ABSTRACT

Title of dissertation: Hierarchical Bayesian Estimation of
Small Area Means Using
Complex Survey Data

Neung Soo Ha, Doctor of Philosophy, 2013

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

In survey data analysis, there are two main approaches -design-based and model-based- for making inferences for different characteristics of the population. A design-based approach tends to produce unreliable estimates for small geographical regions or cross classified demographic regions due to the small sample sizes. Moreover, when there are no samples available in those areas, a design-based method cannot be used. In the case of estimating population characteristics for a small area, model-based methods are used. They provide a flexible modeling method that can incorporate relevant information from similar areas and external databases. To provide suitable estimates, many model building techniques, both frequentist and Bayesian, have been developed, and when the model-based method makes an explicit use of prior distributions on the hyperparameters, inference can be carried out in the Bayesian paradigm. For estimating small area proportions, mixed models are often used because of the flexibility in combining information from different sources and of the tractability of error sources. Mixed models are categorized into two broad classes, area-level and unit-level models, and the use of either model

depends on the availability of information.

Generally, estimation of small area proportions with the hierarchical Bayes(HB) method involves transformation of the direct survey-weighted estimates that stabilizes the sampling variance. Additionally, it is commonly assumed that the survey-weighted proportion has a normal distribution with a known sampling variance. We find that these assumptions and application methods may introduce some complications. First, the transformation of direct estimates can introduce bias when they are back transformed for obtaining the original parameter of interest. Second, transformation of direct estimates can cause additional measures of uncertainty. Third, certain commonly used functions for transformation cannot be used, such as log transformation on a zero survey count. Fourth, applying fixed values for sampling variances may fail to capture the additional variability. Last, assumption of the normality of the model distribution might be inappropriate when the true parameter of interest lies near the extremities (near 0 or 1).

To address these complications, we first expand the Fay-Herriot area-level model for estimating proportions that can directly model the survey-weighted proportions without using any transformation functions. Second, we introduce a logit function for the linking model, which is more appropriate for estimating proportions. Third, we model the sampling variance to capture the additional variability. Additionally, we develop a model that can be used for modeling the survey weighted counts directly.

We also explore a new benchmarking approach for the estimates. Estimates are benchmarked when the aggregate of the estimates from the smaller regions matches that of the corresponding larger region. The benchmarking techniques involve a number of constraints. Our approach introduces a simple method that can be applied to complicated constraints when applying a traditional method may fail. Finally, we investigate the “triple-goal” estimation method that can concurrently achieve the three specific goals relatively well as an ensemble.

Hierarchical Bayesian Estimation of Small Area Means Using Complex Survey Data

by

Neung Soo Ha

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2013

Advisory Committee:
Professor Frank Alt
Dr. Robert Fay
Professor Partha Lahiri, Chair/Advisor
Professor Michael Rendall
Professor Paul Smith

Dedication

This dissertation is dedicated to my parents, Tae Woong Ha and Sun Hwan Oh, and my sisters, Jinsoo and Minsoo Ha.

Acknowledgments

I owe my sincere gratitude to all the people who have made this thesis possible. Without their support, it wouldn't have been possible for my pursuing the degree.

First and foremost, I'd like to express my sincere gratitude to my advisor, Professor Partha Lahiri. It has been a great experience and such a privilege to work closely with a researcher of his pedigree. Throughout the years, he has continuously guided me in my research with his wealth of knowledge and insight. I really appreciate the enormous amount of time and patience he gave to guide me in my research, and without his help, it would never be possible to reach the current stage of my career.

I also want to extend my gratitude for other committee members: Dr. Bob Fay, Dr. Francis Alt, Dr. Michael Rendall, and Dr. Paul Smith. The final version of my thesis has improved greatly because of their constructive comments and analysis. In particular, I want to give special thanks to Dr. Smith for his support and guidance when I first started the graduate program and throughout the Ph.D. study. I also thank Dr. Eric Slud, Dr. Abram Kagan, and Dr. Richard Valliant for their valuable teachings.

For my former colleagues at the National Center for Health Statistics, I want to thank them for their guidance and support during my internship. I truly feel that I have gained very valuable experience for working in the research environment outside of school. My special thanks goes to Dr. Van Parsons and and Dr. Don Malec for providing me with recommendations and general guidance in my research.

I shall also thank the members of my two extended families: the Applied Math and Scientific Computing and the Joint Program in Survey Methodology programs. I want to thank Alverda McCoy for her help from the AMSC program. She has helped me countless times and provided many answers to my questions about the administrative related issues. I also would like to thank Gina Hsu for her help with my JPSM department related topics.

I would like to thank my friends throughout my Ph.D. study: Shu Zhang, Ziliang Li, Ritaja Sur and Anastasia Voulgaraki. Thank you so much for everything that we have shared while we were in the program together. Tong Meng, Rongrong Wang, Ran Ji, Jin Yan, Dave Shaw, and Jong Jun Li, thank you for all your help and bringing cheers to my mundane life. I like to give special thanks to Carolina Franco, Doug Galagate and Jiraphan Suntornchost for just being great friends all around. I would not have made it without their support.

I owe my deepest thanks to my family. I am so thankful that my sister, Minsoo, and her husband, Kyonsik Jun, who have accepted me and let me become a part of their family. They have provided me the greatest support that I can ask for, and my nieces brought me the brightest joys. I also want to thank my other sister and her family in Seoul for their uncompromising affection and support.

I dedicate this dissertation to my mother. She has given me the greatest love that I can ask for. Without her, I do not think it would be possible for me to finish the program. I felt at times that her unconditional love for me was the only thing that drove me during my studies.

I also dedicate this dissertation to my father who is not with me to see this. I wish that he was here, but I am sure that he is looking out for me in the sky and feeling proud of me finishing the Ph.D. program. Without his consent and love, it would not have been possible to come to the U.S. to pursue the doctorate program.

I apologize to those that I should show my gratitude to but who are left out inadvertently. My only excuse is that it is nearly impossible to remember everyone who has contributed in my graduate school life.

Lastly, thank you all!

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Uses of small area statistics

In many instances, various institutions, such as government agencies, use sample surveys to obtain information on a wide range of population characteristics, and these sample surveys are used to provide estimates of population as well as sub-population (domains) characteristics. Examples of domains include a geographical region (a state, county, municipality, etc.), a demographic group (specific age × gender × race), or a cross-classification of geographical regions and socio-demographic groups. Domains (or areas) are classified as large if the samples within the domain produce direct estimates with an adequate precision under the sample design. On the other hand, small domains have inadequate sample sizes to produce reliable direct estimates. Some of the terms for denoting small areas include “sub-domains”, ”sub-state”, or ”counties.”

The estimators obtained from sample surveys are called “direct” estimates, where they are typically design-based, and the inferences of the design-based estimates are based on the probability distribution of the survey design (hence they are design-based estimates) with the population values, y, being fixed. The calculation of design-based estimators involves survey weights; that is, an inverse of the inclusion probability of a unit. For large domains, direct estimators produce reliable estimates under the sampling design; however, the same mode of inference produces unreliable estimates for small areas, and thus it is necessary to use “indirect” estimators that “borrow strength” from related areas to increase the information to produce estimates with better precision. Obtaining the indirect estimators is carried out through a model (implicit/explicit) that utilizes a link between supplementary information and direct estimates in small areas.

More recently, there has been a larger demand, especially at the federal agency level, for small area estimators in many different applications. This is due to the growing use in formulating policies and programs to allocate government funds among different geographical areas. Federal agencies make policy decisions, like social and economic programs, by using the surveys, which are designed for large areas, to allocate programs for smaller local levels. For example, the National Agricultural Statistical Services (NASS) publishes model-based county estimates of crop acreage using satellite data. The county estimates assist the local and federal agricultural authorities in decision making, such as allocation of different subsidy programs, Rao (2003).

In the early 1990s, the U.S. Census Bureau established the Small Area Income and Poverty Estimates (SAIPE) program to provide estimates of income and poverty for administrating federal programs and allocating federal funds to local government. The SAIPE program has been especially focused on the poverty rate and count estimates of school-aged (ages 5-17) children for school districts since 1995. The state and county estimates have been produced using a variance of the model, originally suggested by Fay and Herriot (1979), and for a while the Census have used the Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS) to produce estimates. However, the CPS ASEC contains only about 100,000 nationwide addresses, which is not enough sample size to provide accurate estimates for all school districts in the U.S., the

Ceusus Bureau made a decision in 2006 to use the American Community Survey (ACS) to build a model that relates state and county estimates of income and poverty to other indicators. The ACS has a much large sample size (about 3 million addresses nationwide) than the CPS ASEC. Some of the estimates that the SAIPE produces at county and state levels are: total number of people in poverty, number of children under age 18 in poverty, and median household income. For more information about the history of the SAIPE program, see Bell et al. (2007). Citro and Kalton (2000) provided a review for a variety of uses of these estimates from the SAIPE program.

1.2 Direct Estimation

As mentioned before, sample survey data are used to provide reliable direct estimates of totals or means of the population of interest for large domains. There has been an extensive literature on theories on the direct domain estimation under the design-based mode, see Lohr (1999), Cochran (1977), and Sarndal et al. (1992). In this dissertation, we focus on estimates for small domains (areas).

Let U be a population consisting of N distinct units and y_j be a characteristic of interest for a unit $j, (j = 1, \dots, N)$. The parameter of interest could be the population total, $Y = \sum_{j=1}^{N} y_j$, or the population mean, $\bar{Y} = Y/N$, and the corresponding estimators are denoted as $\hat{Y}$ and $\hat{\bar{Y}}$. The sampling design used to select a sample, s, with a probability, $p(s)$, depends on the design scheme of the survey. The examples of design scheme includes simple random sampling (SRS), stratified simple random sampling, or stratified multistage complex sampling methods.

The inference on Y is made from the y_j value associated with each unit $j \in s$. In the design-based approach, the estimator, $\hat{Y}$, is design-unbiased if $E_p(\hat{Y}) = \sum_s p(s)\hat{Y} = Y$, where the summation is over all possible samples under the specified design. The true design variance of $\hat{Y}$ is denoted as $V_p(\hat{Y}) = E_p(\hat{Y} - E_p(\hat{Y}))^2$, and its corresponding unbiased estimator is denoted as $v(\hat{Y})$, which is equivalent to the sample variance of $\hat{Y}$. The bias of the estimator is defined as $Bias_p(\hat{Y}) = E_p(\hat{Y}) - Y$. The design consistency of $\hat{Y}$ is defined if $\hat{Y}$ is approximately design-unbiased and $V_p(\hat{Y})$ tends to zero, Rao (2003). For more rigorous definition of design-consistency, see section 1.3 of Fuller (2009).

We also construct the estimator of the population size by using the survey weights, $w_j(s), j \in s$. The unit survey weight, w_j, is interpreted as a number of units represented by unit j in the population U. The basic weight is the inverse of the inclusion probability π_j, where $\pi_j = \sum_{s:j\in s} p(s)$. In other words, π_j is a probability that a unit j is included in a selected sample s.

In practice, there are usually two additional steps, after considering the inclusion probabilities, in order to obtain the final survey weights . First, weights are usually adjusted for the nonresponse of a unit since not all sampled units respond. Let I_j be the indicator variable for the unit $j \in s$, then we can denote $P(I_j = 1) = \pi_j$, and let R_j be an indicator variable for whether the unit j responds or not, with $P(R_j = 1) = \phi_j$. If we assume non-informative sampling design; that is the probability of selection is independent of the response. Then the adjusted inclusion probability is defined as: $P(\text{unit} j \text{ is selected and responds}) = \pi_j \phi_j$, and the corresponding nonresponse adjusted weight becomes, $w_j = 1/\pi_j \phi_j$.

The second step for adjustment involves poststratification. Generally, a population can be stratified into groups based on different demographics, such as a race or a gender. Let H denote the total number of different groups and N_h be a known population count in a subgroup h. The sum of the weights for each unit $j \in h$, $\sum_{j \in h} w_j$, should estimate N_h, but usually their values do not match. The poststratification method uses the ratio estimator within each group that adjusts the weights to the true population count. Let w_{hj} be a weight for a jth respondent in a subgroup h, then the poststratified adjusted weight, w^*_{hj}, is defined as:

$$w^*_{hj} = w_{hj} \cdot \frac{N_h}{\sum_{j \in h} w_{hj}}.$$

Depending on the specific survey, the weight can be poststratified by using several demographic groups.

Suppose we are interested in estimating area-level means ,$\bar{Y}_i = \sum_{j=1}^{N_i} y_{ij}/N_i, i = 1, \ldots, m$ for m small areas, and let N_i and n_i be the population count and the sample count for the ith area. Let y_{ij} be the response variable of a characteristic of interest for a unit j in the area i. Let w_{ij} be the corresponding weight for the unit j in the area i. Then the usual survey weighted area-level estimator, $\hat{\bar{Y}}_i$, is denoted as:

$$\hat{\bar{Y}}_i = \frac{\sum_{j=1}^{n_i} w_{ij} y_{ij}}{\sum_{j=1}^{n_i} w_{ij}}, i = 1, \ldots, m, \tag{1.1}$$

The corresponding estimate of the variance $v(\hat{\bar{Y}}_i)$ can be obtained through linearization or resampling methods, such as jackknife and balanced repeated replication (BRR).

There are several advantages for design-based estimates. First, they provide reliable and design-consistent inferences in large samples. Second, they incorporate survey design features, such as strata or clusters, to calculate estimators. Third, because the inference does not involve a model assumption and the design-based estimates are distribution free, so model failures need not be considered.

On the other hand, there are a lot of weaknesses. First, if the randomization distribution or the sampling mechanism, is corrupted by nonsampling errors, such as measurement or systematic errors, the method is not applicable, Kalton (2002). In practice, measurement errors can arise if respondents provides an inaccurate answer, the interviewer records the answer incorrectly, or there are additional types of processing error. A seminal paper on survey measurement errors can be found in Hansen et al. (1961).

Second, when the sample sizes are small or non-existent, design-based methods provide limited guidance and produce imprecise estimators due to the sampling design that aims to provide reliable data for large areas and pays little or no attention to the smaller areas. For example, before the change in the SAIPE program, the CPS ASEC typically sampled only about 1,100 counties out of possible 3,141 counties in the U.S., Bell et al. (2007), and for those counties with no samples, estimates were obtained as pure regression predictions. Additionally, sampling variances for the direct estimates were not available due to the small sample sizes. The design-based method would not be applicable to produce either point estimates nor the variance estimates, and alternative approaches have to be considered.

1.3 Model-based estimation method for small areas

Because of the inadequate direct information, a method for an improved estimation calls for implicit or explicit models. In general, they provide a link between small areas to their related information sources, such as various administrative/census records. Once relevant sources of information are identified, a decision is needed for a method of choosing an appropriate method. Rao (2003) provided various indirect methods with implicit models; however, many of them have shortfalls. For example, the synthetic estimation method, see Gonzales (1973), is based on a very restrictive model even though it produces model-unbiased estimators. The method of composite estimators expands the synthetic estimation method by using a weighted average between the direct estimator and the synthetic estimator. The composite estimators aim to balance the potential bias of the synthetic estimators under model failure against the instability of the direct estimators. However, application of the method is challenging because implementing a proper method for determining weights between the two estimators is difficult, see Rao (2003). In general, the formal evaluation of the implicit model can be problematic because its composition is not clearly laid out.

On the other hand, the use of explicit models have several advantages. First, model diagnostics can be used for model fits and comparisons. Second, explicit models, such as linear mixed models or nonlinear mixed models, can be applied to accommodate complex data structures; and third, well-developed methodologies can be used to obtain accurate inferences on parameter estimates. In particular, mixed effects models are suitable for small area estimation because of their flexibility in combining different sources of information and the tractability of different error sources. Furthermore, mixed models are categorized into two broad classes, area-level and unit-level models, and the use of either model depends on the availability of the information.

When estimates from small areas are aggregated, the overall estimates for a larger geographical area may be quite different from the corresponding direct estimate, with the latter considered reliable. This could especially be true if the survey is designed to achieve specified inferential accuracy at the larger geographical regions, and it can be more severe in the event of model failure since applying model validation can be challenging. One method to avoid this problem of discrepancy at the higher level is to use the “benchmarking” approach: it modifies the model-based estimates, such that their aggregate always matches with the corresponding design-based estimate at a higher level, Pfeffermann and Tiller (2006).

1.3.1 Area-level model

A general structure of the area-level model is composed of two models: sampling and linking. The sampling model accounts for the sampling error of the survey weighted direct estimates and possible errors from the survey design. The linking model interconnects between the population characteristic and the known area-specific auxiliary variables. Since the survey-weighted estimates are used in the area-level model, the estimates are design-consistent. The main difficulty in applying the area-level model is that it requires precise estimates of sampling variances of the survey weighted design-based estimates. It is a challenging problem because the sampling variance estimates become unstable due to the small sample sizes in the areas of interest.

A typical example of a basic area-level model is the Fay-Herriot model, Fay and Herriot (1979). Assume that $\hat{\theta}_i = g(\hat{\bar{Y}}_i)$ is an observation with a transformation function g with related auxiliary data $\boldsymbol{x}_i' = (1, x_{i1}, \ldots, x_{ip})'$, and the corresponding transformed parameter of interest becomes $\theta_i = g(\bar{Y}_i)$. Fay and Herriot used $g(\cdot) = \log$ on the observed values and applied the following model to estimate the per capita income (PCI)for small areas in which the area-level population is less than 1000.

$$\begin{array}{lcll} Level1(sampling\ model) & : & \hat{\theta}_i|\theta_i, \psi_i & \overset{\text{ind}}{\sim} (\theta_i, \psi_i), \\ Level2(prior\ model) & : & \theta_i|x_i', \boldsymbol{\beta}, A & \overset{\text{ind}}{\sim} (\boldsymbol{x}_i'\boldsymbol{\beta}, A), \end{array}$$

where level 1 is used to account for the sampling variability of the direct observed estimate, $\hat{\theta}_i = \log(\hat{\bar{Y}}_i)$ of the true small area means $\theta_i = \log(\mu_i)$, where $\mu_i =$ true per capita income. Level 2 links the true parameter θ_i with auxiliary variables $\boldsymbol{x}_i'$. This model is also called a *matched* model because both level 1 and level 2 can be combined into a single linear mixed model:

$$\hat{\theta}_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \nu_i + \epsilon_i, \theta_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \nu_i$$

where $\nu_i \overset{iid}{\sim} N(0, A), \epsilon_i \overset{ind}{\sim} N(0, \psi_i)$. Parameters $\boldsymbol{\beta}$ and A are called hyperparameters and are generally unknown values, but the sampling variance parameter, ψ_i is assumed known.

Numerous methods have been suggested in order to improve the stability of the sampling variances, Fay and Herriot (1979), Otto and Bell (1995), Wolter (1985), and Gershunskaya and Lahiri (2005). The generalized variance function (GVF), Wolter (1985), is a commonly used method; it uses a mathematical function that describes a relationship between the direct estimate and its corresponding variance. The choice of the function is based on the premise that the relative variance is a decreasing function of the magnitude to the expectation of the estimate. For certain groups of items with *similar* intra-class correlation or design effects, the parameters of the model are estimated through using both direct design-based point estimates and corresponding variance estimates. The fitted value of the variance estimate can be then interpreted as a smoothed estimate of the true sampling variance, see Wolter (1985).

For their model, Fay and Herriot used the log-transformation, $\log(\hat{\bar{Y}}_i) = \hat{\theta}_i$, and obtained an empirical relationship: $CV_i = CV(\hat{\bar{Y}}_i) \approx 3/\sqrt{N_i}$, where CV_i is the estimated coefficient of variation of $\hat{\bar{Y}}_i$ with its variance estimated using standard design-based methods, such as linearlization, jackknife, or balanced repeated replicates (BRR). Furthermore, they made a synthetic assumption that the estimated slope of the model remains the same for all areas and concluded that the true sampling variance of $Var(\hat{\bar{Y}}_i) = \psi_i \mu_i^2$

Well known early area-level models with transformation includes: the Fay and Herriot (1979) (FH) model, the Efron and Morris (1975) (EM) model, and the Carter and Rolph (1974) (CR) model. Both the CR and the EM models can be regarded as particular extensions of the FH model. One potential problem with their methods is the possible bias induced from the back-transformation. The validity of the back-transformation relies on the Taylor expansion that relies on the asymptotic of the sample size. Also, the models fail to incorporate variability from using the estimated sampling variance in place of the true sampling variance. There have been some attempts to incorporate uncertainty in sampling variance estimation through alternate modeling: Arora and Lahiri (1997), and Liu et al. (2007). Chen (2001) used properties of a log-normal distribution for obtaining the Bayes and empirical Bayes estimator of μ_i directly.

1.3.2 Unit-level models

In the unit-level model, it is assumed that some unit-specific and area-specific auxiliary data are available, Moura and Holt (1999). The area-specific random effects term in the model captures the presence of possible correlation between different units in a small area. Typically, it does not incorporate sampling weights nor is it able to incorporate all of the survey design information, such as stratification or clusters. Generally, specific survey design information is not available to the public due to privacy issues and complexity of the design. Thus, the unit-level model estimates are not design-consistent. However, there have been efforts to integrate survey weights to produce design-consistent estimators, Prasad and Rao (1999), You and Rao (2002), Jiang and Lahiri (2006a). The general idea is to define a normalized survey-weighted area-level model from the unit-level model. You and Rao (2002) have developed a model that makes a use of the survey weights that preserves the design consistency and self-benchmarking property. Their approach, however, is only applicable for the unit-level linear mixed model. Jiang and Lahiri (2006a) proposed a general model assisted approach for both continuous and binary response variables, and Lahiri and Mukherjee (2007) have proposed models with the use of survey weights in a hierarchical Bayesian frame work.

A simple example of a unit-level mixed model is the nested error regression model in Battese et al. (1988). They used the model to estimate mean area crop (soybean and corn) production for twelve counties in northern central Iowa, and in the model, they have incorporated random county level effect to include the correlation structure. Their model

(BHF model) is defined as:

$$y_{ik} = \boldsymbol{x}'_{ik}\boldsymbol{\beta} + \nu_i + \epsilon_{ik}, k = 1, \ldots, n_i; i = 1, \ldots, m, \tag{1.2}$$

where y_{ik} is the size of soybean/corn production in the kth segment of the ith county, and $\nu_i \overset{iid}{\sim} N(0, \sigma^2_\nu)$, and $\epsilon_{ik} \overset{iid}{\sim} N(0, \psi)$. The parameter of interest can be defined as $\theta_i = \boldsymbol{x}'_i\boldsymbol{\beta} + \nu_i$, where $\boldsymbol{x}'_i = \sum_{k=1}^{N_i} \boldsymbol{x}_{ik}/N_i$.

Moura and Holt (1999) have extended the BHF model in a random coefficient regression setting, in which both the slope and the intercept are random.

1.3.3 Generalized linear mixed models in small area estimation

When modeling discrete response variables, like binary or count, using linear mixed models can be problematic because they are designed for handling continuous variables, so, the method of generalized linear models (GLM) has been developed to manage discrete variables, McCullagh and Nelder (1983). The GLM model inferences are based on the assumption that the responses are independent; however, this could be disputable in real situations because the observations can be correlated. For example, in multi-stage cluster sampling, the responses of the individuals within the same cluster could be correlated due to the similarity in location. This shortfall in the GLM method led to the development of the generalized linear mixed models (GLMM), McCullagh and Nelder (1983), Jiang and Lahiri (2006a).

The GLMM expands from the GLM by incorporating random effects into the linear predictor. This change allows for the correlation between different units into the context of a broad class of non-normally distributed data. In many cases, survey responses are binary or categorical in nature, so the use of the GLMM method has gained more recognition. For example, Malec et al. (1997) considered a logistic regression model with random regression coefficients to estimate the true area-level proportion using the National Health Interview Survey (NHIS). Other applications of the GLMM for estimating small area proportions can be found in Stroud (1991), Malec et al. (1999), Jiang and Lahiri (2001), and Jiang (2007).

1.3.4 Inference Using Mixed Models

There are two primary approaches for making inferences for small area estimates: the empirical Bayes(EB) approach and the hierarchical Bayes(HB) approach. The goal for both approaches is to approximate the posterior distribution of the parameter of interest. The main difference between the two methods is that, in the EB method, the hyperparameters are estimated by using the traditional method, whereas in the HB method, those hyperparameters are given prior distributions with careful considerations.

Efron and Morris (1975) showed how the Bayes estimators arise in the context of Bayes decision theory. Let $y_i|\theta_i \overset{\text{ind}}{\sim} N(\theta_i, 1),\ i = 1, \cdots, m$. Under the sum of squared error loss (SSEL), the frequentist's risk of $\boldsymbol{y} = (y_1, \cdots, y_m)'$, the maximum likelihood estimator of $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_m)'$, is given by

$$R(\theta, y) = \sum_{i=1}^{m} E[(y_i - \theta_i)^2|\theta_i] = m.$$

It is well-known that for $m \geq 3,\ y$ is inadmissible for θ under SSEL and is uniformly (i.e., for all $\boldsymbol{\theta} \in R^m$, the m-dimensional Euclidian space) inferior to the well-celebrated James-Stein estimator $\hat{\boldsymbol{\theta}}_{JS} = (\hat{\theta}_{1,JS}, \cdots, \hat{\theta}_{m,JS})'$, where

$$\hat{\theta}_{i,JS} = \left(1 - \frac{m-2}{||y||^2}\right) y_i,$$

where $||y|| = \sqrt{\sum_{i=1}^m y_i^2}$ is the norm of y. The above surprising admissibility result was first discovered by Stein (1955), and the James-Stein estimator first appeared in James and Stein (1961).

Efron (1975) proved the following interesting inequality:

$$R(\theta, \hat{\theta}_{JS}) \leq m - \frac{(m-2)^2}{m-2+||\theta||^2}.$$

This inequality provides not only an alternative proof of superiority of the James-Stein estimator over y but also an idea about the amount of gain in using the James-Stein estimator over the maximum likelihood estimator. For example, if $\theta_i = 0\ (i = 1, \cdots, m)$, then

$$R(\hat{\theta}_{JS}, \theta) \leq [m - (m-2)] = 2$$

so that the largest reduction is obtained when $\theta_i = 0\ (i = 1, \cdots, m)$ and m large. In practice, the gain will be smaller because some variation in the θ_i's is expected around zero. Efron and Morris (1973) allowed for variation of θ_i by assuming the following prior distribution:

$$\theta_i \overset{\text{ind}}{\sim} N(0, A),\ i = 1, \cdots, m.$$

Under this prior, they provided a parametric empirical Bayes justification of the James-Stein estimator.

This concept of the Bayesian decision theory provides the background for the constrained Bayes and triple-goal estimators, which will be discussed more thoroughly in later chapters. In discussing those estimators, they are obtained not just through maximizing the Bayes risk but through different constraints.

1.3.5 Discussion and Overview of the Dissertation

In this chapter, we have presented a broad overview of the estimation methods in small areas, their advantages and disadvantages and applications in a variety of settings. We have discussed both design-based and model-based approaches for inferences in small areas. We have covered several methods for the variance component estimation and its importance for inference in obtaining reliable estimates and their measure of uncertainty. In particular, we have discussed small area estimation techniques to produce reliable estimates for area-level proportions. In many instances, normality in modeling is commonly assumed and is applied to survey-weighted proportions with known variance components at the sampling level. In addition, normality is again commonly assumed for the random effects of the area-level or unit-level mixed models. However, these assumptions may not be valid in many cases.

In this dissertation, we develop statistical methodologies for estimating small area means by expanding both basic area-level and unit-level models using complex survey data. For inferences, we specifically focus on the hierarchical Bayesian approach and primarily use the Markov Chain Monte Carlo (MCMC) technique for inferences.

This dissertation is organized as follows. In Chapter 2, we consider three area-level hierarchical Bayesian models to estimate small area proportions using complex survey data. We will provide a description about how the Census Bureau handles the SAIPE program by using a log transformation, and then we will illustrate how our model applies the survey-weighted design estimates directly. We use the MCMC techniques for inferences of our parameter estimates. To evaluate the performance of these models, we have used various model fit and model diagnostic techniques.

In Chapter 3, we expand the area-level model by using the non-Gaussian distribution for discrete count data. The difficulty lies in how to model a weighted survey count in a small area because a weighed count is not always integer-valued. We have made an adjustment to the survey weights and the effective sample size to account for the survey design. We have developed the full Bayesian model to a complex survey and obtained parameter estimates through the MCMC. We also compare our results to estimates derived from other well-studied method, such as the EB method.

In Chapters 4 and 5, our interest expands from obtaining point estimates to evaluating an ensemble of estimates that would satisfy multiple criteria. In Chapter 4, we show that, although posterior estimates reduce variances individually, they show over-shrinkage as an ensemble. We consider a method of finding a new set of Bayes estimators with specific and known benchmarking constraints, with minimizing a specified distance function with a given set of estimators.

In Chapter 5, we explore an ensemble of estimates that satisfy three criteria: rank, empirical histogram, and point estimate. This ensemble of estimates is called triple-goal estimates, Shen and Louis (1998). We expand the triple-goal estimates procedure by first applying a transformation to the observed data. We then apply additional constraints from Chapter 4 such that our new estimators satisfy the benchmarking property.

In Chapter 6, we provide a summary of this dissertation and give directions for future research.

Chapter 2

Hierarchical Bayes Estimation of Small Area Proportions

2.1 Introduction

Let θ_i be the true proportion for the ith small area ($i = 1, \cdots, m$). As noted in Chapter 1, the survey-weighted direct estimate, $\hat{\theta}_i$, is highly unreliable for small areas. In order to improve on the survey-weighted small area proportions, different model-based methods, which combine survey data with related administrative and census records, have been proposed in the literature. In many applications, a transformation function on survey-weighted proportions or counts is used. For example, the U.S. Census Bureau uses logarithmic transformation on the survey-weighted counts or proportions in estimating poverty rates for the U.S. counties, Bell et al. (2007). The log scale of regression variables are used at the prior model level, and they come from many different administrative sources, such as Internal Revenue Service (IRS) tax information, food stamp programs by the USDA Food and Nutrition Service, and Population Estimates Program (PEP) by the Census Bureau. The method for obtaining estimates of small area proportions is straightforward. First, an empirical Bayes estimator of transformed true small area proportion is found and then a simple back-transformation is taken to obtain an estimate of the true small area proportion θ_i.

In many instances, there are still some counties with zero estimate school-age children in poverty even with using the ACS. For example, in the 2005 ACS, 169 (about 5 percent of the 3,141 total counties) had zero counts, Bell et al. (2007), and those counties are excluded from the regression prediction because of the log of zero cannot be taken; instead, the pure regression predictions are used for their area-level estimates.

There are several shortcomings of the Census Bureau's methodology for producing county poverty rate estimates. First, inconsistencies may come from excluding counties with zero survey-weighted counts from the model prediction. Second, although the survey-weighted count is unbiased or approximately unbiased, the logarithmic transformed survey-weighted count is likely to be biased. Third, the optimality property of the empirical Bayesian method is lost when taking the back-transformation. Fourth, the usual method of finding the mean squared error of the back-transformed empirical Bayes estimator is not second-order unbiased.

This chapter is purely application-oriented. In section 2.2, we discuss modeling and different related inferential issues in obtaining improved estimates of smoking prevalances for the U.S. states. In sections 2.3-2.5, we discuss the databases used for estimation, model selection, and estimation, respectively.

2.2 Extensions of the Fay-Herriot Model and the Related Inferences

To estimate the smoking prevalences for the U.S. states, we explore the following two models:

The Normal-Logistic Model:(NL)

For $i = 1, \ldots, m$,

$$\begin{aligned} &\text{Level 1 } (\mathit{Sampling\ model}) &:& \quad \hat{\theta}_i|\theta_i \overset{\text{ind}}{\sim} \mathcal{N}(\theta_i, \psi_i), \\ &\text{Level 2 } (\mathit{Linking\ model}) &:& \quad \text{logit}(\theta_i)|\boldsymbol{\beta}, A \overset{\text{ind}}{\sim} \mathcal{N}(\mathbf{x}_i'\boldsymbol{\beta}, A). \end{aligned} \tag{2.1}$$

The Normal-Logistic Random Sampling Variance Model:(NL$_{rs}$)

For $i = 1, \ldots, m$,

$$\begin{aligned} &\text{Level 1 } (\mathit{Sampling\ model}) &:& \quad \hat{\theta}_i|\theta_i \overset{\text{ind}}{\sim} \mathcal{N}\Big(\theta_i, \psi_i = \frac{\theta_i(1-\theta_i)}{n_i}\text{DEFF}_i\Big), \\ &\text{Level 2 } (\mathit{Linking\ model}) &:& \quad \text{logit}(\theta_i)|\boldsymbol{\beta}, A \overset{\text{ind}}{\sim} \mathcal{N}(\mathbf{x}_i'\boldsymbol{\beta}, A), \end{aligned} \tag{2.2}$$

where DEFF$_i$, the true design effect, is the ratio of the true sampling variance of the survey-weighted proportion under the complex design to the true sampling variance of the un-weighted sample proportion under simple random sampling of size n_i.

Note that the sampling variances ψ_i are not estimable from the area-level data $\{(\hat{\theta}_i, \mathbf{x}_i),\ i = 1, \cdots, m\}$. For standard implementation of the FH and NL models, the sampling variances ψ_i are estimated using additional design information (within small areas), but errors due to estimation of sampling variances are generally ignored in the subsequent inferences. This is a well-known deficiency of the area-level model compared to a unit-level model. In spite of this deficiency, area-level models are widely used in practice since it is usually easier to model aggregate statistics than individual observations, and area level modeling offers a natural way to incorporate survey weights and other survey design properties into the hierarchical model.

The proposed NL_{rs} model in order to partially rectify the problem associated with FH and NL models mentioned in the last paragraph. Although NL_{rs} captures a part of uncertainty due to the estimation of small area sampling variances, the design effects,

DEFF_i, still need to be estimated using the design information. Note that inferences from this model would not incorporate the error due to the estimation of DEFF_i. The idea of separating various design features from the sampling variances of survey statistics can be traced back to Arora and Lahiri (1997) who used a hierarchical random sampling variance model to estimate average expenditure on certain items for small geographical areas using the Consumer Expenditure survey.

2.2.1 Estimation of ψ_i and DEFF$_i$

To use the NL_{rs} model, we need reliable estimates of DEFF$_i$. In survey data, some low population areas have few sampled clusters, thus their design-based estimates of DEFF$_i$ are subject to both biases and instabilities. In this paper, we estimate DEFF$_i$ by $\widehat{\text{deff}}_j^{rgn}$, a design-based estimator of the design effect for the jth larger geographical region, DEFF$_j$, in which the ith small area is located. These estimators are expected to be less variable than the corresponding direct estimators of DEFF$_i$, being based on larger number of sampled clusters. In proposing $\widehat{\text{deff}}_j^{rgn}$, we are implicitly assuming that the true design effects for all the states in a given region are similar.

To implement NL model, we need reliable estimates of the sampling variances ψ_i. The design-based estimates of ψ_i, being based on a few sampled clusters, are generally unreliable and so we do not recommend the use of design-based estimates of ψ_i in the NL model. In an effort to reduce the variability of the sampling variance estimators, we first note

$$\psi_i = \frac{\theta_i(1-\theta_i)}{n_i}\text{DEFF}_i,$$

and then use synthetic estimates of θ_i and DEFF$_i$. As before, we estimate DEFF$_i$ by $\widehat{\text{deff}}_j^{rgn}$, and θ_i by $\hat{\theta}_j^{rgn}$, i.e. the direct survey-weighted proportion estimates for the larger region j in which the small area i lies. We denote

$$\hat{\psi}_i = \frac{\hat{\theta}_j^{rgn}(1 - \hat{\theta}_j^{rgn})}{n_i} \cdot \widehat{\text{deff}}_j^{rgn}, \tag{2.3}$$

to estimate ψ_i in the NL$_1$ model and call the model NL_1. For DEFF$_i = 1$, Carter and Rolph (1974) and Morris (1983) used models similar to NL_1. On the other hand, if we use

$$\hat{\psi}_i = \frac{\hat{\theta}_i^{synth} \cdot (1 - \hat{\theta}_i^{synth})}{n_i} \cdot \widehat{\text{deff}}_j^{rgn}, \tag{2.4}$$

to estimate ψ_i in the NL model, we call the model NL_2. Mohadjer et al. (2007) used a similar model to estimate proportions at the lowest level of literacy for states and counties. Unlike the model NL, the model NL_{rs} is capable of accounting for uncertainty in estimating the sampling variances. As a result, NL_{rs} is expected to reflect more control over the variability than the former model.

2.2.2 Inference on smoking prevalences for small areas

We consider a hierarchical Bayesian (HB) approach to make inferences about the smoking prevalence for the states. We choose weakly informative prior distributions for $\sqrt{A}$ (uniform in a finite interval with large length) and $\boldsymbol{\beta}$ (normal distributions with wide variances), Gelman (2006). All of our data analysis is carried out by the Markov Chain

Monte Carlo (MCMC) method using the well-known WinBUGS package (`http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml`). The readers are referred to Gelman et al. (2004) and Robert and Casella (2004) for details about stochastic sampling methods, such as Metropolis Hastings and Gibbs sampling methods.

2.2.3 The review of the MCMC method

Let $\boldsymbol{\eta} = (\boldsymbol{\theta}, \boldsymbol{\lambda})$ be a vector with parameters of interest $\boldsymbol{\theta}$ and hyperparameters $\boldsymbol{\lambda}$. When obtaining an explicit form for $f(\boldsymbol{\eta}|\text{data})$, the distribution of $\boldsymbol{\eta}$, the Markov chain $(\boldsymbol{\eta}^{\ell}, \ell = 0, 1, 2, \dots,)$ is used. With a starting point $\boldsymbol{\eta}^{(0)}$, the distribution of $\boldsymbol{\eta}^{\ell}$ converges to a unique stationary distribution $\pi(\boldsymbol{\eta})$ that is equivalent to $f(\boldsymbol{\eta}|\text{data})$ after discarding a sufficiently large "burn-in" initial Markov chain samples. To reduce the autocorrelation between the samples, a thinning method is involved, that is to keep every kth draw from each sequence. After the d "burn-in" samples, we treat $\eta^{d+1}, \dots, \eta^{d+T}$ as "independent" samples from the target distribution $f(\boldsymbol{\eta}|\text{data})$. By applying ergodic theory, see Robert and Casella (2004), the average and the variance of the the sample, $(\eta^{d+1}, \dots, \eta^{d+T})$, can be used to approximate the posterior mean, $E(\eta|\text{data})$, and posterior variance, $V(\eta|\text{data})$.

2.2.4 Full conditional distributions for the HB models

Assume that the prior distributions for the model parameters $\boldsymbol{\beta}$ and A are $f(\boldsymbol{\beta}) \propto 1, A \sim Unif(0, L)$. Let $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \dots, \hat{\theta}_m)'$ and $B_i = \frac{\psi_i}{\psi_i + A}$.

The full conditional distribution for NL_1 is given as:

$$\theta_i|\boldsymbol{\beta}, A, \boldsymbol{\theta} \propto \frac{1}{\theta_i(1-\theta_i)\sqrt{A}\sqrt{\psi_i}} \exp\left(-\frac{(\hat{\theta}_i - \theta_i)^2}{2\psi_i} - \frac{(\text{logit}(\theta_i) - \boldsymbol{x}_i'\boldsymbol{\beta})}{2A}\right), \text{ for } \theta_i \in (0,1)$$

$$\boldsymbol{\beta}|\theta_i, A, \hat{\boldsymbol{\theta}} \sim N\left(\sum_{i=1}^{m} (\boldsymbol{x}_i\boldsymbol{x}_i')^{-1} \left(\sum_{i=1}^{m} \boldsymbol{x}_i \text{logit}(\theta_i)\right), A\,(\boldsymbol{x}_i\boldsymbol{x}_i')^{-1}\right), \text{ for } \boldsymbol{\beta} \in \boldsymbol{R}^p$$

$$A|\boldsymbol{\beta}, \theta_i, \hat{\boldsymbol{\theta}} \sim \begin{cases} ING\left(\frac{1}{2}m - 1, \frac{1}{2}\sum_{i=1}^{m}(\text{logit}(\theta_i) - \boldsymbol{x}_i'\boldsymbol{\beta})^2\right) & A \in (0, L) \\ 0 & A \geq L, \end{cases}$$

The full conditional distributions for NL$_{\text{rs}}$ are the same as those of NL except that ψ_i is replaced by $\frac{\theta_i(1-\theta_i)}{n_i}\widehat{deff}_j^{rgn}$ for the distribution of θ_i given other parameters.

Since the full conditional distribution $f(\theta_i|\boldsymbol{\beta}, A, \boldsymbol{\theta})$ is not in explicit form, we need to use the Metropolis-Hastings algorithm for that step. First, we let $\tilde{\theta}_i = \frac{\theta_i}{1-\theta_i}$, then the density of $(\theta_i|\boldsymbol{\beta}, A, \boldsymbol{\theta})$ can be written as:

$$\pi(\tilde{\theta}_i|\boldsymbol{\beta}, A, \boldsymbol{\theta}) \propto h(\tilde{\theta}_i) f(\tilde{\theta}_i|\boldsymbol{\beta}, A),$$

where $f(\tilde{\theta}_i|\boldsymbol{\beta}, A)$ is a log-normal density function given as:

$$f(\tilde{\theta}_i|\boldsymbol{\beta}, A) \propto \frac{1}{\tilde{\theta}_i} exp\left(\frac{(log(\tilde{\theta}_i) - \boldsymbol{x}_i'\boldsymbol{\beta})^2}{2A}\right),$$

and $h(\tilde{\theta}_i)$ is a function given by

$$h(\tilde{\theta}_i) = (1 + \tilde{\theta})^2 exp\left(\frac{\hat{\theta}_i - \frac{\tilde{\theta}_i}{1+\tilde{\theta}_i}}{2\psi_i}\right)$$

We use $f(\tilde{\theta}_i|\boldsymbol{\beta}, A)$ as the "candidate" generating density function in the Metropolis-

Hastings updating step.

To update $\tilde{\theta}_i$, we draw a candidate $\tilde{\theta}_i^{(k+1)}$ from the log-normal density $f(\tilde{\theta}_i|\boldsymbol{\beta}, A)$ with accepting probability α, where α is defined as

$$\alpha(\tilde{\theta}_i^{(k)}, \tilde{\theta}_i^{(k+1)}) = \min\left(h(\tilde{\theta}_i^{(k+1)})/h(\tilde{\theta}_i^{(k)}), 1\right).$$

Drawing samples from $\boldsymbol{\beta}|\tilde{\theta}, A, \hat{\theta}$ and $A|\boldsymbol{\beta}, \tilde{\theta}, \hat{\theta}$ are straightforward since full-conditional distributions have explicit forms.

For implementing the MCMC, we have followed the guidelines given in Gelman et al. (2004) and Cowles and Carlin (1996). We carefully examine the auto-correlation function (ACF), Albert (2007), and trace plots within each chain for all parameters of a given model. We use the Gelman-Rubin potential scale reduction factor and the Geweke diagnostic criterion to determine the number of iterations and burn-in for each chain. In the MCMC, we consider three parallel chains, each with 2.0×10^6 iterations, a burn-in of 7.5×10^5, and thinning at 500 iterations. The maximum of potential scale reduction factor values are displayed in Table 2.1 for all the models they are very close to 1, and so there is no indication that the chains have not converged according to this criterion. The Geweke diagnostic criterion essentially compares means of the first 10 percent against the last 50 percent of the MCMC samples for each chain. The z-scores for hyperparameters from all three chains of each model are given in Table 2.2.

	NL_1	NL_2	NL_{rs}
$\hat{R}$	1.011	1.013	1.006

Table 2.1: Max value of Gelman-Rubin statistic

NL$_1$	β_0	β_1	β_2	β_3	β_4	A
chain 1	-0.55396	1.09104	0.08498	-0.38427	0.75539	0.91984
chain 2	0.5024	1.2851	0.8709	-1.0219	-0.7413	-1.0701
chain 3	-0.55396	1.09104	0.08498	-0.38427	0.75539	0.91984

NL$_2$	β_0	β_1	β_2	β_3	β_4	A
chain 1	0.1602	-1.6886	-0.9258	1.2218	-0.1350	2.0000
chain 2	-1.2510	0.5931	0.8681	-0.5818	1.0890	0.4111
chain 3	-0.0770	-0.0049	0.3157	-0.5080	0.1697	-0.5393

NL$_{rs}$	β_0	β_1	β_2	β_3	β_4	A
chain 1	0.9492	-1.0812	-0.7199	0.3203	-0.6934	-0.4429
chain 2	-1.0532	0.1153	0.0478	0.7524	0.7808	0.0295
chain 3	-0.7625	0.6837	-0.6708	1.0699	0.5077	-0.4421

Table 2.2: Geweke diagnostic: z scores from each chain for all models

2.3 Data Source Description

2.3.1 The National Health and Interview Survey

The National Health Interview Survey, conducted by the U.S. Census Bureau for the National Center for Health Statistics (NCHS), is an annual survey with a state-stratified multi-stage complex design. It includes geographical identifiers, demographic and health related variables. The survey is designed to produce reliable survey-weighted direct estimates and associated design-based standard errors for the nation and four major census regions: (Region 1: Northeast; Region 2: Midwest; Region 3: South; and Region 4: West).

The survey provides data for our study variable, a binary current smoking status for survey respondents. We have analyzed the current smoking status of persons and divided them into two groups: smoking and not smoking. Current smokers were defined as those who had smoked at least 100 cigarettes in their life time, and at the time of interview, reported smoking everyday or some days.

The survey design includes primary sampling units (PSUs) which are individual counties or contiguous groups of counties, and they are sampled without replacement with a probability proportional to their estimated sizes. Within each PSU, the sampling frame is further divides into sub-strata and clusters. We have used in-house data, and so we are able to include survey design features unavailable to the public. In particular, clusters of blocks of housing units are available for variance estimation.

In addition to detailed survey design variables, the data include various geog For more details about the survey, see (http://www.cdc.gov/nchs/nhis/quest_data_related_1997_forward.htm#2008_NHIS).

2.3.2 The American Community Survey

For the area-specific auxiliary variables used in the hierarchical models, we have used data from the 2008 American Community Survey (ACS). The ACS essentially replaces the long form of the decennial census and is the largest household sample survey that the Census Bureau administers. Each year the ACS collects data on various geographic and demographic, socio-economic variables with about 3 million addresses. The ACS contains information similar to that of the Census long form, and it also enables the data users to find various quantities in different geographical groupings. For more information about the ACS, see (http://www.census.gov/acs/www/about_the_survey/american_community_survey/).

Since the auxiliary variables are derived from the ACS, they are subject to sampling errors, an aspect we have not addressed in the paper. But, since the state-wide sample sizes are large, one could ignore sampling variances for such estimates as a good approximation.

2.4 Model Selection

In section 2.3, we have noted that the NL_{rs} model, unlike NL_1 or NL_2, offers a partial remedy to account for uncertainty in estimating sampling variances and thus should provide more accurate credible intervals for smoking prevalence for the states. However, it is instructive to compare these three models using different statistical tools, given the same set of auxiliary variables. In subsection 2.4.1, we discussed selection of a reasonable set of auxiliary variables needed to implement the three models. In subsection 2.4.2, we will compare standard model fit and model selection statistics for the three hierarchical models. Readers are referred to Lahiri (2001) for various model selection methods available for hierarchical models. In subsection 2.4.3, we compare estimates from the three models with the corresponding design-based estimates.

2.4.1 Selection of auxiliary variables for Level 2

There are several state specific auxiliary variables available from the ACS. In order to select a common set of reasonable auxiliary variables for Level 2, we apply standard regression model selection techniques with logit$(\hat{\theta})$ as the dependent variable using data from the 15 largest states. In this preliminary data analysis, we implicitly assume that for these 15 largest states, the sampling variances of survey-weighted estimates are small, so we use standard regression model selection tools. See Jiang et al. (2001) for a similar data analysis.

We select the following state specific auxiliary variables: percent of minority population, poverty rate, percent of population without high school diploma, percent of population age 65 and above. The auxiliary variable poverty percentage was marginally non-significant at 0.1 with a p-value of 0.103. However, including this auxiliary variable results in an increase of the adjusted-R squared value from 0.65 to 0.71 and thus this auxiliary variable is included in the subsequent data analysis.

2.4.2 Comparison of different model fit and model selection statistics for the three models

We first check the Bayesian $p-$values for the three models. Gelman et al. (2004), Datta et al. (1999) and Rao (2003)) discuss the Bayesian $p-$value for checking the adequacy of hierarchical models. The main idea is that if the model fits, then replicated data generated under the model should be similar to observed data. That is, the observed data should look plausible under the posterior predictive distribution. Thus, the technique for checking the model fit is to draw simulated samples from the posterior predictive distribution and compare these to the observed data.

Let y_{obs} denote observed data and y_{new} be predicted data from a distribution, $f(y|\theta)$.

Let functions $f(d(y_{obs},\theta)|y_{obs})$ and $f(d(y_{new},\theta)|y_{obs})$ be the posterior (predictive) distributions of $d(y_{obs},\theta)$, and $d(y_{new},\theta)$, where $d(y,\theta)$ is a χ^2-type discrepancy measure defined as:

$$d(y,\theta) = \sum_{i=1}^{50} (\sigma_i^2)^{-1}(y_i - \theta_i)^2.$$

The parameter σ_i^2 is the true sampling variance for area i. For models NL_1 and NL_2, we replace σ_i^2 by $\hat{\psi}_i$ as given in subsection 2.2.1 For model NL_{rs}, we use $\sigma_i^2 = \frac{\theta_i(1-\theta_i)}{n_i}\widehat{\text{deff}}_i^{rgn}$.

We generate parameters, $\theta^{(\ell)}$, from the posterior distribution, $f(\theta|y_{obs})$, and new data $y^{(\ell)}$ from $f(y|\theta^{(\ell)})$, $\ell = 1, \ldots, B$, where $B(= 7,500)$ is the total number of iterations. Then, we generate two sets of samples, $d(y_{obs}, \theta^{(\ell)})$ and $d(y^{(\ell)}, \theta^{(\ell)})$. These are used to approximate the Bayesian p-value $P\{d(y_{new},\theta) \geq d(y_{obs},\theta)|y_{obs}\}$ by

$$p_B \approx B^{-1} \sum_{\ell=1}^{B} \mathcal{I}\{d(y^{(\ell)}, \theta^{(\ell)}) \geq d(y_{obs}, \theta^{(\ell)})\}, \tag{2.5}$$

where $\mathcal{I}(\cdot)$ is an indicator function.

An extreme value (near 0 or 1) of the Bayesian p-value approximated in (2.5) indicates lack of fit of a given model; whereas for an adequate model, this measure will be close to 0.5. According to this criterion (Bayesian p-values are reported in Table 2.4), none of the three models shows a clear lack-of-fit.

Next we compare three models using the deviance information criterion (DIC) criterion suggested by Spiegelhalter et al. (1998) The DIC is based on the posterior distribution of the deviance statistic,

$$D(\boldsymbol{\theta}) = -2\log f(\boldsymbol{y}|\boldsymbol{\theta}) + 2\log h(\boldsymbol{y}),$$

where $f(\boldsymbol{y}|\boldsymbol{\theta})$ is the likelihood function for the observed data, y_{obs}, given the parameter $\boldsymbol{\theta}$, and $h(\boldsymbol{y})$ is a standardized function of the data alone. The fit of the model is then expressed by the posterior expectation of the deviance, $\bar{D} = E_{\theta|y}(D)$, and the complexity the model is captured by the effective number of parameters p_D,

$$p_D = E_{\theta|y}(D) - D(E_{\theta|y}(\boldsymbol{\theta})) = \bar{D} - D(\bar{\boldsymbol{\theta}}).$$

Then, the deviance information criterion (DIC) is defined as:

$$DIC = \bar{D} + p_D = 2\bar{D} - D(\bar{\boldsymbol{\theta}}).$$

The smaller the values of DIC, the better the model. Table 2.3 displays the DICs values for the three models. The models NL$_1$ and NL$_{rs}$ are very similar and only marginally better than NL$_2$, according to this criterion.

Next we compare the three models using the posterior predictive divergence approach given in Laud and Ibrahim Laud et al. (1995). We approximate the Laud-Ibrahim divergence measure $d(y_{new}, y_{obs}) = E(n^{-1}\|y_{new} - y_{obs}\|^2|y_{obs})$ by $(nB)^{-1}\sum_{\ell}^{B}|y^{(\ell)} - y_{obs}|^2$, where n is the dimension of y_{obs}. The smaller the divergence measure, the better the model. Table 2.3 displays the Laud-Ibrahim divergence measures for the three models. The models NL$_{rs}$ and NL$_1$ are very similar and only marginally better than NL$_2$ according to this criterion.

models	Bayes p-value	DIC	L-I disc measure
NL$_1$	0.4305	-172.69	0.0042
NL$_2$	0.382	-170.81	0.0048
NL$_{rs}$	0.4058	-172.134	0.0043

Table 2.3: Bayesian p-values and Laud-Ibrahim discrepancy measures for different models.

Thus, by using the standard techniques of model fitting and comparison, it is hard to discern the difference between the three models. However, one major advantage of using NL$_{rs}$ over NL$_1$ or NL$_2$ is that it provides an idea of variability of the sampling variance estimates. To illustrate this point we plot the sampling variance estimates in Figure 2.1 for models NL$_1$ and NL$_{rs}$ along with the credible intervals associated with model NL$_{rs}$. On the x-axis, states are ordered by the number of PSUs within each state. These PSUs are clusters defined explicitly for in-house variance estimation. For confidentiality reasons, actual state names are omitted. This rule will be applied for all Figures that show state level information. The graph shows that variability in estimating sampling variances could be substantial for the smaller states. Thus, the model NL$_{rs}$ will provide a better measure of interval estimates of the small area proportions θ_i compared to the other two models.

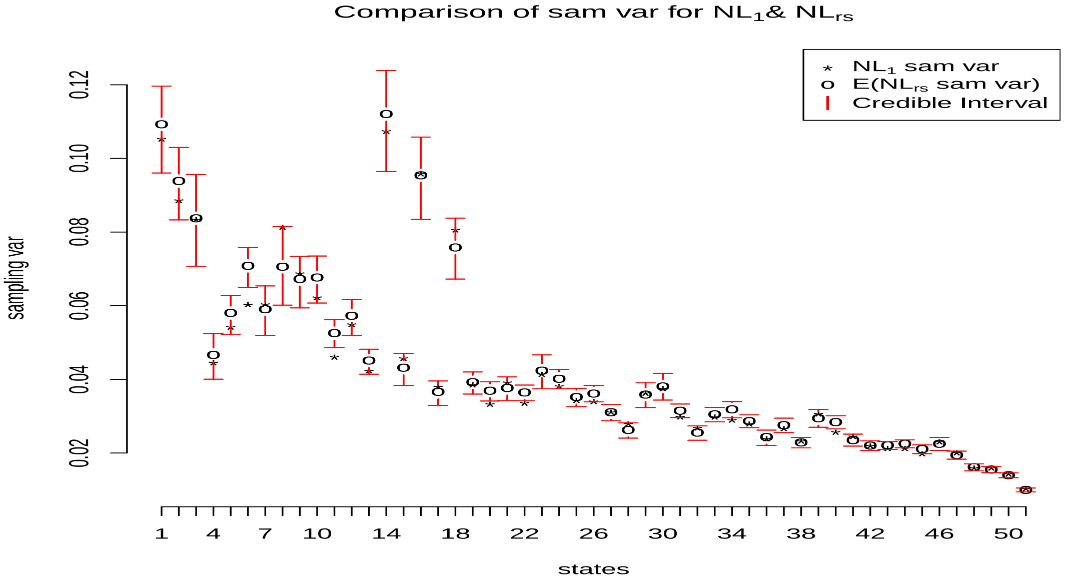


Figure 2.1: Coverage for model NL$_{rs}$

2.4.3 Comparison of different model based estimates with the corresponding design-based estimates

In Figure 2.2, we plot the direct and the hierarchical Bayes estimates of smoking prevalence against states ordered by the number of PSUs. The national smoking average in 2008 is represented by the solid line at 20.6%.

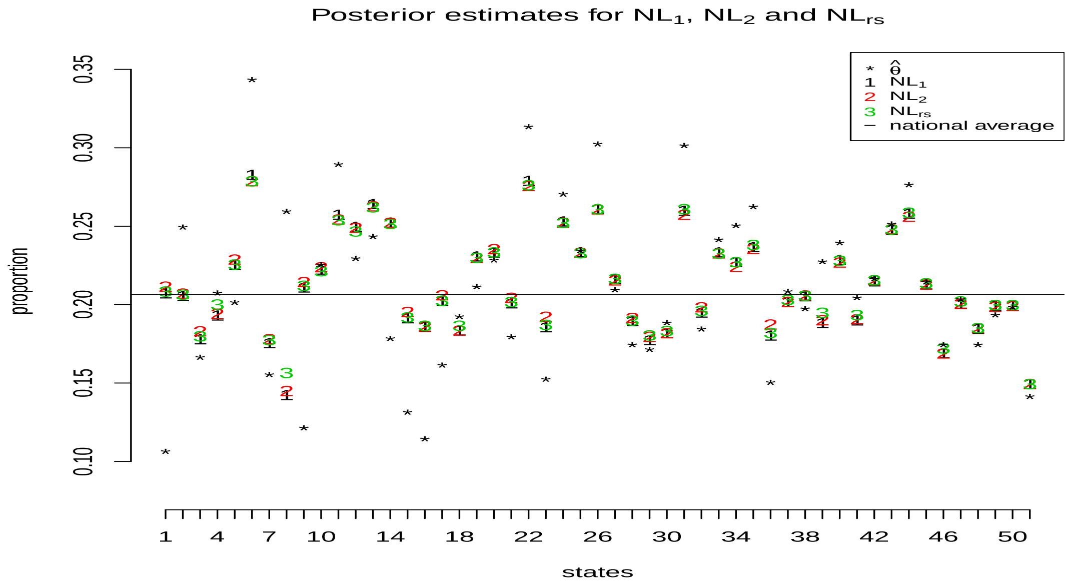


Figure 2.2: Estimates by states

The HB estimates, like the direct estimates, randomly fluctuate around the national estimates and do not show any systematic pattern to indicate possible bias from the modeling. The direct estimates appear to be more variable around the national estimates than the corresponding HB estimates, especially for the group of states with smaller sample size.

The variability among the direct estimates reduces as we move from the left to right side of the graph. For the largest states, direct estimates are very similar to the two hierarchical Bayes estimates. As shown in Figure 2.2, we see the pattern that among smaller states, model-based estimates of both methods are pulled towards the national average.

Since direct design-based estimates at higher levels of aggregations with large sample sizes can be considered accurate, relative differences between design-based and different model-based estimates at higher level can provide a way to compare different models. Since the NHIS was designed to produce reliable estimates at the census regional level, we compare direct estimates with those derived from three models at the census regional level. To be specific, in Table 2.4, we examine the following relative errors (RE) for the four census regions :

$$\text{RE}_j = \left| \frac{\sum_{i \in j} \hat{w}_i \hat{\theta}_i^{ps} - \hat{\theta}_{j,region}}{\hat{\theta}_{j,region}} \right|, j = 1, \ldots, 4,$$

where $\hat{\theta}_{j,region}$ is the census regional design-based direct estimate, $\hat{w}_i$ is the survey weighted population estimate for state i, and $\hat{\theta}_i^{ps}$ is the posterior mean from the area-level HB model at state i. The smaller the value of RB, the better the model is under this assessment criteria. We note that, in Census Region 4, all methods performed well. Overall, the model NL$_{rs}$ performs the best. The models NL$_1$ and NL$_2$ perform equally well; NL$_1$ was slightly better in regions 3 and 4 whereas NL$_2$ was better in regions 1 and 2.

Cen. Rgn.		NL$_1$	NL$_2$	NL$_{rs}$
Rgn 1:	NE	0.0329	0.0307	0.0268
Rgn 2:	MW	0.0372	0.0352	0.0335
Rgn 3:	S	0.0542	0.0554	0.0524
Rgn 4:	W	0.0019	0.0024	0.0010

Table 2.4: Relative errors for three models

2.5 Estimation

From the discussion of subsection 2.4.2, it is clear the model NL$_{rs}$ should provide the most sensible analysis among all the three models considered because it includes the variability at the sampling level in the model, so in this section we compare HB estimates from model NL$_{rs}$ with direct estimates. In Figure 4, we plot the direct and HB estimates along with the credible intervals for states arranged by the number of PSUs. In all but five small states, direct estimates are within the posterior credible intervals, indicating a lack of evidence of possible bias in model NL$_{rs}$.

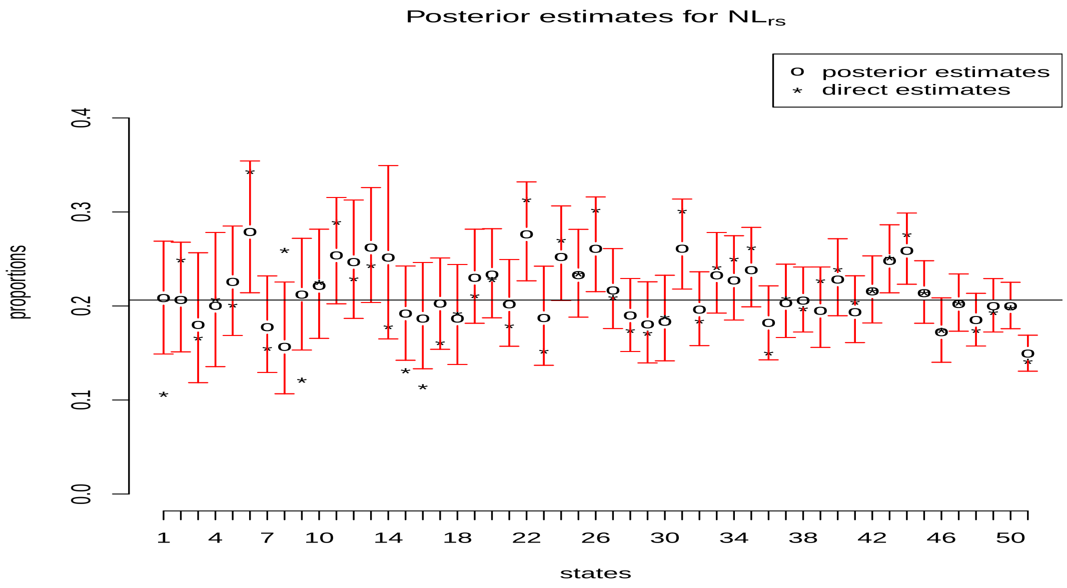


Figure 2.3: Coverage for model NL$_{rs}$

In Figure 2.3, we plot posterior standard errors and the sampling standard errors against states ordered by the number of PSUs. The HB and direct estimates perform very well for the large states. However, the HB estimates display improvement over the direct estimates for the small states by having narrower standard errors.

2.6 Summary

This study provides evidence that hierarchical Bayesian methodology can show improvement over direct design-based estimates; this was illustrated by estimating the smoking prevalence for the U.S. states. We carefully examined different area level hierarchical models and found that the normal-logistic random sampling model has more success in providing better sensible data analysis among the models that were considered. In the future, we would like to extend the hierarchical Bayesian methodology proposed in this paper to estimate smoking prevalence for sub-state areas.

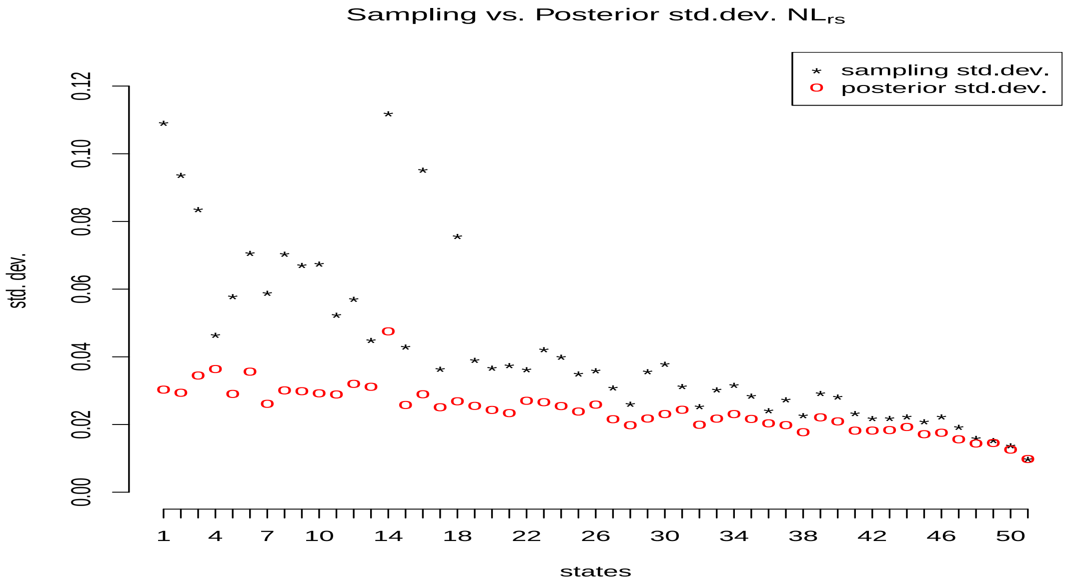


Figure 2.4: NL$_{rs}$Comparison of std dev between two levels

Chapter 3

Hierarchical Bayes Estimation of Poverty Rate

3.1 Introduction

In Chapter 2, we assumed normality for the sampling distribution of the survey-weighted small area proportions and argued that such a model has a number of advantages over a similar model-based on transformed survey-weighted proportions. However, the normality assumption can still be problematic when the true proportions of interest lies near the boundaries (0 or 1). Thus, for cases where the normality assumption is not tenable, it calls for more flexible models.

Dempster and Toberlin (1980) proposed an empirical Bayes method to estimate the Census undercount for local areas using the a logistic regression model. MacGibbon and Tomberlin (1989) considered the following hierarchical logistic regression model:

$$\begin{aligned} y_{ij}|\pi_{ij} &\overset{ind}{\sim} Bernoulli(\pi_{ij}), \\ \text{logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) &= \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \nu_i, \\ \nu_i &\overset{ind}{\sim} N(0, \sigma_\nu^2), \end{aligned} \tag{3.1}$$

where y_{ij} is a binary response variable; $\pi_{ij} = P(y_{ij} = 1|\pi_{ij})$; $\boldsymbol{x}'_{ij}$ is a vector of covariates for unit j in area i, $j = 1, \ldots N_i,\ i = 1, \ldots m$.

The parameter of interest is the true area proportion $P_i = \sum_{j=1}^{N_i} y_{ij}/N_i$. An estimator of P_i is given by $\hat{P}_i = \sum_{j=1}^{N_i} \hat{\pi}_{ij}/N_i$, where $\hat{\pi}_{ij}$ is obtained from (3.1) using either an empirical Bayes (EB) or a hierarchical Bayes (HB) method. Applications of similar models can be found in Wong and Mason (1985) and Tomberlin (1988).

Malec et al. (1997) considered the following random regression coefficient model for binary data. Their model is defined as

$$\begin{aligned}
y_{ij}|\pi_{ij} &\overset{ind}{\sim} Binomial(\pi_{ij}, n_{ij}) \\
\text{logit}(\pi_{ij}) &= \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \boldsymbol{x}_j'\boldsymbol{\beta}; \\
\boldsymbol{\beta}_i &= \boldsymbol{Z}_i\alpha + \boldsymbol{v}_i; \\
\boldsymbol{v}_i &\overset{iid}{\sim} N(0, \boldsymbol{\Sigma}_v),
\end{aligned} \tag{3.2}$$

where y_{ij} is the total number of individuals with the characteristic in class j in the area i with a common probability π_{ij}; $\boldsymbol{x}_j$ is a class-specific covariate vector; $\boldsymbol{Z}_i$ is a $p \times q$ area level covariate matrix, $j = 1, \ldots N_i$, $i = 1, \ldots m$. By applying a hierarchical Bayes (HB) method, they obtained an estimator for the finite population proportion $P = \sum_{i=1}^{m} \sum_{j}^{N_i} y_{ij} / \sum_{i \in I} N_i$ by aggregating up the model-based estimate, $\hat{\pi}_{ij}$. However, this method does not incorporate survey specific features, such as survey weights.

For simple random sampling, unweighted counts are typically modeled by a binomial distribution, but binomial distribution may not fit well for weighted counts, for complex sampling. In section 3.2, we explore a few adjustments on the survey-weighted counts before applying a binomial sampling distribution for the survey weighted counts. In section 3.3, we describe a Bayesian model for survey-weighted counts and the related hierarchical Bayes methodology. In section 3.4, we discuss elements of poverty mapping. In section 3.5, we apply our proposed method in estimating poverty rates for Chilean municipalities. We add discussions in section 3.6.

3.2 Notations and Data Preparation Steps

Consider a population U partitioned into m small areas U_i, where $\bigcup_i U_i = U$. Let y_{ij} denote the binary characteristic of interest (e.g., poverty status) associated with the jth observational unit (e.g., person) in the ith small area ($i = 1, \cdots, m;\ j \in U_i$). Our goal is to estimate the small area proportions

$$P_i = \sum_{j=1}^{N_i} y_{ij}/N_i,$$

where N_i is the number of observational units in the ith small area $(i = 1, \cdots, m)$.

We define ultimate sampled units as the smallest units of the finite population selected by a sampling mechanism. We consider a situation where observational units are nested within ultimate sampled units. Thus, in our case sampled units can be viewed as a cluster of observational units. In some cases, values of the binary variable for the observational units could be the same within a given sampled unit. In other words, observations within a given sampled unit are perfectly correlated. In such cases, we define s_{ij} as the set of observational units in the jth sampled unit in the ith small area and $\tilde{s}_i$ be the set of sampled units in the ith small area ($j \in \tilde{s}_i,\ i = 1, \cdots, m$). Thus $s_i = \bigcup_{h \in \tilde{s}_i} s_{ih}$ denotes a set of observational units in small area i. Let $\tilde{n}_i$ denote the total number of observational units in s_i.

3.2.1 Adjusted survey-weighted counts

We first define the effective sample size as $d_i = \frac{\tilde{n}_i}{\text{deff}}$, where deff is an estimate of the design effect for a large geographical area in which the ith small area is located ($i = 1, \cdots, m$). One could also use an approximation formula by Kish (1965) to obtain deff; however, this approximation is very unstable when the sample size within the cell is very small. In the next subsection, we will further describe a method for obtaining deff. We define $n_i = \lfloor d_i \rfloor$, where $\lfloor d_i \rfloor$ denotes the largest integer less than d_i. Next, we make a simple ratio adjustment to the final survey weights, say w_{ij}^{final}, by the factor $\frac{n_i}{\sum_{j \in s_i} w_{ij}^{\text{final}}}$. Thus, our adjusted weights, w_{ij}, are calibrated such that $\sum_{j \in s_i} w_{ij} = n_i$.

Instead of modeling the original survey-weighted counts $\sum_{j \in s_i} w_{ij}^{\text{final}} y_{ij}$, we consider a model for the adjusted survey-weighted counts $\tilde{y}_i = \sum_{j \in s_i} w_{ij} y_{ij}$. Note that the adjusted survey-weighted counts $\tilde{y}_i$ are generally not integers and so we model $y_i = \lfloor \tilde{y}_i \rfloor$ or $\lfloor \tilde{y}_i \rfloor + 1$.

3.2.2 Estimation of design effect for a large area

The true design effect, denoted by DEFF, is defined as:

$$\text{DEFF} = \frac{V(p_w)}{V_{srs}(p)},$$

where p_w is a survey-weighted proportion; $V(p_w)$ is the true design-based variance of p_w under complex sampling; p is unweighted proportion; $V_{srs}(p)$ is the variance of p under simple random sampling.

To estimate $V(p_w)$ for a large area, we can simply use a standard design-based variance estimation technique. Note that we cannot use standard formula for estimating $V_{srs}(p)$ because the sample is obtained using a complex sample survey design. We now describe a procedure to estimate $V_{srs}(p)$ using complex survey data.

Let s denote a sample of ultimate sampled units in the large area and s_j denote the set of observational units in the jth sampled unit ($j \in s$). Note that we can write the unweighted sample proportion p as

$$p = \frac{\sum_{j \in s} \tilde{n}_j y_j}{\tilde{n}},$$

where $\tilde{n}_j$ is the number of observational units in the jth sampled unit ($j \in s$) and $\tilde{n} = \sum_{j \in s} \tilde{n}_j$.

To approximate $V_{srs}(p)$, we use the following working model often used for simple random sampling:

$$E_m(y_j) = P,\ V_m(y_j) = P(1-P),\ \text{and } \text{Cov}_m(y_j, y_{j'}) = 0,\ \text{ for } j \neq j'.$$

We use the subscript m in expectation, variance and covariance to indicate that they are with respect to the above model.

Under the above SRS model,

$$\begin{aligned}
V_m(p) &= V_m\left(\frac{\sum_{j\in s}\tilde{n}_j y_j}{\tilde{n}}\right)\\
&= \frac{1}{\tilde{n}^2}\sum_{j\in s}\tilde{n}_j^2 V_m(y_j)\\
&= \frac{1}{\tilde{n}^2}\sum_{j\in s}\tilde{n}_j^2 P(1-P)\\
&= \frac{P(1-P)}{\tilde{n}^2}\sum_{j\in s}\tilde{n}_j^2.
\end{aligned}$$

We can obtain a lower bound of the $V_m(p)$ using the Cauchy-Swartz inequality:

$$V_m(p) \geq P(1-P)\cdot\frac{(\sum n_j)^2}{h\tilde{n}^2} = \frac{P(1-P)}{h},$$

where h denotes the number of ultimate sampled units in s.

Thus, a conservative estimate of DEFF, deff, is given by:

$$\frac{v(p_w)}{p_w(1-p_w)/h},$$

where $v(p_w)$ is a standard design-based variance estimate.

3.3 The Binomial-Beta model for survey-weighted counts

We assume the following Binomial-Beta model for the survey-weighted counts:

Model L:

For $i = 1, \ldots, m$,

$$\begin{aligned}
\text{Level 1 :} \qquad & y_i|\pi_i \overset{ind}{\sim} \text{Binomial}(n_i, \pi_i), \\
\text{Level 2 :} \qquad & \pi_i|\beta, \xi \overset{ind}{\sim} \text{Beta}(\mu_i\xi, (1-\mu_i)\xi), \\
& \xi > 0, \text{logit}(\mu_i) = \boldsymbol{x}_i'\boldsymbol{\beta} \qquad (3.3)
\end{aligned}$$

where $\boldsymbol{x}_i$ is a vector of p known predictors; $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression coefficients (intercept included). We also assume that $\boldsymbol{\beta}$ has an improper uniform in R^p and is independent of $1/\xi \sim (0, \infty)$. From an unpublished work of Tak and Morris (2012), it can be shown that the posterior distribution of π_i is proper under this prior distribution.

In the above model, Beta(a, b) is the Beta distribution with the following probability density function:

$$\frac{1}{\text{Beta}(a, b)} x^{a-1}(1-x)^{b-1},\ 0 < x < 1; a, b > 0.$$

Note that the posterior distribution of π_i is given by

$$\pi_i|y, \beta, \xi \overset{ind}{\sim} \text{Beta}(y_i + \mu_i\xi, n_i - y_i + (1-\mu_i)\xi).$$

Then, we have

$$
\begin{aligned}
E(\pi_i|y,\beta,\xi) &= (1-B_{i;L})p_{i;L} + B_{i;L}\mu_i = \pi^B_{i;L}(\beta,\xi) \equiv \pi^B_{i;L}, \\
V(\pi_i|y,\beta,\xi) &= \frac{1}{n_i+\xi+1}\pi^B_{i;L}(1-\pi^B_{i;L}) = V^B_{i;L}(\beta,\xi),
\end{aligned}
$$

where

$$
\begin{aligned}
p_{i;L} &= \frac{y_i}{n_i}, \\
B_{i;L} &= \frac{\xi}{\xi+n_i} = B_{i;L}(\beta,\xi).
\end{aligned}
$$

From Model L (3.4), we can see that Bayesian inference can be made using the following posterior distribution:

$$
f(\pi_1,\ldots,\pi_m|\boldsymbol{y}) \propto \int_{\boldsymbol{\beta}}\int_{\xi} f(\pi_1,\ldots,\pi_m,\boldsymbol{\beta},\xi)d\boldsymbol{\beta}d\xi.
$$

Since the joint distribution $f(\pi_1,\ldots,\pi_m,\boldsymbol{\beta},\xi)$ cannot be expressed in a simple closed form, an approximation is needed. Tak and Morris (2012) used an approximation method to obtain the posterior distribution in which the integration is approximated by Laplace's method.

The posterior distribution of π_i $(i=1,\cdots,m)$ can be generated by:

Step 1: Generate β and ξ using MCMC;

Step 2: Generate $\pi_{i;L}$ $(i=1,\cdots,m)$ from Beta$(y_i+\mu_i\xi, n_i-y_i+(1-\mu_i)\xi)$ using β and ξ from Step 1;

Note that there is another possible model, say Model U, with n_i and y_i replaced by n_i+1 and y_i+1, respectively in Model L. We suggest that the final model is a weighted average model between Model L and Model U.

We define the Model avg as:

Model avg:

$$\pi_i^{avg} = \phi_i \pi_{i;L} + (1 - \phi)\pi_{i;U}, \tag{3.4}$$

where $\phi_i = 1 - (d_i - n_i)$. We can generate π_i^{avg} from each MCMC run of $\pi_{i;L}$ and $\pi_{i;U}$ from Model L and Model U, respectively. The posterior mean, $E(\pi_i|\boldsymbol{y})$, is approximated by the sample mean of the posterior samples and the posterior variance ,$V(\pi_i|\boldsymbol{y})$, is used as a measure of variability.

3.4 Review on Poverty

Poverty is regarded as one of the most serious social problems around the world. International organizations, such as the World Bank and the United Nations, (UN), have developed many programs to confront this challenge. For example, the UN established a program, called the Millennium Development Goal, for developing countries to reduce their extreme poverty rates by 50 % by 2015, `(http://www.un.org/millenniumgoals).` To compare the severity of poverty problems among different countries, the World Bank defines the common international poverty line in absolute terms in dollars. It defines extreme poverty as a person living on less than US$1.25 per day and moderate poverty as a person living with less than $2.00 a day. Despite progress, it's estimated that more than 1.4 billion people, or one quarter of the population of the developing countries, still lived below the extreme poverty line in 2005, Chen and Ravallion (2008).

There are two methods of measuring poverty: absolute and relative. Absolute poverty is defined in terms of the minimal necessary requirements to afford sets of standards, such as food, clothing, health care, and shelter. An example of absolute measurement would be the percentage of the population living with less than an adequate nutritional diet, (approximately 2000-2500 daily calories recommended by the World Health Organization (WHO) and the Food and Agriculture Organization (FAO), Tontisirin and Haen (2001)). One of the advantages of this method is its consistency over time and between different countries. However, it is often criticized that the amount of minimal survival resources may not be the same in all places and across different time periods. For example, a person living in a colder climate requires heat during colder months while a person living in the tropical area does not.

On the other hand, relative poverty is a measure for a person living below a level of relative poverty threshold based on "economic distance." The economic distance is a level of household income that is usually set at a fixed percentage of the national median household income, Ravallion (2010). This means that there will always be a family living in relative poverty by its nature, and this attribute can sometimes lead to some strange results. For example, if the median household income in a wealthy neighborhood earns US$1 million dollars/year, then a family with US$100,000 in that neighborhood would be considered living in poverty. At the other end of the spectrum in a poor neighborhood where the median household income is below the national poverty level, a person with the median income would not be considered poor in this neighborhood.

Each country adopts one method or another to assess its poverty rates by conducting large population household surveys. For example, the U.S. Census Bureau uses the absolute poverty measure to analyze the poverty rates with the American Community Surveys (ACS). It categorizes each person or family into one of 48 possible poverty thresholds to determine the poverty status, (https://www.census.gov/hhes/www/poverty/methods/measure.html). In the United Kingdom, the Office for National Statistics uses the relative poverty rate. In 2006, 60% of the yearly median income was estimated at £12,000 for a 35-hour working week before tax, Cooke and Lawton (2008).

3.4.1 Poverty indicators

In the literature, there are many different indicators intended to summarize poverty of income inequality in one measure. Each of them illustrates one particular aspect of poverty measures. Foster et al. (1984) have given a brief description of the class of poverty indicators. Consider a finite population of size N partitioned into D small areas of sizes $N_1, \ldots, N_D$. Let E_{dj} be a suitable quantitative measure of welfare, such as income or expenditure, for an individual j in a small area d. Let z be a fixed poverty level; that is, the threshold for E_{dj} under which a person is considered as living "under poverty." Then, the family of FGT poverty measures for each small area d is defined as

$$F_{\alpha d} = \frac{1}{N_d} \sum_{j=1}^{N_d} F_{\alpha dj}, d = 1, \ldots, D, \tag{3.5}$$

where

$$F_{\alpha dj} = \left(\frac{z - E_{dj}}{z} \right)^{\alpha} \mathcal{I}(E_{dj} < z), \quad j = 1, \ldots, N_d, \alpha = 0, 1, 2,$$

and $\mathcal{I}(E_{dj} < z) = 1$ means person under poverty $(E_{dj} < z)$ and $\mathcal{I}(E_{dj} < z) = 0$ means otherwise.

For $\alpha = 0$, we get the proportion of individuals under poverty in area d that is equivalent to the simple head count ratio at the area-level. When $\alpha = 1$, the measure F_{1d} is called the poverty gap, and it counts only the fraction $\left(\frac{z-E_{dj}}{z}\right)^1$ in the small area d. In other words, F_{1d} only measures the area mean of the relative distance to the poverty level of each individual. When $\alpha = 2$, the measure F_{2d} is called the poverty severity and large values of F_{2d} can point out areas with severe levels of poverty.

3.4.2 Estimation of the FGT measure

In a survey setting, a direct estimator of $F_{\alpha d}$ for an area d can be defined as:

$$\hat{F}_{\alpha d} = \frac{1}{\hat{N}_d} \sum_{j \in s_d} w_{dj} F_{\alpha dj}, \alpha = 0, 1, 2, d = 1, \ldots, D, \tag{3.6}$$

where w_{dj} is the sampling weight of an individual j from the area d, and $\hat{N}_d = \sum_{j \in s_d} w_{dj}$ is the design-unbiased estimator of N_d. This estimator, $\hat{F}_{\alpha d}$, becomes unreliable when there are limited sample sizes, n_d, for area d.

Molina and Rao (2010) suggested the following method for such cases. Suppose that there is a one-to-one transformation $Y_{dj} = T(E_{dj})$ of the welfare variable, E_{dj}, where Y_{dj} follows a nested error linear regression model, as in the BHF model. Then, equation (3.5) can be written as:

$$F_{\alpha dj} = \left(\frac{z - T^{-1}(Y_{dj})}{z}\right)^{\alpha} \mathcal{I}(T^{-1}(Y_{dj} < z)) := h_{\alpha}(Y_{dj}), j = 1, \ldots, N_d \tag{3.7}$$

Molina and Rao's model does not use survey weights; rather, it predicts for the nonsampled units in the population and combines them with the observed values to estimate the poverty measure. Molina et al. (2012) have used similar modeling method by incorporating the intra-class correlation into their model.

In comparison, our model, **Model avg** (3.4), has incorporated survey weights and some features of the survey design to estimate the area-level poverty, and since we are using the survey weights, our estimates will be approximately design-unbiased.

3.5 Data Analysis

3.5.1 Description of Chilean Poverty Data: CASEN 2009

Our model, **Model avg** 3.4, will be applied to measure the poverty rate in Chile to estimate the area-level poverty rates for the Chilean municipalities (comunas). In Chile, the Ministry of Planning and Cooperation uses the absolute poverty rate and administers a survey called the *Encuesta Nacial de Caracterización Socioeconómica*, (CASEN) to assess poverty for non-institutionalized civilians in the country,
`(http://www.ministeriodesarrollosocial.gob.cl/casen/en/index.html).`

CASEN is a multipurpose survey with two major objectives:

- To characterize the situation of households and the population on issues related to poverty, income distribution and access to welfare programs.
- To estimate coverage, focalization and distribution of the government budget on the main social programs of national coverage.

The Chilean government establishes two poverty boundaries and two baskets of goods associated with the poverty levels. The first level, called the extreme poverty level, is based on the cost of buying a basket of food goods which contains the minimum amount of nutrition required for an average person to sustain health. The second poverty level is based on the consumption values for other non-food related and yet essential goods. The values are calculated by multiplying the cost of basic needs times the Engel coefficient, which is defined as a ratio of food consumption in total consumption. If a person can afford only the first basket of goods, he/ she is poor and if the person cannot afford either basket, he/she is extremely poor, Glasinovic (2010).

The poverty line is further defined for different geographical regions: urban and rural. From 2009 CASEN, the poverty level for an urban zone was drawn at \$64.134 and for a rural zone, it was at \$43.24. For extreme poverty level, the amount at the urban area was \$32.06 and \$24.71 in rural area, (http://www.hogardecristousa.org/v4/?pobreza).

3.5.2 The mode of data collection

CASEN is a survey with a complex design. It samples approximately 75,000 housing units from around 4000 geographic areas called "secciones" which are regarded as the PSUs. Every PSU falls within the boundary called 'comuna'(municipality), and comunas are categorized into two classifications: urban or rural. The PSUs are then grouped into strata on the basis of two geographical classifications: comuna and urban/rural classification. Within each stratum, the selection of PSUs is carried out by using the probability proportion to size (PPS) method, where the size variable is the number of housing units.

Once the housing units within a selected PSU are updated, the second stage of sampling is carried out by selecting a sample of housing units, an average of 16-22 housing units, through a systematic sampling method. That is a procedure that uses a random start and a systematic interval to select the units into the sample. It is assumed that those housing units share the same selection probability. The diagram of the CASEN survey design is given in the Figure 3.1.

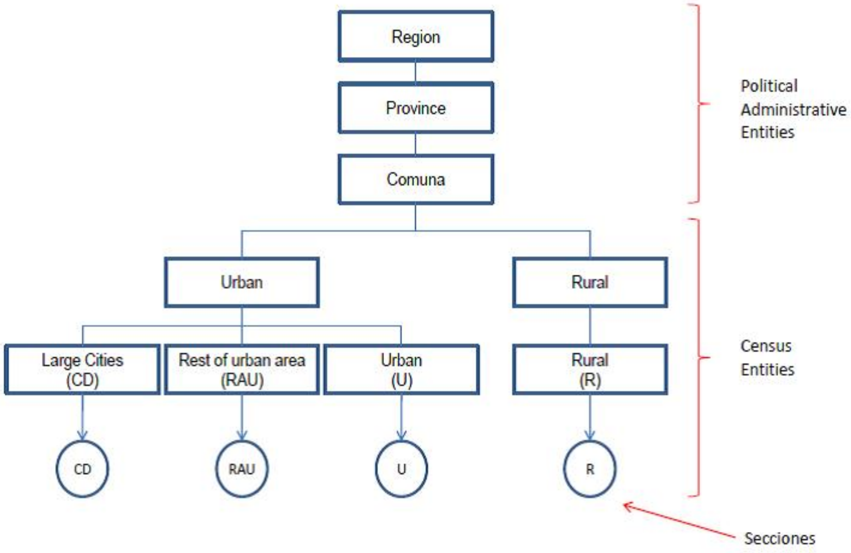


Figure 3.1: CASEN survey design diagram

Within each sampled housing unit, all households within the housing unit are identified and all household members are interviewed. More information about the CASEN can be found at (`http://www.ministeriodesarrollosocial.gob.cl/casen/en/index.html`).

3.5.3 Selection of auxiliary variables

For the **Model avg**, we have used the covariates that are selected from the Chilean government's analysis of the CASEN survey. They have fitted a theoretical regression model and used the backward selection method to choose from many candidate variables. Our choice of covariates are

- historical poverty based on the last three years
- percent of population living in the rural area
- percent of school attendance.

Like in Chapter 2, it should take into account that the sampling errors of covariates are not completely eliminated. However, this aspect will not be discussed further in this section.

3.5.4 Estimation Results

We plot the posterior mean from the **Model avg** (3.4) against the direct design estimates on the 30 smallest comunas and the 30 largest comunas by the sample size.

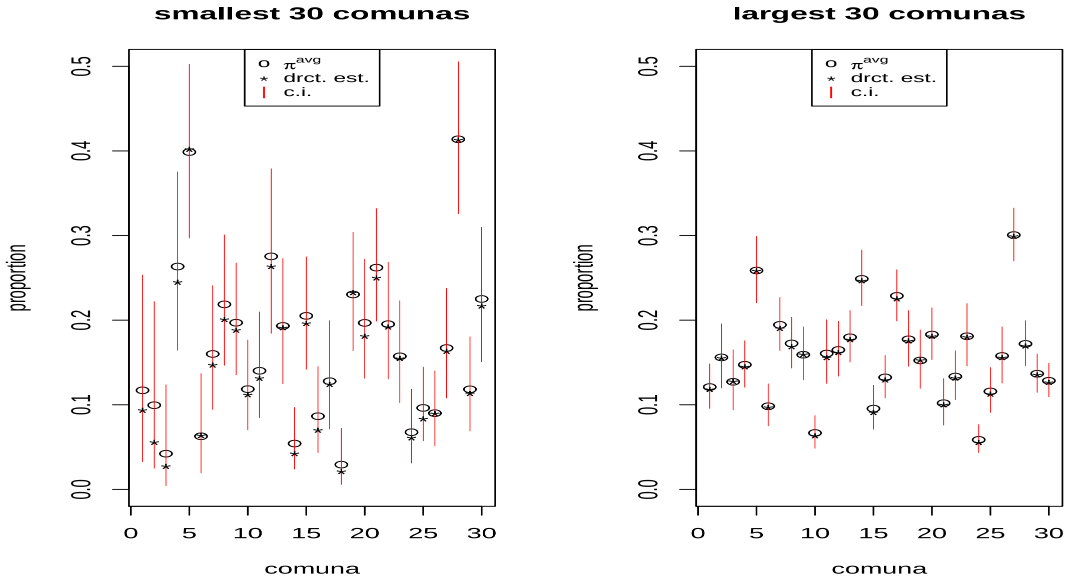
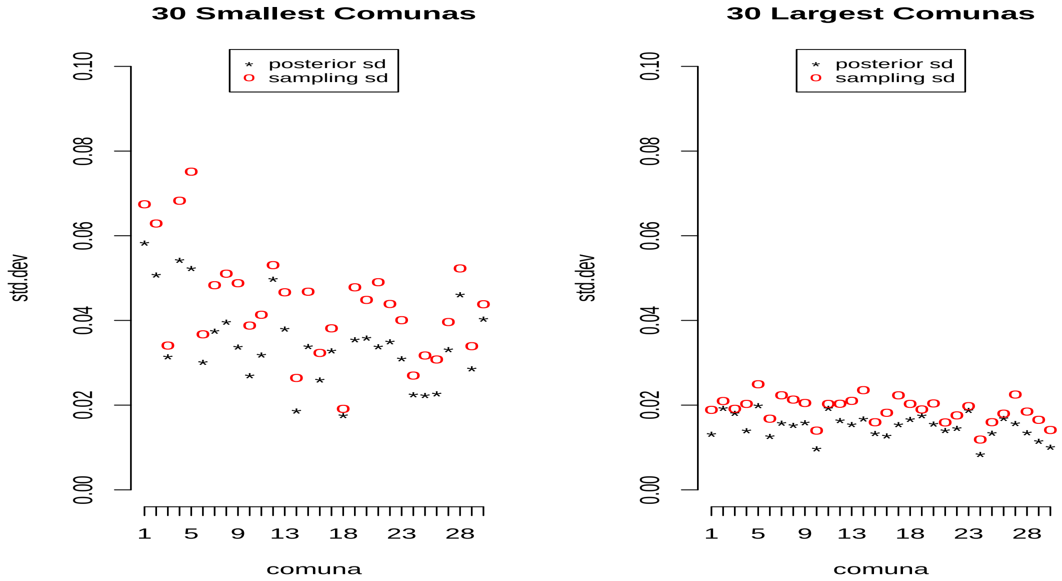


Figure 3.2

We can see from Figure 3.2 that the posterior estimates and the direct estimates are very similar in larger areas. On the other hand, the difference between the two estimates are larger in smaller comunas. As expected, the credible intervals for larger comunas are significantly narrower than those of smaller comunas.

In Figure 3.3, we plot the posterior sampling standard errors and the sampling standard error for the 30 smallest and the 30 largest comunas. It is seen that the standard errors for model-based estimates show an improvement over the sampling standard estimates.


30 Smallest Comunas

30 Largest Comunas

Figure 3.3

3.5.5 Model Diagnostic

To evaluate the possible bias introduced by the model, we use the simple ordinary least squares (OLS) regression method suggested by Brown et al. (2001). We plot the model-based estimates against the direct design estimators. We would expect points randomly scattered around the $45°$ line. If a pattern is shown in the scatter plot, then it may indicate that there is bias due to model mis-specification. We plot the design estimates as X and the model-based estimates as Y, and see how close the regression line, $Y = \alpha X$, is to $Y = X$. We obtain the estimated α value as 0.993 with standard error 0.003. Figure 3.4 shows a scatter plot with the fitted regression line.

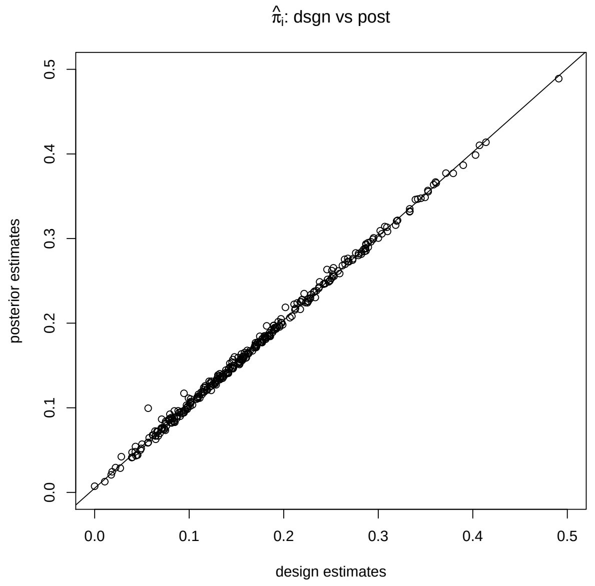


Figure 3.4: Scatter plot with regression line

The regression result shows no significant difference from $Y = X$, and this result may indicate no evidence of bias due to possible model mis-specification.

3.5.6 Model comparison

We will compare our model with the EB estimates, using the Prasad and Rao (1990) method. In the Prasad Rao (PR) method, they have used the hierarchical model in equation (2.3), where hyperparameters are estimated by using the MLE and weighted least squares (WLS) methods. Once estimates are obtained, we calculate their aggregated weighted proportion at the larger regional level and compare them to the corresponding aggregated proportion of the design-based estimates.

Table 3.1 shows the relative errors between two estimators at the regional level. Even though both methods provide good estimates at the regional level, our model-based estimates clearly out-perform those of the EB method in almost all the regions.

region	Model_avg.rgn	EB.rgn
1	0.0038	0.0063
2	0.0112	0.0162
3	0.0171	0.0314
4	0.0109	0.0175
5	0.0126	0.0100
6	0.0117	0.0266
7	0.0063	0.0323
8	0.0115	0.0177
9	0.0296	0.0026
10	0.0176	0.0331
11	0.0172	0.0033
12	0.0077	0.0221
13	0.0217	0.0052
14	0.0093	0.0548
15	0.0125	0.0008

Table 3.1: Relative Error at the region between Model-avg estimates and EB estimates

3.6 Summary

In this section, we have expanded the area-level model for discrete count data. Generally, modeling the survey weighted counts is a challenging problem because of they are non-integer values. This section provides a guideline for making appropriate adjustments for survey weighted counts, and we also obtained estimates for area-level proportions without using the normality assumption. Unlike some other previous work on poverty, we incorporated the survey weights and survey design features into the model, so that our estimates are design consistent. We also have shown that estimates from our model performed better than the EB based estimates, but further analysis could be explored. Also, additional analysis for the model fit can be examined.

Chapter 4

Constrained Bayes Estimates

4.1 Introduction

Louis (1984) proposed a constrained empirical Bayes method for a special case of the Fay-Herriot model with $x_i^T\beta = \mu,\ D_i = D\ (i = 1, \cdots, m)$. Louis showed that even though posterior means are optimal under the sum of the square error loss function (SSEL), they overshrink in the following sense:

$$E\left(\sum_{i=1}^{m}(\theta_i - \bar{\theta})^2|\boldsymbol{y}\right) \geq \sum_{i=1}^{m}\left(\hat{\theta}_i^B - \bar{\hat{\theta}}^B\right)^2,$$

where $\hat{\theta}_i^B = E[\theta_i|\boldsymbol{y};(\mu, A)]$ and $\bar{\hat{\theta}}^B = m^{-1}\sum_{k=1}^{m}\hat{\theta}_i^B$ with equality holding if and only if all $((\theta_i - \bar{\theta}), \ldots, (\theta_m - \bar{\theta}))$ have degenerate posteriors. Ghosh (1992) has provided a proof, and its corresponding weighted version is presented by Frey and Cressie (2003).

In order to address this overshrinking problem, Louis (1984) developed the concept of constrained Bayes (CB) estimation. The CB estimators are obtained by minimizing the posterior risk:

$$E\left[\sum_{k=1}^{m}(\theta_k - \bar{\theta})^2|y;(\mu, A)\right],$$

under the sum of square error loss function, subject to the following two constraints:

$$\begin{aligned}\sum_{k=1} \hat{\theta}_k &= E[\sum_{k=1} \theta_k|y;(\mu,A)] \qquad (4.1)\\ \sum_{k=1}^{m}(\hat{\theta}_k-\bar{\hat{\theta}})^2 &= E\left[\sum_{k=1}^{m}(\theta_k-\bar{\theta})^2|y;(\mu,A)\right], \qquad (4.2)\end{aligned}$$

where $\bar{\theta}=\sum_{k=1}^{m}\theta_k/m$, and $\bar{\hat{\theta}}=1/m\sum_{k=1}^{m}\hat{\theta}_k$. The Bayes estimator satisfies the first constraint but not the second one. Louis (1984) used Langrange's method of undetermined multipliers to arrive at the following constrained Bayes estimator of θ_i:

$$\hat{\theta}_i^{CB}=w\hat{\theta}_i^{pm}+(1-w)\hat{\bar{\hat{\theta}}}^{pm}, \qquad (4.3)$$

where $\hat{\theta}_i^{pm}=E(\theta_i|\boldsymbol{y})$, and $\hat{\bar{\hat{\theta}}}^{pm}=1/m\sum_{i=1}^{m}\hat{\theta}_i^{pm}$. The weight w is defined as

$$w=\left(1+\frac{m^{-1}\sum_{i=1}^{m}V(\theta_i|\boldsymbol{y})}{m^{-1}\sum_{i=1}^{m}(\hat{\theta}_i^{pm}-\hat{\bar{\hat{\theta}}}^{pm})^2}\right)^{1/2},$$

where $V(\theta_i|\boldsymbol{y})$ is the posterior variance of the ith parameter in the ensemble. The constrained Bayes estimator, $\hat{\theta}_i^{CB}$, involves unknown hyperparameters μ and A. Replacing these unknown hyperparameters by their estimates one obtains constrained empirical Bayes estimator of θ_i. Lahiri (1990) extended the constrained empirical estimation to estimate finite population means under a robust Bayesian model that does not require full specification of the sampling and prior distribution. He showed that under certain regularity conditions, constrained empirical Bayes estimators approach to the corresponding constrained Bayes estimator when m is large. In a constrained hierarchical Bayes approach, we put priors, usually flat priors, on the hyperparameters and obtain the desired estimator by a similar constrained optimization process. Posterior risks, means, variances, etc. are to be interpreted with respect to the posterior distribution of $\boldsymbol{\theta}$ given data y, integrating out μ and A. The formula is exactly the same as the constrained Bayes formula given above. For constrained hierarchical Bayes estimation for a fairly general model, see Ghosh (1992)

4.2 Constrained Bayes estimate: Benchmarked extension

Now we are concerned with the extension where the estimators satisfy certain constraints, such as the aggregate mean. Let $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_1, \cdots, \tilde{\theta}_m)'$ be an estimator of $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_m)'$, where θ_i denotes true total of area i ($i = 1, \cdots, m$). There are different choices for $\tilde{\boldsymbol{\theta}}$ (e.g., direct, synthetic, composite, empirical Bayes, hierarchical Bayes, triple goal etc.). Let $\bar{\theta}_w = \sum_{k=1}^{m} w_k \theta_k$ denote the true average for a large area level covering all of the m small areas, where $\sum_{k=1} w_k = 1$. An example of w_k is the population proportion, $w_k = N_k / \sum N_k = N_k/N$, where $N_k =$ kth area population, and $N =$ total population.

Our task is to find a new set of estimates, $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_m)'$, such that they satisfy the following set of benchmarking constraints:

$$\begin{aligned} \boldsymbol{G}'\hat{\boldsymbol{\theta}} =& \boldsymbol{c} \\ \hat{\boldsymbol{\theta}}\boldsymbol{H}_\ell\hat{\boldsymbol{\theta}} =& d_\ell ; (\ell = 1, \cdots, L), \end{aligned} \tag{4.4}$$

where $\boldsymbol{G}$ and $\boldsymbol{H}_\ell$'s are known matrices and $\boldsymbol{c}$ and $\boldsymbol{d} = (d_1, \ldots d_L)'$ are known vectors of constants. Note that the first constraint is concerned with the first moment and the second constraint is for the second moment of the estimator.

4.2.1 Discussion about the constraints

In this subsection, we will further discuss about the choice of the first constraint, $\boldsymbol{c}$. The constraint for the first moment of the estimators can be interpreted as a typical benchmarking problem. Generally, benchmarking problems can be categorized into two types: external and internal benchmarking, Bell et al. (2013). External benchmarking involves calibrating the estimates to agree with estimates from external data sources that are drawn independently of the survey. On the other hand, internal benchmarking involves calibrating small area estimates to a higher level aggregates obtained from the same survey.

External benchmarking has been commonly used in economic time series estimation setting where a monthly or a quarterly economic survey estimates are calibrated to the corresponding annual survey estimates, Dagum and Cholette (2006). External benchmark can be defined as either a constant or random, which is independent of the direct estimator, Bell et al. (2013) and Rendall et al. (2009).

Internal benchmarking has been explored in many literatures, Pfeffermann and Barnard (1991), Wang et al. (2008), and Pfeffermann and Tiller (2006). In most of them, it usually begins with best linear unbiased estimators for the small areas, and then they are modified to the higher level estimates to produce the benchmarked estimators. Pfeffermann and Barnard (1991) and Wang et al. (2008) have considered a single constraint type problem, and Bell et al. (2013) have expanded a solution to the problem by using general quadratic loss function with multiple constraints. See Pfeffermann et al. (2013)

for reviews of different methods of internal benchmarking.

4.2.2 Illustrative examples

We now consider a few examples.

Example 1:

Suppose we have the following single constraint:

$$\sum_{k=1}^{m} w_k \hat{\theta}_k = c, \tag{4.5}$$

where c is a reliable estimate from an external source or from the same survey. There is no constraint for the second moment; thus, we can express constraint (4.5) as:

$$\sum_{k=1}^{m} w_k \hat{\theta}_k = (w_1, \dots, w_m)' \begin{pmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_m \end{pmatrix} = c \tag{4.6}$$

$$= \boldsymbol{w}'\hat{\boldsymbol{\theta}} = \boldsymbol{G}'\hat{\boldsymbol{\theta}}, \tag{4.7}$$

where $\boldsymbol{w} = (w_1, \dots, w_m)', \boldsymbol{G} = \boldsymbol{w}$.

Example 2:

If we add another constraint for the spread between estimates, then our constraints become:

$$\sum_{k=1}^{m} w_k \hat{\theta}_k = c$$

$$\sum_{k=1}^{m} w_k (\hat{\theta}_k - \bar{\hat{\theta}}_w)^2 = d,$$

where $\bar{\hat{\theta}}_w = \sum_{k=1}^{m} w_k \hat{\theta}_k$, and the first constraint is equivalent to that of Example 1. The second constraint can be rewritten as:

$$\begin{aligned}
&\sum_{k=1}^{m} w_k (\hat{\theta}_k - \bar{\hat{\theta}}_w)^2 \\
=&\sum_{k=1}^{m} w_k \hat{\theta}_k^2 - (\sum_{k=1}^{m} w_k \hat{\theta}_k)^2 \\
=&\hat{\boldsymbol{\theta}}' \text{diag}(w_1, \ldots, w_m) \hat{\boldsymbol{\theta}} - (\hat{\boldsymbol{\theta}}' \boldsymbol{w})^2 \\
=&\hat{\boldsymbol{\theta}}' \{\text{diag}(w_1, \ldots, w_m) - \boldsymbol{w}\boldsymbol{w}'\} \hat{\boldsymbol{\theta}} \\
=&\hat{\boldsymbol{\theta}}' \boldsymbol{H} \hat{\boldsymbol{\theta}},
\end{aligned}$$

where $\boldsymbol{H} = \text{diag}(w_1, \ldots, w_m) - \boldsymbol{w}\boldsymbol{w}'$, and $\text{diag}(w_1, \ldots, w_m)$ represents a $m \times m$ matrix in which its ith element is w_i.

Example 3:

Suppose we are interested in estimating means for $m = IJ$ cells in a two-way contingency table (e.g., by age-group by race). We assume reliable estimates of the margins are available. A natural set of constraints is given by:

$$
\begin{aligned}
&(i) \sum_{i=1}^{I} \sum_{j=1}^{J} w_{ij}\hat{\theta}_{ij} = c \\
&(ii) \sum_{j=1}^{J} w_{ij}\hat{\theta}_{ij} = c_{i+}, \ \ \forall i \\
&(iii) \sum_{i=1}^{I} w_{ij}\hat{\theta}_{ij} = c_{+j}, \ \ \forall j,
\end{aligned}
$$

The first constraint is the same as in Example 1.

The second constraint can be rewritten as:

$$
\begin{aligned}
\sum_{j=1}^{J} w_{ij}\hat{\theta}_{ij} &= (w_{i1}, \ldots, w_{iJ}) \begin{pmatrix} \hat{\theta}_{i1} \\ \vdots \\ \hat{\theta}_{iJ} \end{pmatrix} \\
&= \boldsymbol{w}'_{i.}\hat{\theta}_{i.} = c_{i+}; \forall i, \qquad (4.8)
\end{aligned}
$$

where $\boldsymbol{w}_{i.} = (w_{i1}, \ldots, w_{iJ})$ and $\hat{\boldsymbol{\theta}}_{i.} = (\hat{\theta}_{i1}, \ldots, \hat{\theta}_{iJ})'$. Constraint (iii) can be written in an equivalent way as in constraint (ii).

Example 4

Let θ_{ij} be a small area parameter of interest for the jth state within the ith census division ($i = 1, \cdots, m;\ j = 1, \cdots, n$). We assume that sample sizes in the divisions are small as well, so we need to consider the indirect method for both division and states within division. Suppose $\hat{\theta}_{ij}$ denote an indirect estimator (e.g., HB, triple-goal, EB, etc.) of θ_{ij}. We need to adjust $\tilde{\theta}_{ij}$ so that they add up to the national level direct (or some other reliable estimate), but at the same time capture both within division and between division variabilities.

Then, the following constraints can illustrate our problem of interest:

$$
\begin{aligned}
&(i)\ \sum_{i=1}^{m}\sum_{j=1}^{n} w_{ij}\hat{\theta}_{ij} = c \\
&(ii)\ \sum_{j=1}^{n} w_{ij}(\hat{\theta}_{ij} - \bar{\hat{\theta}}_{iw})^2 = d_{1i}\ \forall i \\
&(iii)\ \sum_{i=1}^{m} w_{i+}(\bar{\hat{\theta}}_{iw} - \bar{\hat{\theta}}_{w})^2 = d_2,
\end{aligned}
$$

where $c,\ d_{1i},\ d_2$ are pre-specified, and

$$
\begin{aligned}
&\bar{\hat{\theta}}_{iw} = \frac{\sum_{j=1}^{n} w_{ij}\hat{\theta}_{ij}}{w_{i+}}, \\
&w_{i+} = \sum_{j=1}^{n} w_{ij} \\
&\bar{\hat{\theta}} = \sum_{i=1}^{m} w_{i+}\bar{\hat{\theta}}_{iw} \\
&\sum_{i=1}^{m}\sum_{j=n}^{J} w_{ij} = 1
\end{aligned}
$$

Constraint (i) can be written as:

$$
\sum_{i=1}^{m}\sum_{j=1}^{n} w_{ij}\hat{\theta}_{ij} = (w_{11}, \dots w_{mn}) \begin{pmatrix} \hat{\theta}_{11} \\ \vdots \\ \hat{\theta}_{mn} \end{pmatrix}, \tag{4.9}
$$

$$
= \boldsymbol{w}'\hat{\boldsymbol{\theta}} = c \tag{4.10}
$$

Constraint (ii) becomes:

$$\begin{aligned}\sum_{j=1}^{n} w_{ij}(\hat{\theta}_{ij} - \bar{\hat{\theta}}_{iw})^2 =& \sum_{j=1}^{n} w_{ij}\hat{\theta}_{ij}^2 + \sum_{j=1}^{n} w_{ij}\bar{\hat{\theta}}_{iw}^2 - 2\sum_{j=1}^{n} w_{ij}\hat{\theta}_{ij}\bar{\hat{\theta}}_{iw} \\ =& \sum_{j} w_{ij}\hat{\theta}_{ij}^2 - \frac{(\sum_j w_{ij}\hat{\theta}_{ij})^2}{w_{i+}} \\ =& \hat{\boldsymbol{\theta}}_i' \text{diag}(w_{i1}, \ldots, w_{in})\hat{\boldsymbol{\theta}}_i - \frac{\hat{\boldsymbol{\theta}}_i'\boldsymbol{w}_i\boldsymbol{w}_i'\hat{\boldsymbol{\theta}}_i}{w_{i+}} \\ =& \hat{\boldsymbol{\theta}}_i'\boldsymbol{H}_{1i}\hat{\boldsymbol{\theta}}_i,\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{H}_{1i} = \text{diag}(w_{i1}, \ldots, w_{in}) - \frac{\boldsymbol{w}_i\boldsymbol{w}_i'}{w_{i+}} \\ \boldsymbol{w}_i = (w_{i1}, \ldots, w_{in})' \\ \hat{\boldsymbol{\theta}}_i = (\hat{\theta}_{i1}, \ldots, \hat{\theta}_{in})'\end{aligned}$$

Finally, constraint (iii) becomes:

$$\begin{aligned}\sum_{i} w_{i+}(\bar{\hat{\theta}}_{iw} - \bar{\hat{\theta}}_w)^2 = \sum_{i} w_{i+}\bar{\hat{\theta}}_{iw}^2 - \bar{\hat{\theta}}_w^2 \\ =\bar{\hat{\boldsymbol{\theta}}}_w'\{\text{diag}(w_{1+}, \ldots, w_{i+}) - (\boldsymbol{w}_{.+}\boldsymbol{w}_{.+}')\}\bar{\hat{\boldsymbol{\theta}}}_w,\end{aligned}$$

where $\boldsymbol{w}_{.+} = (w_{1+}, \ldots w_{m+})'$, and $\bar{\hat{\boldsymbol{\theta}}}_w = (\bar{\hat{\theta}}_{1w}, \ldots, \bar{\hat{\theta}}_{mw})'$

4.3 A new approach for complex benchmarking constraints

In this section, we propose a general method that can produce a solution to the problem in equation (4.3). Under our proposed approach, we determine our target estimate, $\hat{\boldsymbol{\theta}}$, by making it closer to the starting estimator, $\tilde{\boldsymbol{\theta}}$, by using a model: $\hat{\boldsymbol{\theta}} = \boldsymbol{X}\boldsymbol{\Phi}$, where $\boldsymbol{X} = \boldsymbol{X}(\tilde{\boldsymbol{\theta}})$, such that $\boldsymbol{\Phi}$ satisfies the following system of equations:

$$\boldsymbol{A}\boldsymbol{\Phi} = \boldsymbol{b}, \tag{4.11}$$

We now illustrate how our method works for the examples in Section (4.2).

Example 1:

We begin with the starting estimator: $\tilde{\theta}_k$ = posterior mean, and we are interested in finding a a new estimator $\hat{\theta}_k$, such that $\sum_{k=1}^{m} w_k\hat{\theta}_k = c$. We let c as a flexible but a known value. For example, it can be defined as: $c = \sum_{k=1}^{m} w_k\bar{y}_k, w_k = N_k/\sum_{k=1}^{m} N_k$. Note that c can be regarded as a reliable internal benchmarking constraint. Our goal is to express $\hat{\theta}_i$ as a function of a starting estimate, $\tilde{\theta}_i$. By using equation (4.4) for constraints, our problem is to find β such that

$$\hat{\theta}_k = \beta\tilde{\theta}_k, \tag{4.12}$$

Using the constraint, we can easily solve for β:

$$\sum_{k=1}^{m} w_k\hat{\theta}_k = c$$

$$\Leftrightarrow \sum_{k=1}^{m} w_k\beta\tilde{\theta}_k = c$$

$$\Leftrightarrow \beta = \frac{c}{\sum_{k=1}^{m} w_k\tilde{\theta}_k}$$

Thus, the estimator with the benchmarking property is given as:

$$\hat{\theta}_i = \frac{c}{\sum_{k=1}^{m} w_k\tilde{\theta}_k}\tilde{\theta}_i$$

Example 2:

Instead of assuming that $\hat{\theta}_i = \beta\tilde{\theta}_i$, we can expand it by adding an intercept; that is:

$$\hat{\theta}_i = \alpha + \beta(\tilde{\theta}_i - \bar{\tilde{\theta}}_w), \tag{4.13}$$

where $\bar{\tilde{\theta}}_w = \sum_{k=1}^m w_k\tilde{\theta}_k$

With the first constraint, we see that

$$\begin{aligned} \sum w_i\hat{\theta}_{k=1}^m =& \alpha + \beta\sum_{i=1}^m w_i(\tilde{\theta}_i - \bar{\tilde{\theta}}_w) = c \\ \Leftrightarrow \alpha =& c \end{aligned}$$

With the second constraint, one can choose $d = E\{\sum_{k=1}^m w_k(\theta_k - \bar{\theta}_w)^2|data\}$ or any other appropriate choice of d. Then we have

$$\begin{aligned} \sum_{k=1}^m w_k(\hat{\theta}_k - \bar{\tilde{\theta}}_w)^2 = d \\ \Leftrightarrow \sum_{k=1}^m w_k\{\alpha + \beta(\tilde{\theta}_k - \bar{\tilde{\theta}}_w) - \alpha\}^2 = d \\ \Leftrightarrow \beta^2\sum_{k=1}^m w_k(\tilde{\theta}_k - \bar{\tilde{\theta}}_w)^2 = d \\ \Leftrightarrow \beta^2 = \frac{d}{\sum_{k=1}^m w_k(\tilde{\theta}_k - \bar{\tilde{\theta}}_w)^2}. \end{aligned}$$

Then equation (4.13) becomes:

$$\hat{\theta}_i = c + \sqrt{\frac{d}{\sum_{k=1}^m w_k(\tilde{\theta}_k - \bar{\tilde{\theta}}_w)^2}}(\tilde{\theta}_i - \bar{\theta}_w) \tag{4.14}$$

Example 3:

Assume $\hat{\theta}_{ij} = \mu + \alpha_i(\bar{\tilde{\theta}}_{i\cdot} - \bar{\tilde{\theta}}_{\cdot\cdot}) + \beta_j(\bar{\tilde{\theta}}_{\cdot j} - \bar{\tilde{\theta}}_{\cdot\cdot})$, for $i = 1, \ldots I, j = 1, \ldots, J$

We define:

$$\tilde{\theta}_{i\cdot} = \frac{\sum_j w_{ij}\tilde{\theta}_{ij}}{w_{i+}} \qquad \tilde{\theta}_{\cdot j} = \frac{\sum_i w_{ij}\tilde{\theta}_{ij}}{w_{+j}} \qquad \bar{\tilde{\theta}}_{\cdot\cdot} = \sum_i \sum_j w_{ij}\tilde{\theta}_{ij}$$

$$e_{ij} = w_{ij}(\bar{\tilde{\theta}}_{i\cdot} - \bar{\tilde{\theta}}_{\cdot\cdot}) \qquad e_{i+} = w_{i+}(\bar{\tilde{\theta}}_{i\cdot} - \bar{\tilde{\theta}}_{\cdot\cdot})$$

$$f_{ij} = w_{ij}(\bar{\tilde{\theta}}_{\cdot j} - \bar{\tilde{\theta}}_{\cdot\cdot}) \qquad f_{+j} = w_{+j}(\bar{\tilde{\theta}}_{\cdot j} - \bar{\tilde{\theta}}_{\cdot\cdot})$$

Then constraints (i), (ii), and (iii) can be rewritten as:

$$\begin{aligned} &\mu + \sum_i e_{i+}\alpha_i + \sum_j f_{+j}\beta_j = c, \\ &w_{i+}\mu + e_{i+}\alpha_i + \sum_j f_{ij}\beta_j = c_{i+}, i = 1, \ldots I \\ &w_{+j}\mu + \sum_i e_{ij}\alpha_i + f_{+j}\beta_j = c_{+j}, j = 1, \ldots J \end{aligned} \tag{4.15}$$

Equation 4.15 can be rewritten in the following matrix form:

$$\begin{pmatrix} 1 & e_{1+} & \dots & e_{I+} & f_{+1} & \dots & f_{+J} \\ w_{1+} & e_{1+} & 0 & 0 & f_{11} & \dots & f_{1J} \\ \vdots & 0 & \ddots & 0 & \vdots & \ddots & \vdots \\ w_{I+} & 0 & 0 & e_{I+} & f_{I1} & \dots & f_{IJ} \\ w_{+1} & e_{11} & \dots & e_{I1} & f_{+1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 & \ddots & \vdots \\ w_{+J} & e_{1J} & \dots & e_{IJ} & 0 & \dots & f_{+J} \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_I \\ \beta_1 \\ \vdots \\ \beta_J \end{pmatrix} = \begin{pmatrix} c \\ c_{1+} \\ \vdots \\ c_{I+} \\ c_{+1} \\ \vdots \\ c_{+J} \end{pmatrix}, \qquad (4.16)$$

This becomes in a partitioned matrix form,

$$\Leftrightarrow \left(\begin{array}{c|c|c} 1 & \boldsymbol{e}'_{+} & \boldsymbol{f}'_{+} \\ \hline \boldsymbol{w}_{.+} & \text{diag}(e_{1+}, \dots, e_{I+}) & \boldsymbol{f} \\ \hline \boldsymbol{w}_{+.} & \boldsymbol{E} & \text{diag}(f_{+1}, \dots, f_{+J}) \end{array} \right) \begin{pmatrix} \mu \\ \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix} = \begin{pmatrix} c \\ \boldsymbol{c}_{.+} \\ \boldsymbol{c}_{+.} \end{pmatrix}, \quad (4.17)$$

where

$$\begin{aligned} \boldsymbol{e}'_{+} &= (e_{1+}, \dots, e_{I+}) & \qquad \boldsymbol{c}_{.+} &= (c_{1+}, \dots c_{I+})' \\ \boldsymbol{f}'_{+} &= (f_{+1}, \dots, f_{+J}) & \qquad \boldsymbol{c}_{+.} &= (c_{+1}, \dots c_{+J})' \\ \boldsymbol{E} &= ((e_{ij})) & \qquad \boldsymbol{w}_{.+} &= (w_{1+}, \dots w_{I+})' \\ \boldsymbol{f} &= ((f_{ij})) & \qquad \boldsymbol{w}_{+.} &= (w_{+1}, \dots w_{+J})' \end{aligned}$$

We can see that equation (4.17) is in the form of equation (4.11), $\boldsymbol{A\Phi} = \boldsymbol{b}$. We can solve for $\boldsymbol{\Phi} = \boldsymbol{A}^{-1}\boldsymbol{b}$ to find $\hat{\boldsymbol{\theta}} = \boldsymbol{X}(\tilde{\boldsymbol{\theta}})\boldsymbol{\Phi}$.

Example 4:

Let $\hat{\theta}_{ij} = \alpha + \beta(\bar{\tilde{\theta}}_i - \bar{\tilde{\theta}}_w) + \gamma_i(\tilde{\theta}_{ij} - \bar{\tilde{\theta}}_{iw})$. The first constraint becomes:

$$\sum_i \sum_j w_{ij} \left(\alpha + \beta(\bar{\tilde{\theta}}_i - \bar{\tilde{\theta}}_w) + \gamma_i(\tilde{\theta}_{ij} - \bar{\tilde{\theta}}_{iw})\right) = c$$
$$\Leftarrow \alpha \sum_i \sum_j w_{ij} + \beta \sum_i \sum_j (\bar{\tilde{\theta}}_i - \bar{\tilde{\theta}}_w) + \sum_i \gamma_i \sum_j w_{ij}(\tilde{\theta}_{ij} - \bar{\tilde{\theta}}_{iw}) = c, \tag{4.18}$$

The second term of the equation (4.18) becomes:

$$\begin{aligned}
&\sum_i \sum_j w_{ij}\bar{\tilde{\theta}}_i - (\sum_i \sum_j w_{ij})\bar{\tilde{\theta}}_w \\
&= \sum_i \bar{\tilde{\theta}}_i \sum_j w_{ij} - \bar{\tilde{\theta}}_w \\
&= \sum_i \sum_j \frac{w_{ij}\tilde{\theta}_{ij}}{\sum_j w_{ij}} \sum_j w_{ij} - \bar{\tilde{\theta}}_w \\
&= \bar{\tilde{\theta}}_w - \bar{\tilde{\theta}}_w \\
&= 0,
\end{aligned}$$

The third term becomes:

$$\begin{aligned}
&\sum_j w_{ij}(\tilde{\theta}_{ij} - \bar{\tilde{\theta}}_{iw}) \\
&= \sum_j w_{ij}\tilde{\theta}_{ij} - \bar{\tilde{\theta}}_{iw} \sum_j w_{ij} \\
&= 0,
\end{aligned}$$

Thus, from the first constraint, we obtain:

$$\alpha \sum_i \sum_j w_{ij} = c$$

.

$$\Leftrightarrow \alpha = c$$

In the second constraint, we have:

$$\sum_j w_{ij}(\hat{\theta}_{ij} - \bar{\hat{\theta}}_{iw})^2 = d_{1i},$$

and $\bar{\hat{\theta}}_{iw}$ can be expressed as:

$$\begin{aligned}
\bar{\hat{\theta}}_{iw} &= \frac{\sum_j w_{ij}\hat{\theta}_{ij}}{\sum_j w_{ij}} \\
&= \frac{\sum_j w_{ij}(\alpha + \beta(\bar{\tilde{\theta}}_{iw} - \bar{\tilde{\theta}}_w) + \gamma_i(\tilde{\theta}_{ij} - \bar{\tilde{\theta}}_{iw}))}{\sum_j w_{ij}} \\
&= \alpha + \beta(\bar{\tilde{\theta}}_{iw} - \bar{\tilde{\theta}}_w) + \gamma_i \frac{\sum_j w_{ij}(\tilde{\theta}_{ij} - \bar{\tilde{\theta}}_{iw})}{\sum_j w_{ij}} \\
&= \alpha + \beta(\bar{\tilde{\theta}}_{iw} - \bar{\tilde{\theta}}_w) + \gamma_i \left(\frac{\sum_j w_{ij}\tilde{\theta}_{ij}}{\sum_j w_{ij}} - \bar{\tilde{\theta}}_{iw} \right) \\
&= \alpha + \beta(\bar{\tilde{\theta}}_{iw} - \bar{\tilde{\theta}}_w) + \gamma_i(\bar{\tilde{\theta}}_{iw} - \bar{\tilde{\theta}}_{iw}) \\
&= \alpha + \beta(\bar{\tilde{\theta}}_{iw} - \bar{\tilde{\theta}}_w)
\end{aligned}$$

Then, the second constraint reduces to:

$$\begin{aligned}
&\sum_j w_{ij}(\alpha + \beta(\bar{\tilde{\theta}}_{iw} - \bar{\tilde{\theta}}_w) + \gamma_i(\tilde{\theta}_{ij} - \bar{\tilde{\theta}}_{iw}) - \alpha - \beta(\bar{\tilde{\theta}}_{iw} - \bar{\tilde{\theta}}_w)) = d_{1i} \\
&\Leftrightarrow \gamma_i^2 \sum_j w_{ij}(\tilde{\theta}_{ij} - \bar{\tilde{\theta}}_{iw})^2 = d_{1i} \\
&\Leftrightarrow \gamma_i^2 = \frac{d_{1i}}{\sum_j w_{ij}(\tilde{\theta}_{ij} - \bar{\tilde{\theta}}_w)^2}
\end{aligned}$$

In the third constraint, we have

$$\sum_i w_{i+}(\bar{\hat{\theta}}_{iw} - \bar{\hat{\theta}}_w)^2 = d_2,$$

where $\bar{\hat{\theta}}_w = \alpha$ by the first constraint. Then, we have:

$$\begin{aligned}
&\sum_i w_{i+}(\alpha + \beta(\bar{\tilde{\theta}}_{iw} - \bar{\tilde{\theta}}_w) - \alpha)^2 = d_2 \\
&\Leftrightarrow \beta^2 \sum_i w_{i+}(\bar{\tilde{\theta}}_{iw} - \bar{\tilde{\theta}}_w)^2 = d_2 \\
&\Leftrightarrow \beta^2 = \frac{d_2}{\sum_i w_{i+}(\bar{\tilde{\theta}}_{iw} - \bar{\tilde{\theta}}_w)^2},
\end{aligned}$$

Then $\hat{\theta}_{ij}$ becomes,

$$\hat{\theta}_{ij} = c + \sqrt{\frac{d_2}{\sum_i w_{i+}(\bar{\tilde{\theta}}_{iw} - \bar{\tilde{\theta}}_w)^2}}(\bar{\tilde{\theta}}_i - \bar{\tilde{\theta}}_w) + \sqrt{\frac{d_{1i}}{\sum_j w_{ij}(\tilde{\theta}_{ij} - \bar{\tilde{\theta}}_w)^2}}(\tilde{\theta}_{ij} - \bar{\tilde{\theta}}_{iw}) \qquad (4.19)$$

4.4 Some optimality properties

Solutions obtained in Section 4.3 have the following optimality properties in certain benchmarking conditions.

Result 1: Suppose we want to find $\hat{\boldsymbol{\theta}}$ such that $\sum_{j=1}^{m}(\hat{\theta}_j - \tilde{\theta}_j)^2$ is minimized subject to the following single constraint:

$$\hat{\theta}. \equiv \sum_{j=1}^{m} \hat{\theta}_j \quad = \quad c_1,$$

where c_1 is a known constants and $\bar{\hat{\theta}} = m^{-1}\sum_{j=1}^{m}\hat{\theta}_j$. In practice, we can choose $c_1 = \sum_{j=1}^{m} y_j$, the design-based estimator, or $c_1 = \sum_{j=1}^{m}\hat{\theta}_j^B$, the Bayes estimator of $\theta. = \sum_{j=1}^{m}\theta_j$.

Result 2: Suppose we want to find $\hat{\boldsymbol{\theta}}$ such that $\sum_{j=1}^{m}(\hat{\theta}_j - \tilde{\theta}_j)^2$ is minimized subject to the following two constraints:

$$\hat{\theta}. \equiv \sum_{j=1}^{m} \hat{\theta}_j \quad = \quad c_1, \tag{4.20}$$

$$\sigma^2_{\hat{\theta}} \equiv \sum_{j=1}^{m}(\hat{\theta}_j - \bar{\hat{\theta}})^2 \quad = \quad c_2^2, \tag{4.21}$$

where c_2 is known constants.For c_2^2 we can choose $E\left[\sum_{j=1}^{m}(\theta_j - \bar{\theta})^2 | y\right]$, the Bayes estimator or some design-based estimator of $\sigma^2_{\theta} = \sum_{j=1}^{m}(\theta_j - \bar{\theta})^2$.

The solution to both the results can be obtained by Lagrange's method of undetermined multipliers. Consider the following objective function:

$$Q(\hat{\boldsymbol{\theta}}; \lambda_1, \lambda_2) = \sum_{j=1}^{m} (\hat{\theta}_j - \tilde{\theta}_j)^2 + 2\lambda_1(\hat{\theta}. - c_1) + \lambda_2(\sigma^2_{\hat{\theta}} - c_2^2).$$

The equation

$$\frac{\partial Q}{\partial \lambda_1} = 0$$

yields

$$\hat{\theta}_i = \frac{\tilde{\theta}_i - \lambda_1 + m^{-1}\lambda_2 c_1}{1 + \lambda_2}, \ \forall i \tag{4.22}$$

From 4.20, we get

$$\lambda_1 = m^{-1}(\tilde{\theta}. - c_1),$$

and hence

$$\hat{\theta}_i = (1 + \lambda_2)^{-1}(\tilde{\theta}_i - \bar{\tilde{\theta}}) + m^{-1}c_1.$$

Thus, using 4.21, we have

$$(1 + \lambda_2)^2 = \frac{\sigma^2_{\tilde{\theta}}}{c_2^2}.$$

Thus the new constrained estimator is given by:

$$\hat{\theta}_i = \frac{c_2}{\sigma_{\tilde{\theta}}}(\tilde{\theta}_i - \bar{\tilde{\theta}}) + \frac{c_1}{m}.$$

Interestingly, this estimator yields the same estimators considered in Ghosh (1992) and Datta et al. (2009) if the starting estimator is the posterior mean.

However, applying Lagrange's method is challenging for certain benchmarking problems, such as constraint conditions in example 3 and example 4 in section 4.2.

4.5 Data Application

In this section, we apply our methods to two data sets. The first data set is the baseball data described in Efron and Morris (1975) where the true value is known. The second data set is the monthly U.S. unemployment rate from the Current Population Survey (CPS) from January, 2009 to December, 2012. After we apply models on each data set and obtain posterior means, we use them as our starting estimator to produce other estimates.

For our estimators, we use the following notations:

- $\hat{\pi}_i^{pm}$: posterior estimates,
- $\hat{\pi}_i^{cb}$: constrained Bayes, Ghosh (1992)
- $\hat{\pi}_i^{bm1}$: benchmarking estimator with one constraint Datta et al. (2009)
- $\hat{\pi}_i^{bm_r}$: benchmarking estimator with one constraint, equation (4.12)
- $\hat{\pi}_i^{bm2}$: benchmarking estimator with two constraints, Datta et al. (2009).

For benchmarking estimators, we need to define the constraints: For the baseball example, the sample average or the average of the posterior means will be used. For the second constraint, the following identity will be used.

Let H denote the second constraint, then

$$\begin{aligned}
H =& E\{\sum w_i(\pi_i - \bar{\pi}_w)^2|\bar{\boldsymbol{y}}\} \\
=& E\{\sum w_i(\pi_i^2 - 2\pi_i\bar{\pi}_w + \bar{\pi}_w^2)|\bar{\boldsymbol{y}}\} \\
=& E\{\sum w_i(\pi_i^2 - \bar{\pi}_w^2)\} \\
=& \sum w_i \left(E(\pi_i^2|\bar{\boldsymbol{y}}) - E((\sum w_i\pi_i)^2|\bar{\boldsymbol{y}})\right) \\
=& \sum w_i \left((Var(\pi_i|\bar{\boldsymbol{y}}) + (E(\pi_i|\bar{\boldsymbol{y}}))^2) - (Var(\sum w_i\pi_i|\bar{\boldsymbol{y}}) + (E(\pi_i|\bar{\boldsymbol{y}}))^2)\right).
\end{aligned} \tag{4.23}$$

Note that

$$Var(\sum w_i\pi_i|\bar{\boldsymbol{y}}) = \sum w_i^2 Var(\pi_i|\bar{\boldsymbol{y}}) + \sum_{i \neq j} w_i w_j Cov(\pi_i, \pi_j|\bar{\boldsymbol{y}}),$$

and if we assume that $Cov(\pi_i, \pi_j|\bar{\boldsymbol{y}})$ is negligible, then equation (4.23) can be written as

$$H = \sum w_i Var(\pi_i|\bar{\boldsymbol{y}}) + \sum w_i(\hat{\pi}_i^{pm})^2 - \sum w_i(\sum w_i^2 Var(\pi_i|\bar{\boldsymbol{y}})) - \sum w_i(\sum w_i\hat{\pi}_i^{pm})^2 \tag{4.24}$$

4.5.1 Baseball analysis

In this subsection, we are using the batting average data described in Efron and Morris (1975) in Table 4.1. They have selected the batting averages of 18 major league baseball players in the 1970 season. Each player had batted 45 times and their batting averages are recorded up to that point. By using only this data, Efron and Morris wanted to predict each player's batting average for the remainder of the 1970 season.

player	$\hat{\pi}$	prev.avg	prev.at.bats	avg.1970
A	0.27	0.12	51	0.22
B	0.16	0.25	3514	0.18
C	0.31	0.25	2244	0.28
D	0.20	0.26	3210	0.28
E	0.40	0.31	8142	0.35
F	0.36	0.28	4826	0.28
G	0.33	0.26	1139	0.24
H	0.29	0.25	2753	0.27
I	0.18	0.26	86	0.30
J	0.22	0.26	2281	0.26
K	0.38	0.30	7542	0.31
L	0.22	0.23	291	0.22
M	0.24	0.28	5658	0.27
N	0.22	0.25	2065	0.30
O	0.31	0.24	454	0.27
P	0.24	0.24	1967	0.23
Q	0.22	0.26	1216	0.26
R	0.22	0.27	888	0.25

Table 4.1: Batting average data: The second column is the batting average with 45 at-bats and the last column is the actual batting average for the entire 1970 season.

Let y_i denote the number of hits in $n = 45$ at-bats, and $\hat{\pi}_i = y_i/n$ be the batting average of ith player. To predict the season batting average, Efron and Morris (1975) considered the following variance stabilizing transformation :

- $\hat{\theta}_i = \sqrt{n}\,\text{arcsin}(2\hat{\pi}_i - 1)$,
- $\theta_i = \sqrt{n}\,\text{arcsin}(2\pi_i - 1)$,
- π_i : True season bating average.

Note that using the Taylor series expansion, $V(\hat{\theta}_i|\theta_i) \approx 1$ for large m. Efron and Morris (1975) considered the following two-level Bayesian model:

The Efron-Morris Model

$$\begin{aligned} &level1,\text{(sampling distribution)} \quad &:\hat{\theta}_i|\theta_i \quad &\overset{ind}{\sim} N(\theta_i, 1), i = 1, \ldots, m, \\ &level2,\text{(prior distribution)} \quad &:\theta_i|\beta, A \quad &\overset{iid}{\sim} N(\boldsymbol{x}_i'\boldsymbol{\beta}, A), \end{aligned} \tag{4.25}$$

where x_i is a previous season batting average, β and τ^2 are unknown and independent hyperparameters with non-informative prior distribution; $f(\boldsymbol{\beta}) \propto 1$, and , $f(\tau^2) = (0, \infty)$.

Once we have obtained the MCMC samples, we use back transformation to get the inference about $f(\pi_i|\bar{y})$. We compare our result with the true value; that is the season end batting average of each player illustrated in column (4) of Table 4.1.

players	$\hat{\pi}_i^{pm}$	$\hat{\pi}_i^{cb}$	$\hat{\pi}_i^{bm1}$	$\hat{\pi}_i^{bm_r}$	$\hat{\pi}_i^{bm2}$
A	-0.004	-0.046	0.023	-0.003	-0.045
B	0.056	0.033	0.083	0.057	0.033
C	-0.005	0.001	0.022	-0.003	0.004
D	-0.025	-0.035	0.002	-0.024	-0.033
E	-0.042	-0.002	-0.015	-0.040	0.005
F	0.006	0.027	0.033	0.008	0.032
G	0.040	0.052	0.067	0.042	0.056
H	0.002	0.004	0.029	0.003	0.007
I	-0.056	-0.072	-0.029	-0.054	-0.071
J	-0.005	-0.013	0.022	-0.004	-0.012
K	-0.003	0.029	0.023	-0.002	0.036
L	0.024	0.010	0.051	0.025	0.011
M	0.001	0.005	0.028	0.003	0.007
N	-0.042	-0.051	-0.015	-0.040	-0.050
O	-0.004	0.002	0.023	-0.002	0.004
P	0.024	0.017	0.051	0.025	0.019
Q	-0.002	-0.009	0.025	-0.000	-0.008
R	0.010	0.007	0.037	0.011	0.008

Table 4.2: The deviation of estimators from the true value for different estimators.

The Table 4.2 and Figure 4.1 show the deviation, which is calculated by $(\theta_i - \hat{\theta}_i|data)$, for different estimates from the true value for each player. In Figure 4.1, the reference line is drawn at zero horizontal line, and we can see that $\hat{\theta}_i^{bm_r}$ is near the zero reference line more often than other estimators, and $\hat{\theta}_i^{bm1}$ is furthest from the reference line.

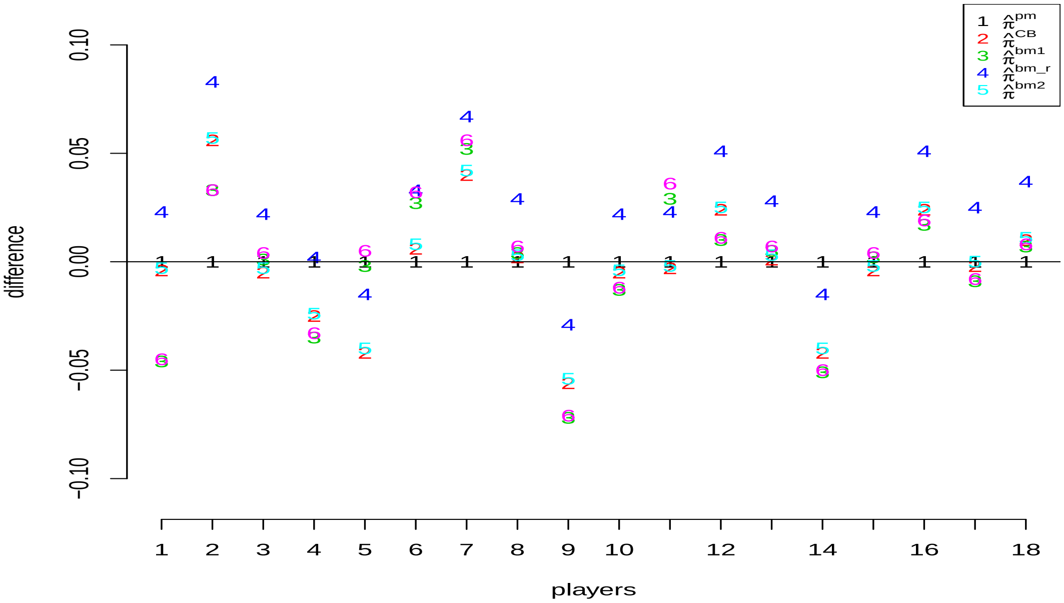


Figure 4.1: Deviation from the true average

$\hat{\pi}_i^{pm}$	$\hat{\pi}_i^{cb}$	$\hat{\pi}_i^{bm1}$	$\hat{\pi}_i^{bm_r}$	$\hat{\pi}_i^{bm2}$
0.013	0.017	0.025	0.013	0.018

Table 4.3: Sum of the square deviations

We also calculated the sum of the square deviations, $\sum_i(\hat{\pi}_i - \pi_i)^2$, for all estimators. As expected, $\hat{\pi}_i^{pm}$ performs well since it is obtained by minimizing the squared error loss given the data. The benchmarked estimators are not as optimal as in terms of Datta et al. (2009) when the estimators are benchmarked, but adding the second constraint makes some improvements.

We would also like to compare different estimators as an ensemble. Let R1 $= \frac{\sum_i \hat{\pi}_i}{\sum_i \pi_i}$, and R2 $= \frac{1/m\sum_i(\hat{\pi}_i - \bar{\hat{\pi}})^2}{1/m\sum_i(\pi_i - \bar{\pi})^2}$, where $\hat{\pi}_i$ denotes different types of estimator, $\bar{\hat{\pi}}$ be the mean of the estimator, π_i be the true average, and $\bar{\pi}$ be the mean of the true average. We can see that R1 is the ratio of the aggregated sum between estimators and the true value and R2 is the corresponding ratio of the sample variances. For both ratios, we prefer values near 1.

	R1	R2
$\hat{\pi}^{pm}$	0.995	0.333
$\hat{\pi}^{cb}$	0.992	1.179
$\hat{\pi}^{bm1}$	1.097	0.333
$\hat{\pi}^{bm_r}$	1.001	0.337
$\hat{\pi}^{bm2}$	1.001	1.300

Table 4.4: Ratio of different estimators

In Table 4.4, we can see that all estimators perform very well for R1; however, when we look at R2 only $\hat{\pi}^{bm2}$ and $\hat{\pi}^{cb}$ perform reasonably well because they include the constraint for the second moment of the estimator. Additionally, it's shown clearly that posterior estimate and benchmarked estimates with just one constraint display overshrinkage problem.

4.5.2 Unemployment rate data analysis

In the second data analysis, we use the monthly CPS to estimate the unemployment rate for each state from January, 2009 to December 2012. The CPS is conducted by the Census Bureau and its monthly sample comprises of about 72,000 housing units and is collected for about 729 areas consisting of more than 1,000 counties covering every state and the District of Columbia. More information about the CPS can be found (`http://www.bls.gov/cps/`).

The unemployment rate is one of the five key economic indicators published by the Bureau of Labor Statistics (BLS) and represents the number of unemployed as a percent of the labor force. The BLS publishes monthly unemployment rate estimates for the entire U.S. and its different demographic and geographic subdomains using sophisticated time-series model, Pfeffermann et al. (2013). The unemployment estimates are made for all 50 states and the District of Columbia, all metropolitan statistical areas (MSA), all counties (cities and towns of New England), and all cities with population 25,000 or greater. The local unemployment rates are used in regional planning and fund allocation under various federal assistance programs. The BLS administers extensive research for estimating unemployment, and information about their methodology can be found in Bell and Hillmer (1990) and Pfeffermann and Tiller (2002). For more information about the BLS's current ongoing research for benchmarking estimators, see `(http://www.bls.gov/lau/).`

For our data analysis, let $\hat{\pi}_i$ be the direct survey weighted estimate of the unemployment rate for area i $(i = 1, \cdots, m)$. Let $\hat{\theta}_i = \arcsin(\sqrt{\hat{\pi}_i})$ be the transformation for the direct estimate $\hat{\pi}_i$ and $n_i^{eff} = n_i/deff$ be the effective sample size for area i, and $deff$ is the design effect estimate at the national level. Then we apply the Carter-Rolph model for $\hat{\theta}_i$:

The Carter-Rolph Model:

Level 1: $\hat{\theta}_i \mid \theta_i \overset{ind}{\sim} (\theta_i, 1/(4n_i^{eff}))$,

Level 2: $\theta_i \overset{ind}{\sim} [\mu, A]$,

where the hyperparameters, μ, A, are given vague prior distributions. Like before, we will use back-transformation to produce posterior estimates for π_i and then obtain other estimators.

For our analysis, we first look at the aggregate sum of estimates at the national level for each month from January 2009 to December 2012.

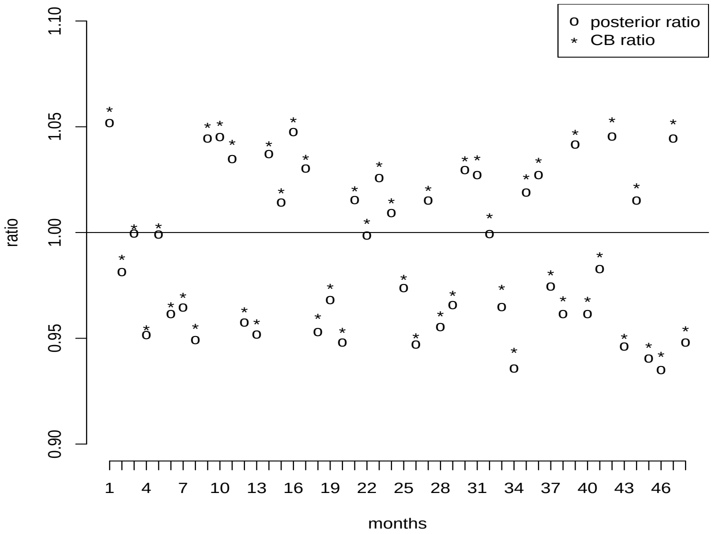


Figure 4.2: Ratio between the national estimate and other estimates

Figure 4.2 shows the ratio between the national estimates and the aggregated estimates for 48 months. The ratios are defined as:

$$\text{posterior ratio} = \frac{\sum_{i=1}^{51} N_i \pi_i^{pm}}{\text{total number of unemployed}}$$

$$\text{CB ratio} = \frac{\sum_{i=1}^{51} N_i \hat{\pi}_i^{cb}}{\text{total number of unemployed}},$$

where N_i, $\hat{\pi}_i^{pm}$ and $\hat{\pi}_i^{cb}$ are the number of persons in the labor force, posterior means and constrained Bayes estimates of unemployment rate for the ith state ($i = 1, \cdots, 51$), and the reference line is drawn at 1. It's clear from the picture that monthly aggregated estimates both posterior means and the CB estimates generally differ from the national level estimates.

There are two perspectives to explain reasons for the discrepancy between ratios to the reference line: methodology and model. In general, we expect that the aggregated model-based estimates would not be perfectly equivalent of the national estimates. Additionally, our model is very crude that we did not use any covariates into the model. With a more sophisticated model, we would expect the difference between the ratio from the reference line to be smaller.

We also observe that the CB ratio for each month is always slightly higher than the corresponding posterior means, and both estimates are within 5% of the national estimates. We did not consider other estimates since they are already benchmarked at the national level.

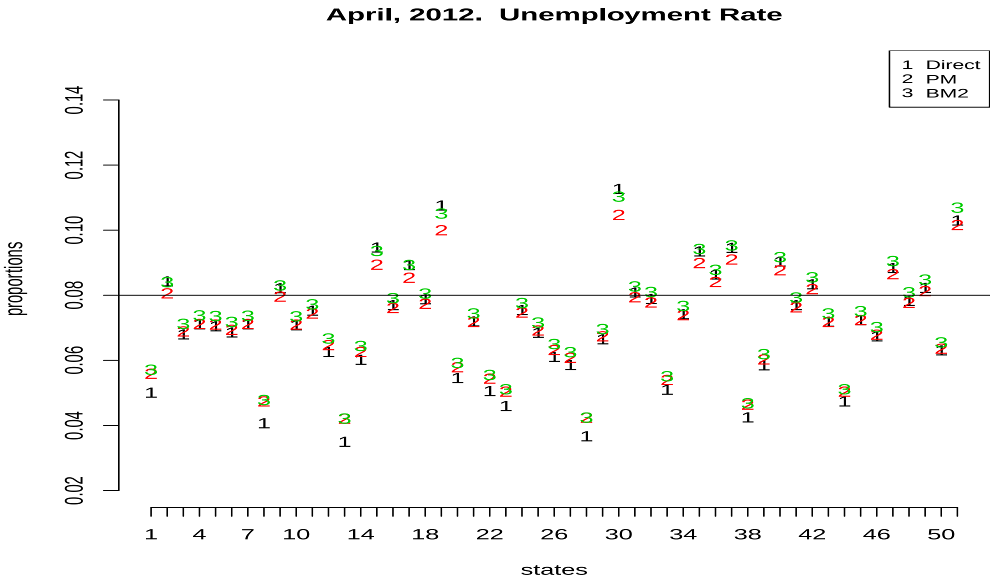


Figure 4.3: states are arranged in the order of effective sample size

Figure 4.3 shows the unemployment rates for each state for April, 2012. The national average, at 8%, is drawn as the reference line. Since the model uses no covariates, the predicted estimates for smaller states are near the unweighted sample average. Also, we can attain the same conclusion as in Chapter 2 that the model-based estimates are closer to the design-based estimates for larger states than for smaller states.

	Apr.2009	Apr.2010	Apr.2011	Apr.2012
$\hat{\pi}$	0.96	1.05	0.96	0.97
$\hat{\pi}^{pm}$	0.95	1.05	0.96	0.96
$\hat{\pi}^{cb}$	0.96	1.05	0.96	0.97
$\hat{\pi}^{bm1}$	1.00	1.00	1.00	1.00
$\hat{\pi}^{bm_r}$	1.00	1.00	1.00	1.00
$\hat{\pi}^{bm2}$	1.00	1.00	1.00	1.00

Table 4.5: average ratio, $\hat{\pi}$ = direct estimates

Table 4.5 shows the ratio between the national rates and the national aggregate of different estimators. We can see that both $\hat{\pi}^{cb}$ and $\hat{\pi}^{pm}$ behave very similarly.

	Apr.2009	Apr.2010	Apr.2011	Apr.2012
$\hat{\pi}$	1.09	1.07	1.05	1.08
$\hat{\pi}^{pm}$	0.89	0.86	0.86	0.84
$\hat{\pi}^{cb}$	1.02	1.00	1.03	1.04
$\hat{\pi}^{bm1}$	1.01	0.75	1.06	1.02
$\hat{\pi}^{bm_r}$	0.99	0.78	0.94	0.91

Table 4.6: ratio of the sample variance

Table 4.6 shows a ratio between sample variance of different estimators and that of $\hat{\pi}^{bm2}$. We can see that $\hat{\pi}^{cb}$ preserves its variance over different months; whereas, other estimators show over-shrinkage problems in some months.

4.6 Discussion

In this chapter, we have explored the constrained Bayes estimates with multidimensional conditions. We have shown that we can obtain a new set of estimators from an existing set by minimizing the distance given a set of conditions, where the condition are defined by the user or known values. In general, the solution is produced by using the Lagrange's method. However, applications of the Lagrange's method appears to be complicated as the constraints become more elaborate. With the assumption of a linear relationship between two sets of estimators, we have shown that the new set of estimators are relatively easy to obtain. We have applied the new methods to a data set with known true values and a data set with unknown values, and our results illustrate that some of the new estimates improve the over-shrinkage problem.

Chapter 5

Triple-goal Estimation With Benchmarking Constraints

5.1 Introduction

In many data analysis applications, we are not only interested in estimating individual area-level parameters but also in reporting an ensemble of ranked estimates or finding a set of estimates whose values that exceed a pre-specified threshold. For example, the goal can be estimating the performance evaluation, like the rank, among different companies, Landrum et al. (2000). Reporting an ensemble of estimates can also provide useful interpretations in disease mapping to ascertain the variation in disease rates for different geographical regions, Conlon and Louis (1999), Devine and Louis (1994).

There are a number of papers on the estimation of parameters for individual small area, Rao (2003) and Jiang and Lahiri (2006b), a histogram of small area parameters, Lahiri (1990), Louis (1984), or ranking small area parameters, Laird and Louis (1989), Morris and Christiansen (1996). However, there is little research about finding a set of estimates that would optimally satisfy multiple criteria all at once because finding an optimal estimators depends on the definition of the loss function. If individual specific parameters are of interest, posterior means are the optimal choice. If the ranks of parameters are the target, the conditional expected ranks are the optimal, but ranking posterior means can perform poorly, Goldstein and Spiegelhalter (1996). If the feature of interest is the histogram or the empirical distribution function (EDF) of the parameters, then the conditional expected EDF is optimal, but the histogram is overdispersed and that of the posterior means of the parameters is underdispersed, Ghosh (1992). From administrative point of view, reporting several ensembles for all different situations would be inefficient and may cause inconsistencies.

While there does not exist a set of point estimates that simultaneously optimize all of these criteria (Gelman and Price (1999)), Shen and Louis (1998) developed an interesting method, called ”triple-goal” estimation method, which produces estimates that perform reasonably well with respect to all three criteria.

The triple-goal estimation method involves the following three goals:

Goal 1: Produce element-specific point estimates with “optimality” qualities for the region of interest;

Goal 2: Obtain an ensemble of point estimates that best approximate the histogram of the true parameter ensemble, Louis (1984);

Goal 3: Rank within a selected ensemble.

In Section 5.2, we extend the triple-goal estimation for complex surveys and apply the methodology developed in Chapter 4 to benchmark triple-goal estimates. We analyze the baseball and unemployment rate data in Sections 5.3. and 5.4, respectively.

5.2 A Benchmarked Triple-Goal Estimation Procedure for Survey Data

Let $\hat{\pi}_i$ be the survey-weighted direct estimate of the true proportion π_i for the ith small area $(i = 1, \cdots, m)$. We are interested in producing an ensemble of triple-goal estimators of $\boldsymbol{\pi} = (\pi_1, \cdots, \pi_m)$. Let $\hat{\theta}_i = \arcsin(\sqrt{\hat{\pi}_i})$. We consider the following Bayesian model:

Model:

For $i = 1, \cdots, m$,

$$\begin{aligned} &(i) && \hat{\theta}_i|\theta_i \overset{\text{ind}}{\sim} N(\theta_i, \psi_i), \\ &(ii) && \theta_i|\beta, A \overset{\text{ind}}{\sim} N(x_i^T\beta, A), i = 1, \cdots, m; \\ &(iii) && f(\beta, A) \propto 1. \end{aligned}$$

In the above, $\psi_i = \frac{1}{4n_i}$, where n_i is the effective sample size for the ith small area. In this chapter, we use $n_i = \frac{\tilde{n}_i}{\text{deff}}$, where $\tilde{n}_i$ is the sample size for area i and deff is an estimate of design effect for a large area that covers the small area i.

The procedure for obtaining triple-goal estimators follows along the line of (Shen and Louis (1998)), which is described below:

First, we need to obtain an estimate of the empirical distribution function (EDF) of $\boldsymbol{\pi}$. The EDF of π_i is defined as:

$$F_m(t) = \frac{1}{m}\sum_{i=1}^{m} \mathcal{I}\{\pi_i \le t\}, \tag{5.1}$$

where $t \in \mathbb{R}$ and $\mathcal{I}$ is the indicator function. Under the following integrated squared error loss function:

$$\text{ISEL}(F_m, \tilde{F}_m) = \int \left[F_m(t) - \tilde{F}_m(t)\right]^2 dt, \tag{5.2}$$

the Bayes estimator of EDF is given by

$$\hat{F}_m(t) = E\left[F_m(t)|\hat{\theta}\right] = \frac{1}{m}\sum_{i=1}^{m} P(\pi_i \leq t|\hat{\theta}). \tag{5.3}$$

Second, we need to obtain rank of the parameter ensemble P. The rank is defined as

$$R_i = \text{rank}(\pi_i) = \sum_{j=1}^{m} \mathcal{I}\{\pi_i \geq \pi_j\}. \tag{5.4}$$

Under the rank squared error loss(RSEL), defined as

$$\text{RSEL}(\boldsymbol{R}, \tilde{\boldsymbol{R}}) = \frac{1}{m}\sum_{i=1}^{m}(R_i - \tilde{R}_i)^2, \tag{5.5}$$

the Bayes estimator of R_i is given by

$$\bar{R}_i = E(R_i|\hat{\theta}) = \sum_{j=1}^{m} P(\pi_i \geq \pi_j|\hat{\theta}) \tag{5.6}$$

The $\bar{R}_i$'s are not integers in general; however, it is easy to transform them in order and denote it as:

$$\hat{R}_i = \text{rank}(\bar{R}_i|\boldsymbol{R}), i = \ldots, m. \tag{5.7}$$

Finally, we generate an ensemble of point estimates, conditional on the optimal estimate of the ensemble EDF, $\hat{F}_m$, and the optimal estimate of the ranks, $\hat{R}_i$. Furthermore, the added constraint that $\hat{F}_m$ is a discrete distribution with at most m mass points, then the estimator is defined as:

$$\hat{\hat{\pi}}_i^{TG} = \hat{F}_m^{-1}\left(\frac{2\hat{R}_i - 1}{2m}\right), i = 1, \ldots, m. \tag{5.8}$$

We propose obtaining the estimators by using the Gibbs sampler, Gelfand and Smith (1990). To implement the Gibbs sampler, we obtain the full conditional under the hierarchical model as:

(a) $\theta_i|\beta, A, \hat{\theta} \overset{\text{ind}}{\sim} N\left[(1 - B_i)\hat{\theta}_i + B_i x_i'\beta, \frac{\psi_i A}{A+\psi_i}\right],\ i = 1, \cdots, m$

(b) $\beta|\theta, A, \hat{\theta} \sim N\left[(X^T X)^{-1} X^T \theta, A(X^T X)^{-1}\right]$

(c) $A|\beta, \theta, \hat{\theta} \sim IG\left[\frac{1}{2}\sum(\theta_i - x_i^T\beta)^2, \frac{m-2}{2}\right],$

where $B_i = \frac{\psi_i}{A+\psi_i},\ (i = 1, \cdots, m)$ and IG represents an inverted Gamma distribution. Then we apply the following algorithm.

Gibbs Sampling Algorithm:

(i) Draw $\theta_i^{(1)},\ i = 1, \cdots, m$, from (a), using $\beta^{(0)}$ & $A^{2(0)}$ as starting values. Obtain $\pi_i^{(1)} = \sin^2\left(\theta_i^{(1)}\right),\ i = 1, \cdots, m.$

(ii) Draw $\beta^{(1)}$ from (b) using $\theta^{(1)}$ & $A^{2(0)}$.

(iii) Draw $A^{2(1)}$ from (c), using $\theta^{(1)}$ & $\beta^{(1)}$.

The steps (i)-(iii) complete one cycle. Perform a large number of cycles. The simulated samples after deleting the first t "burn-in" samples, i.e.

$$\left\{\beta^{(t+r)}, A^{2(t+r)}, \pi^{(t+r)},\ r = 1, \cdots, R\right\}$$

are considered as R simulated samples from the posterior distribution of β, A, P.

The posterior density of π is approximated by

$$\left\{\pi^{(t+r)},\ r = 1, \cdots, R\right\}.$$

In particular, we need the following approximations:

$$\hat{F}_m(t) \approx \frac{1}{m}\sum_{i=1}^{m}\left\{\frac{1}{R}\sum_{r=1}^{R}\mathcal{I}\left[\pi_i^{(r)} \le t|\hat{\theta}\right]\right\} \tag{5.9}$$

$$\bar{R}_i = \approx \sum_{j=1}^{m}\left\{\frac{1}{R}\sum_{r=1}^{R}\mathcal{I}\left[\pi_i^{(r)} \le \pi_j^{(r)}|\hat{\theta}\right]\right\} \tag{5.10}$$

Finally, we apply the methodology developed in Chapter 4 to obtain benchmarked triple-goal estimates of $\pi_i\ (i = 1, \cdots, m), \hat{\pi}_i^{TG_b}$.

5.3 Baseball Data Analysis

In this section, we use the baseball data described in Chapter 4 to evaluate different estimators of ranks, EDFs , and point estimates of individual parameters.

For our estimators we use the following notations:

- $\hat{\pi}^{pm}$: posterior estimate
- $\hat{\pi}^{TG}$: triple-goal estimate
- $\hat{\pi}^{TG_r}$: triple-goal estimate with the ratio benchmarking constraint
- $\hat{\pi}^{cb}$: constrained Bayes, Ghosh (1992)
- $\hat{\pi}^{bm2}$: constrained Bayes, Datta et al. (2009)

We compare different estimators by the following four different summary statistics:

- Sum of Squared Error Loss (SSEL): $\frac{1}{m}\sum_{i=1}^{m}(\hat{\pi}_i - \pi_i)^2$
- Integrated Squared Error Loss (ISEL): $\int \left[F_m(t) - \tilde{F}_m(t)\right]^2 dt$
- Ratio between Posterior Sample Variance (RPSV): $\frac{\sum_{i=1}^{m}(\hat{\pi}_i - \bar{\hat{\pi}})^2}{\sum_{i=1}^{m}(\pi_i - \bar{\pi})^2}$
- Rank Squared Error Loss (RSEL): $\frac{1}{m}\sum_{i=1}^{m}(\hat{R}_i - R_i)^2$,

where $\bar{\pi}_i$ is the average of true π_i's.

The SSEL produces a summary statistic from each estimate, $\hat{\pi}_i$, against the corresponding true value, π_i. As an aggregate sum, there is little difference among all estimators. This result coincides with the result in Chapter 4 that the posterior mean performs better than other estimators. However, when we consider the ISEL, which measures the squared error loss between ECDF of two parameters, it shows that the both triple-goal estimators outperform other estimators. The advantage of the triple-goal estimators is more clear in Figure 5.1 that the histogram from the triple-goal estimates is closer in shape to the true histogram than any other estimators. Even after we benchmark our estimators, the histogram of our final benchmarked triple-goal estimates still retain the shape of the triple-goal estimates.

It's clear from RPSV that compared to other estimators, the hierarchical Bayes estimator,$\hat{\pi}^{pm}$, has an over-shrinkage problem. In other words, each estimate $\hat{\pi}_i$ does not deviate too much from its average value.

The RSEL, rank squared error loss, shows the summary static between the estimated rank against the true rank. The TG estimators perform better than other estimators since the estimators are obtained by optimizing under the RSEL function. Note that the RSEL value for $\hat{\pi}^{pm}$ is the same as that of $\hat{\pi}^{cb}$. It's always true that the ranks based on the cb estimates are always identical with the ranks based on the posterior means, Shen and Louis (1998).

	$\hat{\pi}^{pm}$	$\hat{\pi}^{TG}$	$\hat{\pi}^{TG_r}$	$\hat{\pi}^{cb}$	$\hat{\pi}^{bm2}$
SEL	0.01335	0.01581	0.01590	0.01724	0.01820
ISEL	0.00137	0.00023	0.00024	0.00074	0.00066
RPSV	0.33311	1.18150	1.19587	1.17878	1.29976
RSEL	26.66667	23.44444	23.44444	26.66667	26.66667

Table 5.1: Summary statistics for different estimators

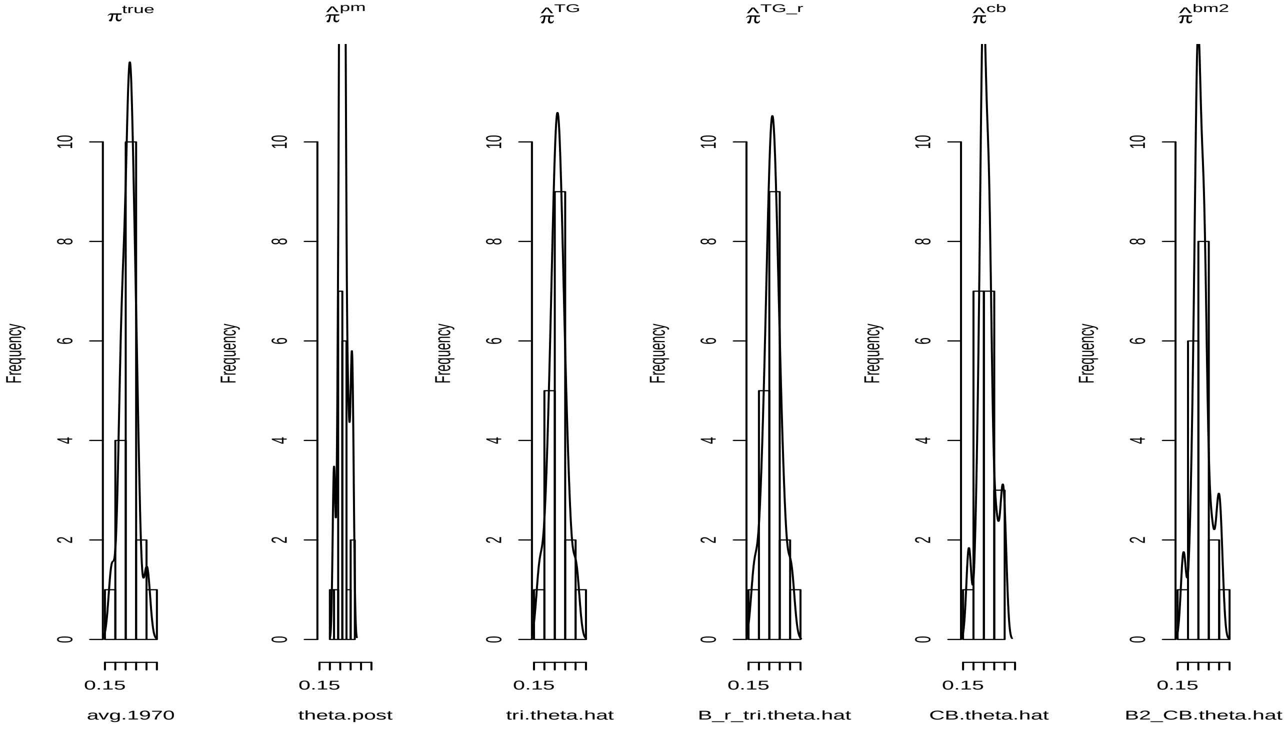


Figure 5.1: Histogram of different estimators

5.4 Estimation of unemployment rates for US states

In this section, we analyze the same CPS unemployment rate data we used in Chapter 4. We first begin with investigating benchmarking properties of the hierarchical Bayes (i.e., the posterior mean) and triple-goal estimates for 48 months of data. That is, we investigate how close the hierarchical Bayes and the triple-goal estimates are to the survey-weighted national estimates for a given month, when aggregated over all the 50 states and the District of Columbia.

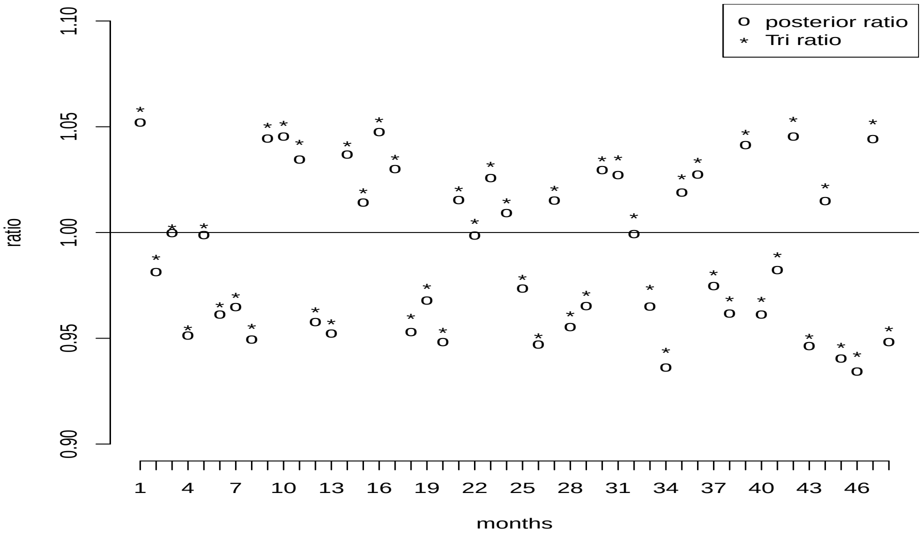


Figure 5.2: Ratio between national estimates and other estimates

Figure 5.2 displays the following ratios:

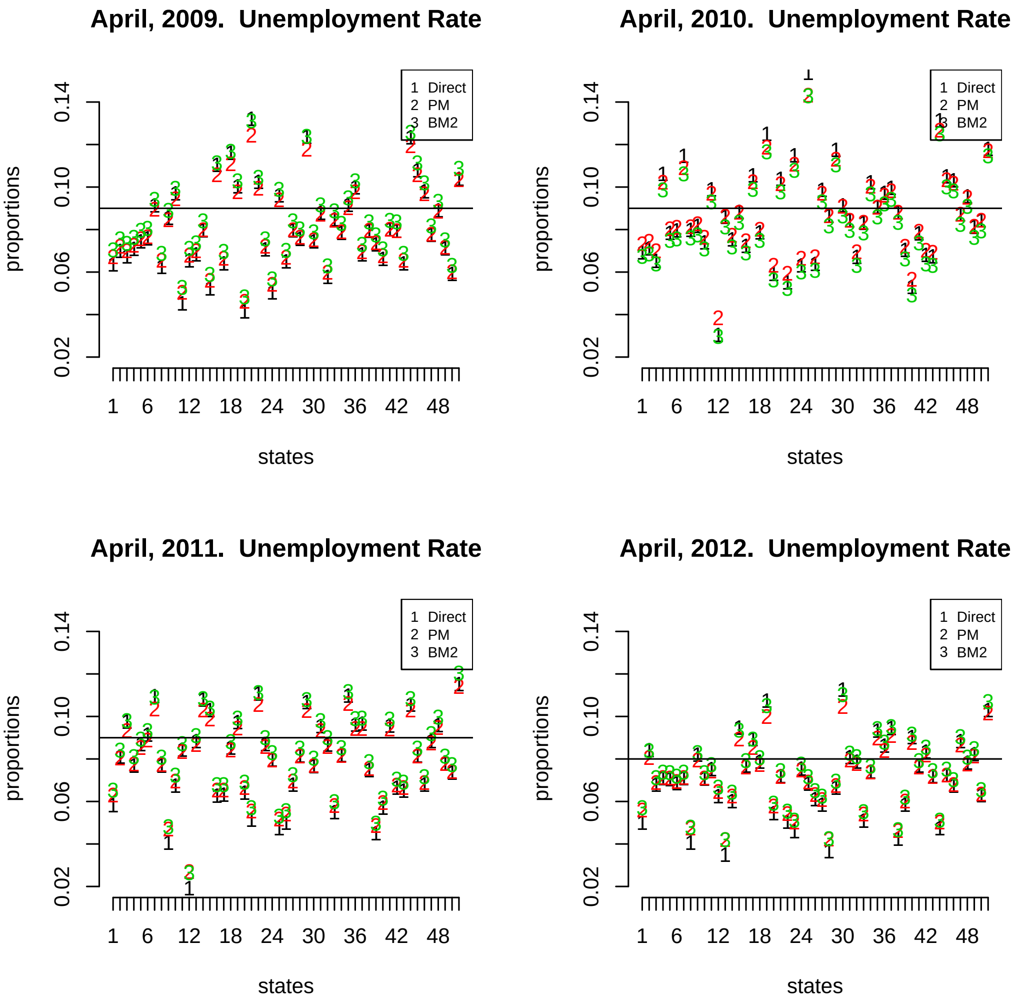
$$\text{posterior ratio} = \frac{\sum_{i=1}^{51} N_i \hat{\pi}_i^{pm}}{\text{total number of unemployed}}$$

$$\text{Tri ratio} = \frac{\sum_{i=1}^{51} N_i \hat{\pi}_i^{TG}}{\text{total number of unemployed}},$$

where N_i, $\hat{\pi}_i^{pm}$ and $\hat{\pi}_i^{TG}$ denote the number of persons in the labor force, hierarchical

Bayes and triple-goal estimates of unemployment rate for the ith state ($i = 1, \cdots, 51$). If a ratio lies on the reference line at 1, the corresponding state estimates are perfectly benchmarked. With respect to the benchmarking criteria, both estimates are behaving similarly and they do not show any systematic pattern over different months. From the figure, we observe that triple-goal estimates are always slightly higher than the corresponding hierarchical Bayes estimates. Generally, they are both hierarchical Bayes and triple-goal estimates are within 5% of the national estimates. Needless to say by construction our benchmarked triple-goal estimates are perfectly benchmarked.

In Figure 5.3, we plot different estimates of unemployment rates for 50 states and the District of Columbia for the month of April over four consecutive years. In each graph, the reference line corresponds to the national estimate. The states are arranged in increasing order of the effective sample sizes. For the larger states, different estimates are closer than the small states.


April, 2009. Unemployment Rate

April, 2010. Unemployment Rate

April, 2011. Unemployment Rate

April, 2012. Unemployment Rate

Figure 5.3: Unemployment rate for each state

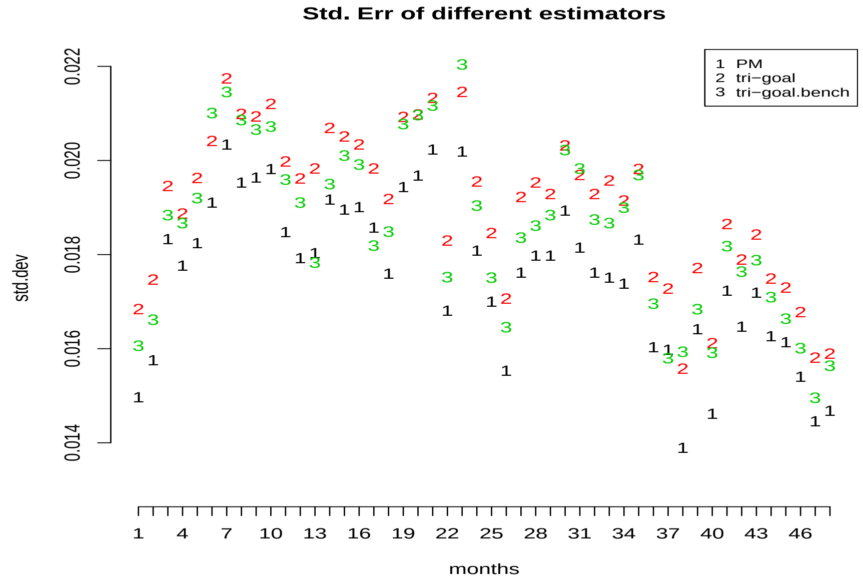


Figure 5.4: Monthly Std. Err. of estimates

When we compare the variability of the estimates across states, we find the hierarchical Bayes estimates vary the least among all the other estimates, supporting our theory. It is shown in Figure 5.4, in which we plotted standard errors of the 51 states each month for posterior means, triple goal estimators and benchmarked triple goal estimators. The picture clearly supports our claim that the posterior mean has the lowest standard errors. It's interesting to note that the benchmarked triple goal has standard errors between the other two estimates in most cases, and the difference between triple-goal and benchmarked triple-goal estimators is closer than that between posterior and benchmarked triple-goal estimators.

5.5 Discussion

In this chapter, we have explored the triple goal estimates developed by Shen and Louis (1998) and expanded it by putting benchmarking conditions as constraints. Previously in Chapter 4, our starting estimator was the posterior mean, but in this chapter, we have used the triple-goal estimators as our starting estimator. From our result, we can see that the benchmarked triple goal estimators are still successful at preserving the triple-goal properties while maintaining the benchmarked property.

Chapter 6

Future Research

We would like to conduct our future research in two different directions. In Chapter 4 and 5 of this dissertation, we have explored different benchmarking estimation methods. However, like other articles in this area of research, we have not explored any method for measuring their uncertainty. Since proposed benchmarked methods are not fully Bayesian, it does not seem reasonable to use posterior variance as an uncertainty measure. One possibility is to estimate the mean squared error (MSE) of the benchmarked estimator. Even though there are many different methods for estimating the MSE, the parametric bootstrap appears to be most promising. Chatterjee et al. (2008) described such method in the context of constructing confidence intervals based on empirical best predictors (EBLUP). While it seems straightforward to apply their method to estimate MSE of the benchmarked estimator, the theoretical properties of parametric bootstrap for benchmarked estimators are unknown. This could be an interesting research area to pursue.

Second, throughout this dissertation, we have used different area-level models for generating our estimates. We would like to explore benchmarking for unit-level models, similar to the BHF model. The challenge would be to find an appropriate model that captures different salient features of the complex survey design. Datta and Ghosh (1991) proposed a general Bayesian framework for linear mixed models with particular emphasis on small area estimation. For the future research, we would like to expand their method for non-linear unit level models with bench-marking properties.

Bibliography

Albert, J. (2007). *Bayesian Computation with R*. Springer, New York, NY.

Arora, V. and Lahiri, P. (1997). On the superiority of the Bayesian method over the BLUP in small area estimation problems. *Statistical Sinica*, 7:pp. 1053–1063.

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. *Journal of the American Statistical Association*, 83:28–36.

Bell, W., Basel, W., Cruse, C., Dalzell, L., Maples, J., Ohara, B., and Powers, D. (2007). Use of Data to Produce SAIPE Model-Based Estimates of Poverty for Counties. Research report series, U.S. Census Bureau.

Bell, W. R., Datta, G. S., and Ghosh, M. (2013). Benchmarking small area estimators. *Biometrika*, 100:189–202.

Bell, W. R. and Hillmer, S. C. (1990). The time series approach to estimation for repeated surveys. *Survey Methodology*, 16:195–215.

Brown, G., Chambers, R., Heady, P., and Heasman, D. (2001). Evaluation of small area estimation methods - An application to unemployment estimates from the UK LFS. In *Proceedings of Statistics Canada Symposium 2001 Achieving Data Quality in a Statistical Agency: A methodological Perspective*.

Carter, G. M. and Rolph, J. E. (1974). Empirical bayes methods applied to estimating fire alarm probabilities. *Journal of the American Statistical Association*, 69(348):pp. 880–885.

Chatterjee, S., Lahiri, P., and Li, H. (2008). Parametric bootstrap approximation to the disribution of EBLUP, and related prediction intervals in linear mixed models. *Annals of Statistics*, 36:1221–1245.

Chen, S. (2001). *Empirical best prediction and hierarchical Bayes methods in small area estimation*. PhD thesis, University of Nebraska, Lincoln.

Chen, S. and Ravallion, M. (2008). The developing world is poorer than we thought, but no less successful in the fight against poverty. Policy Research Working Paper Series 4703, The World Bank.

Citro, C. and Kalton, G. (2000). Small-Area Income and Poverty Estimates: Priorities for 2000 and Beyond. In *Proceedings of the Survey Research Methods Section, ASA*.

Cochran, W. G. (1977). *Sampling Techniques*. Wiley, New York.

Conlon, E. M. and Louis, T. A. (1999). Addressing multiple goals in evaluating region-specific risk using Bayesian methods. In *In; Disease Mapping and Risk Assessment for Pulbic Health*, pages 31–47. Wiley, Chichester.

Cooke, G. and Lawton, K. (2008). Woking out of poverty. Technical report, Institute for Public Policy Research.

Cowles, M. K. and Carlin, B. P. (1996). Markov Chain and Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, 91:883–904.

Dagum, E. B. and Cholette, P. A. (2006). *Benchmarking, Temporal Distribution, and Reconciliation Methods for Time Series*. Springer, New York, NY.

Datta, G., Lahiri, P., T., M., and Lu, K. (1999). Hierarchical Bayes Estimation of Unemplyment Rates for the States of the U.S. *Journal of the American Statistical Association*, 94:1074–1082.

Datta, G. S. and Ghosh, M. (1991). Bayesian prediction in linear models: Application to small area estimation. *Annals of Statistics*, 19:1748–1770.

Datta, G. S., Ghosh, M., Steorts, R., and Maples, J. (2009). Bayesian Benchmarking with Applications to Small Area Estimation. Research report series, The U.S. Census Bureau.

Dempster, A. P. and Toberlin, T. J. (1980). The analhysis of Census undercount from a postenumeration survey. In *Proceedings of the Survey Research Methods Section, ASA.*, pages 88–94.

Devine, O. J. and Louis, T. A. (1994). A constrained empirical bayes estimator for incidence rates in areas with small populations. *Statistics in Medicine*, 13:1119–1133.

Efron, B. (1975). Biased versus unbiased estimation. *Advances in Mathematics*, 16:117–277.

Efron, B. and Morris, C. (1973). Stein's estimation rule and ints competitors -an empirical Bayes approach. *Journal of the American Statistical Association*, 68:117–130.

Efron, B. and Morris, C. (1975). Data analysis using stein's estimator and its generalizations. *Journal of the American Statistical Association*, 70(350):pp. 311–319.

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.

Foster, J., Greer, J., and Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica*, 52(3):pp. 761–766.

Frey, J. and Cressie, N. (2003). Some results on constraned Bayes estimators. *Statistics and Probablity Letters*, 65:pp. 389–399.

Fuller, W. (2009). *Sampling Statistics*. Springer, New York, NY.

Gelfand, A. and Smith, A. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:267–277.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models.

Bayesian Analysis, 1:515–533.

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2004). *Bayesian Data Analysis*. Chapman and Hall/CRC, New York, NY.

Gelman, A. and Price, P. N. (1999). All maps of parameter estimates are misleading. *Statistics in Medicine*, 18:3221–3234.

Gershunskaya, J. B. and Lahiri, P. (2005). Variance estimation for domains in the U.S. current employment statistics program. In *Proceedings of the Survey Research Methods Section, ASA*, pages 3044–3051.

Ghosh, M. (1992). Constrained Bayes estimation with applications. *Journal of the American Statistical Association*, 87:533–540.

Glasinovic, V. (2010). The Politics of Poverty Measurement: The Chilean Case. Technical report, Sociedad Chilena de Políticas Públicas.

Goldstein, H. and Spiegelhalter, D. J. (1996). League tables and their limitations: statistical isses in comparisons of institutional performance. *Journal of Royal Statistical Society, Series A*, 159:385–409.

Gonzales, M. E. (1973). Use and evaluation of synthetic estimation. In *Proceedings of the Social Statistics Section, ASA*, pages 33–36.

Hansen, M., Hurwitz, W., and Bershad, M. (1961). Measurement Errors in Censuses and Surveys. *Bulletin of the International Statistical Institute*, 38:pp. 359–374.

James, W. and Stein, C. (1961). Estimation with Quadratic Loss. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, pages 361–379.

Jiang, J. (2007). *Linear and Generalized Linear Mixed Models and Their Applications*. Springer, New York, NY.

Jiang, J. and Lahiri, P. (2001). Empirical best prediction for small area inference with binary data. *Annals of Institute of Statistical Mathematics*, 53:217–243.

Jiang, J. and Lahiri, P. (2006a). Estimation of finite population domain means: A model-assisted empirical best prediction approach. *Journal of the American Statistical Association*, 101:301–311.

Jiang, J. and Lahiri, P. (2006b). Mixed model prediction and small area estimation. *Test*, 15:111–999.

Jiang, J., Lahiri, P., Wan, S., and Wu, C. (2001). Jackknifying in the Fay-Herriot model with an example. In *Proceedings of the Seminar on Funding Opportunity in Survey Research Council of Professional Associations on Federal Statistics*.

Kalton, G. (2002). Models in the Practice of Survey Sampling. *Journal of Official Statistics*, 18(2):pp. 129–154.

Kish, L. (1965). *Survey Sampling*. Wiley, New York, NY.

Lahiri, P. (1990). Adjusted Bayes and Emprical Bayes estimation in population sampling. *Sankhya*, 52:50–66.

Lahiri, P. (2001). *Model Selection*, volume 38. IMS Lecture Notes/Monograph.

Lahiri, P. and Mukherjee, K. (2007). Hierarchical bayes estimation of small area means under generalized linear models and design consistency. *Annals of Statistics*, 35:724–737.

Laird, N. M. and Louis, T. A. (1989). Empirical Bayes ranking methods. *Journal of Educational Statistics*, 14:29–46.

Landrum, M. B., Bronskill, S. E., and Normand, S. (2000). Analytic methods for constructing cross-sectional profiles of health care providers. *Health Services Outcomes Research Methodology*, 1:23–47.

Laud, P. W., Wisconsin, M. C., and Ibrahim, J. G. (1995). Predictive model selection.

Journal of the Royal Statistical Society, Ser. B, 57:247–262.

Liu, B., Lahiri, P., and Kalton, G. (2007). Hierarchical Bayes Modleing of Survey-Weighted Small Area Proportions. In *Proceedings of the Survey Research Methods Section, ASA*.

Lohr, S. (1999). *Sampling, Design and Analysis*. Duxbury, Pacific Grove, CA.

Louis, T. (1984). Estimating a population of parameter vallues using Bayes and empirical Bayes methods. *Journal of the American Statistical Association*, 79:393–398.

MacGibbon, B. and Tomberlin, T. (1989). Small area estimates of proportions via empirical Bayes techniques. *Survey Methodology*, 15:237–252.

Malec, D., Davis, W., and Cao, X. (1999). Small area estimates of overweight prevalence using sample selection adjustment. *Statistics in Medicine*, 18:3189–3200.

Malec, D., Sedransk, J., Moriarity, C., and LeClere, F. (1997). Small area inference for binary variables in National Health Interview Survey. *Journal of the American Statitical Association*, 92:815–826.

McCullagh, P. and Nelder, J. (1983). *Generalized Linear Models*. Chapman and Hall/CRC, New York, NY.

Mohadjer, L., Rao, J. N. K., Liu, B., Krenzke, T., and van de Kerckhov, W. (2007). Hierarchical Bayes small are estimates of adlt literacy using unmatched sampling and linking models. In *Proceedings of the Survey Research Methods Section, ASA.*, pages 3203–3210.

Molina, I., Nandram, B., and Rao, J. N. K. (2012). Hierarchical Bayes Small Area Estimation of General Parameters with Application to Poverty Indicators. *Biometrika*, 98:1–24.

Molina, I. and Rao, J. N. K. (2010). Small area estimation of poverty indicators. *The*

Canadian Journal of Statistics, 38(3):pp. 369–385.

Morris, C. (1983). Parametric empirical Bayes inference: Theory and applications. *Journal of the American Statistical Association*, 78:47–55.

Morris, C. N. and Christiansen, C. (1996). Hierarchical models for ranking and for identifying extremes with applications. *Bayes Statistics*, 5:277–297.

Moura, F. and Holt, D. (1999). Small area estimation using multilevel models. *Survey Methodology*, 25:73–80.

Otto, M. and Bell, W. (1995). Sampling Error Modelling of Poverty and Income Statistics for States. In *Proceedings of the Government Statics Section, ASA*.

Pfeffermann, D. and Barnard, C. H. (1991). Some new estimators for small area means with application to the assessment of farmland values. *Journal of Business and Economic Statistics*, 9:73–83.

Pfeffermann, D., Sikov, A., and Tiller, R. (2013). Single and two-stage cross-sectional and time series benchmarking procedures for small area estimation.

Pfeffermann, D. and Tiller, R. (2002). State Space Modelling with Correlated Measurements with Application to Small Area Estimation Under Benchmark Constraints. In *State Space and Unobserved Components Models in Honour of Professor J. Durbin*.

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

Prasad, N. G. N. and Rao, J. N. K. (1990). The estimation of mean squared errors of small area estimators. *Journal of the American Statistical Association*, 85:163–171.

Prasad, N. G. N. and Rao, J. N. K. (1999). On Robust Small Area Estimation Using a Simple Random Effects Model. *Survey Methodology*, 25:67–72.

Rao, J. N. K. (2003). *Small Area Estimation*. Wiley Series in Survey Methodology, Hoboken, NJ.

Ravallion, M. (2010). Poverty Lines across the World. Policy Research Working Paper Series 5284, The World Bank.

Rendall, M., Handcock, M., and Jonsson, S. (2009). Bayesian estimation of Hispanic fertility hazards from survey and population data. *Demography*, 46:65–83.

Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer, New York, NY.

Sarndal, C. E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer-Verlag, New York.

Shen, W. and Louis, T. (1998). Triple-goal estimates in two-stage hierarchical models. *Journal of Royal Statistical Society, Series B*, 60:455–471.

Spiegelhalter, D. J., Best, N. G., and Carlin, B. P. (1998). Bayesian deviance, the effective number of parameters, and the comparison of arbitrarily complex models. Technical report.

Stein, C. (1955). Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, pages 197–206.

Stroud, T. W. F. (1991). Hierarchical Bayes predicative means and variances with application to sample survey inference. *Communications in Statistics- Theory and Methods*, 20:13–36.

Tak, H. S. and Morris, C. (2012). Binomial regression interactive multilevel modeling.

Tomberlin, T. J. (1988). Predicting accident frequencies for drivers classified by two factors. *Journal of the American Statistical Association*, 83:309–321.

Tontisirin, K. and Haen, H. d. (2001). *Human energy requirement*. Report of a Joint FAO/WHO/ UNU Expert Consultation.

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

Wolter, K. (1985). *Introduction to Variance Estimation*. Springer, New York, NY.

Wong, G. Y. and Mason, W. M. (1985). The hierarchical logistic regression model for multilevel analysis. *Journal of the American Statistical Association*, 80:513–524.

You, Y. and Rao, J. (2002). A pseudo-empirical best linear unbiased prediction approach to small area estimation using survey weights. *Canadian Journal of Statistics*, 30:431–439.