(.040)	.251(.040)	.254(.039)	.253(.039)	.253(.039)
UCH	.247(.037)	.249(.037)	.243(.036)	.242(.045)	.241(.044)	.243(.045)	.244(.042)	.244(.042)	.245(.043)
UIL	.255(.033)	.253(.033)	.252(.033)	.254(.040)	.256(.040)	.254(.039)	.256(.037)	.257(.038)	.259(.038)
UIH	.237(.031)	.237(.031)	.236(.031)	.242(.036)	.241(.037)	.240(.036)	.241(.035)	.241(.035)	.240(.035)
UNL	.278(.043)	.280(.042)	.274(.042)	.273(.053)	.275(.053)	.272(.052)	.272(.049)	.276(.050)	.276(.051)
UNH	.279(.050)	.277(.050)	.272(.052)	.302(.066)	.308(.069)	.305(.071)	.277(.057)	.270(.056)	.271(.055)
RIL	.256(.050)	.253(.048)	.251(.051)	.273(.070)	.267(.068)	.269(.074)	.247(.055)	.248(.056)	.248(.055)
RIH	.184(.041)	.184(.041)	.185(.043)	.139(.052)	.136(.052)	.136(.055)	.158(.049)	.164(.050)	.166(.047)
RNL	.161(.030)	.164(.031)	.164(.032)	.147(.034)	.147(.035)	.146(.035)	.160(.034)	.160(.035)	.160(.034)
RNH	.204(.033)	.205(.032)	.203(.032)	.196(.039)	.197(.037)	.194(.039)	.202(.037)	.201(.036)	.201(.037)

Note: Here, in the notation a(b), a is the posterior finite population mean (i.e. the estimate of the finite population crime proportion) and b is the standard deviation of a.

To fit the SSM and the SPM, we assign the same prior distributions described in subsection 3.1. In order to obtain Table 3.1, we generate three chains of 16000 iterations each for both SSM and SPM with $\omega = 1/5,\ 1/10,\ 1/50$, respectively, "burn in" the first 1000 and then thinned the chain by taking every 15th sample value to obtain 1000 samples. As for the ISM, we can get 1000 independent iterations directly.

Table 3.1 presents the estimates of the three models and the standard deviations of posterior finite population proportions as well as a sensitive analysis of ω. Apparently in the table, SSM can increase the estimates from ISM by as much as 30%, which is much more significant compared with pattern-mixture model. Also, the efficiency of the SSM estimator relative to other 2 models is always greater than 1. Furthermore, all three estimators are not sensitive to the choice of ω. With nothing to lose, we can fix $\omega = 1/5$ in the following studies.

The SPM estimator, which also has the "pooling" effect, obviously out-performs the original ISM estimator which merely takes advantage of the data from the individual area itself. However, it pales by comparison to the SSM estimator in terms of both "pooling" effect and efficiency.

Table 3.2: Comparison of two sets of priors

Area	Previous Priors	Flat Priors	ISM with Flat Priors
UCL	.251(.032)	.245(.029)	.248(.040)
UCH	.247(.037)	.240(.031)	.242(.045)
UIL	.255(.033)	.246(.030)	.254(.040)
UIH	.237(.031)	.230(.028)	.242(.036)
UNL	.278(.043)	.275(.036)	.273(.053)
UNH	.279(.050)	.263(.046)	.302(.066)
RIL	.256(.050)	.246(.045)	.273(.070)
RIH	.184(.041)	.184(.035)	.139(.052)
RNL	.161(.030)	.162(.027)	.147(.034)
RNH	.204(.033)	.198(.029)	.196(.039)

Note: Here, we fix $\omega = 1/5$. In the notation a(b), a is the posterior finite population mean (i.e. the estimate of the finite population crime proportion) and b is the standard deviation of a.

3.3 Prior Distribution Choices and Convergence Tests

As we have stated in the Introduction, as long as we have several small area datasets, we can see the "pooling" effect between the datasets by the application of small-area models. However, it is generally a tough question to pick suitable priors for SSM to get the best "pooling" effect based on the information of the finite population. The prior choice in Bayesian statistics has produced much discussion and is different based on the settings. In this section, we compare the performance of SSM estimator based on two different sets of non-informative priors: the priors discussed in section 3.1 and the flat priors.

Table 3.2 shows that flat priors seem better than the previous priors in terms of the "pooling data" effect and efficiency. The good choices of priors always incorporate the underlying information of the finite population crime characteristic, responding characteristic and the sample size. One perspective on varying the priors for $\mu_1,\ \mu_2,\ \tau_1,\ \tau_2$ is that it allows us to know how much additional information we

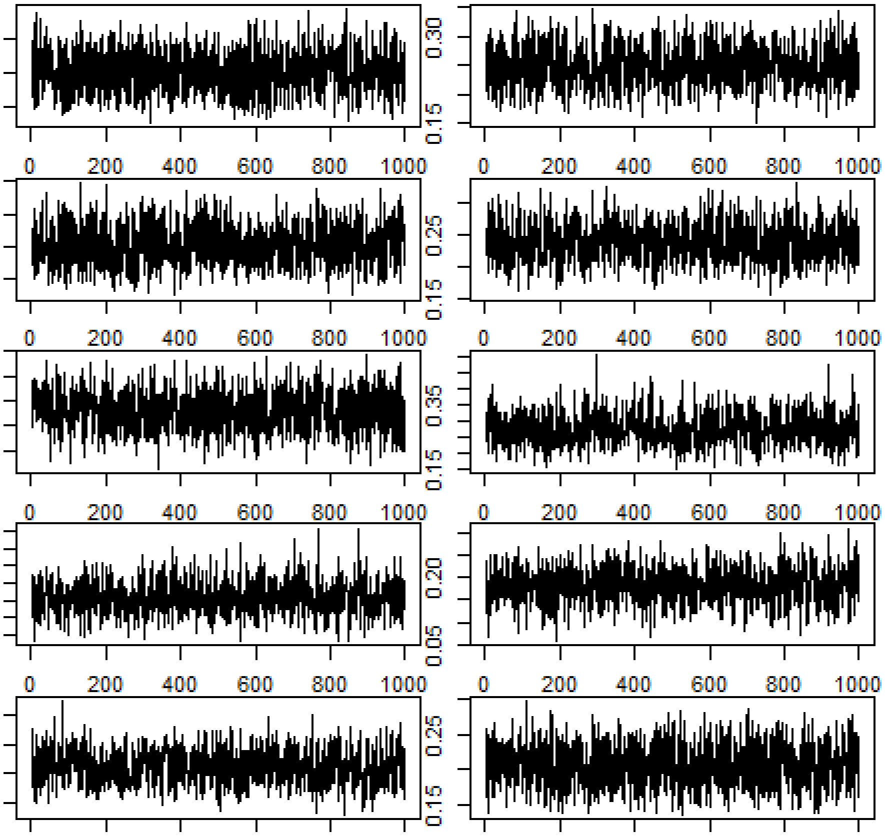

Figure 3.1: Trace plots of the finite population proportions of 10 different areas

need to make the analysis accurate.

Throughout this work, the convergence of MCMC in the proposed model was assessed by trace plots as this approach can be run using the sampling outputs, reducing the additional calculation.

Figure 3.1 shows the trace plots of the finite population proportions of 10 different areas. The plots indicate the convergence of the Markov chains. Moreover, the trace plots of other parameters are all graphed in the same way, which give no indication against the convergence. Since too many trace plots should be tested, we will not show the rest of them here. Furthermore, all of the samples of parameters have been tested by Geweke's diagnostics. All shows that the convergence was achieved. We will only show the results of the predicted finite population proportions

Table 3.3: Geweke's Test P-values

Area	Small-Area Selection	Individual-Area Selection	Small-Area Pattern-mixture
UCL	.814	.760	.593
UCH	.197	.818	.795
UIL	.695	.805	.125
UIH	.348	.423	.958
UNL	.973	.238	.564
UNH	.415	.629	.839
RIL	.964	.244	.328
RIH	.886	.889	.991
RNL	.124	.083	.268
RNH	.097	.639	.911

based on these three models in Table 3.3. In addition, except that area UNH and RNL has a effective sample size 892 and 823 respectively, the posterior samples of finite population proportions obtained in each of the other eight areas has no autocorrelation.

Chapter 4

Simulation Studies

We run a simulation study to compare the performance of the previously discussed three models. We want to investigate the difference between the direct estimation and pooled estimation based on the simulated data. More intuitively, we hope to show that the SSM estimator is more accurate and provides higher precision than other two estimators.

The simulation step is as follows. First, in each area, we fix n_i unchanged and generate z_i by $z_i \mid p_i, n_i \sim \text{Binomial}(n_i, p_i)$, where $\mathbf{p}$=(.25,.25,.25,.24,.28,.27, .25,.19,.16,.20), which is the posterior mean of $\mathbf{p}$ by the SSM. Next, we generate y_i by $y_i \mid \gamma_i, \pi_i, z_i \sim \text{Binomial}(z_i, \gamma_i\pi_i)$, where $\boldsymbol{\pi} = (.91, .91, .91, .93, .89, .90, .91, .93, .95, .93)$ and $\boldsymbol{\gamma} = (.84, .81, .86, .89, .79, .85, .81, .50, .72, .77)$, which are respectively the posterior mean of $\boldsymbol{\pi}$ and $\boldsymbol{\gamma}$ by the SSM as well. Also, we generate $r_i - y_i$ by $r_i - y_i \mid \pi_i, z_i, n_i \sim \text{Binomial}(n_i - z_i, \pi_i)$. In this way. we have performed 1000 simulated runs, which gives us 1000 simulated samples in the same structure as the NCS data discussed previously in section 2.

We fit each of the 1000 simulated runs using these 3 models in the exact same manner as described above in section 3 and use the flat priors for $\mu_1, \mu_2, \tau_1, \tau_2$. For

Table 4.1: Compared Results of Simulation Studies

Area	RB			PRMSE			WID			COV		
	SSM	ISM	SPM	SSM	ISM	SPM	SSM	ISM	SPM	SSM	ISM	SPM
UCL	-.0048	-.0191	-.0698	.0436	.0485	.0475	.1229	.1334	.1317	.969	.981	.979
UCH	.0007	-.0298	-.0159	.0472	.0535	.0518	.1334	.1475	.1444	.956	.964	.965
UIL	.0011	-.0400	.0053	.0433	.0474	.0467	.1201	.1298	.1287	.952	.968	.970
UIH	.0091	.0160	.0164	.0417	.0454	.0447	.1169	.1275	.1263	.947	.956	.964
UNL	-.0084	-.0398	-.0314	.0548	.0629	.0603	.1551	.1734	.1668	.958	.973	.966
UNH	.0701	.1374	.0352	.0779	.0931	.0801	.2003	.2488	.2093	.875	.912	.893
RIL	.0779	.1034	.0255	.0718	.0933	.0811	.1985	.2516	.2086	.852	.913	.888
RIH	.0108	-.2264	-.1168	.0581	.0788	.0672	.1580	.1786	.1719	.905	.836	.882
RNL	-.0031	-.1070	-.0449	.0412	.0463	.0437	.1137	.1172	.1192	.935	.878	.934
RNH	.0047	-.0463	-.0188	.0416	.0466	.0451	.1178	.1271	.1261	.960	.956	.972

Note: SSM, ISM, SPM denotes small-area selection model, individual-area selection model, small-area pattern-mixture model respectively.

each of the 1000 simulated runs, we performed the Geweke test of stationarity, calculated the autocorrelations and the effective sample sizes. The performance of the Gibbs sampler was satisfactory in most cases. This computation procedures take less than 10 hours running in a personal computer.

We calculated the posterior means (PM), posterior standard deviations (PSD), the 95% HPD intervals and its width. Moreover, we computed the relative bias, $RB = (PM - T)/T$, where T denotes the true finite population proportion and $T = \mathbf{z}./\mathbf{n}$. Also, we computed the posterior root mean squared error, $PRMSE = \sqrt{(PM - T)^2 + PSD^2}$. Regarding intervals, we calculated their average width (WID) and the coverage of the true proportion (COV).

Table 4.1 and Appendix A present the compared results of the simulation. Apparently, SSM estimator out-performs both ISM and SPM estimators in almost all the areas in terms of RB, PRMSE and WID, which means SSM estimator is more accurate and more precise. However, the SSM estimator does not overwhelm the other two estimators in terms of 95% HPD interval coverage of the true proportions. The other two estimators exhibit better performance in COV possibly due to the less WID of SSM estimator.

In all, from these results, we conclude that the performance of the proposed small-area selection model is obviously better than the individual-area selection model and the small-area pattern-mixture model. The "pooling" effect of the proposed model helps small areas to "borrow strength" from other areas, which makes the proposed estimator more accurate and precise.

Chapter 5

Conclusion and Future Outlook

We have considered the problem on estimating the proportion of people victimized by crimes based on the data with nonresponse from NCS survey. Our approach is an application of the Bayesian hierarchical nonresponse model in Nandram and Choi (2002). Based on the real-data experiment and the simulated results, we can show that, apparently, there is an improvement in precision and accuracy of the proposed estimator. Also, we have smaller relative bias, smaller posterior root mean squared error and better width for the 95% HPD intervals in simulation study. Therefore, we conclude the proposed model is significantly better than the other two models, i.e. the individual-area selection model and the small-area pattern-mixture model.

Note that in section 3.3 we showed the completely non-informative priors are better than the Jeffrey's priors, which are still non-informative, in estimation of finite population proportions. An interesting extention is how to accommodate the underlying population characteristic structure assumption by the choice of the prior distribution. Possibly we can incorporate the prior information of $\mathbf{p}$, $\boldsymbol{\pi}$ and the survey sample sizes to find some satisfying informative priors.

Appendix A

Simulation Results

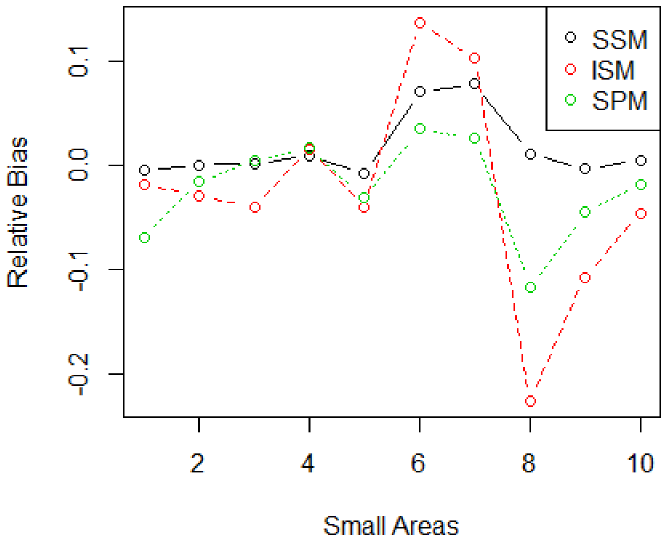

Figure A.1: Relative Bias

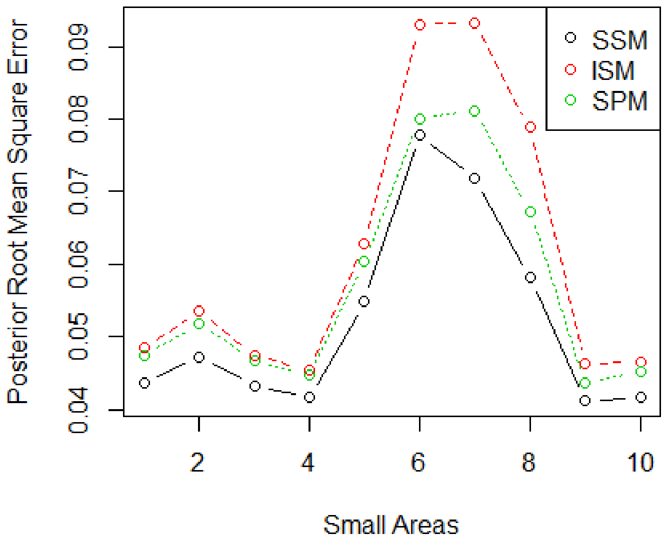

Figure A.2: Posterior Root Mean Squared Error

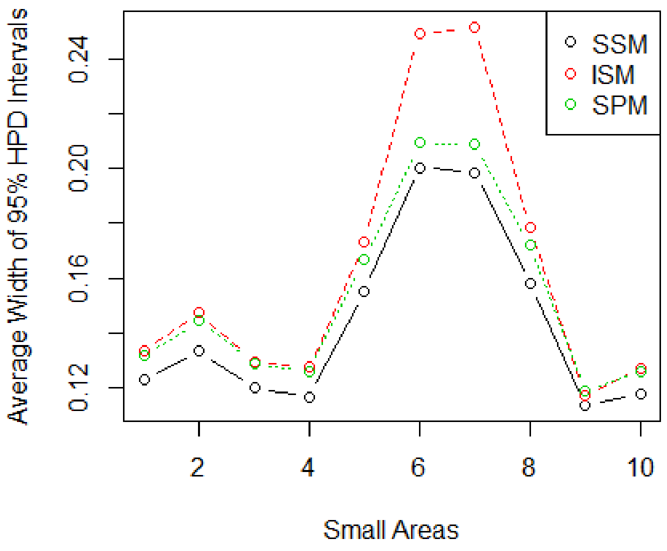

Figure A.3: Average Width of 95% HPD Interval

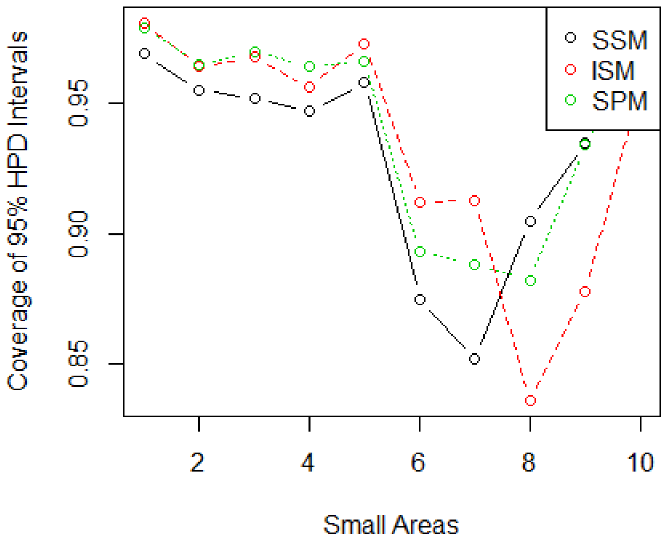

Figure A.4: Coverage of the True Proportion

Bibliography

[1] Baker, S. G. (1995), “Marginal Regression for Repeated Binary Data with Outcome Subject to Non-Ignorable Non-Response,” *Biometrics*, 51, 1042-1052.

[2] Berger, J. (2006), “The case for objective Bayesian analysis,” *Bayesian Analysis*, 1, 385-402.

[3] Efron, B. (2013), “A 250-year argument: Belief, behavior, and the bootstrap,” *Bulletin of the American Mathematical Society*, 50, 129-146.

[4] Ekholm, A., and Skinner, C. (1998), “The Muscatine children’s obesity data reanalysed using pattern mixture models,” *Applied Statistics*, 47, 251-263.

[5] Little, R.J. (1993), “Pattern-Mixture Models for Multivariate Incomplete Data,” *Journal of the American Statistical Association*, 88, 125-134.

[6] Nandram, B., and Choi, J.W. (2000), “Bayesian Inference for a Proportion Under Nonignorable Nonresponse,” technical report, National Center for Health Statistics.

[7] Nandram, B., and Choi, J.W. (2002), “Hierarchical Bayesian Nonresponse Models for Binary Data from Small Areas with Uncertainty about Ignorability,” *Journal of the American Statistical Association*, 97, 381-388.

[8] Nandram, B., and Sedransk, J. (1993), "Bayesian Predictive Inference for a Finite Population Proportion: Two-Stage Cluster Sampling," *Journal of the Royal Statistical Society*, Ser. B., 55, 399-408.

[9] Rubin, D. B. (1976), "Inference and Missing Data," *Biometrika*, 63, 581-590.

[10] Rubin, D. B. (1987), "Multiple Imputation for Nonresponse in Surveys," New York: Wiley.

[11] Stasny, E. A. (1990), "Symmetry in Flows Among Reported Victimization Classifications with Nonresponse," *Survey Methodology*, 16, 305-330.

[12] Stasny, E. A. (1991), "Hierarchical Models for the Probabilities of a Survey Classification and Nonresponse: An Example From the National Crime Survey," *Journal of the Royal Statistical Society*, 86, 296-303.