

**UNIVERSITÀ
DEGLI STUDI
DI PADOVA**

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Statistiche
SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE STATISTICHE
CICLO XXIII

A Flexible Characterization of Models for Small Area Estimation: Theoretical Developments and Applications

Direttore della Scuola: Prof. ssa ALESSANDRA SALVAN

Supervisore: Prof. NICOLA TORELLI

Co-supervisore: Prof. GAURI. S. DATTA

Dottorando: ANTONY KIBIRA WANJOYA

Data: Febbraio 14, 2011.

“... to the acknowledgement of the mystery of God, and of the Father, and of Christ; In whom are hid all the treasures of wisdom and knowledge”.

Colossians 2:2-3.

Abstract

The demand for reliable small area estimates derived from survey data has increased greatly in recent years due to, among other things, their growing use in formulating policies and programs, allocation of government funds, regional planning, small area business decisions and other applications. Traditional area-specific (direct) estimates may not provide acceptable precision for small areas because sample sizes are seldom large enough in many small areas of interest. This makes it necessary to borrow information across related areas through indirect estimation based on models, using auxiliary information such as recent census data and current administrative data. Methods based on models are now widely accepted. The principal focus of this thesis is the development of a flexible modeling strategy in small area estimation with demonstrations and evaluations using the 1989 United States census bureau median income dataset.

This dissertation is divided into two main parts, the first part deals with development of the proposed model and comparison of this model to the standard area-level Fay-Herriot model through the empirical Bayes (EB) approach. Results from these two models are compared in terms of average relative bias, average squared relative bias, average absolute bias, average squared deviation as well as the empirical mean square error. The proposed model exhibits remarkably better performance over the standard Fay-Herriot model.

The second part represents our attempt to construct a hierarchical Bayes (HB) approach to estimate parameters for the proposed model, with implementation carried out by Markov chain Monte Carlo (MCMC) techniques. MCMC was implemented via the Gibbs sampling algorithm using *R* software package. We used several graphical tools to assess convergence and determine the length of the burn-in period. Results from the two models are compared in terms of average relative bias, average squared relative bias and average absolute bias. Our empirical results highlight the superiority of using the proposed model over the Fay-Herriot model. However, the advantage of the proposed model comes at a price since its implementation is mildly more difficult than the Fay-Herriot model.

Riassunto

L'esigenza di stime affidabili per piccole aree tratte da sondaggi è cresciuta notevolmente negli ultimi anni, grazie all'aumento del loro utilizzo nella formulazione delle politiche, nella ripartizione dei fondi statali, nella pianificazione regionale, nelle applicazioni business e in altre applicazioni. Le tradizionali stime specifiche per l'area (stime dirette) potrebbero non fornire una precisione accettabile, perché la numerosità campionaria in molte delle piccole aree d'interesse potrebbe essere ridotta o nulla. Questo rende necessario sfruttare le informazioni dalle zone simili, tramite una stima indiretta basata sui modelli per informazioni ausiliarie come i dati dei censimenti o i dati amministrativi. I metodi basati sui modelli sono ora piuttosto diffusi. L'attenzione principale di questa tesi è sviluppare una strategia di modellazione flessibile nella stima di piccole aree, e la sua valutazione utilizzando il Censimento negli Stati Uniti sul reddito mediano, del 1989.

Questa dissertazione è composta di due parti : la prima tratta lo sviluppo del modello e il confronto del modello proposto con il modello standard di Fay-Herriot tramite l'approccio di Bayes empirico. I risultati per questi due modelli sono stati confrontati in termini del bias relativo medio, del bias quadratico medio, del bias medio assoluto, della deviazione quadratica media ed inoltre in termini del errore quadratico medio empirico. Il modello proposto dimostra un rendimento assai migliore rispetto al modello standard di Fay-Herriot.

La seconda parte presenta il nostro tentativo di costruire un approccio di Bayes Gerarchico per la stima dei parametri del modello proposto, con l'attuazione delle tecniche di Markov Chain Monte Carlo (MCMC). MCMC è stato utilizzato tramite l'algoritmo di campionamento Gibbs, utilizzando il software *R*. I risultati dai due modelli sono stati confrontati in termini di bias relativo medio, bias relativo quadratico medio e il bias assoluto medio. I nostri risultati empirici sottolineano la superiorità del modello proposto rispetto al modello Fay-Herriot. Tuttavia, il vantaggio del modello proposto è limitato visto che la sua attuazione è leggermente più complicata rispetto al modello di Fay-Herriot.

Acknowledgements

During the period of my study, I have had cause to be grateful for the advice, support and understanding of many people. In particular I would like to express my boundless gratitude to my supervisor Prof. Nicola Torelli for helping me get into the problem of small area estimation and arranging valuable international contacts. He has guided me in my research through the use of his profound knowledge, insightful thinking, thoughtful suggestions as well as his unflagging enthusiasm. I really appreciate the enormous amount of time and patience he has spent on me. This work has also benefited greatly from much advice, and many useful discussions with Prof. Gauri S. Datta and Prof. Abhyuday Mandal of the University of Georgia, USA.

As well as the people mentioned above, my special thanks go to the director of the PhD school, Prof. Alessandra Salvan for her consistent support and encouragement throughout my study and members of department of statistics for the interesting lectures in fundamental statistical courses they offered to us. A mention must also go to my wonderful classmates, current and former students of the PhD school who provided pleasant atmosphere and useful distraction during my stay in Padua.

I gratefully acknowledge the financial support from Fondazione Cassa di Risparmio di Padova e Rovigo (CARIPARO), Italy as well as study leave granted to me by the Department of Statistics and Actuarial Science, Jomo Kenyatta University of Agriculture and Technology, Kenya.

I am also very grateful to Joyce, Njeri, Reuben and Dr. Mwangi, for being the surrogate family during the six months I stayed in Athens, GA, USA. and hopefully for their continued moral support thereafter.

Finally but certainly not least, I want to remember my wife Elizabeth, our daughter Rita, my Late Dad and the entire family in Kenya whose unlimited support, prayers, and encouragement carried me through times of despondency and threatening despair.

Antony Kibira Wanjoya

Padua

February 14, 2011.

To my wife Elizabeth and our daughter Rita

In memory of my parents

To my grandmother Beth, my source of inspiration.

Contents

Abstract	ii
Riassunto	iii
Acknowledgements	iv
List of Figures	x
List of Tables	xi
List of Abbreviations	xiii
1 Introduction	1
1.1 Overview	1
1.2 Main Contributions of the Thesis	4
2 Approaches to Small Area Estimation	6
2.1 Introduction	6
2.2 Direct Estimation	6
2.2.1 Horvitz-Thompson (HT) Estimator	7
2.2.2 Generalized Regression (GREG) Estimator	8
2.2.3 Modified Direct Estimator	9
2.3 Indirect Estimation	9
2.3.1 Synthetic Estimation	10
2.3.2 Composite Estimation	11
2.3.3 Basic Area Level Model	12

5.2.2	Gibbs Sampling Algorithm	57
5.3	Proposed Model	58
5.3.1	Posterior Distributions	58
5.4	Simulation Study	62
5.4.1	Empirical Comparision of Hierarchical Bayes Estimates . . .	64
5.5	Simulation Results	65
5.5.1	Simulation: Data generated using Fay-Herriot model	66
5.5.2	Simulation: Data generated using proposed model when p=0.25.	67
5.5.3	Simulation: Data generated using proposed model when p=0.5.	69
5.5.4	Simulation: Data generated using proposed model when p=0.75.	70
5.6	Discussion of results	72
6	Concluding Remarks and Future Research	73
A	Program Documentation	76
A.1	Computational Aspects	76
A.2	Hierarchical Bayes Approach	76
	References	83

List of Figures

4.1	Plots of empirical values of MSE using double-bootstrap approach: Comparison of the Fay-Herriot model and the proposed model at $p = 0.25, 0.5$ and 0.75	52
5.1	Data simulated from the Fay-Herriot model: Histograms, trace plots and autocorrelation plots for β_0, β_1, A and p for the last 15,000 after a burn-in of 5,000 iterations. The point * in the histograms plots denotes the true value of the corresponding parameter.	66
5.2	Data simulated from the proposed model when $p = 0.25$: Histograms, trace plots and autocorrelation plots for β_0, β_1, A and p for the last 15,000 after a burn-in of 5,000 iterations.	68
5.3	Data simulated from the proposed model when $p = 0.5$: Histograms, trace plots and autocorrelation plots for β_0, β_1, A and p for the last 15,000 after a burn-in of 5,000 iterations. The point * in the histograms plots denotes the true value of the corresponding parameter.	69
5.4	Data simulated from the proposed model when $p = 0.75$: Histograms, trace plots and autocorrelation plots for β_0, β_1, A and p for the last 15,000 after a burn-in of 5,000 iterations.	71

List of Tables

4.1	Empirical Comparison of EB Estimates under Fay-Herriot Model (FH) and Proposed Model (PM) using Four Different Criteria. . . .	51
4.2	Empirical Comparison of EB Estimates under Fay-Herriot Model (FH) and Proposed Model (PM) using Four Different Criteria assuming that $p = 0.25$	51
4.3	Empirical Comparison of EB Estimates under Fay-Herriot Model (FH) and Proposed Model (PM) using Four Different Criteria assuming that $p = 0.5$ and 0.75	51
5.1	Empirical Comparison of HB estimates under Fay-Herriot model (FH) and proposed model (PM) using four different criteria. Data simulated from the Fay-Herriot model	66
5.2	Posterior summary statistics for data simulated from Fay-herriot model.	67
5.3	Empirical Comparison of HB estimates under Fay-Herriot model (FH) and proposed model (PM) when data is simulated from the proposed model assuming that $p = 0.25$	67
5.4	Posterior summary statistics for data simulated from the proposed model ($p = 0.25$).	68
5.5	Empirical Comparison of HB estimates under Fay-Herriot model (FH) and proposed model (PM) when data is simulated from the proposed model assuming that $p = 0.5$	69

5.6	Posterior summary statistics for data simulated from the proposed model ($p = 0.50$).	70
5.7	Empirical Comparison of HB estimates under Fay-Herriot model (FH) and proposed model (PM) when data is simulated from the proposed model assuming that $p = 0.75$	70
5.8	Posterior summary statistics for data simulated from the proposed model ($p = 0.75$).	71

List of Abbreviations

AAB	A verage A bsolute B ias
AIC	A kaike I nformation C riterion
ARB	A verage R elative B ias
ASD	A verage S quared D eviation
ASRD	A verage S quared R elative D eviation
BIC	B ayesian I nformation C riterion
BUGS	B ayesian I nference U sing G ibbs S ampling
DIC	D eviance I nformation C riterion
EB	E mpirical B ayes
FH	F ay H erriot
GREG	G eneralised R egression
HB	H ierarchical B ayes
MCMC	M arkov C hain M onte C arlo
ML	M aximum L ikelihood
MSE	M ean S quare E rror
REML	R esctricted M aximum L ikelihood
SAE	S mall A rea E stimation
SD	S tandard D eviation
SE	S tandard E rror
SMAS	S mall A rea S tatistics

Chapter 1

Introduction

1.1 Overview

In recent years, the statistical technique of small area estimation (SAE) has been a very hot topic, and there is an ever-growing demand for reliable estimates of small area populations of all types. Reliable estimates of the population of small areas are important for several reasons. These important estimates are used for, among other things, determination of state funding allocations, determination of exact boundaries for schools and voting districts, administrative planning, disease mapping and marketing guidance.

For historical and/or administrative reasons a country is usually divided into regions (municipalities) and these are further divided into districts (counties) or even further (streets or parts of streets). These smaller units of a country in demographical and epidemiological studies are normally referred to as *small areas*.

Data users and analysts have got their appetites whetted by richer (inter) national databases. Naturally, they seek further detail to extend their understanding of the data, this means that *small area data*, are needed. Such data may include labour force, income levels, unemployment, employment, marriages, etc. subdivided into geographical areas.

From these data, for instance, economic and social development indices could be calculated for each area. Development indices, which are mainly needed for the development projects, may be calculated using data on income, employment and unemployment levels which are to be supplied annually for each small area. Unemployment rates which may also be used for the distribution of unemployment benefits may be calculated from labour force survey.

In the medical area, the counts of a disease prevalence or incidence for each small area may be kept in addition to the expected counts of the disease. The expected counts are obtained by calculating what we would have often obtained if there was an equal risk applying to the whole geographical area, but taking into account the age by sex structure of the study population. Simple statistics such as the ratios and differences of the observed and expected counts can easily be computed to give the relative risk of the disease for each small area.

In recent years, the demand for *small area statistics* abbreviated as SMAS is increasing and this trend is likely to continue for many years. The reason for this continuous increase in the demand for SMAS are not far fetched and we will explain this below.

Traditionally, the national well-being of the population of a nation is measured by the income per capita. But the nation's per capita income per se does not indicate that every member of that nation enjoys the same level of economic and social development. Typically, within a nation, *small areas*, in the sequel referred to as *subregions* will deviate from this national average. If some subregions have much lower average per capita income, this causes concern and perhaps asks for remedial actions. The need to identify such subregions, to develop and stimulate policy solutions and to administer and evaluate the effects of programs on these subregions all imply a need for data and statistical information at the relevant geographical and demographic levels. The SMAS support the realisation of such objectives.

Further, it is common knowledge that many business decisions depend on local social, economic and environmental conditions. Statistics such as the percentage of

the population that is economically active and its further subdivision into males and females, the average monthly earnings per employee and electricity consumption in kilowatts per capita can be beneficial for entrepreneurs in their business decision taking process.

Politicians also show an increasing interest in obtaining statistical information for geographical areas. For instance when they want to assess the impact of certain medical policies or legislations on their constituents, they need (geographical) health information on the inhabitants. For instance, the Regional Industrial Development Program (RIDP) of Statistics Canada's small area data program requires the calculation of a "development index" for each of 260 Census Divisions (Brackstone, 1987). This index, an example of SMAS, is then used to rank the Census Divisions into four classes that qualify for successively higher maximum levels of government assistance for approved industrial development projects.

In the recent years, the increasing demand for SMAS has been recognized and strategies have been developed in several countries to achieve high quality SMAS. We now discuss what in the past affected most the availability of the SMAS. In the past, data and statistics were respectively collected and computed on national and provincial levels. For that reason, the volume of data and the statistics required did not warrant high speed and large computers. However, things are quite different these days. The subdivision of a nation into subregions and the extra details required to compare each subregion to national and provincial data, necessarily create the need for voluminous data sets. Since the subregions are heterogeneous, small area data collected over such regions have the characteristic of being heterogeneous as well. We can classify the factors that influence heterogeneity in the data into (1) uncontrollable and (2) controllable factors. Here by uncontrollable factors, we are referring to factors like: (i) environmental factors, (ii) exposure factors, (iii) demographic factors and (iv) differences among individuals. The controllable factors are for example those related to recording activities.

Luckily, things have improved considerably in recent years (at least in some western countries). For instance, we experience in the computer industry, a tremendous advance in software and hardware. Further, there has been much statistical research which resulted in new modelling methods that can accommodate the heterogeneous nature of small area data.

This prelude has introduced the idea of small area estimation and has highlighted that concepts of small area estimation can be applied to a wide variety of situations. The work presented in this thesis can be divided into two broad categories, the first, dealing with empirical Bayes approach with inclination towards small area estimation. The second category involves use of hierarchical Bayes method in small area estimation. The rest of this thesis is therefore structured in the following way. Chapter 2 gives an overview of approaches to small area estimation. In chapter 3, general concepts pertaining to frequentist and Bayesian methods are briefly reviewed. A new model for small area estimation is developed and compared to standard area level model (Fay and Herriot, 1979) through the empirical Bayes approach in chapter 4 . In chapter 5, a hierarchical Bayesian approach is developed to estimate parameters for the proposed model, with implementation carried out by Markov chain Monte Carlo techniques. The thesis ends with a discussion of the methods presented and possible future research projects which have been inspired by this work in chapter 6.

1.2 Main Contributions of the Thesis

Overall it is the aim of this dissertation to demonstrate the effectiveness of use of a new flexible methodology in small area estimation. We have considered both the area-level and unit-level models. We have formulated a new flexible small area model by incorporating a tuning parameter into the random part of standard area-level (Fay-Herriot) model as well as the standard unit-level small area model. Bayesian inference in small area estimation has received a renewed research effort.

This dissertation embodies three research contributions as well as a discussion of extensions of these.

The first contribution addresses the performance of the proposed flexible model as compared with standard area-level Fay-Herriot model in the context of empirical Bayes (EB) approach.

The second contribution develops tools to assess the performance of our proposed model in a hierarchical Bayes (HB) framework. The novelty here is that we have developed a flexible way to handle random effects in an area-level small area estimation model. The third contribution constructs tools for the empirical Bayes (EB) framework for the proposed flexible unit-level model as well as standard unit-level model.

Chapter 2

Approaches to Small Area Estimation

2.1 Introduction

In this chapter, we briefly review the approaches to small area estimation. Traditionally there are two types of small area estimation namely direct and indirect estimation. The direct small area estimation is based on survey design and includes the Horvitz - Thompson (HT) estimator, generalised regression (GREG) estimator and modified direct estimator. On the other hand, indirect approaches are mainly based on different statistical models and techniques. Implicit model based approaches include synthetic and composite estimations; whereas explicit models are categorized as area level and unit level models.

2.2 Direct Estimation

Direct estimates are classical design-based estimators that are obtained by applying survey weights to the sample units in each small area (Saei and Chambers, 2003). Direct estimators only rely on the sample obtained from the survey. For direct estimation all small areas must be sampled in order to produce these kinds

of estimates. Although it is rare, when survey samples are large enough to cover all the study areas with sufficient data in each area, different estimators can be developed. However, a major disadvantage of such estimators is that unacceptably large standard errors may result: this is especially true if the sample size within the small area is small. The following two estimators are common in direct estimation.

2.2.1 Horvitz-Thompson (HT) Estimator

The Horvitz-Thompson is the simplest direct estimator. A finite survey population Ω , consists of N distinct elements identified through the labels $\{1, 2, \dots, k, \dots, N\}$ contains E_i ($E_i \subset \Omega$) subpopulation for a small area i ($1, 2, \dots, m$) with N_i elements and $\sum_{i=1}^m N_i = N$. Consider a sample s ($s \subseteq \Omega$) drawn from Ω with a given probability design $p(\cdot)$, and s_i $s_i \subset s$ is the set of individuals that have been selected in the sample from the small area i . Suppose the inclusion probability $\pi_k = Pr(k \in s)$ is a strictly positive and known quantity. For the elements $k \in s_i$, let (y_{ik}, x_{ik}) be a set of sample observations, where y_{ik} is the value of variable of interest for the k^{th} unit in the small area i and x_{ik} is a vector of auxiliary information associated with y_{ik} .

Now, if Y_i and X_i represent the target variable and the available covariates for a small area i , then the Horvitz-Thompson estimator (Cochran, 1977) of the population total for i^{th} small area can be defined as

$$\hat{Y}_{i,HT} = \sum_{k \in s_i} d_{ik} y_{ik}$$

where $d_{ik} = 1/\pi_{ik}$ ($k \in s_i \subset s$) are design weights depending on the given probability sampling design $p(\cdot)$.

It is worth noting that in principle the HT estimator is not designed to use auxiliary information or covariates. However it is possible to consider auxiliary information to evaluate this estimator (Sarndal et al., 1992). When $\pi_{ik} > 0, \forall k \in E_i$ and there is sufficient sample size available at i^{th} small area, the HT estimator is

unbiased but not efficient. In the context of small area estimation problems, with an inadequate sample, HT estimator can be biased and more unreliable.

2.2.2 Generalized Regression (GREG) Estimator

The generalized regression (GREG) estimator is obtained by combining information about the auxiliary variables (X_i), with the individual sample information from the survey data, and can be defined as

$$\hat{Y}_{i,GREG} = X_i' \hat{\beta} + \sum_{k \in s_i} d_{ik} (y_{ik} - x_{ik}' \hat{\beta}),$$

$$\hat{\beta} = \left(\sum_{k \in s_i} d_{ik} x_{ik} x_{ik}' \right)^{-1} \left(\sum_{k \in s_i} d_{ik} x_{ik} y_{ik} \right).$$

where $\hat{\beta}$ is the sample weighted least square estimates of generalized regression (Sarndal et al., 1992, Rao, 2003) and other notations have similar definition to section 2.1.1.

GREG estimator could be negative in some small areas when the linear regression overestimates the variable of interest as well as when there is no sample observation in the small area i . A comprehensive discussion about the GREG estimator in the context of small area estimation is given by Rao (1999b). The GREG estimator is approximately design-unbiased for small area estimation but not consistent because of high residuals. This estimator can take more general form (by choosing the revised weights, which is the product of design weight and estimation weight, instead of design weights) to provide estimates for all target variables under different small areas, and hence the GREG estimator ensures consistency of results of different areas when aggregated over different variables of interest (see Rao, 2003, pg. 13).

2.2.3 Modified Direct Estimator

The modified direct estimator can be used to improve estimators reliability by borrowing strength for estimating the regression coefficient over small areas. If auxiliary information X_i in the i^{th} domain is available, then a modified direct estimator of population total is given by

$$\hat{Y}_{i,MD} = \hat{Y}_i + (X_i - \hat{X}_i)' \hat{\beta},$$

where \hat{Y}_i and \hat{X}_i are the HT estimators of the target variable Y_i and covariates X_i respectively for the small area i , and $\hat{\beta} = \left(\sum_{k \in s_i} d_k x_k x_k' \right)^{-1} \left(\sum_{k \in s_i} d_k x_k y_k' \right)$ is the overall sample weighted least square estimates of regression coefficients.

The modified direct estimator is approximately design-unbiased as the overall sample size increases, even if the regional sample size is small. Although the modified direct estimator borrows strength for estimating the overall regression coefficients, it does not increase the effective sample size, unlike indirect small area estimators (Rao, 2003). This estimator is also referred to in Woodruff (1966) , and Battese et al. (1988) as the modified GREG estimator or the “survey regression” estimator.

2.3 Indirect Estimation

Indirect or model-based small area estimators rely on statistical models to provide estimates for all small areas. Once the model is chosen, its parameters are estimated using the data obtained in the survey. An important issue in indirect small area estimation is that auxiliary information or covariates are needed. In this section we review four statistical techniques of indirect estimation.

2.3.1 Synthetic Estimation

The name synthetic estimator derives from the fact that these estimators borrow strength by synthesising data from many different areas. Gonzalez (1973) defines an estimator as synthetic when a reliable direct estimator for a large area is used to derive an indirect estimator for a small area belonging to the large area under the assumption that all small areas have the similar characteristics as the large area. In addition, Levy (1979) and Rao (2003) provide extensive overviews on various synthetic estimation approaches and its application in small areas estimation.

The synthetic estimators are derived by partitioning the whole population into a series of mutually exclusive and exhaustive cells and deriving the estimate as sum of products. Suppose the whole sample domain Y' is partitioned in Y_j 's large domains and a reliable direct estimate \hat{Y}_j of the j^{th} domain total can be obtained from the survey data; the small area, i , may cut across j so that $Y_j = \sum_i Y_{ij}$, where Y_{ij} is the total for cell (i,j) . Further let X_{ij} be auxiliary information total available for the i^{th} small area within j^{th} . Then synthetic estimator of small area total $Y_i = \sum_j Y_{ij}$ can be defined as

$$\hat{Y}_{i,S} = \sum_j (X_{ij}/X_{.j})\hat{Y}_j,$$

where $X_{.j} = \sum_i X_{ij}$ (Ghosh and Rao, 1994, see). This estimator is also known as the ratio-synthetic estimator. The most commonly used regression-synthetic estimator of the i^{th} area population total can be defined as

$$\hat{Y}_{i,S} = X'_i\hat{\beta},$$

where X_i is the known total of available auxiliary information in a small area i and β denotes estimates of population regression coefficients. Synthetic estimator performs well only if the true regression coefficient happens to be close to $\hat{\beta}$

in all considered areas. Otherwise the synthetic estimator can be highly biased in some areas. The assumption that the area-specific regression coefficients are approximately equal is what Rao (2003) calls implicit modelling.

2.3.2 Composite Estimation

In the practise of survey statistics, as the sample size in a small area increases, a direct estimator becomes more desirable than a synthetic estimator. This holds whether or not the surveys are designed to produce estimates for small areas. In other words, when area level sample sizes are relatively small the synthetic estimator outperforms the traditional direct estimator, whereas when the sample sizes are large enough the direct estimator outperforms the synthetic estimator. A weighted sum of these two estimators would be an ideal alternative to choosing one over the other in order to balance their degree of bias, and this type of estimator is called composite estimator.

According to Ghosh and Rao (1994), the composite estimator of the population total Y_i for a small area i can be defined as

$$\hat{Y}_{i,C} = \phi_i \hat{Y}_{i,D} + (1 - \phi_i) \hat{Y}_{i,S},$$

where $\hat{Y}_{i,D}$ and $\hat{Y}_{i,S}$ denote direct and synthetic estimator of Y_i respectively. ϕ_i is a suitably chosen weight that lies between 0 and 1. Schaible (1978) provides a detailed discussion on choice of weights for the composite estimator. The choice of optimal weight (ϕ_i^{opt}) can be obtained by minimising the mean square error (MSE) of the composite estimator, $\hat{Y}_{i,C}$, with respect to ϕ_i (Ghosh and Rao, 1994, Rao, 2003). This yields

$$\phi_i^{opt} = \frac{MSE(\hat{Y}_{i,S})}{MSE(\hat{Y}_{i,D}) + MSE(\hat{Y}_{i,S})}.$$

A number of estimators proposed in literature have the form of composite estimators, for instance the James-Stein estimator proposed by James and Stein (1961) which considers common weight ϕ . Efron and Morris (1975) have generalised the James-Stein estimator. Composite estimators are biased and they may have improved precision depending on the selection of the weight.

2.3.3 Basic Area Level Model

Small area estimators are based on area level computations if the models link small area information on response variable to area-specific auxiliary variables. One of the most widely used area level models for small area estimation was proposed by Fay and Herriot (1979). According to the Fay-Herriot model, a basic area level model assumes that the small area parameter of interest θ_i is related to area-specific auxiliary data x_i through a linear model

$$\theta_i = x_i^T \beta + \nu_i, \quad i = 1, \dots, m. \quad (2.1)$$

where m is the number of small areas, $\beta = (\beta_1, \dots, \beta_p)'$ is $p \times 1$ vector of regression coefficients, and the ν_i 's are area-specific random effects assumed to be independent and identically distributed (iid) with $E(\nu_i) = 0$ and $var(\nu_i) = A$. The assumption of normality may also be included.

The area level model assumes that there exists a direct survey estimator y_i for the small area parameter θ_i such that

$$y_i = \theta_i + e_i, \quad i = 1, \dots, m. \quad (2.2)$$

where the e_i is the sampling error associated with the direct estimator y_i . We also assume that the e_i 's are independent normal random variables with mean $E(e_i|\theta_i) = 0$ and sampling variance $var(e_i|\theta_i) = D_i$. Combining models (2.1) and (2.2) yields the area level linear mixed model

$$y_i = x_i^T \beta + \nu_i + e_i. \quad (2.3)$$

Note that model (2.3) involves both model based random errors ν_i and design-based sampling errors e_i . It takes the sampling design into account through the survey weights in the direct estimator y_i . Rao (1999a) indicates that the success of small area estimation through a model based approach largely depends on getting good auxiliary information that leads to small model variance A , relative to known or estimated sampling variance D_i .

2.3.4 Basic Unit Level Model

The basic unit level model is based on unit level auxiliary variables. These variables are related to the unit level values of response through a nested error linear regression model. This type of model can be represented by the following equation

$$y_{ij} = x_{ij}^T \beta + \nu_i + e_{ij}. \quad (2.4)$$

where y_{ij} is the response of unit j , $j = 1, 2, \dots, N_i$, in area i , $i = 1, 2, \dots, m$. x_{ij} is the vector of auxiliary variables, β is the vector of regression parameters, ν_i is the random effect of area i and e_{ij} the individual unit error term. The area effects ν_i are assumed independent with mean zero and variance σ_u^2 . The errors e_{ij} are independent with mean zero and variance σ_e^2 . In addition, the ν_i 's and e_{ij} 's are assumed to be independent.

The nested error unit level regression model (2.4) was first used to model county crop areas in North-Central Iowa, USA (Battese et al., 1988).

2.4 Discussion

In this chapter, we have looked at the various approaches of small area estimation that have been described in the literature. In the practice of survey statistics the direct design-based estimators have been highly appreciated, because they are at least approximately design unbiased even in complex samplings designs. The model-based estimators, which rest on sampling from an infinite population characterized by a stochastic model, can suffer from severe bias if the model is not correct. In addition, they may be biased and inconsistent with respect to the chosen design.

In small area estimation the question is often about the trade off between bias and variance. With small sample sizes the unbiasedness of the direct estimators may be of no practical value due to large variance of the estimator. The model-based estimators are prone to bias, but they have the advantage of small variances compared to the design-based estimators. There is evidence that the model-based small area estimators outperform the direct estimators with respect to the estimation accuracy measured with mean squared error (MSE) (Torabi and Rao, 2008). This is possibly why the model-based approach is widely accepted as the framework for small area estimation.

Chapter 3

Review: Linear Mixed Models and Bayesian Methods

3.1 Introduction

In this chapter we present an overview of linear mixed model as well as its characteristics within the Bayesian framework. It is not intended to be an all-encompassing exposition on the subject, the rationale is to briefly explore the methods used for parameter estimation throughout the thesis. In addition, it serves as an introduction to the models used in chapter 4 and 5. The model-based small area estimation largely employs linear mixed models involving random area effects. The auxiliary variables are introduced in the fixed part of the model as covariates.

Linear mixed models have a wide range of applications. In particular, the ability to predict linear combination of fixed and random effects is one the more attractive properties of such models. In a series of papers, Henderson (1975) developed the best linear unbiased prediction (BLUP) method for mixed models. However, the BLUP methods described in Henderson (1975) assumed that the variances associated with random effects in the mixed model (the variance components)

are known. In practice such variance components are unknown and have to be estimated from the data. There are several methods for estimating variance components reviewed in Harville (1977). The predictor obtained from the BLUP when unknown variance components are replaced by associated estimators is called the empirical best linear unbiased predictor (EBLUP) and is described in Robinson (1991).

Mixed models have been used to improve estimation of small area characteristics of small area based on survey sampling or census data by Fay and Herriot (1979), Ghosh and Rao (1994), Rao (1999b) and Pfeiffermann (2002). In this context, estimation of a small area mean is equivalent to prediction of the realization of the unobservable random area effect in a linear mixed model for the superpopulation distribution of the variable defining this mean. See Valliant et al. (2000) for a discussion of the distinction between the traditional interpretation of model-based prediction and its application in survey sampling.

In addition to EBLUP, empirical Bayes (EB) and hierarchical Bayes (HB) estimation and inference methods have been also applied to small area estimation. Under the EB approach, Bayes estimation and inferential approaches are used in which posterior distributions are estimated from data. Under the HB approach, unknown model parameters (including variance components) are treated as random, with values drawn from specified prior distributions. Ghosh and Rao (1994) review the application of these estimation methods in small area estimation. You and Rao (2000) have used HB methods to estimate small area means under random effect models.

3.2 Linear Mixed Effects Models

A linear mixed effect model is a model that contains both the fixed and random effects. That is

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \mathbf{e} \quad (3.1)$$

where \mathbf{Y} is a $(n \times 1)$ vector of responses, \mathbf{X} is a $(n \times p)$ design matrix that characterizes the systematic part of the response, β is a $(p \times 1)$ vector of parameters referred to as fixed effects, that complete the characterization of the systematic part of the response. \mathbf{Z} is a $(n \times k)$ design matrix that characterizes random variation in the response attributable to among-unit sources. \mathbf{b} is a $(k \times 1)$ vector of random effects and \mathbf{e} is a $(n \times 1)$ vector of within-unit deviations denoting variation due to sources like measurement error.

The random effects \mathbf{b} and the residuals \mathbf{e} can come from any probability distribution, but they are commonly assumed to be independent. Another common assumption about these distributions is that they are both normal with mean zero, that is

$$\begin{bmatrix} \mathbf{b} \\ \mathbf{e} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix} \right)$$

where \mathbf{G} is a $(k \times k)$ covariance matrix that defines the among-unit sources and \mathbf{R} is a $(n \times n)$ covariance matrix that defines the within-unit sources. The covariance matrices \mathbf{G} and \mathbf{R} are functions of a set of variance parameters (σ) . Therefore we can write $\mathbf{G} = \mathbf{G}(\sigma)$, $\mathbf{R} = \mathbf{R}(\sigma)$ and $\Sigma = \Sigma(\sigma)$. With these assumptions, we have

$$\begin{aligned} E(\mathbf{y}) &= E(\mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \mathbf{e}) \\ &= \mathbf{X}\beta \quad \text{and} \end{aligned}$$

$$\begin{aligned} \text{var}(\mathbf{y}) &= \text{var}(\mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \mathbf{e}) \\ &= \text{var}(\mathbf{Z}\mathbf{b}) + \text{var}(\mathbf{e}) \\ &= \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R} \\ &= \Sigma \end{aligned}$$

The model with the above assumptions on \mathbf{b} and \mathbf{e} implies that \mathbf{Y} are multivariate normal random vectors of dimension n with a particular form of covariance matrix. That is

$$\mathbf{Y} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}). \quad (3.2)$$

3.2.1 Estimation of Model Parameters

ANOVA Method

In variance component models, where all the parameters in σ are variances, Analysis of Variance (ANOVA) method or the Henderson 3 method (Henderson, 1953) based on equating the ANOVA mean squares to their expected values and solving the estimates from the resulting system of equations can be used for estimating σ . The Henderson 3 method (also known as method of fitting constants) has been suggested for use in small area models (Prasad and Rao, 1990). ANOVA methods are non-iterative and therefore easy to implement, yield unbiased variance estimates and they do not require normality of random effects.

The regression coefficients $\boldsymbol{\beta}$ can be estimated by the generalized least squares:

$$\boldsymbol{\beta}_{GLS} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y} \quad (3.3)$$

The estimator $\boldsymbol{\beta}_{GLS}$ is the best linear unbiased estimator of $\boldsymbol{\beta}$. If the covariance matrix $\boldsymbol{\Sigma} = \mathbf{ZG}(\sigma)\mathbf{Z}^T + \mathbf{R}(\sigma)$ is unknown, it is replaced with its estimate $\hat{\boldsymbol{\Sigma}} = \mathbf{ZG}(\hat{\sigma})\mathbf{Z}^T + \mathbf{R}(\hat{\sigma})$, where $\hat{\sigma}$ is obtained via the ANOVA method, for example. Anova methods are only applicable to limited choice of models. In order to overcome this, we need to use maximum likelihood (ML) and restricted maximum likelihood (REML) estimation methods, which are applicable to more general models and boast of attractive properties like consistency, efficiency and asymptotic normality of the estimators (see (Searle et al., 1992)).

Maximum Likelihood Estimation (ML) and Residual Maximum Likelihood Estimation (REML)

In standard mixed model where both residual and random effects are normally distributed, the marginal distribution is also normal (Verbeke and Molenberghs, 1997) with functional form

$$f(\mathbf{y}; \beta, \Sigma) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{X}\beta)' \Sigma^{-1} (\mathbf{y} - \mathbf{X}\beta) \right\} \quad (3.4)$$

This marginal likelihood for standard normal mixed model can be maximized with respect to fixed effects β and the variance parameters σ . Maximization for β yields the generalized least squares estimate

$$\beta = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{Y}$$

which is unbiased for β . The maximum likelihood estimates of the variance parameters are biased downwards (Verbeke and Molenberghs, 1997, 2000). The bias arises because estimation of the fixed effects is not taken into consideration when estimating the variance parameters. An unbiased method for estimating the variance parameters is obtained if the observed data are partitioned by a linear transformation. This idea was introduced by Patterson and Thompson (1971) and is called restricted maximum likelihood (REML).

3.2.2 Prediction of Random Effects

As noted in equation (3.2), the distribution of the observed outcomes \mathbf{y} is not dependent on the observed random effects \mathbf{b} . However, the \mathbf{b} do appear in the linear model (3.1) and may be of interest for interpretation. These effects are drawn from a known distribution, with mean zero and variance defined by the (estimated) variance parameters. For this reason, prediction of the random effects

once the variance parameters are known is different from estimation of fixed effects and variance components.

The predictions should be unbiased. That is, they have zero expected value and their variance is equal to the relevant variance component. Another desirable attribute that could be imposed is that the predictions have minimum variance of all unbiased predictors. These requirements, along with linearity, lead to a class of predictors known as best linear unbiased predictors (BLUPs). For a detailed discussion on the use of BLUPs see Robinson (1991) and Searle et al. (1992).

The minimum variance requirement of BLUPs implies that the distance between the unbiased predictors and the true effects is minimized amongst all choices of predictors. This leads to the BLUPs being unbiased predictors that minimize the mean squared error (MSE) between the true and predicted effects. Therefore

$$\begin{aligned}
 \text{MSE} &= E \left((b - \tilde{b}(\mathbf{y}))^2 \right) \\
 &= E \left(E \left((b - \tilde{b}(\mathbf{y}))^2 \mid \mathbf{y} \right) \right) \\
 &= E \left(\text{var}(b \mid \mathbf{y}) + E^2(b \mid \mathbf{y}) - 2E(b \mid \mathbf{y})\tilde{b}(\mathbf{y}) + \tilde{b}^2(\mathbf{y}) \right) \\
 &= E(\text{var}(b \mid \mathbf{y})) + E \left((E(b \mid \mathbf{y}) - \tilde{b}(\mathbf{y}))^2 \right).
 \end{aligned}$$

showing that MSE is minimized if $\tilde{b} = E(b \mid \mathbf{y})$.

Under the assumptions of model (3.1), \mathbf{b} and \mathbf{y} have a joint multivariate normal distribution

$$\begin{bmatrix} \mathbf{b} \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{X}\beta \end{bmatrix}, \begin{bmatrix} \mathbf{G} & \mathbf{GZ}' \\ \mathbf{ZG} & \Sigma \end{bmatrix} \right) \quad (3.5)$$

Using (3.5), we have that the expression of conditional expectation of \mathbf{b} given data \mathbf{y} is

$$\begin{aligned}
 E(\mathbf{b} \mid \mathbf{y}) &= E(\mathbf{b} + \text{cov}(\mathbf{b}, \mathbf{y})[\text{cov}(\mathbf{y})]^{-1}(\mathbf{y} - E(\mathbf{y}))) \\
 &= \mathbf{GZ}'\Sigma^{-1}(\mathbf{y} - \mathbf{X}\beta).
 \end{aligned} \quad (3.6)$$

This linear function in \mathbf{y} is the best linear predictor of \mathbf{b} . When β is replaced by its estimator $\hat{\beta}$ in 3.6, we get

$$\tilde{\mathbf{b}} = \mathbf{GZ}'\Sigma^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta}). \quad (3.7)$$

$\tilde{\mathbf{b}}$ is the best linear unbiased predictor (BLUP) for \mathbf{b} . In cases where Σ is not known, one forms the “approximate” BLUP for \mathbf{b} as

$$\hat{\mathbf{b}} = \mathbf{GZ}'\hat{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta}). \quad (3.8)$$

This predictor (3.8) is often referred to as the empirical Bayes estimator for \mathbf{b} . The conditional expectation $E(\mathbf{y}_i|\mathbf{b}_i) = \mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i$. The BLUP for $\mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i$ is

$$\mathbf{X}_i\hat{\beta} + \mathbf{Z}_i\mathbf{GZ}'_i\hat{\Sigma}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\hat{\beta}). \quad (3.9)$$

The predictor (3.9) has the interpretation that it may be rewritten in the form of a weighted average combining information from individual i only and information from the population. The BLUP for $\mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i$ may be viewed as borrowing strength across individuals to get the best prediction for individual i .

In the context of small area estimation, the concept of borrowing strength is closely related to the concept of shrinkage (Longford, 2005). When the data from a region is small, giving weak information on that region, the regional estimate is strengthened by supplementing the regional data with the global data. The smaller the regional data, the more weight the global information gets in the estimation. This is equivalent to strong shrinkage towards the global estimate. When the regional data are large, the local information receives more weight compared to the global information, and thus there will be less shrinkage in the regional estimate.

3.3 The Bayesian Paradigm

Bayesian methods provide a way to reason coherently about the world around us, in the face of uncertainty. Bayesian approaches are based on a mathematical handling of uncertainty, initially proposed by Bayes and Laplace in the 18th century and further developed by statisticians and philosophers in the 20th century. Before proceeding, it is useful to present the fundamental ideas behind the Bayesian approach to statistical analysis.

Suppose you are interested in estimating parameter θ from data $\mathbf{y} = \{y_1, \dots, y_n\}$ by using a statistical model described by a density $p(\mathbf{y}|\theta)$. Further, let $\pi(\theta)$ denote the prior distribution that expresses your beliefs about the parameter before you examine the data. Bayesian philosophy rests on use of Bayes' theorem, which enables one to combine the prior distribution and the model to obtain the posterior distribution, $p(\theta|\mathbf{y})$ in the following way:

$$p(\theta|\mathbf{y}) = \frac{p(\theta, \mathbf{y})}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\theta)\pi(\theta)}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\theta)\pi(\theta)}{\int p(\mathbf{y}|\theta)\pi(\theta)d\theta}$$

The quantity

$$p(\mathbf{y}) = \int p(\mathbf{y}|\theta)\pi(\theta)d\theta$$

is the normalizing constant of the posterior distribution. This quantity $p(\mathbf{y})$ is sometimes called the marginal distribution of the data. Simply put, Bayes' theorem tells you how to update existing knowledge with the new information. You begin with a prior belief $\pi(\theta)$, and after learning information from data \mathbf{y} , you update the belief about θ and obtain $p(\theta|\mathbf{y})$. These are the essential elements of the Bayesian approach to data analysis.

The use of Bayesian methods in applied problems has exploded during the 1990s. The availability of fast computing machines was combined with a group of iterative simulation methods known as Markov chain Monte Carlo (MCMC) algorithms

that greatly aided the use of complex Bayesian models. The idea behind MCMC is to produce approximate samples from the posterior distribution of interest, by generating a Markov chain which has the posterior as its limiting distribution. This revolutionary approach to Monte Carlo was originated in the particle Physics literature in Metropolis et al. (1953). It was then generalised by Hastings (1970) to a more statistical setting. However, it was Gelfand and Smith (1990) that introduced MCMC methods to mainstream statistics and since then, the use of Bayesian methods for applied statistical modelling has increased rapidly.

A comprehensive account of MCMC-related issues and the advances in statistical methodology generated by using this set of computational tools until 1995 is provided in Gilks et al. (1996). In an introductory technical level, Congdon (2001, 2003) describes the analysis of a wide range of statistical models using Bayesian inference using Gibbs sampling (BUGS), freely available software for Bayesian inference using MCMC, see Spiegelhalter et al. (1996).

3.3.1 Markov Chain Monte Carlo Sampling

The Markov chain Monte Carlo (MCMC) method is a simulation method for sampling from posterior distributions and computing posterior quantities of interest. MCMC methods sample successively from a target distribution. The Markov chain method has been quite successful in modern Bayesian computing. Only in simple Bayesian models can analytical forms of the posterior distributions be recognized and summarize inferences directly. In complex models, posterior densities are too difficult to work with directly. With the MCMC method, it is possible to generate samples from a posterior density $p(\theta|\mathbf{y})$ and to use samples to approximate expectations of quantities of interest. Most importantly, if the simulation algorithm is implemented correctly, the Markov chain is guaranteed to converge to the target distribution $p(\theta|\mathbf{y})$, regardless of where the chain was initialized. Furthermore, if the chain is run for a long time, one can recover $p(\theta|\mathbf{y})$ to any precision. Also, the

simulation algorithm is easily extensible to models with a large number of parameters or high complexity. We present the basic Metropolis(-Hastings) algorithm and the Gibbs sampler.

The Metropolis Algorithm

The Metropolis algorithm can be used to obtain random samples from any complicated target distribution of any dimension that is known up to a normalizing constant. Suppose you want to obtain T samples from a univariate distribution with probability density function $f(\theta|\mathbf{y})$. Suppose θ^t is the t^{th} sample from f . To use the Metropolis algorithm, you need to have an initial value θ^0 and a symmetric proposal density $q(\theta^{t+1}|\theta^t)$ (meaning that the likelihood of jumping to θ^{t+1} from θ^t is the same as the likelihood of jumping back to θ^t from θ^{t+1}).

For the $(t + 1)$ th iteration, the algorithm generates a sample from $q(\cdot|\cdot)$ based on the current sample θ^t , and it makes a decision to either accept or reject the new sample. If the new sample is accepted, the algorithm repeats itself by starting at the new sample. If the new sample is rejected, the algorithm starts at the current point and repeats. The Metropolis algorithm can be summarized (Tierney, 1994, Gelman et al., 2004) as follows:

1. Set $t = 0$. Choose a starting point θ^0 . This can be an arbitrary point as long as $f(\theta^0|\mathbf{y}) > 0$.
2. Generate a new sample, θ_{new} , by using the proposal distribution $q(\cdot|\theta^t)$.
3. Calculate the ratio of densities and get the following quantity:

$$r = \min \left\{ \frac{f(\theta_{new}|\mathbf{y})}{f(\theta^t|\mathbf{y})}, 1 \right\}$$

4. Sample u from the uniform distribution $\mathcal{U}(0, 1)$.
5. Set $\theta^{t+1} = \theta_{new}$ if $u < r$; otherwise set $\theta^{t+1} = \theta^t$

6. Set $t = t + 1$. If $t < T$, the number of desired samples, return to step 2. Otherwise stop.

This algorithm defines a chain of random variates whose distribution will converge to the desired distribution $f(\theta|\mathbf{y})$, and so from some point forward, the chain of samples is a sample from the distribution of interest. In Markov chain terminology, this distribution is called the stationary distribution of the chain. For more detailed descriptions and proofs of the Metropolis algorithm, see Roberts (1996) and Liu (2001).

The Metropolis-Hastings Algorithm

The Metropolis-Hastings (MH) algorithm is a more general form of the Metropolis algorithm that was proposed by Hastings (1970). The MH algorithm uses an asymmetric proposal distribution, that is, there is no requirement that $q(\theta^{t+1}|\theta^t) \neq q(\theta^t|\theta^{t+1})$. The difference in its implementation comes in calculating the ratio of densities (Robert and Casella, 2004, Gelman et al., 2004). The acceptance ratio is now:

$$r = \min \left\{ \frac{f(\theta_{new}|\mathbf{y})q(\theta^t|\theta_{new})}{f(\theta^t|\mathbf{y})q(\theta_{new}|\theta^t)}, 1 \right\}$$

Allowing an asymmetric proposal distribution can be useful in increasing the speed of the random walk. The Metropolis algorithm can be extended to a high-dimensional problem by using a multivariate version of proposal distribution $q(\cdot|\cdot)$. Chib and Greenberg (1995) provide a useful tutorial on the algorithm.

Gibbs Sampler

The Gibbs sampler, named by Geman and Geman (1984), is a special case of the Metropolis sampler in which the proposal distributions exactly match the posterior conditional distributions and the proposals are accepted 100% of the time. Gibbs sampling requires decomposition of the joint posterior distribution into full

conditional distributions for each parameter in the model and then sampling from them. The sampler can be efficient when the parameters are not highly dependent on each other and the full conditional distributions are easy to sample from. When parameters are highly correlated with each other, one might experience poor mixing or slow convergence of the Markov chain because the chain slowly traverses the parameter space.

Suppose $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ is the parameter vector, $p(\mathbf{y}|\theta)$ is the likelihood, and $\pi(\theta)$ is the prior distribution. The full posterior conditional distribution is

$$\pi(\theta_i|\theta_j, i \neq j) \propto p(\mathbf{y}|\theta)\pi(\theta).$$

The Gibbs sampler works as follows:

1. Set $t = 0$, and choose an arbitrary initial value of $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_k^{(0)})$.
2. Generate each component of θ as follows:
 - draw $\theta_1^{(t+1)}$ from $\pi(\theta_1|\theta_2^{(t)}, \dots, \theta_k^{(t)}, \mathbf{y})$
 - draw $\theta_2^{(t+1)}$ from $\pi(\theta_2|\theta_1^{(t+1)}, \theta_3^{(t)}, \dots, \theta_k^{(t)}, \mathbf{y})$
 - ...
 - draw $\theta_k^{(t+1)}$ from $\pi(\theta_k|\theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_{k-1}^{(t+1)}, \mathbf{y})$
3. Set $t = t + 1$. If $t < T$, the number of desired samples, return to step 2. Otherwise stop.

In order to compute the posterior statistics with sufficient precision, the Monte Carlo process needs a large of iterations (T). It is a common practice to discard the initial portion of a Markov chain sample so that the effect of initial values on the posterior inference is minimized. More details on the choice of starting values and the number of burn-in iterations are discussed by Gelman and Rubin (1992) and Gelfand (2000).

3.3.2 Assessing MCMC Coverage

Simulation-based Bayesian inference requires using simulated draws to summarize the posterior distribution or calculate relevant posterior quantities of interest. There are usually two issues. First, one has to decide whether the Markov chain has reached its stationary or the desired posterior distribution. Second, you have to determine the number of iterations to keep after the Markov chain has reached stationarity. Convergence diagnostics help to resolve these issues. See Cowles and Carlin (1996), Brooks and Roberts (1998) and Gelman et al. (2004) for detailed discussions about convergence diagnostics. Assessment of convergence involves the use of visual analysis (e.g. autocorrelation and trace plots) as well as statistical diagnostic tests.

Visual Analysis via Autocorrelation and Trace Plots

Trace plots of samples versus the iteration index can be very useful in assessing convergence. The trace plots informs whether the chain has converged to its stationary distribution. A trace plot and autocorrelation function plot can also tell you whether the chain is mixing well, a chain that mixes well traverses its posterior space rapidly. The aspects of stationarity recognizable from the trace plot are a relatively constant mean and variance.

MCMC Convergence Diagnostics

The convergence diagnostics of Gelman and Rubin (1992), Brooks and Gelman (1997) currently are the most popular amongst the statistical community, at least in part because computer programs for their implementation are available from their creators. In addition to Gelman and Rubin diagnostics, we briefly discuss the methods of Geweke (1992), Raftery and Lewis (1992, 1996) and of Heidelberger and Welch (1981, 1983).

Gelman and Rubin method (Gelman and Rubin, 1992) is based on analysing multiple simulated MCMC chains by comparing the variances within the chain and the variance between chains. Large deviation between these two variances indicates nonconvergence.

Define $\{\theta^t : t = 1, 2, \dots, n\}$ to be a collection of a single Markov chain output. The Parameter θ^t is the t th sample of the Markov chain. Suppose you have M parallel MCMC chains initialized from various parts of the target distribution. Each chain is of length n after discarding the burn-in, the simulations are labelled as $\{\theta_m^t : t = 1, 2, \dots, n; m = 1, 2, \dots, M\}$. The between-chain variance B and the within-chain variance W are calculated as

$$B = \frac{n}{M-1} \sum_{m=1}^M (\bar{\theta}_m - \bar{\theta})^2, \quad \text{where} \quad \theta_m = \frac{1}{n} \sum_{t=1}^n \theta_m^t, \quad \bar{\theta} = \frac{1}{M} \sum_{m=1}^M \bar{\theta}_m$$

$$W = \frac{1}{M} \sum_{m=1}^M s_m^2, \quad \text{where} \quad s_m^2 = \frac{1}{n-1} \sum_{t=1}^n (\theta_m^t - \bar{\theta}_m)^2$$

The posterior marginal variance, $\text{var}(\theta|\mathbf{y})$, is a weighted average of W and B . The estimate of the variance is

$$\hat{V} = \frac{n-1}{n}W + \frac{M+1}{nM}B$$

If all M chains have reached the target distribution, the posterior variance estimate should be very close to the within-chain variance W . Therefore the ratio \hat{V}/W should be close to unity. The square root of this ratio is called the potential scale reduction factor (PSRF). Large PSRF indicates that the between chain variance is substantially greater than the within chain variance, so that longer simulation is needed. If the PSRF is close to 1, you conclude that each of the M chains has stabilized.

The Geweke test (Geweke, 1992) compares values in the early part of the Markov chain to those in the latter part of the chain in order to detect failure of convergence. Geweke (1992) recommends the use of methods from spectral analysis to assess convergence of the Gibbs sampler when the intent of the analysis is to estimate the mean of some function of the parameters being simulated. Disadvantages of Geweke's method include that it is sensitive to the specification of the spectral window. In addition, while his diagnostic is quantitative, Geweke does not specify a procedure for applying it but instead leaves that to the experience and subjective choice of the statistician.

Heidelberger and Welch test (Heidelberger and Welch, 1981, 1983) consists of two parts: a stationary portion test and a half-width test. The stationarity test assesses the stationarity of a Markov chain by testing the hypothesis that the chain comes from a covariance stationary process. The half-width test checks whether the Markov chain sample size is adequate to estimate the mean values accurately.

If the interest lies in posterior percentiles, you want a diagnostic test that evaluates the accuracy of the estimated percentiles. The Raftery-Lewis test (Raftery and Lewis, 1992, 1996) is designed for this purpose. Raftery and Lewis emphasize that being able to pin down the accuracy of the estimation of quantiles is very useful since they are at the heart of density estimation as well as providing robust estimates of center and spread of a distribution.

3.4 Model Selection for Small Area Estimation

Auxiliary data (predictor variables) play an important role in small area estimation. The choice of small area models depends on the availability of auxiliary data and the relationship between these data and the variables of interest at the small area level. Auxiliary data are often obtained from various administrative and census records. Interest lies in borrowing strength from these auxiliary data to boost accuracy of the small area estimates.

When a large pool of potential auxiliary variables is available, the selection of a subset of suitable auxiliary variables is necessary in many small area estimation projects. Model selection techniques may be applied to select the best set of auxiliary variables. With the classical modeling approach, the Akaike information criterion (AIC) and Bayesian information criterion (BIC) are commonly used for model selection purposes. For detailed information on these criteria and how to use them in small area context, refer to Rao, 2003, pg. 105-107. For the HB approach, a commonly used model selection criterion is the deviance information criterion (DIC) proposed by Spiegelhalter et al. (2002).

DICs are comparable only over models with the same observed data. Let θ , \mathbf{y} and $p(\mathbf{y}|\theta)$ denote the unknown parameters of the model, the data and the likelihood respectively and let $f(\mathbf{y})$ be a function of the data alone. Then the deviance is defined as:

$$D(\theta) = -2\log[p(\mathbf{y}|\theta)] + 2\log[f(\mathbf{y})]. \quad (3.10)$$

The second term in the deviance involves \mathbf{y} only and cancels out when comparing deviances from different models. The posterior mean of the deviance, $\bar{D} = E[D(\theta)|\mathbf{y}]$, is a measure of goodness of fit of the model. The larger the value of \bar{D} , the poorer the fit. The measure of effective number of parameters of a Bayesian model is computed as:

$$P_D = \bar{D} - \bar{D}(\bar{\theta}), \quad (3.11)$$

where $\bar{\theta} = E[\theta|\mathbf{y}]$ is the posterior mean of the parameters. The larger the p_D , the easier it is for the model to fit the data. The DIC is calculated as:

$$DIC = P_D + 2\bar{D} - \bar{D}(\bar{\theta}). \quad (3.12)$$

The model with the smallest DIC is judged to be the model that would best predict a replicate dataset of the same structure as the one observed. Incorporating p_D in the DIC calculation penalizes complex models. DIC assumes that the posterior mean is a good measure of the parameter. For more details on DIC, see Spiegelhalter et al. (2002) and Gelman et al. (2004).

Chapter 4

Empirical Bayes Approach

4.1 Introduction

In this chapter, we present the empirical Bayes (EB) approach in the estimation of parameters. The EB method requires a direct estimate of the characteristic of interest and variables correlated with that characteristic to be available. The direct estimate of the characteristic is regressed on these correlated variables and for each observation a regression estimate of the characteristic is calculated in addition to the direct estimate. The EB estimator is then a weighted average of the direct estimate and the linear regression estimate of the characteristic. The weights on the two estimates depend on the sampling variance of the direct estimates and how well the model fits the data.

The organization of this chapter is as follows. In section 4.1 we present the standard area-level model. A new model is described in section 4.2. The setup of a simulation study to assess the performance of the proposed model is presented in section 4.3. We wrap up with the results of the simulation study and a discussion of results in section 4.4.

4.2 Fay-Herriot Model

According to the well-known Fay-Herriot model (Fay and Herriot, 1979) in small area estimation, a basic area level model assumes that the small area parameter of interest θ_i is related to area specific auxiliary data x_i through a linear model

$$\theta_i = x_i^T \beta + \nu_i, \quad i = 1, \dots, m. \quad (4.1)$$

where m is the number of small areas, $\beta = (\beta_1, \dots, \beta_p)'$ is $p \times 1$ vector of regression coefficients, and the ν_i 's are area-specific random effects assumed to be independent and identically distributed (iid) with $E(\nu_i) = 0$ and $var(\nu_i) = A$. The assumption of normality may also be included.

The area level model assumes that there exists a direct survey estimator y_i for the small area parameter θ_i such that

$$y_i = \theta_i + e_i, \quad i = 1, \dots, m, \quad (4.2)$$

On combining equations (4.1) and (4.2), we obtain

$$y_i = x_i^T \beta + \nu_i + e_i, \quad i = 1, \dots, m, \quad (4.3)$$

where e_i is the sampling error associated with the direct estimator y_i . We also assume that the e_i 's are independent normal random variables with mean $E(e_i|\theta_i) = 0$ and sampling variance $var(e_i|\theta_i) = D_i$. We further assume that $D_i (i = 1, \dots, m)$ are known.

4.2.1 Estimation of the parameters of the Fay-Herriot Model

The Bayes estimator of θ_i under the Fay-Herriot model (Fay and Herriot, 1979) is developed under the following models:

$$y_i|\theta_i \sim \mathcal{N}(\theta_i, D_i)$$

and the prior distribution of $\theta_i, i = 1, \dots, m$ is

$$\theta_i \sim \mathcal{N}(x_i^T \beta, A)$$

Equivalently, we can write

$$f(y_i|\theta_i) = \frac{1}{\sqrt{2\pi D_i}} \exp\left(-\frac{1}{2D_i}(y_i - \theta_i)^2\right) \quad \text{and} \quad \pi(\theta_i) = \frac{1}{\sqrt{2\pi A}} \exp\left(-\frac{1}{2A}(\theta_i - x_i^T \beta)^2\right)$$

Therefore from the Bayesian models we have that

$$f(\mathbf{y}, \theta|\beta, A) = \prod_{i=1}^m \frac{1}{\sqrt{2\pi A}} \exp\left(-\frac{1}{2A}(\theta_i - x_i^T \beta)^2\right) \cdot \frac{1}{\sqrt{2\pi D_i}} \exp\left(-\frac{1}{2D_i}(y_i - \theta_i)^2\right). \quad (4.4)$$

where $\mathbf{y} = (y_1, y_2, \dots, y_m)^T$ and $\theta = (\theta_1, \theta_2, \dots, \theta_m)^T$.

On looking at the exponential parts in equation (4.4) without the $(-1/2)$ factor, we see that

$$\begin{aligned} \frac{1}{D_i}(y_i - \theta_i)^2 + \frac{1}{A}(\theta_i - x_i^T \beta)^2 &= \left(\frac{1}{D_i} + \frac{1}{A}\right) \theta_i^2 - 2\left(\frac{y_i}{D_i} + \frac{x_i^T \beta}{A}\right) \theta_i + k_i \\ &= \left(\frac{1}{D_i} + \frac{1}{A}\right) \left\{ \theta_i - \left(\frac{y_i}{D_i} + \frac{x_i^T \beta}{A}\right) / \left(\frac{1}{D_i} + \frac{1}{A}\right) \right\}^2 + k_i \end{aligned}$$

where k_i is a constant independent of θ_i . Thus

$$\theta_i|\beta, A, y_i \sim \mathcal{N}\left(x_i^T \beta + \frac{A}{A + D_i}(y_i - x_i^T \beta), \frac{AD_i}{A + D_i}\right) \quad (4.5)$$

The best linear predictor of $\theta_i = x_i^T \beta + \nu_i$ is $E[\theta_i | y_i] = E[x_i^T \beta + \nu_i | y_i]$. Under normality, y_i and θ_i have a joint normal distribution with means $x_i^T \beta$ and $x_i^T \beta$. Also, $\text{var}(y_i) = D_i + A$; $\text{var}(\theta_i) = \text{var}(x_i^T \beta + \nu_i) = A$ and $\text{Cov}(y_i, x_i^T \beta + \nu_i) = A$. Hence based on this formulation and when β and A are known, we can view that the Bayes predictor of θ_i , $\hat{\theta}_i^B$ is

$$\begin{aligned} \hat{\theta}_i^B &= E[\theta_i | y_i, \theta, A] \\ &= E[x_i^T \beta + \nu_i | y_i, \beta, A] \\ &= x_i^T \beta + \frac{A}{A + D_i} (y_i - x_i^T \beta) \\ &= B_i x_i^T \beta + (1 - B_i) y_i. \end{aligned}$$

where $B_i = D_i / (D_i + A)$, $i = 1, 2, \dots, m$. However, in practice both β and A are unknown and are replaced by $\hat{\beta}$ and \hat{A} respectively in $\hat{\theta}_i^B$ to get the empirical Bayes predictor $\hat{\theta}_i^{EB}$

$$\begin{aligned} \hat{\theta}_i^{EB} &= x_i^T \hat{\beta} + \frac{\hat{A}}{\hat{A} + D_i} (y_i - x_i^T \hat{\beta}) \\ &= \hat{B}_i x_i^T \hat{\beta} + (1 - \hat{B}_i) y_i. \end{aligned}$$

We estimate parameters β and A from the marginal distribution of y_i . Marginally,

$$y_i \sim \mathcal{N}(\theta_i, D_i + A)$$

For known A , the best linear unbiased estimator of β is the weighted least squares estimator $\tilde{\beta}$. To estimate β and A , we iteratively solve the equations (Fay and Herriot, 1979)

$$\tilde{\beta} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} y, \quad \Sigma = \text{diag}(D_1 + A, \dots, D_m + A). \quad (4.6)$$

and

$$\sum_{i=1}^m (y_i - x_i^T \tilde{\beta})^2 / (D_i + A) = m - r(X). \quad (4.7)$$

where $r(X)$ denotes the rank of design matrix X assumed to be equal to number of model parameters p and less than m and equation (4.7) is obtained via method of moments (details of this derivation are provided in Appendix A).

4.3 Proposed Area-Level Model

In the new model, we assume that there exists a direct survey estimator y_i for the small area parameter θ_i such that

$$y_i = \theta_i + e_i, \quad (4.8)$$

and the small area parameter of interest θ_i is related to area specific auxiliary data x_i through a linear model

$$\theta_i = x_i' \beta + \delta_i \nu_i, \quad i = 1, \dots, m. \quad (4.9)$$

On combining equations (4.8) and (4.9), we obtain

$$y_i = x_i^T \beta + \delta_i \nu_i + e_i, \quad i = 1, \dots, m, \quad (4.10)$$

where e_i is the sampling error associated with the direct estimator y_i . We also assume that the e_i 's are independent normal random variables with mean $E(e_i|\theta_i) = 0$ and sampling variance $\text{var}(e_i|\theta_i) = D_i$. We further assume that $D_i (i = 1, \dots, m)$ are known. The model specification is completed by defining a model tuning parameter δ_i which can take a value of zero or one in a Bernoulli fashion. ν_1, \dots, ν_m and $\delta_1, \dots, \delta_m$ are assumed to be independent.

In the new model (4.10), we have taken a new approach in constructing the random part of the standard Fay-Herriot model (4.3) by introducing a tuning parameter δ_i , this parameter takes a value of zero (random part in model (4.3) vanishes) when enough data is available in the i th small area to perform ordinary linear

regression. On the other hand, the tuning parameter takes a value of one when there is insufficient data in the i th small area necessitating incorporation of a random effect for the corresponding small area.

Given that $\delta_i = 1$, $\nu_i \sim \mathcal{N}(0, A)$, $pr(\delta_i = 1) = p$ and assuming that A , p , and β are known, the Bayes predictor of θ_i becomes:

$$\begin{aligned}\hat{\theta}_i^B &= E(\theta_i|\mathbf{y}) \\ &= x_i^T \beta + E[\delta_i \nu_i | \mathbf{y}] \\ &= x_i^T \beta + E[E(\delta_i \nu_i | \delta_i, \mathbf{y}) | \mathbf{y}] \\ &= x_i^T \beta + E[\nu_i | \delta_i = 1, \mathbf{y}] \cdot pr[\delta_i = 1 | p, \beta, \mathbf{y}]\end{aligned}$$

On observing that

$$y_i | \nu_i, \delta_i = 1, \beta \sim \mathcal{N}(x_i^T \beta + \nu_i, D_i); \quad \nu_i | \delta_i = 1 \sim \mathcal{N}(0, A);$$

and

$$\nu_i | \delta_i = 1, \beta, \mathbf{y} \sim \mathcal{N}\left(\frac{A}{A + D_i}(y_i - x_i^T \beta), \frac{AD_i}{A + D_i}\right);$$

The Bayes predictor of θ_i becomes

$$\begin{aligned}\hat{\theta}_i^B &= x_i^T \beta + \frac{A}{A + D_i}(y_i - x_i^T \beta) \cdot pr(\delta_i = 1 | p, \beta, A, \mathbf{y}) \\ &= x_i^T \beta + \frac{A}{A + D_i}(y_i - x_i^T \beta) \cdot \hat{p}_i(p, \beta, A);\end{aligned}$$

the probability $\hat{p}_i(p, \beta, A) = pr(\delta_i = 1 | p, \beta, A, \mathbf{y})$ is derived by observing that

$$\begin{aligned}pr(\delta_i = 1 | \mathbf{y}, A) &= \frac{pr(\delta_i = 1, y_i)}{f(y_i)} \\ &= \frac{f(y_i | \delta_i = 1)pr(\delta_i = 1)}{f(y_i | \delta_i = 1)pr(\delta_i = 1) + f(y_i | \delta_i = 0)pr(\delta_i = 0)}.\end{aligned}$$

But

$$y_i | \delta_i = 1 \sim \mathcal{N}(x_i^T \beta, A + D_i) \quad \text{and} \quad y_i | \delta_i = 0 \sim \mathcal{N}(x_i^T \beta, D_i).$$

Therefore,

$$\hat{p}_i(p, \beta, A) = \frac{\frac{1}{\sqrt{2\pi(A+D_i)}} \exp\left(-\frac{(y_i - x_i^T \beta)^2}{2(A+D_i)}\right) \times p}{\frac{1}{\sqrt{2\pi(A+D_i)}} \exp\left(-\frac{(y_i - x_i^T \beta)^2}{2(A+D_i)}\right) \times p + \frac{1}{\sqrt{2\pi D_i}} \exp\left(-\frac{(y_i - x_i^T \beta)^2}{2D_i}\right) \times (1-p)}$$

Hence the marginal density of y_i , $f(y_i)$, is:

$$f(y_i) = \frac{p}{\sqrt{2\pi(A+D_i)}} \exp\left(-\frac{(y_i - x_i^T \beta)^2}{2(A+D_i)}\right) + \frac{(1-p)}{\sqrt{2\pi D_i}} \exp\left(-\frac{(y_i - x_i^T \beta)^2}{2D_i}\right)$$

The empirical Bayes predictor ($\hat{\theta}_i^{EB}(\hat{\beta}, \hat{p}, \hat{A}; y_i)$) of θ_i can be obtained by estimating the parameters β , p and A from the marginal distribution of Y_1, \dots, Y_m :

$$f(\mathbf{y} | \beta, p, A) = \prod_{i=1}^m \left[\frac{p}{\sqrt{2\pi(A+D_i)}} \exp\left(-\frac{(y_i - x_i^T \beta)^2}{2(A+D_i)}\right) + \frac{(1-p)}{\sqrt{2\pi D_i}} \exp\left(-\frac{(y_i - x_i^T \beta)^2}{2D_i}\right) \right]$$

Note that $A = 0$ will lead to $\hat{p} = 0$. On the other hand, $p = 0$ will make the estimation of A impossible. So we shall assume that $p > 0$ and $A > 0$.

4.3.1 Estimation of the Parameters of the Proposed Model.

Under the following models

$$\begin{aligned} Y_i | \delta_i, p, \beta, A &\sim N(x_i' \beta, D_i + \delta_i A), i = 1, \dots, m \\ \delta_i | p &\sim \text{Bernoulli}(p) \end{aligned}$$

We can estimate A from

$$Y^T \left(\Sigma^{-1} - \Sigma^{-1} X (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} \right)^{-1} Y = m - r(X)$$

where

$$\begin{aligned}\Sigma &= D + A\Delta \\ D &= \text{diag}(D_1, \dots, D_m) \\ \Delta &= \text{diag}(\delta_1, \dots, \delta_m)\end{aligned}$$

The probability $pr(\delta_i = 1|A, \beta, p, \mathbf{y})$ is obtained from

$$P(\delta_i = 1|A, \beta, p, \mathbf{y}) = \frac{p}{p + (1-p)\sqrt{\frac{D_i+A}{D_i}} \exp\left\{-\frac{(y_i - x_i^T \beta)^2 A}{2D_i(D_i+A)}\right\}} \quad (4.11)$$

and the regression coefficients are obtained by using the relation

$$\beta = \left(X^T \Sigma^{-1} X\right)^{-1} X^T \Sigma^{-1} \mathbf{y}.$$

We start with an initial guess of the parameters A_0, β_0 and p_0 and repeat the following iterations $k = 1, \dots, N$ times. Now, using equation (4.11), we draw R random sample samples of δ 's. Then we get

$$\hat{p}^{(k)} = \frac{1}{m} \sum_{i=1}^m P(\delta_i = 1|A_{(k-1)}, \beta_{(k-1)}, p_{(k-1)}, \mathbf{y})$$

and $\hat{A}^{(k)}$ can be obtained by solving the equation

$$\frac{1}{R} \sum_{j=1}^R Y^T \left(\Sigma_j^{-1} - \Sigma_j^{-1} X (X^T \Sigma_j^{-1} X)^{-1} X^T \Sigma_j^{-1} \right) Y = m - r(X)$$

where $\Sigma_j = D + A\Delta_j (j = 1, \dots, R)$. Finally $\hat{\beta}^{(k)}$ can be obtained by solving

$$\hat{\beta}^{(k)} = \left\{ \sum_{j=1}^R \left(X^T \hat{\Sigma}_{j(k)}^{-1} X \right) \right\}^{-1} \left(\sum_{j=1}^R X^T \hat{\Sigma}_{j(k)}^{-1} \mathbf{y} \right),$$

where $\hat{\Sigma}_{j(k)} = D + \hat{A}^{(k)} \Delta_j$.

4.4 Unit-level Small Area Model

4.4.1 Introduction

In this section, we develop the empirical Bayes (EB) approach in the estimation of parameters of the standard unit-level model and our new model. A unit level model can be used when unit-specific response variables are available in each small area. This class of models can incorporate auxiliary information at both the unit and area level (Moura and Holt, 1999). The area-specific random effect terms in unit level models can capture the correlation possibly present among the sample units within a sample area. The main advantage of unit level models is that they can incorporate all possible sources of uncertainty; in particular, they can capture the uncertainty due to estimation of the sampling variances. Unit level models are not always easy to apply in small area estimation problems because practitioners need to find auxiliary variables appearing in the sample (at unit-level) and in administrative registers (at area level). Confidentiality of unit-level data puts more obstacles to the use of these models.

4.4.2 Unit Level Standard Model

In the standard unit-level model, it is assumed that unit-specific auxiliary data $X_{ij} = (x_{ij1}, \dots, x_{ijp})^T$ are available for each population element j in each small area i . Further variable of interest y_{ij} , is assumed to be related to X_{ij} through a nested error linear regression model

$$y_{ij} = x_{ij}^T \beta + \nu_i + e_{ij}. \quad (4.12)$$

where y_{ij} is the response of unit j , $j = 1, 2, \dots, N_i$, in area i , $i = 1, 2, \dots, m$. x_{ij} is the vector of auxiliary variables, β is the vector of regression parameters, the random effect ν_i represents the effect of area characteristics that are not accounted for by the auxiliary variables X_{ij} and e_{ij} the individual unit error term. The area

effects ν_i are assumed independent with mean zero and variance σ_u^2 , The errors e_{ij} are independent with mean zero and variance σ_e^2 . In addition, the ν_i 's and e_{ij} 's are assumed to be independent.

The small area parameter of interest, θ_i , may be approximated by

$$\theta_i = \bar{X}_i^T \beta + \nu_i, \quad i = 1, \dots, m. \quad (4.13)$$

assuming that N_i is large, where \bar{X}_i is the vector of population means of the x_{ij} for the i th area, that is, $\bar{X}_i = \sum_{j=1}^{N_i} x_{ij}/N_i$. The sample data $\{y_{ij}, x_{ij}, j = 1, \dots, n_i, i = 1, \dots, m\}$ are assumed to obey model (4.12), i.e.,

$$y_{ij} = x_{ij}^T \beta + \nu_i + e_{ij}, \quad j = 1, \dots, n_i, \quad i = 1, \dots, m. \quad (4.14)$$

where n_i is the sample size in the i th small area. This implies that the selection bias is absent. For a proof of this absence of selection bias, see (Rao, 2003, pg. 79). Assuming that $\nu_i \sim \mathcal{N}(0, \sigma_u^2), i = 1, \dots, m$ and expressing $y_{ij}|\beta, \nu_i, \sigma_e^2 \sim \mathcal{N}(x_{ij}^T \beta + \nu_i, \sigma_e^2), j = 1, \dots, n_i, i = 1, \dots, m$, we have the distribution of ν_i from the unit-level model (4.14) conditional on $y_{ij}, \beta, \sigma_u^2$, and σ_e^2 as

$$\nu_i | y_{ij}, \beta, \sigma_u^2, \sigma_e^2 \sim \mathcal{N} \left(B_i (\bar{y}_i - \bar{x}_i^T \beta), B_i \frac{\sigma_e^2}{n_i} \right)$$

where $B_i = \sigma_u^2 / (\sigma_u^2 + n_i^{-1} \sigma_e^2)$ and (\bar{y}_i, \bar{x}_i) are the sample means for the i th area. Hence,

$$E[\nu_i | y_{ij}, \beta, \sigma_u^2, \sigma_e^2] = B_i (\bar{y}_i - \bar{x}_i^T \beta). \quad (4.15)$$

Using equation (4.13) and (4.15), we obtain the the Bayes predictor of θ_i , θ_i^B as

$$\begin{aligned} \theta_i^B &= E[\theta_i | y_{ij}, \beta, \sigma_u^2, \sigma_e^2] \\ &= \bar{X}_i^T \beta + B_i (\bar{y}_i - \bar{x}_i^T \beta) \\ &= (1 - B_i) \bar{X}_i^T \beta + B_i [\bar{y}_i + (\bar{X}_i - \bar{x}_i)^T \beta]. \end{aligned} \quad (4.16)$$

Replacing β , σ_u^2 and σ_e^2 by $\hat{\beta}$, $\hat{\sigma}_u^2$ and $\hat{\sigma}_e^2$ respectively we obtain the empirical Bayes predictor of θ_i , θ_i^{EB}

$$\theta_i^{EB} = (1 - \hat{B}_i)\bar{X}_i^T \hat{\beta} + \hat{B}_i[\bar{y}_i + (\bar{X}_i - \bar{x}_i)^T \hat{\beta}], \quad (4.17)$$

where $\hat{B}_i = \hat{\sigma}_u^2 / (\hat{\sigma}_u^2 + n_i^{-1} \hat{\sigma}_e^2)$.

4.4.3 Proposed Unit-Level Model

In the new unit-level model, we assume that unit-specific auxiliary data $X_{ij} = (x_{ij1}, \dots, x_{ijp})^T$ are available for each population element j in each small area i . Further variable of interest y_{ij} , is assumed to be related to X_{ij} through a nested error linear regression model

$$y_{ij} = x_{ij}^T \beta + \delta_i \nu_i + e_{ij}. \quad (4.18)$$

where y_{ij} is the response of unit j , $j = 1, 2, \dots, N_i$, in area i , $i = 1, 2, \dots, m$. x_{ij} is the vector of auxiliary variables, β is the vector of regression parameters, the random effect ν_i represents the effect of area characteristics that are not accounted for by the auxiliary variables X_{ij} and e_{ij} the individual unit error term. $\delta_i, i = 1, \dots, m$ are independent Bernoulli random variables. The area effects ν_i are assumed independent with mean zero and variance σ_u^2 , the errors e_{ij} are independent with mean zero and variance σ_e^2 . In addition, the ν_i 's and e_{ij} 's are assumed to be independent.

The small area parameter of interest, θ_i , is approximated by

$$\theta_i = \bar{X}_i^T \beta + \delta_i \nu_i, \quad i = 1, \dots, m. \quad (4.19)$$

assuming that N_i is large, where \bar{X}_i is the vector of population means of the x_{ij} for the i th area, that is, $\bar{X}_i = \sum_{j=1}^{N_i} x_{ij}/N_i$. The sample data $\{y_{ij}, x_{ij}, j = 1, \dots, n_i, i = 1, \dots, m\}$ are assumed to obey model (4.18), i.e.,

$$y_{ij} = x_{ij}^T \beta + \delta_i \nu_i + e_{ij}, \quad j = 1, \dots, n_i, \quad i = 1, \dots, m. \quad (4.20)$$

On assuming that β , σ_u^2 , σ_e^2 and p (where $pr(\delta_i = 1) = p$) are known and that $\nu_i \sim \mathcal{N}(0, \sigma_u^2)$, $i = 1, \dots, m$ when $\delta_i = 1$, we get the Bayes predictor of θ_i from the unit-level model (4.20) as

$$\begin{aligned} \hat{\theta}_i^B &= E(\theta_i | \mathbf{y}) \\ &= X_i^T \beta + E[\delta_i \nu_i | \mathbf{y}] \\ &= X_i^T \beta + E[E(\delta_i \nu_i | \delta_i, \mathbf{y}) | \mathbf{y}] \\ &= X_i^T \beta + E[\nu_i | \delta_i = 1, \mathbf{y}] \cdot pr[\delta_i = 1 | \mathbf{y}] \end{aligned} \quad (4.21)$$

On observing that

$$y_{ij} | \nu_i, \delta_i = 1, \beta \sim \mathcal{N}(x_{ij}^T \beta + \nu_i, \sigma_e^2); \quad \nu_i | \delta_i = 1 \sim \mathcal{N}(0, \sigma_u^2);$$

and

$$\nu_i | \delta_i = 1, \beta, \sigma_u^2, \sigma_e^2, \mathbf{y} \sim \mathcal{N}\left(B_i(\bar{y}_i - \bar{x}_i^T \beta), B_i \frac{\sigma_e^2}{n_i}\right);$$

where $B_i = \sigma_u^2 / (\sigma_u^2 + n_i^{-1} \sigma_e^2)$ and (\bar{y}_i, \bar{x}_i) are the sample means for the i th area.

The Bayes predictor of θ_i becomes

$$\hat{\theta}_i^B = \bar{X}_i^T \beta + B_i(\bar{y}_i - \bar{x}_i^T \beta) \cdot pr(\delta_i = 1 | p, \beta, \mathbf{y}) \quad (4.22)$$

the probability $pr(\delta_i = 1|p, \beta, A, \mathbf{y})$ is derived by observing that

$$\begin{aligned} pr(\delta_i = 1|\mathbf{y}) &= \frac{pr(\delta_i = 1, y_{ij})}{f(y_{ij})} \\ &= \frac{f(y_{ij}|\delta_i = 1)pr(\delta_i = 1)}{f(y_{ij}|\delta_i = 1)pr(\delta_i = 1) + f(y_{ij}|\delta_i = 0)pr(\delta_i = 0)}. \end{aligned} \tag{4.23}$$

where

$$y_{ij}|\delta_i = 1 \sim \mathcal{N}(x_{ij}^T\beta, \sigma_u^2 + \sigma_e^2) \quad \text{and} \quad y_{ij}|\delta_i = 0 \sim \mathcal{N}(x_{ij}^T\beta, \sigma_e^2).$$

4.5 Simulation Study

Our simulation study is based on the 1989 United States Census Bureau four-person families median income dataset. The United States Department of Health and Human services provides energy assistance to low-income families. Eligibility for the program is determined by a formula where the most important variable is an estimate of the current median income of four-person families by states (the 50 states and District of Columbia). The variables in this dataset includes: \mathbf{y}_i : State median income based on the current year from the current population survey (CPS) data, \mathbf{x}_i : Adjusted state median income for the current year and \mathbf{D}_i : Sampling variance associated with y_i .

We generate data using two different set ups. In the first case, we use the Fay-Herriot model as defined by equations (5.1) and (5.2) to generate data as follows:

- (i) Generate values of θ and y from the Fay-Herriot model where

$$\theta_i = x_i^T\beta + \nu_i, \quad y_i = \theta_i + e_i, \quad i = 1, \dots, m$$

- (ii) Use the Fay-Herriot model to fit the generated data and estimate the parameters β , θ and A .

- (iii) Use the proposed model to fit the generated data to estimate estimate the parameters β , p and A . The estimated values of parameters θ are obtained using parameters β , θ , p , A and equation (5.5).

Under the modeling assumptions of Fay-Herriot model, we have that

$$\nu_i \sim \mathcal{N}(0, 5 \times 10^6), \quad e_i \sim \mathcal{N}(0, D_i) \text{ and } \beta = (4700, 0.7)^T,$$

where $D = \text{diag}(D_1, D_2, \dots, D_m)$.

In the second set-up, we use the proposed model as defined by equations (5.3) and (5.4) to generate data as follows:

- (i) Generate values of θ and y using the proposed model where

$$\theta_i = x_i^T \beta + \delta_i \nu_i, \quad y_i = \theta_i + e_i, \quad i = 1, \dots, m$$

- (ii) Fit the generated data to the Fay-Herriot model to estimate the parameters β , θ and A .
- (iii) Use the proposed model to fit the generated data to estimate estimate the parameters β , θ , p and A . The estimated values of parameters θ are obtained using parameters β , θ , p , A and equation (5.5).

In the porposed model, we assume that $\nu_i \sim \mathcal{N}(0, 5 \times 10^6)$, $e_i \sim \mathcal{N}(0, D_i)$ and $\beta = (4700, 0.7)^T$,

where $D = \text{diag}(D_1, D_2, \dots, D_m)$. Three different datasets are generated using the proposed model because the values of δ_i are generated from Bernoulli distribution with sucess probability p taking values of 0.25, 0.5 and 0.75.

4.5.1 Performance Criteria

Estimation of the variance components is relatively easy whereas the assessment of the uncertainty due to the estimation is a big challenge. Extensive references on assessing this uncertainty can be found in the SAE literature. For example,

Prasad and Rao (1990) used a mean squared error (MSE) criterion to measure uncertainty of EBLUP under a general linear mixed model, and proposed an approximation to the MSE using the Taylor series method under normality assumption of the Fay-Herriot model. Lahiri and Rao (1995) demonstrated the robustness of this approximation against non-normality. More recently, Jackknife methods (e.g. Lohr and Rao (2009)) and parametric bootstrap methods (Lahiri (2003), Hall and Maiti (2006)) have been proposed to estimate the MSE. For extensive reviews on resampling methods, we refer to the paper by Gershunskaya et al. (2009).

In this study, the performance of the empirical Bayes (EB) estimates obtained under the two models is examined from two general standpoints: the accuracy of the point estimates and the MSE of the estimates. The former is considered through the relative errors and absolute relative errors of the EB estimates and the empirical MSE is estimated following the double bootstrap approach proposed by Hall and Maiti (2006). A comparison of MSE values will be made between the Fay-Herriot model and the proposed model.

4.5.1.1 Empirical Comparison of Empirical Bayes Estimates

Empirical comparison is considered through the relative errors and absolute relative errors of the EB estimates. The EB estimates are compared according to four different criteria recommended by the panel on small area estimates of population and income set up by the United States committee on National Statistics in 1978 (Ghosh et al., 1996, Datta et al., 2002). We compare the EB estimates on the basis of average relative bias, average squared relative bias, average absolute bias and average squared deviation.

Suppose e_{iTR} denotes the true median income for the i th state, and e_i is any estimate of e_{iTR} , $i = 1, \dots, m$. Then

1. Average relative bias (ARB)

$$ARB = \frac{1}{m} \sum_{i=1}^m \left| \frac{e_i - e_{iTR}}{e_{iTR}} \right|$$

2. Average squared relative bias (ASRB)

$$ASRB = \frac{1}{m} \sum_{i=1}^m \left(\frac{e_i - e_{iTR}}{e_{iTR}} \right)^2$$

3. Average absolute bias (AAB)

$$AAB = \frac{1}{m} \sum_{i=1}^m |e_i - e_{iTR}|$$

4. Average squared deviation (ASD)

$$ASD = \frac{1}{m} \sum_{i=1}^m (e_i - e_{iTR})^2$$

4.5.1.2 Estimation of MSE Using Double Bootstrap Technique for the Fay-Herriot Model

After obtaining the parameter estimates $\hat{\xi} = (\hat{A}, \hat{\beta})$ for the Fay-Herriot model from the simulated data, we need to gauge the performance of the Fay-Herriot model by estimating the mean-squared error of small area parameter θ_i^{EB} . We achieve this by using the double bootstrap technique of Hall and Maiti (2006). This technique enables us to obtain a bias-corrected, mean-squared error estimator of θ_i^{EB} . We generate 500 bootstrap samples at each level of the double bootstrap as described in the algorithm below

Double Bootstrap Algorithm for the Fay-Herriot Model

1. Do this $b_1 = 1, \dots, B_1$ times

generate $v_{b_1}^*$ from $N(0, \hat{A})$

calculate $\theta_{b_1}^*(\hat{\xi}) = X\hat{\beta} + v_{b_1}^*$

generate $y_{b_1}^* = \theta_{b_1}^*(\hat{\xi}) + e_{b_1}^*$ where $e_{b_1}^* \sim N(0, D)$

get $\hat{\xi}_{b_1}^* = (\hat{A}_{b_1}^*, \hat{\beta}_{b_1}^*)$ from the bootstrap data $(y_{b_1}^*, X, D)$

For each b_1 , do this $b_2 = 1, \dots, B_2$ times

generate $v_{b_2}^{**}$ from $N(0, \hat{A}_{b_1}^*)$

calculate $\theta_{b_2}^{**}(\hat{\xi}_{b_1}^*) = X\hat{\beta}_{b_1}^* + v_{b_2}^{**}$

generate $y_{b_2}^{**} = \theta_{b_2}^{**}(\hat{\xi}_{b_1}^*) + e_{b_2}^{**}$ where $e_{b_2}^{**} \sim N(0, D)$

get $\hat{\xi}_{b_2}^{**} = (\hat{A}_{b_2}^{**}, \hat{\beta}_{b_2}^{**})$ from the bootstrap data $(y_{b_2}^{**}, X, D)$

calculate $\hat{\theta}_{b_2}^{**EBP} = X\hat{\beta}_{b_2}^{**} + \frac{\hat{A}_{b_2}^{**}}{\hat{A}_{b_2}^{**} + D} (y_{b_2}^{**} - X\hat{\beta}_{b_2}^{**})$

calculate $\hat{\theta}_{b_1}^{*EBP} = X\hat{\beta}_{b_1}^* + \frac{\hat{A}_{b_1}^*}{\hat{A}_{b_1}^* + D} (y_{b_1}^* - X\hat{\beta}_{b_1}^*)$

2. Now calculate the first phase bootstrap estimator (\widehat{MSE}_1) and second phase bootstrap estimator (\widehat{MSE}_2)

$$\widehat{MSE}_1 = \frac{1}{B_1} \sum_{b_1=1}^{B_1} \left(\hat{\theta}_{b_1}^{*EBP} - \theta_{b_1}^*(\hat{\xi}) \right)^2$$

$$\widehat{MSE}_2 = \frac{1}{B_1} \sum_{b_1=1}^{B_1} \frac{1}{B_2} \sum_{b_2=1}^{B_2} \left(\hat{\theta}_{b_2}^{**EBP} - \theta_{b_2}^{**}(\hat{\xi}_{b_1}^*) \right)^2$$

$$\text{and finally, } \widehat{MSE} = \begin{cases} 2\widehat{MSE}_1 - \widehat{MSE}_2, & \text{if } \widehat{MSE}_1 > \widehat{MSE}_2 \\ \widehat{MSE}_1 \exp \left\{ \frac{\widehat{MSE}_1 - \widehat{MSE}_2}{\widehat{MSE}_2} \right\}, & \text{otherwise.} \end{cases}$$

4.5.2 Estimation of MSE Using Double Bootstrap Technique for the Proposed Model

We use the parameter estimates $\hat{\xi} = (\hat{A}, \hat{p}, \hat{\beta})$ obtained from the proposed model in the estimation of the mean-squared error of small area parameter θ_i^{EB} using the double bootstrap technique of Hall and Maiti (2006). We generate 500 bootstrap samples at each level of the double bootstrap as described in the algorithm below

Double Bootstrap Algorithm for the Proposed Model

1. Do this $b_1 = 1, \dots, B_1$ times

generate $\delta_{b_1}^*$ from Bernoulli (\hat{p})

generate $v_{b_1}^*$ from $N(0, \hat{A})$

calculate $\theta_{b_1}^*(\hat{\xi}) = X\hat{\beta} + \delta_{b_1}^* v_{b_1}^*$

generate $y_{b_1}^* = \theta_{b_1}^*(\hat{\xi}) + e_{b_1}^*$ where $e_{b_1}^* \sim N(0, D)$

get $\hat{\xi}_{b_1}^* = (\hat{A}_{b_1}^*, \hat{\beta}_{b_1}^*, \hat{p}_{b_1}^*)$ from the bootstrap data $(y_{b_1}^*, X, D)$

For each b_1 , do this $b_2 = 1, \dots, B_2$ times

generate $\delta_{b_2}^{**}$ from Bernoulli ($\hat{p}_{b_1}^*$)

generate $v_{b_2}^{**}$ from $N(0, \hat{A}_{b_1}^*)$

calculate $\theta_{b_2}^{**}(\hat{\xi}_{b_1}^*) = X\hat{\beta}_{b_1}^* + \delta_{b_2}^{**} v_{b_2}^{**}$

generate $y_{b_2}^{**} = \theta_{b_2}^{**}(\hat{\xi}_{b_1}^*) + e_{b_2}^{**}$ where $e_{b_2}^{**} \sim N(0, D)$

get $\hat{\xi}_{b_2}^{**} = (\hat{A}_{b_2}^{**}, \hat{\beta}_{b_2}^{**}, \hat{p}_{b_2}^{**})$ from the bootstrap data $(y_{b_2}^{**}, X, D)$

calculate $\hat{\theta}_{b_2}^{**EBP} = X\hat{\beta}_{b_2}^{**} + \frac{\hat{A}_{b_2}^{**}}{\hat{A}_{b_2}^{**} + D} (y_{b_2}^{**} - X\hat{\beta}_{b_2}^{**}) \widehat{prob}_{b_2}^{**}$

calculate $\hat{\theta}_{b_1}^{*EBP} = X\hat{\beta}_{b_1}^* + \frac{\hat{A}_{b_1}^*}{\hat{A}_{b_1}^* + D} (y_{b_1}^* - X\hat{\beta}_{b_1}^*) \widehat{prob}_{b_1}$

The definition of probability \widehat{prob} is given by equation (4.11).

2. Now calculate the first phase bootstrap estimator (\widehat{MSE}_1) and second phase bootstrap estimator (\widehat{MSE}_2)

$$\widehat{MSE}_1 = \frac{1}{B_1} \sum_{b_1=1}^{B_1} \left(\widehat{\theta}_{b_1}^{*EBP} - \theta_{b_1}^*(\widehat{\xi}) \right)^2$$

$$\widehat{MSE}_2 = \frac{1}{B_1} \sum_{b_1=1}^{B_1} \frac{1}{B_2} \sum_{b_2=1}^{B_2} \left(\widehat{\theta}_{b_2}^{**EBP} - \theta_{b_2}^{**}(\widehat{\xi}_{b_1}^*) \right)^2$$

$$\text{and finally, } \widehat{MSE} = \begin{cases} 2\widehat{MSE}_1 - \widehat{MSE}_2, & \text{if } \widehat{MSE}_1 > \widehat{MSE}_2 \\ \widehat{MSE}_1 \exp \left\{ \frac{\widehat{MSE}_1 - \widehat{MSE}_2}{\widehat{MSE}_2} \right\}, & \text{otherwise.} \end{cases}$$

4.6 Simulation Results

TABLE 4.1: Empirical Comparison of EB Estimates under Fay-Herriot Model (FH) and Proposed Model (PM) using Four Different Criteria.

Data Simulated from the Fay-Herriot Model				
Model	Average relative deviation	Average squared relative deviation	Average absolute deviation	Average squared deviation
FH	1329642	0.00340	899.41	0.0454
PM	1368768.47	0.00328	875.46	0.0439

TABLE 4.2: Empirical Comparison of EB Estimates under Fay-Herriot Model (FH) and Proposed Model (PM) using Four Different Criteria assuming that $p = 0.25$.

Data Simulated from the Proposed Model				
$p = 0.25$				
	Average relative	Average squared	Average absolute	Average squared
FH	638664.3	0.00177	590.53	0.0301
PM	445436.78	0.00082	452.45	0.0217

TABLE 4.3: Empirical Comparison of EB Estimates under Fay-Herriot Model (FH) and Proposed Model (PM) using Four Different Criteria assuming that $p = 0.5$ and 0.75 .

Data Simulated from the Proposed Model				
$p = 0.50$				
Model	Average relative deviation	Average squared relative deviation	Average absolute deviation	Average squared deviation
FH	1083228	0.00285	802.63	0.0412
PM	732942.42	0.00188	702.07	0.0348
$p = 0.75$				
FH	1371545	0.00369	925.40	0.0468
PM	893110.28	0.00208	729.75	0.0359

For each of the simulation settings described in Section 4.4, it is investigated how the small area parameters θ would be affected under the two models. We have compared the results of Fay-Herriot model and the proposed model on the basis of average absolute relative bias, average squared relative bias, average absolute bias and average squared deviation under two different simulation settings when the data is simulated from the area-level Fay-Herriot model as well as the area-level proposed model. Table (4.1) to (4.3) reports the values of these measures of empirical comparison.

The results in tables (4.1) to (4.3) appear to suggest a better fit for the proposed method compared to the standard area-level (Fay-Herriot) model. The percentage improvement in terms of average absolute relative bias, average squared relative bias, average absolute bias and average squared deviation ranges from 5% to 22%.

We now present the results of empirical comparison of the Fay-Herriot model and the proposed model in terms of empirical MSE obtained using the double-bootstrap approach.

FIGURE 4.1: Plots of empirical values of MSE using double-bootstrap approach: Comparison of the Fay-Herriot model and the proposed model at $p = 0.25, 0.5$ and 0.75

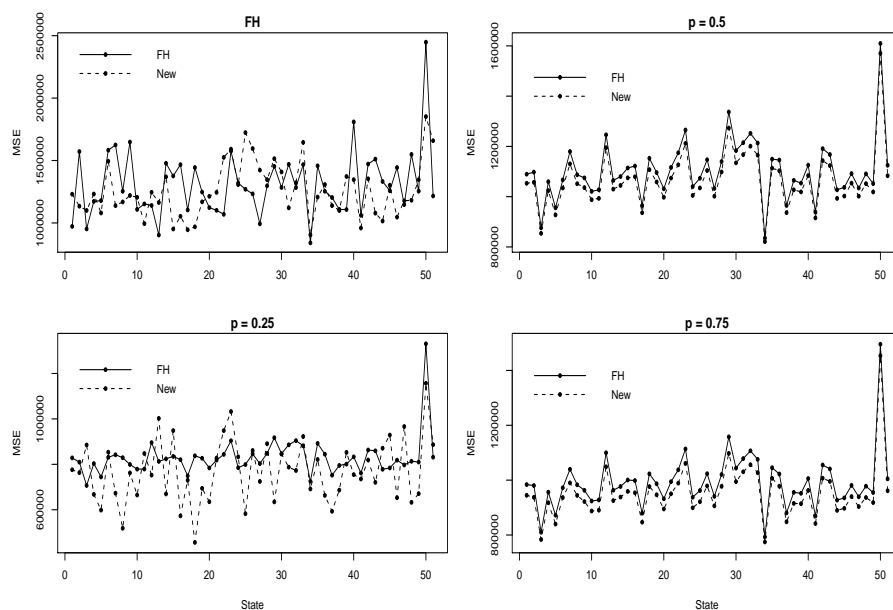


Figure (4.1) provides a visual comparison of the Fay-Herriot model and the proposed model in terms of empirical MSE obtained via the double-bootstrap approach. The vertical axis on the charts gives the empirical value of MSE over 500 double-bootstrap iterations, while the horizontal axis gives the state (small area) indicator.

We see that the proposed model is markedly better than the fay-Herriot model at $p = 0.5$ and $p = 0.75$ in all the small areas whereas the porposed model potrays better performance than the standard Fay-Herriot model for the majority of the small areas at $p = 0.25$. However, for the data that was simulated from the Fay-Herriot model, we do not seem to witness a clear winner between the two competing models in terms of empirical MSE as is evident in the top-left panel in figure (4.1).

Chapter 5

Hierarchical Bayes Approach

5.1 Introduction

In this chapter, we place the Fay-Herriot model and the proposed model into the hierarchical Bayes (HB) framework. Estimates of the posterior mean and the corresponding credible interval of the model parameters are obtained using the Gibbs sampling method. The full Bayesian approach requires the determination of the joint posterior distribution of all the model parameters given the data. The posterior distribution is known to be proportional to the product of the data likelihood and the prior distribution of all model parameters.

5.2 Area-level Model

According to the Fay-Herriot model (Fay and Herriot, 1979) in small area estimation, a basic area level model assumes that the small area parameter of interest θ_i is related to area specific auxiliary data x_i through a linear model

$$\theta_i = x_i^T \beta + \nu_i, \quad i = 1, \dots, m, \quad (5.1)$$

where m is the number of small areas, $\beta = (\beta_1, \dots, \beta_p)'$ is $p \times 1$ vector of regression coefficients, and the ν_i 's are area-specific random effects assumed to be independent and identically distributed (iid) with $E(\nu_i) = 0$ and $\text{var}(\nu_i) = A$. The assumption of normality may also be included.

The area level model assumes that there exists a direct survey estimator y_i for the small area parameter θ_i such that

$$y_i = \theta_i + e_i, \quad i = 1, \dots, m, \quad (5.2)$$

where the e_i is the sampling error associated with the direct estimator y_i . We also assume that the e_i 's are independent normal random variables with mean $E(e_i|\theta_i) = 0$ and sampling variance $\text{var}(e_i|\theta_i) = D_i$.

5.2.1 Bayesian formulation

We present the Fay-Herriot model in a hierarchical Bayes framework using the following distributions:

Conditional on the θ_i, β , and A , the direct survey estimate y_i is normally distributed with mean θ_i and variance D_i , where $D_i (i = 1, 2, \dots, m)$ are known. i.e.

$$y_i \sim \mathcal{N}(\theta_i, D_i)$$

Conditional on β and A , θ_i is normally distributed with mean $X_i^T \beta$ and variance A , where \mathbf{X}_i are known design matrices.

$$\theta_i \sim \mathcal{N}(X_i^T \beta, A)$$

We assume that marginally β and A are mutually independent with prior distribution given as $\beta \propto 1$.

The joint posterior distribution of parameters given data is:

$$\pi(\theta_1, \dots, \theta_m, \beta, A | \mathbf{y}) \propto \exp\left(-\sum_{i=1}^m \frac{(y_i - \theta_i)^2}{2D_i}\right) \exp\left(-\sum_{i=1}^m \frac{(\theta_i - X_i^T \beta)^2}{2A}\right) A^{-\frac{m}{2}} \pi(\beta, A)$$

where

$$\pi(\beta, A) = \frac{\bar{D}}{(A + \bar{D})^2} \text{ and } \bar{D} = \frac{\sum_{i=1}^m D_i}{m}.$$

We are interested in estimating the small area parameter θ_i . In the hierarchical Bayes analysis, θ_i is estimated by its posterior mean $E(\theta_i | y)$ and the associated uncertainty is measured by the posterior variance. We use Gibbs sampling (Gelman and Rubin, 1992, Gelfand, 2000) to obtain the posterior mean and posterior variance of θ_i . In order to implement the Gibbs sampler we generate samples from the full conditional distributions of the parameters β , A and θ_i . The full conditional distributions are as outlined below:

1. The conditional distribution of θ given β , A and data \mathbf{y} is multivariate normal:

$$\theta_i | \mathbf{y}, \beta, A \sim \mathcal{N}\left(\frac{D_i x_i^T \beta + A y_i}{A + D_i}, \frac{A D_i}{A + D_i}\right)$$

Equivalently, we can write the conditional distribution of θ given β , A and data \mathbf{y} as:

$$\theta_i | \mathbf{y}, \beta, A \sim \mathcal{N}(B_i x_i^T \beta + (1 - B_i) y_i, (1 - B_i) D_i),$$

where

$$B_i = \frac{D_i}{D_i + A}$$

2. The conditional distribution of β given θ , A and data \mathbf{y} is multivariate normal:

$$\beta | \mathbf{y}, \theta, A \sim \mathcal{N}((X^T X)^{-1} X^T \theta, A(X^T X)^{-1})$$

3. The conditional distribution of A given θ , β and data \mathbf{y} is:

$$\pi(A|\beta, \theta, \mathbf{y}) \propto \exp\left(-\sum_{i=1}^m \frac{(\theta_i - X_i^T \beta)^2}{2A}\right) A^{-\frac{m}{2}} \frac{1}{(A + \bar{D})^2}$$

5.2.2 Gibbs Sampling Algorithm

The resulting Markov chain Monte Carlo (MCMC) algorithm is outlined as follows:

Step 1: Set $k = 1$ and initialize parameters β , A , and θ and as outlined below:

$$\beta_{(1)} = (X^T D^{-1} X)^{-1} X^T D^{-1} y, \quad D = \text{diag}(D_1, D_2, \dots, D_m), \quad A_{(1)} = \bar{D}/2$$

and

$\theta_{(1)}$ is a vector of normal random variates with mean $(D_i x_i^T \beta_{(1)} + A_{(1)} y_i) / (A_{(1)} + D_i)$ and variance $A_{(1)} D_i / (A_{(1)} + D_i)$.

For $k = 2, \dots, M$, implement steps (2) to (4).

Step 2: Generate $A_{(k)}$ from inverse gamma distribution (3). It is assumed that $g = 1/A_{(k)}$ has gamma distribution with parameters $a = m/2$ and $b = 2/\sum(\theta_i - X_i^T \beta)^2$ (using the parameterization that has mean ab and variance ab^2).

Use accept-reject algorithm.

- (i) Draw g from $\text{Gamma}(m/2, 2/\sum(\theta_i - X_i^T \beta)^2)$.
- (ii) Set $A = g^{-1}$ and calculate $T = A/(A + \bar{D})^2$.
- (iii) Draw $U \propto \mathcal{U}(0, 1)$. If $U/4\bar{D} < T$ then $A_{(k)} = A$ otherwise go back to (i).

Step 3: Generate $\theta_{(k)}$ using conditional distribution (2) using $\beta_{(k-1)}$ and $A_{(k)}$.

Step 4: Generate $\beta_{(k)}$ from multivariate normal distribution (2) using $\theta_{(k)}$ and $A_{(k)}$.

5.3 Proposed Model

In this section, we develop the hierarchical Bayes framework for the proposed model. We consider a basic area level model such that the small area parameter of interest θ_i is related to area specific auxiliary data x_i through a linear model

$$\theta_i = x_i^T \beta + \delta_i \nu_i, \quad i = 1, \dots, m. \quad (5.3)$$

where m is the number of small areas, $\delta_1, \dots, \delta_m$ are iid Bernoulli random variables (i.e. $\delta_i \sim \text{Bern}(p)$), $\beta = (\beta_1, \dots, \beta_p)'$ is $p \times 1$ vector of regression coefficients, and the ν_i 's are area-specific random effects assumed to be independent and identically distributed (iid) with $E(\nu_i) = 0$ and $\text{var}(\nu_i) = A$. The assumption of normality may also be included. Here ν_1, \dots, ν_m and $\delta_1, \dots, \delta_m$ are assumed to be independent.

In this area level model we further assume that there exists a direct survey estimator y_i for the small area parameter θ_i such that

$$y_i = \theta_i + e_i, \quad i = 1, \dots, m, \quad (5.4)$$

where the e_i is the sampling error associated with the direct estimator y_i . We also assume that the e_i 's are independent normal random variables with mean $E(e_i|\theta_i) = 0$ and known sampling variance $\text{var}(e_i|\theta_i) = D_i$.

5.3.1 Posterior Distributions

Under the modelling assumptions in equation (5.3) and (5.4), the joint posterior density of parameters given data is:

$$f(\beta, \delta, p, A | \mathbf{y}) \propto \prod_{i=1}^m p^{\delta_i} (1-p)^{1-\delta_i} \cdot \prod_{i=1}^m \left\{ (A + D_i)^{-\frac{\delta_i}{2}} \exp\left(-\frac{\delta_i (y_i - x_i^T \beta)^2}{2(A + D_i)}\right) \right\} \cdot \prod_{i=1}^m \left\{ D_i^{-\frac{(1-\delta_i)}{2}} \exp\left(-\frac{(1-\delta_i)(y_i - x_i^T \beta)^2}{2D_i}\right) \right\} \cdot \pi(\beta) \cdot \pi(p) \cdot \pi(A)$$

which simplifies to

$$f(\beta, \delta, p, A, \mathbf{y}) \propto p^{\sum \delta_i} (1-p)^{m-\sum \delta_i} \cdot \prod_{i=1}^m \left\{ (A + D_i)^{-\frac{\delta_i}{2}} \exp\left(-\frac{\delta_i (y_i - x_i^T \beta)^2}{2(A + D_i)}\right) \right\} \cdot \prod_{i=1}^m \left\{ D_i^{-\frac{(1-\delta_i)}{2}} \exp\left(-\frac{(1-\delta_i)(y_i - x_i^T \beta)^2}{2D_i}\right) \right\} \cdot \pi(\beta) \cdot \pi(p) \cdot \pi(A)$$

To complete the Bayesian model specification, the prior distributions are assigned as follows:

$$\pi(p) = \text{uniform}(0, 1), \quad \pi(A) = \frac{\bar{D}}{(A + \bar{D})^2} \quad \text{and} \quad \bar{D} = \frac{\sum_{i=1}^m D_i}{m}$$

Again, like in the standard Fay-Herriot model, our main interest lies in the estimation of the small area parameter θ_i . The Bayes predictor of parameter θ_i is

$$\begin{aligned} \hat{\theta}_i^B &= x_i^T \beta + \frac{A}{A + D_i} (y_i - x_i^T \beta) \cdot pr(\delta_i = 1 | \beta, A, p, \mathbf{y}) \\ &= x_i^T \beta + \frac{A}{A + D_i} (y_i - x_i^T \beta) \cdot \hat{p}_i; \end{aligned} \tag{5.5}$$

where $\hat{p}_i = pr(\delta_i = 1 | \beta, A, p, \mathbf{y})$.

We use Gibbs sampling (Gelman and Rubin, 1992, Gelfand, 2000) to generate samples from the full conditional distributions of the parameters β , A , and p .

The full conditional distributions of these parameters are outlined below:

1. The conditional distribution of β given δ , A and p and data \mathbf{y} is

$$\pi(\beta|\delta, p, A, \mathbf{y}) \propto \exp\left(-\frac{1}{2} \sum_{i=1}^m \frac{\delta_i}{A + D_i} (y_i - x_i^T \beta)^2\right) \times \exp\left(-\frac{1}{2} \sum_{i=1}^m \frac{(1-\delta_i)}{D_i} (y_i - x_i^T \beta)^2\right)$$

Therefore the conditional distribution of β given δ , p , A and \mathbf{y} is multivariate normal with mean-vector $H^{-1}g$ and covariance matrix H^{-1} , i.e.

$$\beta|\mathbf{y}, \delta, p, A \sim \mathcal{N}(H^{-1}g, H^{-1})$$

where

$$H = \sum_{i=1}^m \left\{ \frac{\delta_i}{A + D_i} + \frac{(1 - \delta_i)}{D_i} \right\} x_i x_i^T,$$

and

$$g = \sum_{i=1}^m \left\{ \frac{\delta_i}{A + D_i} + \frac{(1 - \delta_i)}{D_i} \right\} x_i y_i, \quad x_i = (1, x_{i1})^T.$$

2. The conditional distribution of $\delta_i = 1$ given β , p , A and \mathbf{y} is

$$pr(\delta_i = 1|p, \beta, A, \mathbf{y}) = \frac{p}{p + (1 - p) \sqrt{\frac{A+D_i}{D_i}} \exp\left\{-\frac{(y_i - x_i^T \beta)^2}{2} \left(\frac{1}{D_i} - \frac{1}{A+D_i}\right)\right\}}$$

3. The conditional distribution of p given β , δ , A and \mathbf{y} is

$$p|\beta, \delta, A, \mathbf{y} \sim \text{Beta}\left(\sum_{i=1}^m \delta_i + 1, m - \sum_{i=1}^m \delta_i + 1\right)$$

4. The conditional distribution of A given β , p , δ and \mathbf{y} is

$$\Pi(A|\beta, \delta, p, \mathbf{y}) \propto \exp\left(-\frac{1}{2} \sum_{i=1}^m \frac{\delta_i}{A + D_i} (y_i - x_i^T \beta)^2\right) \cdot \prod_{i=1}^m (A + D_i)^{-\frac{\delta_i}{2}} \cdot \pi(A)$$

we allow

$$\pi(A) = \frac{\bar{D}}{(A + \bar{D})^2} \text{ and } \bar{D} = \frac{\sum_{i=1}^m D_i}{m}$$

Gibbs Sampling Algorithm: Proposed model

Step 1: Set $k = 1$ and initialize parameters β , p , A , and δ and as outlined below:

$$\beta_{(1)} = (X^T D^{-1} X)^{-1} X^T D^{-1} y,$$

where $D = \text{diag}(D_1, D_2, \dots, D_m)$ and $X = (1, \mathbf{x})^T$.

$$p_{(1)} = 0.2,$$

$$A_{(1)} = \bar{D}/2 \text{ and}$$

$\delta_{(1)}$ is a vector of Bernoulli random variates with success probability p_1 .

For $k = 2, 3, \dots, N$, implement steps (2) to (5).

Step 2: Generate $A_{(k)}$ from conditional distribution (4).

In order to sample from the full conditional distribution of A given β , p , δ and \mathbf{y} , we use the accept-reject algorithm. The target density is

$$f(A) = \frac{\prod_{i=1}^m \left\{ \exp\left(-\frac{1}{2}(y_i - x_i^T \beta)^2 \left(\frac{1}{A + D_i} - \frac{1}{A + \bar{D}}\right)\right) \left(\frac{A + D_i}{A + \bar{D}}\right)^{-\frac{1}{2}}\right\}^{\delta_i}}{(A + \bar{D})^{2 + \frac{1}{2} \sum \delta_i} \exp\left(-\frac{\sum \delta_i (y_i - x_i^T \beta)^2}{2(A + \bar{D})}\right)}$$

The proposal density is

$$g(A) = \frac{1}{(A + \bar{D})^{2 + \frac{1}{2} \sum \delta_i}} \exp\left(-\frac{\sum \delta_i (y_i - x_i^T \beta)^2}{2(A + \bar{D})}\right) \text{ and}$$

$$M = \max_A \frac{f(A)}{g(A)}.$$

In order to stabilize the calculations, we define $\psi(A) = \log \frac{f(A)}{g(A)}$ and then evaluate $\psi(A)$ at $A = 10^{-10}, 10^{-9}, \dots, 10^{10}$ and take the maximum of $\exp \psi(A)$ to approximate M .

Accept-Reject Algorithm:

$$(a) \sum \delta_i = 0, \quad A_{(k)} = \frac{1 - U}{U} \bar{D} \text{ where } U \sim \text{uniform}(0, 1)$$

$$(b) \sum \delta_i > 0$$

$$(i) \text{ Draw } U \text{ from } \text{uniform}(0, 1)$$

- (ii) Draw R from truncated gamma distribution with shape = $\frac{1}{2} \sum \delta_i + 1$
and scale = $\frac{2}{\sum \delta_i (y_i - x_i^T \beta)^2}$
- (iii) $R < \frac{1}{\bar{D}}$, then take $A_{(k)} = \frac{1}{R} - \bar{D}$, otherwise draw another sample R .
- (iv) If $\frac{f(A_{(k)})}{M \cdot g(A_{(k)})} > U$ then take $A_{(k)}$ as a sample from f , otherwise go back to (ii) and draw another sample from truncated gamma distribution.

Step 3: Generate $\beta_{(k)}$ from multivariate normal distribution (1).

Step 4: Generate $\delta_{(k)}$ using conditional distribution (2) and parameters $\beta_{(k)}$, $A_{(k)}$, and $p_{(k-1)}$.

Step 5: Generate $p_{(k)}$ from beta distribution defined in (3) and values of $\delta_{(k)}$

5.4 Simulation Study

Our simulation study is based on the 1989 United States Census Bureau four-person families median income dataset. The United States Department of Health and Human services provides energy assistance to low-income families. Eligibility for the program is determined by a formula where the most important variable is an estimate of the current median income of four-person families by states (the 50 states and District of Columbia). The variables in this dataset includes: \mathbf{y}_i : State median income based on the current year from the current population survey (CPS) data, \mathbf{x}_i : Adjusted state median income for the current year and \mathbf{D}_i : Sampling variance associated with y_i .

We generate data using two different set ups. In the first case, we use the Fay-herriot model as defined by equations (5.1) and (5.2) to generate data as follows:

- (i) Generate values of θ and y from the Fay-Herriot model where

$$\theta_i = x_i^T \beta + \nu_i, \quad y_i = \theta_i + e_i, \quad i = 1, \dots, m$$

- (ii) Use the Fay-Herriot model to fit the generated data and estimate the parameters β , θ and A .
- (iii) Use the proposed model to fit the generated data to estimate estimate the parameters β , p and A . The estimated values of parameters θ are obtained using parameters β , θ , p , A and equation (5.5).

Under the modeling assumptions of Fay-Herriot model, we have that

$$\nu_i \sim \mathcal{N}(0, 5 \times 10^6), \quad e_i \sim \mathcal{N}(0, D_i) \text{ and } \beta = (4700, 0.7)^T,$$

where $D = \text{diag}(D_1, D_2, \dots, D_m)$.

In the second set-up, we use the proposed model as defined by equations (5.3) and (5.4) to generate data as follows:

- (i) Generate values of θ and y using the proposed model where

$$\theta_i = x_i^T \beta + \delta_i \nu_i, \quad y_i = \theta_i + e_i, \quad i = 1, \dots, m$$

- (ii) Fit the generated data to the Fay-Herriot model to estimate the parameters β , θ and A .
- (iii) Use the proposed model to fit the generated data to estimate estimate the parameters β , θ , p and A . The estimated values of parameters θ are obtained using parameters β , θ , p , A and equation (5.5).

In the proposed model, we assume that $\nu_i \sim \mathcal{N}(0, 5 \times 10^6)$, $e_i \sim \mathcal{N}(0, D_i)$ and $\beta = (4700, 0.7)^T$,

where $D = \text{diag}(D_1, D_2, \dots, D_m)$. Three different datasets are generated using the proposed model because the values of δ_i are generated from Bernoulli distribution with success probability p taking values of 0.25, 0.5 and 0.75.

5.4.1 Empirical Comparison of Hierarchical Bayes Estimates

In this simulation study the performance of the hierarchical Bayes (HB) estimators is examined from the accuracy of the point estimates standpoints. This is considered through the relative errors and absolute relative errors of the HB estimates. The HB estimates are compared according to four different criteria recommended by the panel on small area estimates of population and income set up by the United States committee on National Statistics in 1978 (Ghosh et al., 1996, Datta et al., 2002). We compare the HB estimates on the basis of average relative bias, average squared relative bias, average absolute bias and average squared deviation.

Suppose e_{iTR} denotes the true median income for the i th state, and e_i is any estimate of e_{iTR} , $i = 1, \dots, m$. Then

1. Average relative bias (ARB)

$$ARB = \frac{1}{m} \sum_{i=1}^m \left| \frac{e_i - e_{iTR}}{e_{iTR}} \right|$$

2. Average squared relative bias (ASRB)

$$ASRB = \frac{1}{m} \sum_{i=1}^m \left(\frac{e_i - e_{iTR}}{e_{iTR}} \right)^2$$

3. Average absolute bias (AAB)

$$AAB = \frac{1}{m} \sum_{i=1}^m |e_i - e_{iTR}|$$

4. Average squared deviation (ASD)

$$ASD = \frac{1}{m} \sum_{i=1}^m (e_i - e_{iTR})^2$$

5.5 Simulation Results

We report on a simulation study that investigates the performance of the proposed model as compared to the Fay-Herriot model. For both models (Fay-Herriot model and the proposed model), we initialize the model parameters and run an Markov Chain Monte Carlo (MCMC) chain until convergence is attained. We checked the convergence of the chain under each of the two models via trace plots and autocorrelation plots of parameters β , p , and A . The details of the initial values of parameters β , p , and A can be found in section 5.4. Each chain was ran for a total of 20,000 iterations whereby the first 5,000 iterations were taken as burn-in and the other 15,000 iterations used for convergence check and to compute posterior inference for the parameters of interest.

We present in tables to the results of empirical comparison of the Fay-Herriot model and the proposed model under four different simulation settings. Also presented in figures to are the histograms, autocorrelation plots and trace plots for the estimated parameters β , p , and A . For the posterior estimates of the model parameters, we computed the means and the corresponding 95% credible intervals for each parameter. These posterior inference statistics are presented in table 5.2, 5.4, 5.6 and 5.8.

5.5.1 Simulation: Data generated using Fay-Herriot model

In this subsection, we present the results after analysing the data simulated from the Fay-herriot model. We ran one long chain with length 20,000 iterations. We discard the first 5,000 iterations. The trace plots, histograms and autocorrelation plots for β , p , and A are presented in figure (5.1). The results of empirical comparison of the estimates under the two models are presented in table (5.1).

TABLE 5.1: Empirical Comparison of HB estimates under Fay-Herriot model (FH) and proposed model (PM) using four different criteria. Data simulated from the Fay-Herriot model

Data Simulated from the Fay-Herriot Model				
Model	Average relative deviation	Average squared relative deviation	Average absolute deviation	Average squared deviation
FH	1330049.85	0.0034	899.92	0.0454
PM	1356812.87	0.0035	908.61	0.0459

FIGURE 5.1: Data simulated from the Fay-Herriot model: Histograms, trace plots and autocorrelation plots for β_0 , β_1 , A and p for the last 15,000 after a burn-in of 5,000 iterations. The point * in the histograms plots denotes the true value of the corresponding parameter.

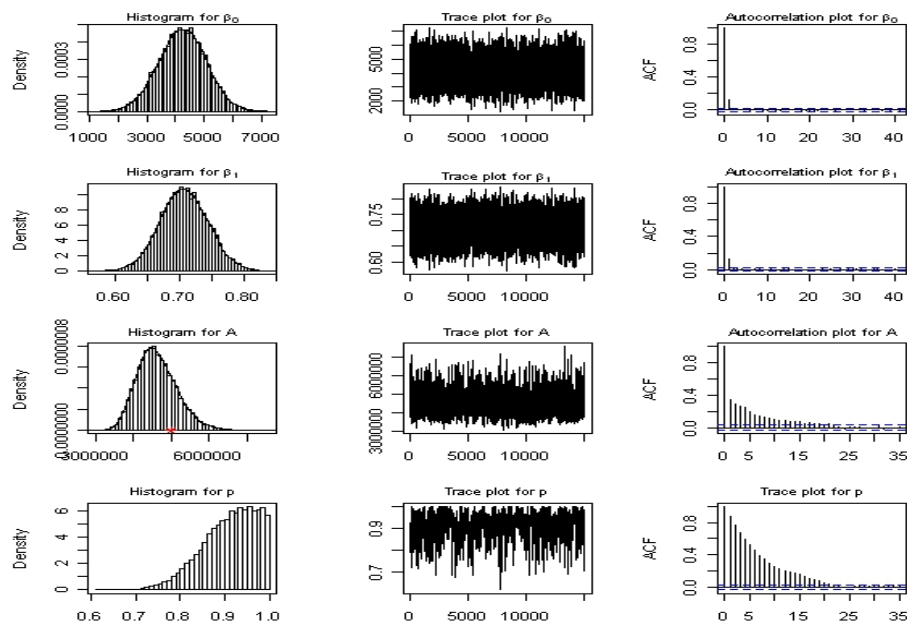


TABLE 5.2: Posterior summary statistics for data simulated from Fay-herriot model.

	Fay-Herriot Model			Proposed Model		
	mean	95% HDR		mean	95% HDR	
		2.5%	97.5%		2.5%	97.5%
β_0	4340.41	2674.31	5917.17	4242.24	2588.28	5852.46
β_1	0.70	0.63	0.78	0.71	0.64	0.78
A	4202862.21	3592770.13	4894461.15	4627162.51	3718889.96	5811476.73
p	-	-	-	0.91	0.77	0.98

From figure (5.1), we can see that all the chains are stabilized and well mixed. The results in table (5.1) suggest that both models are equally good, no model seems to be a clear winner when the data is generated from the Fay-Herriot model. This motivates us to argue that our proposed model is ideal even in the case where standard Fay-Herriot model would seem more appropriate. The posterior estimates are presented in table (5.2) and we see that we obtain good estimates for all the parameters.

5.5.2 Simulation: Data generated using proposed model when $p=0.25$.

TABLE 5.3: Empirical Comparison of HB estimates under Fay-Herriot model (FH) and proposed model (PM) when data is simulated from the proposed model assuming that $p = 0.25$.

Data Simulated from the Proposed Model				
$p = 0.25$				
Model	Average relative deviation	Average squared relative deviation	Average absolute deviation	Average squared deviation
FH	1116510.22	0.0046	598.28	0.0306
PM	896793.43	0.0029	482.75	0.0254

FIGURE 5.2: Data simulated from the proposed model when $p = 0.25$: Histograms, trace plots and autocorrelation plots for β_0 , β_1 , A and p for the last 15,000 after a burn-in of 5,000 iterations.

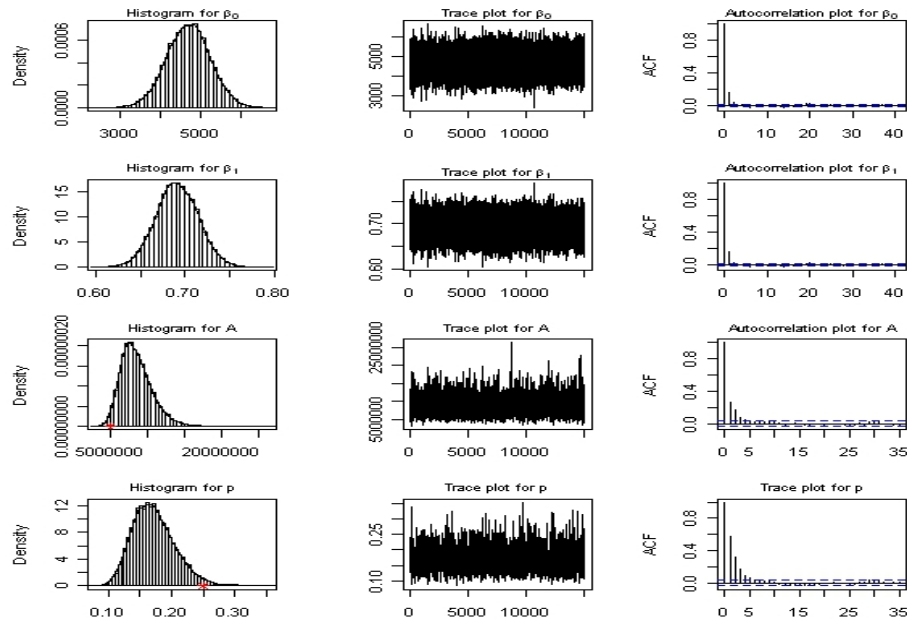


TABLE 5.4: Posterior summary statistics for data simulated from the proposed model ($p = 0.25$).

	Fay-Herriot Model			Proposed Model		
	mean	95% HDR		mean	95% HDR	
		2.5%	97.5%		2.5%	97.5%
β_0	4720.50	3472.51	5944.72	4689.33	3641.17	5715.50
β_1	0.68	0.63	0.74	0.69	0.65	0.74
A	1515104.18	1193255.21	1882942.25	8533315.14	5169677.99	13549118.54
p	-	-	-	0.17	0.12	0.25

The resultant trace plots, histograms and autocorrelation plots for β , p , and A are presented in figure (5.2). From this figure, we can see that all the chains are stabilized and well mixed and the autocorrelations appear to dampen quickly. Although the statistical results of posterior estimates are somewhat worse as seen in table (5.4), the results of empirical comparison of the two models are presented in table (5.3). These results suggest that the proposed model offers better estimates than the Fay-Herriot model in this case.

5.5.3 Simulation: Data generated using proposed model when $p=0.5$.

TABLE 5.5: Empirical Comparison of HB estimates under Fay-Herriot model (FH) and proposed model (PM) when data is simulated from the proposed model assuming that $p = 0.5$.

Data Simulated from the Proposed Model				
$p = 0.50$				
Model	Average relative deviation	Average squared relative deviation	Average absolute deviation	Average squared deviation
FH	1532593.12	0.0058	1165.56	0.0603
PM	1288522.80	0.0036	830.08	0.0423

FIGURE 5.3: Data simulated from the proposed model when $p = 0.5$: Histograms, trace plots and autocorrelation plots for β_0 , β_1 , A and p for the last 15,000 after a burn-in of 5,000 iterations. The point * in the histograms plots denotes the true value of the corresponding parameter.

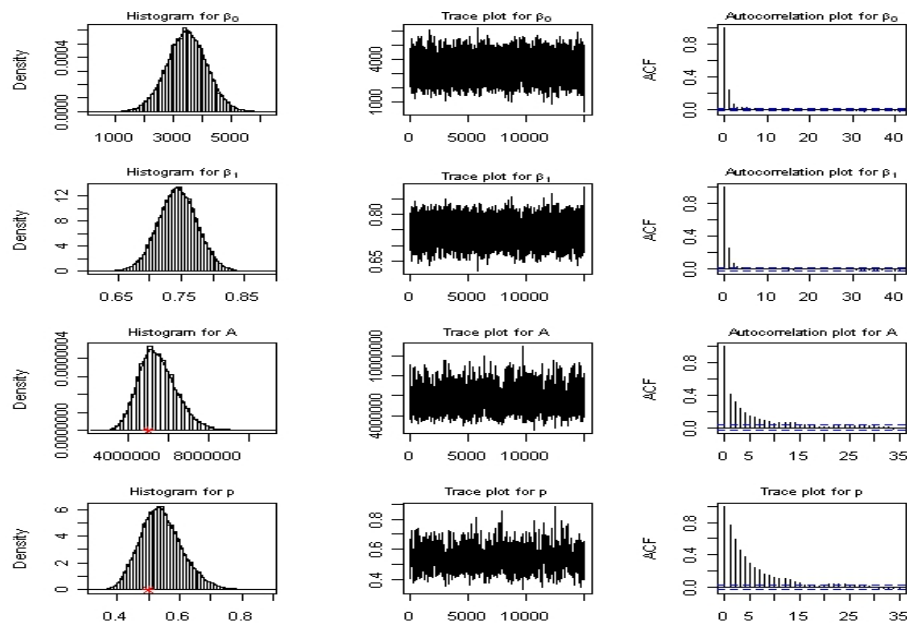


TABLE 5.6: Posterior summary statistics for data simulated from the proposed model ($p = 0.50$).

	Fay-Herriot Model			Proposed Model		
	mean	95% HDR		mean	95% HDR	
		2.5%	97.5%		2.5%	97.5%
β_0	3264.57	1813.34	4697.09	3439.07	2097.21	4780.66
β_1	0.75	0.69	0.82	0.74	0.68	0.80
A	2910852.29	2403098.07	3462882.55	5494823.55	3930702.02	7588448.63
p	-	-	-	0.54	0.42	0.69

The trace plots, histograms and autocorrelation plots for β , p , and A are presented in figure (5.3). The trace plots in figure (5.3) indicates that the chains appears to have reached a stationary distribution. It also has good mixing. The autocorrelation plots indicates low autocorrelation and appear to dampen quickly. This provides added evidence of the convergence of the Markov chain. The results presented in table (5.5) indicate that the proposed model is consistently superior to the Fay-Herriot model. The statistical results are presented in table (5.6). From this table, we see that we obtain better results using the proposed model than the standard fay-Herriot model.

5.5.4 Simulation: Data generated using proposed model when $p=0.75$.

TABLE 5.7: Empirical Comparison of HB estimates under Fay-Herriot model (FH) and proposed model (PM) when data is simulated from the proposed model assuming that $p = 0.75$.

Data Simulated from the Proposed Model				
$p = 0.75$				
Model	Average relative deviation	Average squared relative deviation	Average absolute deviation	Average squared deviation
FH	1979706.56	0.0074	1308.36	0.0679
PM	1499093.99	0.0041	969.79	0.0491

FIGURE 5.4: Data simulated from the proposed model when $p = 0.75$: Histograms, trace plots and autocorrelation plots for β_0 , β_1 , A and p for the last 15,000 after a burn-in of 5,000 iterations.

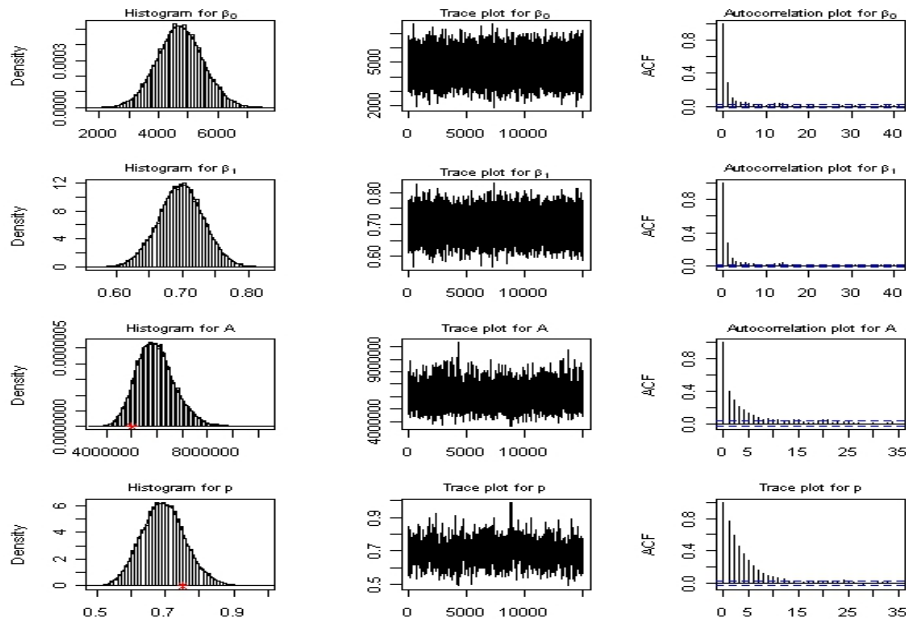


TABLE 5.8: Posterior summary statistics for data simulated from the proposed model ($p = 0.75$).

	Fay-Herriot Model			Proposed Model		
	mean	95% HDR		mean	95% HDR	
		2.5%	97.5%		2.5%	97.5%
β_0	5158.88	3531.58	6783.82	4723.46	3218.03	6250.50
β_1	0.68	0.61	0.75	0.70	0.63	0.76
A	4143076.00	3505689.03	4861893.09	5953074.54	4621120.87	7602451.25
p	-	-	-	0.69	0.58	0.83

The trace plots, histograms and autocorrelation plots for β , p , and A are presented in figure (5.4). The trace plots in figure (5.3) indicates that the chains appears to have reached a stationary distribution. It also has good mixing. The autocorrelation plots indicates low autocorrelation and efficient sampling. The results presented in table (5.7) indicate that the proposed model is consistently superior to the Fay-Herriot model. The statistical results are presented in table (5.8). From this table, we see that we obtain better results using the proposed model than the standard fay-Herriot model.

5.6 Discussion of results

In this chapter, we provide information that allows one to assess the relative performance of the proposed model when compared to the standard Fay-Herriot model. We have implemented the hierarchical Bayes (HB) method of estimation. The empirical results described in section 5.5 are evidence that our proposed model has promise as an advancement over the standard Fay-Herriot model, in particular it performed reasonably well in terms of .

An advantage of the HB approach is that the inferences are “exact” unlike the EB approach, and it can handle complex small area models using MCMC methods, but it needs specification of a prior on the model parameters. The HB approach is attractive but caution should be exercised when using MCMC methods. For example, MCMC algorithms could lead to seemingly reasonable inferences about a non existent posterior distribution. This happens when the posterior distribution is improper, yet the Gibbs conditional distributions used in generating the MCMC samples are proper Hobert and Casella (1996).

Chapter 6

Concluding Remarks and Future Research

This concluding chapter summarises the work presented in this thesis and suggests some future directions. In Chapter 1, we introduced the idea of small area estimation and highlighted that concepts of small area estimation can be applied to a wide variety of situations. Chapter 2 gives an overview of approaches to small area estimation. In chapter 3, we concentrated on review of the general concepts pertaining to frequentist and Bayesian methods. Our new model for small area estimation is developed and compared to standard area level model (Fay and Herriot, 1979) through the empirical Bayes (EB) approach in chapter 4 . In chapter 5, a hierarchical Bayes (HB) approach is developed to estimate parameters for the proposed model, with implementation carried out by Markov chain Monte Carlo (MCMC) techniques. MCMC was implemented via the Gibbs sampling algorithm using *R* software package. We used several graphical tools to assess convergence and determine the length of the burn-in period.

We have examined the performance of our proposed model in the estimation of median income of four person families based on the 1989 United States census bureau dataset. Our simulation studies established clearly the superiority of using the proposed model over the standard Fay-Herriot model. However, the advantage

of the proposed model comes at a price since the its implementation is mildly more difficult than the Fay-Herriot model.

We have proposed an EB estimator of the median income of four-person families at the state level. Compared to the corresponding HB procedure, it saves a tremendous computing time and is very simple to implement.

We are aware of the limitations of this study. We have confined our attention to the framework of linear mixed models with normal random terms namely area-level random effects and/or the sampling errors of the direct survey estimates. The theory of general linear mixed models relies largely on the normality of random effects (McCulloch and Searle, 2001). In real life, however, the normality of area effects is not always justified and heavy-tailed distributions and asymmetric distributions are frequently encountered in empirical studies (see Azzalini (1985, 1986) and Hampel et al. (1986)) and therefore it would be a rewarding topic for future research to assess the performance of our proposed model under non-normal features related to skewness, kurtosis and heavy tails. We also intend to further explore the method we proposed for unit-level small area estimation via a simulation study using a unit-level data.

In addition, it may happen that a relatively few observations do not follow a small area model that adequately explains bulk of data. Outliers may occur in individual observations. Such outlying observations may adversely affect estimation of model parameters. The problem of investigating sensitivity of our approach to outliers is an interesting avenue of future research and it is probably the one, where the next step will be taken.

In this study, our estimation of prediction mean square error(MSE) in the empirical bayes approach was based on double-bootstrap procedure. A possible disadvantage of this bootstrap method is that the bias of the MSE estimator may be sensitive to the choice of number of bootstrap samples, B . It may be advisable to study sensitivity as B changes.

The approach proposed in this work will hopefully broaden the applicability of the small area estimation. As with any applied statistical method, practical experience will be required to assess its final merit.

Appendix A

Program Documentation

A.1 Computational Aspects

A number of different computer packages, including R, WinBUGS and SAS are used to prepare data or fit different models. Appendix B has listings of selected but important *R* programs developed using the hierarchical Bayes Approach. These do not include all data preparation programs. Inevitably, as software develops new functionality is introduced and it may become simpler to build the models in a different package.

Use the provided programs with love and caution. Please advise me if you find any mistakes.

A.2 Hierarchical Bayes Approach

The main *R* program for the hierarchical Bayes approach using the proposed model.

```
#load library  
library(mvtnorm)  
library(foreign)
```

```
#####
#           #
# Read in the US median income dataset           #
# generated by FH model and new model           #
# using p = 0.25,0.5 and 0.75                 #
#           #
#####

# data generated by new model: p = 0.25

usincome=read.dbf('E:/Toni/PhD/PhD project/DATA/usincomeD225.dbf')

attach(usincome)

m = length(x)

# D matrix

d = diag(D)

# X = Design matrix

X = as.matrix(data.frame(int=rep(1,m),x))

# Y = response matrix

Y = as.matrix(y)

# Number of iterations

nsim = 20000

# Values used in computation of M

M.a = 10^c(-10:-1,1:10)

#####
# #
#           Intial Values #
# #
#####

# Vector Beta

beta = array(0,c(nsim,2))

beta0 = solve(t(X)%*%solve(d)%*%X)%*%t(X)%*%solve(d)%*%Y

beta[1,] = t(beta0)

# Intial Value of A

A0 = meanD/2

sigma = array(0,c(nsim,1))

sigma[1] = A0
```

```

# Intial Value of P

P0 = 0.6

P = array(0,c(nsim,1))

P[1] = P0

#####
# #
# Intial values of Delta from Bernoulli #
# distribution with success probability P0 #
# #
#####

delta0 = rbinom(m,1,P0)

Delta = array(0,c(nsim,m))

Delta[1,] = delta0

#####
# #
# This function calculates the conditional #
# probability of delta_i=1 given beta, A, p and data #
# #
#####

fdelta = function(y,A,D,p,X,beta)
{
  term = sqrt((A+D)/D)*exp(-((y-X%*%beta)^2)*(1/D-1/(A+D))/2)
  alpha = p/(p+(1-p)*term)
  return(alpha)
}

#####
# #
# A function to Generate blank plot #
# #
#####

blankplot <- function()
{
  plot(0,0,type="n",xlim=c(0,1),ylim=c(0,1),xaxt ="n",
yaxt="n",xlab="",ylab="",bty="n")
}

# K Number of chains

K = 3

# Matrix to store values of ASD,ASRB,AAB and ARB

store = matrix(0,nrow=4,ncol=K)

```

```

rownames(store) = c("ASD", "ASRB", "AAB","ARB")

for(n in 1:K){

#####
# #
# Gibbs Sampling Algorithm #
# #
#####

for(k in 2:nsim)
{
#####
# #
# Generate a blank plot #
# #
#####

blankplot()

counter = round(k*100/nsim,1)

text(0.5,0.9,"Percentage of simulations completed is",
cex=1.4,col="blue")

text(0.4,0.6,counter,cex=7,col="blue")

#####
# #
#           Generating A           #
# #
# #
#####

# Ratio of Target and Proposal density

fAgA = function(sigma,delta,beta)
{
prod((exp(-0.5*((y-X%%beta)^2)*(1/(sigma+D)-1/(sigma+meanD))))*
((sigma+D)/(sigma+meanD))^(0.5) )^(delta))
}

fAgA.v <- Vectorize(fAgA,'sigma')

#####
# #
# Function to calculate M #
# #
#####

log.psi = function(A)
{
exp(-0.5*sum(Delta[k-1,]*(((Y-X%%beta[k-1,])^2)*
(meanD-D))/((A+D)*(A+meanD))+log(A+D)-log(A+meanD))))
}

```

```

log.psi.v = Vectorize(log.psi,"A")

M = max(log.psi.v(M.a))

if(sum(Delta[k-1,]) > 0)
{
T = 0

#####
#      #
# Acception-Rejection Algorithm #
#      #
#####

while (runif(1) > T)
{

R = 1

while(R > 1/meanD)
{

# Draw a random sample from gamma distribution
# Parameters of gamma distribution

sh = 1+(sum(Delta[k-1,])/2)

sc = 2/(t(Y-X%*%beta[k-1,])%*%diag(Delta[k-1,])%*%(Y-X%*%beta[k-1,]))

R = rgamma(1, shape=sh, scale=sc)
}

# Calculate the value of A

A = 1/R-meanD

# Calculate the ratio T

T = fAgA(A,Delta[k-1,],beta[k-1,])/M

}
sigma[k] = A
}else
{
U = runif(1)
sigma[k]=(1-U)*meanD/U
}

#####
#      #
#      Generating Beta from Multivariate      #
#      Normal distribution                      #
#      #
#####

# Variance-covariance matrix and Mean Vector

```

```

ratio1 = Delta[k-1,]/(sigma[k]+D)
ratio2 = (1-Delta[k-1,])/D
# covariance matrix for beta
H = solve(t(X)%*%diag(ratio1)%*%X+t(X)%*%diag(ratio2)%*%X)
g = t(X)%*%diag(ratio1)%*%Y+t(X)%*%diag(ratio2)%*%Y
#mean-vector for beta
meanbeta = H%*%g
beta[k,] = rmvnorm(1, mean =meanbeta, sigma=H)

#####
#           #
#   Generating delta using   #
#   function fdelta         #
#           #
#####

alph = fdelta(y=y,A=sigma[k],beta=beta[k,],D=D,p=P[k-1],X=X)
Delta[k,] = ifelse(runif(m)<=alph,1,0)

#####
#           #
# Generating P from beta distribution #
#           #
#####

# Parameters of beta distribution
shape1 = 1+sum(Delta[k,])
shape2 = m+1-sum(Delta[k,])
P[k] = rbeta(1, shape1=shape1, shape2=shape2)

} # end of gibbs sampling

#####
#   Calculating values of p(delta=1)#
#           #
#####

pdelta1 = array(0,c(nsim,m))

for(i in 1:nsim)
{
pdelta1[i,] = P[i]/(P[i]+(1-P[i]))*(sqrt((sigma[i]+D)/D))
*exp(-((y-X%*%beta[i,])^2/2)*(1/D-1/(sigma[i]+D))))
}

```

```
}

pdel = rowMeans(pdeltal)

#####
#   Calculating values of theta   #
#   #                               #
#####

  thetnw = array(0,c(nsim,m))

  for(i in 1:nsim)
  {
thetnw[i,] = X%*%beta[i,]+pdel[i]*sigma[i]*(y-X%*%beta[i,])/(sigma[i]+D)
  }

burnin = 5000

thetanw = colMeans(thetnw[burnin:nsim,])

ASD.NM = mean ((thetanw-theta)^2)
ASRB.NM = mean (((thetanw-theta)/theta)^2)
AAB.NM = mean (abs(thetanw-theta))
ARB.NM = mean (abs((thetanw-theta)/theta))

store[,n] = c(ASD.NM, ASRB.NM, AAB.NM,ARB.NM)

}
```


References

- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12:171–178.
- Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica*, 46:199–208.
- Battese, G. E., Harter, R. M., and Fuller, W. A. (1988). An error-components model for prediction of crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83:28–36.
- Brackstone, J. G. (1987). *Small Area Data: Policy issues and technical challenges*. In *Small Area Statistics*. New York: John Wiley and Sons, pp 3-20.
- Brooks, S. P. and Gelman, A. (1997). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7:434–455.
- Brooks, S. P. and Roberts, G. O. (1998). Assessing convergence of markov chain monte carlo algorithms. *Statistics and Computing*, 8:319–335.
- Chib, S. and Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The American Statistician*, 49:327–335.
- Cochran, W. G. (1977). *Sampling Techniques*. New York, John Wiley and Sons Inc., 3rd edition.
- Congdon, P. (2001). *Bayesian Statistical Modelling*. John Wiley & Sons, Chichester.

- Congdon, P. (2003). *Applied Bayesian Modelling*. John Wiley & Sons, Chichester.
- Cowles, M. K. and Carlin, B. P. (1996). Markov chain monte carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, 91:883–904.
- Datta, G. S., Lahiri, P., and Maiti, T. (2002). Estimation of median income of four-person families by state using timeseries and cross-sectional data. *Journal of Statistical Planning and Inference*, 102:83–97.
- Efron, B. and Morris, C. (1975). Data analysis using stein’s estimator and its generalizations. *Journal of the American Statistical Association*, 70(350):311–313.
- 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*, 85:398–409.
- Gelfand, A. E. (2000). Gibbs sampling. *Journal of the American Statistical Association*, 96:1300–1304.
- Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 83:398–409.
- Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2004). *Bayesian Data Analysis*. London, Chapman & Hall, 2nd edition.
- Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7:457–472.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distribution, and the bayesian restoration of images. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 6:721–741.
- Gershunskaya, J., Jiang, J., and Lahiri, P. (2009). *Resampling methods in surveys*. In: Handbook of Statistics 29B; Sample Surveys: Inference and Analysis, North Holland.

- Geweke, J. (1992). *Evaluating the Accuracy of Sampling-Based approaches to the Calculation of Posterior Moments*. In J.M. Bernardo, J.O. Berger, A.P. Dawiv, and A.F.M.Smith, *Bayesian Statistics*, volume 4, 169-193. Oxford University Press, Oxford.
- Ghosh, M., Nangia, N., and Kim, D. (1996). Estimation of median income of four-person families: a bayesian time series approach. *Journal of the American Statistical Association*, 91:1423–1431.
- Ghosh, M. and Rao, J. N. K. (1994). Small area estimation: An appraisal. *Statistical Science*, 9(1):55–76.
- Gilks, W., Spiegelhalter, D., and Richardson, S. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.
- Gonzalez, M. E. (1973). Use and evaluation of synthetic estimates. In *Proceedings of the Social Statistics Section*, American Statistical Association. USA, pp. 33-36.
- Hall, P. and Maiti, T. (2006). Nonparametric estimation of mean-squared prediction error in nested-error regression models. *Annals of Statistics*, 34:1733–1750.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). *Robust Statistics: The approach based on influence function*. New York: John Wiley and Sons.
- Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72:320–340.
- Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57:97–109.
- Heidelberger, P. and Welch, P. D. (1981). A spectral method for confidence interval generation and run length control in simulations. *Communication of the ACM*, 24:233–245.

- Heidelberger, P. and Welch, P. D. (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, 31:1109–1144.
- Henderson, C. R. (1953). Estimation of variance and covariance components. *Biometrics*, 9:226–252.
- Henderson, C. R. (1975). Best linear unbiased estimation and prediction under selection model. *Biometrics*, 31:423–447.
- Hobert, J. P. and Casella, G. (1996). The effect of improper priors on gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association*, 91:1461–1473.
- James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proceedings of the Fourth Berkeley Symposium on Mathematics and Probability*, Berkeley. University of California Press.
- Lahiri, P. (2003). On the impact of bootstrap in survey sampling and small-area estimation. *Statistical Science*, 18:199–210.
- Lahiri, P. and Rao, J. N. K. (1995). Robust estimation of mean squared error of small area estimators. *Journal of the American Statistical Association*, 82:758–766.
- Levy, P. S. (1979). Small area estimation—synthetic and other procedures, 1968-1979. Technical Report In Synthetic Estimates for Small Areas: Statistical Workshop Papers and Discussion, Washington, D.C., National Institute on Drug Abuse Vol.24.
- Liu, J. S. (2001). *Monte Carlo Strategies in Scientific Computing*. Springer-Verlag, New York.
- Lohr, S. and Rao, J. N. K. (2009). Jackknife estimation of mean squared error of small area predictors in nonlinear mixed models. *Biometrika*, 96(2):457–468.
- Longford, N. T. (2005). *Missing Data and Small Area Estimation*. Springer, New York.

- McCulloch, C. E. and Searle, S. R. (2001). *Generalized, Linear and Mixed Models*. John Wiley & Sons, New York.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092.
- Moura, F. and Holt, D. (1999). Small area estimation using multilevel models. *Survey Nethodology*, 25:73–80.
- Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58:545–554.
- Pfeffermann, D. (2002). Small area estimation- new developments and directions. *International Statistical Review*, 70:125–143.
- Prasad, N. G. N. and Rao, J. N. K. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, 85:163–171.
- Raftery, A. E. and Lewis, S. M. (1992). One long run with diagnostics: Implementation strategies for markov chain monte carlo. *Statistical Science*, 7:493–497.
- Raftery, A. E. and Lewis, S. M. (1996). *The Number of Iterations, Convergence Diagnostics and Generic Metropolis Algorithms*. In W.R. Gilks, D.J. Spiegelhalter, and S. Richardson, Markov Chain Monte Carlo in Practice. Chapman & Hall, London.
- Rao, J. N. K. (1999a). Some current trends in sample survey theory and methods (with discussion). *Sankhya: The Indian Journal of Statistics*, 61(1):1–57.
- Rao, J. N. K. (1999b). Some recent advances in model-based small area estimation. *Survey Methodology*, 25:175–186.
- Rao, J. N. K. (2003). *Small Area Estimation*. New Jersey, John Wiley & Sons, Inc.

- Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. New York, Springer-Verlag, 2nd edition.
- Roberts, G. O. (1996). *Markov Chain Concepts Related to Sampling Algorithms*. In W.R. Gilks, D.J. Spiegelhalter, and S. Richardson, *Markov Chain Monte Carlo in Practice*, 45-58. Chapman & Hall, London.
- Robinson, G. K. (1991). That blup is a good thing: The estimation of random effects. *Statistical Science*, 6:15–51.
- Saei, A. and Chambers, R. (2003). Small area estimates: A review of methods based on the application of mixed models. Technical Report Southampton Statistical Sciences Research Institute, Methodology Working Paper, M03/16.
- Sarndal, C. E., Swensson, B., and Wretman, J. (1992). *Model assisted survey sampling*. New York, Springer-Verlag.
- Schaible, W. A. (1978). Choosing weights for composite estimators for small area statistics. In *Proceedings of the Section on Survey Research Methods*, American Statistical Association. pp. 741-746.
- Searle, S. R., Casella, G., and McCulloch, C. E. (1992). *Variance Components*. John Wiley & Sons, New York.
- Spiegelhalter, D., Thomas, A., Best, N., and Gilks, W. (1996). *BUGS: Bayesian Inference Using Gibbs Sampling, Version 0.60*. MRC Biostatistics Unit, Cambridge.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and der Linde, A. V. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B*, 64:583–616.
- Tierney, M. A. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics*, 22(4):1701–1762.
- Torabi, M. and Rao, J. N. K. (2008). Small area estimation under a two-level model. *Survey Methodology*, 34:11–17.

-
- Valliant, R., Dortman, A. H., and Royall, R. M. (2000). *Finite Population Sampling and Inference*. John Wiley & Sons, New York.
- Verbeke, G. and Molenberghs, G. (1997). *Linear Mixed Models in Practice. A SAS-Oriented Approach*. Springer-Verlag, New York.
- Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer-Verlag, New York.
- Woodruff, R. (1966). Use of a regression technique to produce area breakdowns of the monthly national estimates of retail trade. *Journal of the American Statistical Association*, 61:496–504.
- You, Y. and Rao, J. N. K. (2000). Hierarchical bayes estimation of small area means using multi-level models. *Survey Methodology*, 26:173–181.

Antony Kibira Wanjoya

CURRICULUM VITAE

Contact information

Department of Statistical Sciences,
University of Padua,
via Cesare Battisti 241,
35121 Padova. Italy.
Tel. +39 049 827 4174
e-mail: wanjoya@stat.unipd.it

Dept. of Stats. and Actuarial Science,
Jomo Kenyatta University,
P.O. Box 62000-00200,
Nairobi, Kenya.
Tel. +254 722 336139
e-mail: awanjoya@gmail.com

Education

January 2008 – December 2010

PhD student in Statistical Sciences

Università degli Studi di Padova, Italy.

Expected completion: April 2011.

October 2001 – September 2002

MSc degree in Biostatistics

Limburgs Universitair Centrum, Diepenbeek, Belgium.

Title of dissertation: “Assessing the effects of ketamine on bispectral index during propofol anesthesia ”

October 2000 – July 2001

MSc degree in Applied Statistics

Limburgs Universitair Centrum, Diepenbeek, Belgium.

Title of dissertation: “Validation of surrogate markers using bootstrap approach ”

March 1996 – December 1999

BSc degree in Mathematics and Computer Science (first class honours)

Jomo Kenyatta University of Agriculture and Technology, Kenya.

Title of dissertation: “Robustness of standard deviation against representative outliers”

Other training

November 2007

Certificate in Project Monitoring and Evaluation
International Training Center, AMREF
Nairobi, Kenya.

May 2006

Certificate in Monitoring and Evaluation
Center for Finance and Project management
Nairobi, Kenya.

Work experience

December 2007 – present

Tutorial Fellow.

Department of Statistics and Actuarial Science,
Jomo Kenyatta University of Agriculture and Technology, Kenya.

January 2003 – December 2007

Assistant Biostatistician.

International Center for Insect Physiology and Ecology, Nairobi, Kenya.

June 2002 – September 2002

Summer Intern.

Department of Anesthesiology, Gent university hospital, Gent, Belgium.

Visiting periods

November 2009 – May 2010

Statistics department, University of Georgia
Athens, Georgia, USA.
Supervisor: Prof. Gauri Datta.

Awards and scholarships

January 2008 to December 2010

PhD Scholarship: Fondazione Cassa di Risparmio di Padova e Rovigo (CARI-PARO), Italy.

October 2000 to September 2002.

MSc Scholarship: Vlaamse Interuniversitair Raad (VLIR), Flemish Inter-University Council, Belgium.

June 1997

Owen Humpreys prize, Department of Mathematics and Computer Science, Jomo Kenyatta University of Agriculture and Technology, Kenya.

Computer skills

- Operating systems: Linux and Windows.
- Proficient: SAS, R, S-plus, SPSS and L^AT_EX.
- Intermediate: Stata, Epi-info and Microsoft Office suite.

Language skills

Language	Listening	Reading	Writing	Speaking
English	good	good	good	good
Swahili	good	good	good	fluent
Italian	fair	basic	basic	basic

Research interests

- Design and analysis of experiments
- Statistical methods in medical research and entomology
- Survey sampling
- Multi level modelling

- Bayesian inference
- Small area estimation

Publications

Refereed conference proceedings

Wanjoya, A., Torelli, N., & Datta, G. (2011). Small area estimation: A flexible area-level model. To appear in the *Proceedings of the 5th JKUAT Scientific, Technological and Industrialisation Conference “Science, Technology and Innovation for Industrialization and Sustainable Development”*, November, 2010.

Articles in journals

Migiro, L.N., Maniania, N.K., Chabi-Olaye, A., **Wanjoya, A.** & Van den Berg, J. (2011). Effect of Infection by *Metarhizium anisopliae* (Hypocreales: Clavicipitaceae) on the Feeding and Oviposition of the Pea Leafminer *Liriomyza huidobrensis* (Diptera: Agromyzidae) on Different Host Plants. *Biological Control*, 56(2), 179-183.

Githinji E.K., Irungu, L.W., Tonui, W.K., Rukunga, G.M., Mutai, C., Muthaura, C.N., Lugalia, R., Gikandi, G., Wainaina, C.W., Ingonga, J.M. & **Wanjoya, A.K.** (2010). In Vitro Effects of *Warburgia Ugandensis*, *Psiadia Punctulata* and *Chasmanthera Dependens* on *Leishmania Major* Promastigotes. *African Journal of Traditional, Complementary and Alternative Medicines*, 7 (3), 264 - 275.

Howard, A.F.V., Adongo, E.A., Hassanali, A., Omlin, F.X., **Wanjoya, A. K.**, Zhou, G. & Vulule, J. (2009). Laboratory Evaluation of the Aqueous Extract of *Azadirachta indica* (Neem) Wood Chippings on *Anopheles gambiae* s.s. (Diptera: Culicidae) Mosquitoes. *Journal of Medical Entomology*, 46(1), 107-114.

Bugeme, D.M., Knapp M., Hamadi, I.B., **Wanjoya, A. K.**, & Nguya, K.M. (2009). Influence of Temperature on Virulence of Fungal Isolates of *Metarhizium anisopliae* and *Beauveria bassiana* to the Two-Spotted Spider Mite *Tetranychus urticae*. *Mycopathologia*, 167, 221-227.

Mburu, D.M., Ochola, L., Maniania, N.K., Njagi, P.G.N., Gitonga, L.M., Ndung'u, M.W., **Wanjoya, A.K.** & Hassanali, A. (2009). Relationship between virulence and repellency of entomopathogenic isolates of *Metarhizium anisopliae* and *Beauveria bassiana* to the termite *Macrotermes michaelseni*. *Journal of Insect Physiology*, 55, 774-780.

Calatayud, P.A., Guénégo, H., Ahuya, P., **Wanjoya, A.K.**, Le Rü, B., François, J.S. & Brigitte Frérot (2008). Flight and oviposition behaviour of the African stem borer, *Busseola fusca*, on various host plant species. *Entomologia Experimentalis et Applicata*, 129, 348-355.

Calatayud, P.A., Ahuya, P., **Wanjoya, A.K.**, Le Rü, B., Silvain, J.F. & Frérot, B. (2008). Importance of plant physical cues in host acceptance for oviposition by *Busseola fusca*. *Entomologia Experimentalis et Applicata*, 126 (3), 233-243.

Khan, Z., Midega, C.O., Hassanali, A., Pickett, P., Wadhams, L. & **Wanjoya, A.K.** (2006). Management of witchweed, *Striga hermonthica*, and stemborers in sorghum, *Sorghum bicolor*, through intercropping with greenleaf desmodium, *Desmodium intortum*. *International Journal of Pest Management*, 52(4), 297 - 302.

Conference and seminar presentations

Small area estimation: A flexible model. *Seminari intermedi dottorandi XXIII cycle*, Department of Statistical Sciences, University of Padua, Padua, Italy, June 04, 2010.

Small area estimation: An application of a flexible Area-Level model. *The 5th JKUAT Scientific, Technological and Industrialization Conference*, Nairobi, Kenya, November 17-19, 2010.

Service to the Community

Volunteer for the Sagana Wildlife Conservation Group (SAWC); since 2002. The mission of SAWC is to assist in the management and conservation of wildlife in Mt. Kenya forest, Kenya, through working with local communities in education, wildlife conflict resolution and habitat restoration.