ESTIMATION OF GOVERNMENT EMPLOYMENT USING MULTIVARIATE HIERARCHICAL BAYES MODELING

by

ANDREW JOHN SCOTT

(Under the Direction of Gauri S. Datta)

ABSTRACT

Small area estimation is a topic of considerable interest for many agencies wishing to use data collected at the national level to accurately estimate statistics specific to subsets of the general population. Our approach used a lognormal multivariate hierarchical Bayes (MVHB) mixed model to generate improved estimates of small area totals for Census of Governments data drawn through stratified random sampling (SRS) and skewed data drawn through probability proportional to size (PPS) sampling. Additionally, we explored the inclusion of design weights as model covariates for PPS-sampled data. Gibbs samplers were developed to provide estimates of posterior mean and variance. We show that the MVHB models produced improved estimates with decreased posterior variance, when compared to their univariate counterparts for both the PPS and SRS samples.

INDEX WORDS: Multivariate hierarchical Bayes, lognormal models, Gibbs sampling, multivariate mixed linear model, small area estimation, design weights, probability proportional to size.

ESTIMATION OF GOVERNMENT EMPLOYMENT USING MULTIVARIATE HIERARCHICAL BAYES MODELING

By

ANDREW JOHN SCOTT

B.A., The University of Georgia, 2009

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2014

ESTIMATION OF GOVERNMENT EMPLOYMENT USING MULTIVARIATE HIERARCHICAL BAYES MODELING

By

ANDREW JOHN SCOTT

Major Professor:	Gauri S. Datta
Committee:	Kimberly Love-Myers Abhyuday Mandal

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2014

Acknowledgements

I would like to express my thanks to my committee: Doctors Love-Myers and Mandal for their guidance in ensuring the quality of this work, and to Doctor Datta for the patience, guidance, and incomparable expertise that made this possible.

I would also like to express my thanks to my family, an endless source of love and support, and to the Cooks, for making me feel like family — I do not know of a better compliment.

Contents

List of Tables **vi**

List of Figures **vii**

1 Introduction and Literature Review **1**

1.1 Problem Background 1

1.2 Literature and Methods 2

1.3 Organization 4

2 The Data **6**

2.1 Sample Design 6

2.2 Subsampling 8

3 The Models **10**

3.1 Notation 10

3.2 The Multivariate Hierarchical Bayes Model 12

3.3 The Univariate Hierarchical Bayes Model 20

4 Simulation and Results **23**

4.1 The Gibbs Sampler 23

4.2 Posterior Mean and Variance 25

4.3 Sample 1 (PPS) Results 26

4.4 Sample 2 (SRS) Results 32

5 Conclusion **36**

A Example Census of Governments Data **37**

B Implementing the Gibbs Sampler in R **38**

Bibliography **41**

List of Tables

4.1 PPS Sample Relative Posterior Variance 30

4.2 PPS ARE Results 30

4.3 SRS ARE Results 34

4.4 SRS Sample Relative Posterior Variance 34

A.1 Example Data from the Combined 2002, 2007 Census of Governments 37

List of Figures

2.1 Sample Size by Small Area, Independent Samples Taken from the 2007 Census of Governments 9

3.1 2007 Full- & Part-Time Employment Partial Correlations, by Small Area and 2002 Full-Time Employment 11

4.1 PPS Relative Posterior Variance, PPS Sample, Unweighted Models 27

4.2 PPS Sample Relative Posterior Variance, Weighted Models 28

4.3 PPS Sample Relative Posterior Variance by Sample Size, Unweighted Models 29

4.4 PPS Sample Relative Posterior Variance by Sample Size, Weighted Models . 29

4.5 PPS Sample Relative ARE by Sample Size, Unweighted Models 31

4.6 PPS Sample Relative ARE by Sample Size, Weighted Models 31

4.7 SRS Sample Relative Posterior Variance 33

4.8 SRS Sample Relative Posterior Variance by Sample Size 33

4.9 SRS Sample Relative ARE by Sample Size 35

Chapter 1: Introduction and Literature Review

1.1 Problem Background

Organizations are often interested in using survey data to provide reliable estimates of population characteristics for small areas. A small area refers to a subset of the general population surveyed, defined by a number of characteristics including, but not limited to, geographic classifications (states, counties, districts, etc.), organizational classifications (fire departments, hospitals, etc.), and demographic classifications (race, gender, ethnicity, age, etc.). Most surveys are designed to provide estimates for larger populations, without respect to the small areas of which they are comprised. Thus, inferring information on small area population characteristics using design-based estimates (informed only by the methodology of the initial survey) will have limited reliability due to the small sample size of each small area. Improved estimates may be obtained, as they are in the following discussion, from methods that 'borrow strength' from other small areas and related covariates available in the survey data and auxiliary variables available from other records, often prior censuses or surveys.

This discussion outlines and describes the results of an empirical study involving data obtained in the U.S. Census Bureau's Annual Survey of Public Employment and Payroll (ASPEP). This nationwide survey provides information on all state and local governments in the United States and is the only source of information by program function and full-time/part-time employment status. Program function refers to the description of the function of each governmental unit (firefighting, financial administration, streets and highways, sewerage, etc.). The U.S. Census Bureau is required to conduct a full Census of Governments (CoG) every five years (in years ending in 2 or 7). The ASPEP is an annual extension of that effort. Our study focuses specifically on the data from the 2002 and 2007 censuses. The ASPEP and Census of Governments records deal in individual city, county, township, special district, and school district governments across the United States (noting that some states may have neither townships, special districts, nor both). Per the decision by the Census Bureau that cities and townships are similar enough, their strata are for all intents and purposes combined. The employment data collected for each government are further subclassified by program function into governmental units. Hawaii and the District of Columbia are excluded, having a relatively small number of governments permits the sampling of all governments in the ASPEP, see Barth, Cheng, and Hogue [2009] and U.S [2006].

Our estimation procedure is based on the application of a multivariate hierarchical Bayes regression model. Our small areas are defined by program function within forty-nine U.S. states.

Analysis is performed with respect to small areas defined by each governmental unit's state and program function classification in 2007. The survey selection procedure is summarized in Section 2.1.

1.2 Literature and Methods

For survey data of the type generated by the ASPEP, the default purely design-based approach to estimation is through a Horvitz-Thompson (HT) estimator. Given a parameter of interest and a small area, say the total number of full-time employees and the firefighting program function for the state of Missouri, the HT estimator weights the full-time employee count of each sampled unit within the small area by the inverse of its probability of inclusion in the sample. An application of this approach is used in [Cheng, Corcoran, Barth, and Hogue, 2009, Section 2.1]. Design-based estimators are described in increasingly greater detail in Rao [2003], which provides an excellent account of small-area estimation in general, and Cochran [1977].

The discussion of [Cheng et al., 2009, Section 4] demonstrates the advantage of their model-based estimator in its ability to provide more accurate estimates than a purely design-based estimator. Note that the model-assisted approach of Cheng et al. [2009] incorporates the design-weights into the estimation of parameters, unlike an exclusively model-dependent approach.

In Battese, Harter, and Fuller [1988], a small-area estimation procedure is introduced that employs a nested-error linear mixed model. This model fits a components of variance error structure, which assumes the random errors within a small area are correlated and models the small area means as a linear function of fixed and random small area effects.

The more recent discussion of Cheng, Tran, Hogue, and Lahiri [2013] illustrates the use of a unit-level mixed lognormal model on the employment data in small area estimation. Their mixed model is applied to a sub-sample of the 2007 Census of Governments drawn through simple random sampling. Hence, design weight is a non-issue in their estimation procedure as all sampled units would clearly have identical weights. Their proposed lognormal mixed model provided 'generally better' estimates than both an HT model and their application of the nested-error linear mixed model introduced by Battese et al. [1988].

The application of a multivariate nested error regression model to small area estimation is outlined in Datta, Day, and Basawa [1999], which derived the empirical best linear unbiased predictor (EBLUP) for multivariate hierarchical small area models related to the production of small area estimates. These EBLUP estimates are quite accurate in the multivariate case and superior to comparable univariate estimates (per the simulation results of [Datta et al., 1999, Section 5]), as well as very good approximations to hierarchical Bayes estimates.

The hierarchical Bayes (HB) approach, as applied to small area estimation, is detailed first in Datta and Ghosh [1991]. Datta et al. [1999] outlines the empirical Bayes prediction methodology. This approach is expanded in Datta, Day, and Maiti [1998], which extends the modeling strategy to a hierarchical Bayes treatment and again demonstrating the potential improvement of estimates generated by a multivariate model over those generated by a univariate model, provided a degree of correlation in the variates of the multivariate model, as in the ASPEP/CoG.

Faced with unreliability when using direct numerical integration to provide estimates, as in Datta et al. [1998], we use the Gibbs sampler, a special case of the Markov Chain Monte Carlo method of integration, to generate estimates as described in Section 4.1. A general introduction to the Gibbs sampler is presented in Casella and George [1992], the application of the Gibbs sampler to small area estimation problems is covered in Rao [2003], the Gibbs sampler and other Markov Chain Monte Carlo methods in the Bayesian context are discussed in Lee [2004], and specific examples of applications are found in Holmes and Held [2006].

As an alternative to the standard method of design-based estimation, the inclusion of sample weights as a covariate in a linear model is presented in Zheng and Little [2005] to good result. The approach they present deals with the case of a sample drawn through the probability proportional to size method, as in the ASPEP.

However the use of log-transformed data in this method of estimation is not without its perils, as noted in [Chandra and Chambers, 2011, Section 5]. Though their effort to develop a reliable bias correction is neither suited for hierarchical Bayes estimation nor adapted for multivariate modeling, we do make an effort to incorporate a design-weighted component into our model in order to correct for this inherent bias. Specifically, we consider the case in which the design weights are included as a covariate, per the discussion of Zheng and Little [2005], which elaborates on prediction under probability proportional to size sampling via a penalized spline nonparametric regression model.

One of the main goals of this discussion is to apply the hierarchical Bayes proposal of Datta et al. [1998] to the lognormally-distributed and inherently multivariate government employment data and to produce estimates superior to the univariate model of Cheng et al. [2013]. We test our proposed multivariate hierarchical Bayes models on a sample of the government employment data drawn from the 2007 Census of Governments data. The sample is drawn with the intention of simulating an ASPEP for the 2007 year.

1.3 Organization

Chapter 2 offers a description of the ASPEP/CoG data, delineating the variables used in the analysis, their more salient properties, and the small areas analysed. Chapter 3 presents our proposed multivariate hierarchical Bayes (MVHB) model, as well as a competing model, specifically, the model proposed by Cheng et al. [2013] as a univariate case of the hierarchical Bayes model (UVHB). Estimation via the implementation of a Gibbs sampler is detailed with results in chapter 4 while chapter 5 brings the analyses to conclusion.

Chapter 2: The Data

The Annual Survey of Public Employment and Payroll (ASPEP) and Census of Governments (CoG) data details full- and part-time employment and payroll for all fifty U.S. states plus the District of Columbia down to the individual government unit level. Governmental units classify groups government employees by the state, type of government, and program function to which they belong. Each government is of only one type: city, county, township, special district, or school district. Note also that some states do not contain some governmental types, such as townships or special districts. Individual governments are composed of governmental units, groups of government employees with a common program function. Program function classifies employees and their associated payroll under one of twenty-nine categories including but not limited to: airports, corrections, firefighters, local libraries, parks & recreation, welfare, and sewage. Within the data, program function is referred to as 'itemcode'; the two terms are equivalent for the analyses of this document. Selection for inclusion in the ASPEP is performed at the government level; a case would never occur in which the firefighters of one particular government would be included in a sample but the same unit's local libraries would be excluded.

Our primary variable of interest in the following discussion is the number of full-time employees for each small area, defined by all valid combinations of state and program function. Our proposed multivariate model jointly estimates full- and part-time employment as well as total pay for the 2007 year, employing the full- and part-time employment and total pay from the previous census (2002) as explanatory variables. Classification of government units is described in greater detail in U.S [2006].

2.1 Sample Design

Per Barth et al. [2009], the ASPEP is taken via a two stage sampling approach. Sampling is performed within strata defined by valid combinations of state and type of government, of which there are one hundred and eighty-seven. First, an initial sample is drawn using probability proportional to size (PPS) sampling, see Cochran [1977], with total pay (the sum of full- and part-time payroll for each strata) as the size variable. Under PPS sampling, the probability of a government's inclusion in the sample, π_{hj}, is defined as:

$$\pi_{hj} = \begin{cases} n_h t_{hj}/T_h & \text{for} \quad n_h t_{hj} \leq T_h \\ 1 & \text{for} \quad n_h t_{hj} > T_h \end{cases} \tag{2.1}$$

where the state and government type strata are indexed by h, each containing n_h governments, and the individual governments are indexed by j. Under this scheme of indexing the units, the size of government j in stratum h is designated t_{hj} and the sum of the sizes of all governments in stratum h is designated T_h.

Second, the governments within each strata are classified into large and small substrata using the cumulative square root of frequency (CSRF) method, outlined in Cochran [1977], and a sub-sample of individual governments is taken from the stratum comprised of the smaller units. This subsample from the smaller units is drawn through simple random sampling.

Prior to sub-sampling in this manner, small governments, which contributed little to overall employment and payroll totals were over-represented in the sample. Note that this sampling methodology does not apply to all units in the population. Some governments are guaranteed to be included in the ASPEP based on criteria independent of the PPS and CSRF routine. These are designated as 'certainty units'. These units comprise a non-trivial percentage of the population of governments as of 2011 and of the ASPEP 2011 sample; while 'certainty units' comprised only 6% of the population of Governments in 2011, they were represented by 45% of the ASPEP sample.

A more in-depth description of the ASPEP sampling methodology is provided in [Barth et al., 2009, Section 4]. PPS and stratified sampling are discussed more completely in

Cochran [1977]. Ten rows from the combined 2002 and 2007 Census of Governments data are presented in Appendix A for illustration.

2.2 Subsampling

Two samples were drawn independently from the 2007 Census of Governments data to allow comparison of analysis results to known values. The state of Hawaii and the District of Columbia were excluded from the sample: their small number of governments permits all units in those states to be sampled in the ASPEP. "Sample 1" is based on the PPS sampling scheme used in the ASPEP, with sample sample sizes made proportional to those used in the ASPEP 2011 survey. "Sample 2" uses the simple random sampling within the strata defined by states and types of governments to draw samples from the 2007 Census of Governments, with sample sizes per strata chosen such that the same proportion of the universe of governments was sampled in our 2007 sample as was sampled in the 2011 ASPEP. After sampling, the units are classified by small area (each valid combination of state and program function), and units from areas that either have all or no units sampled are discarded from that sample. Sample sizes after this process are summarized in Appendix A.

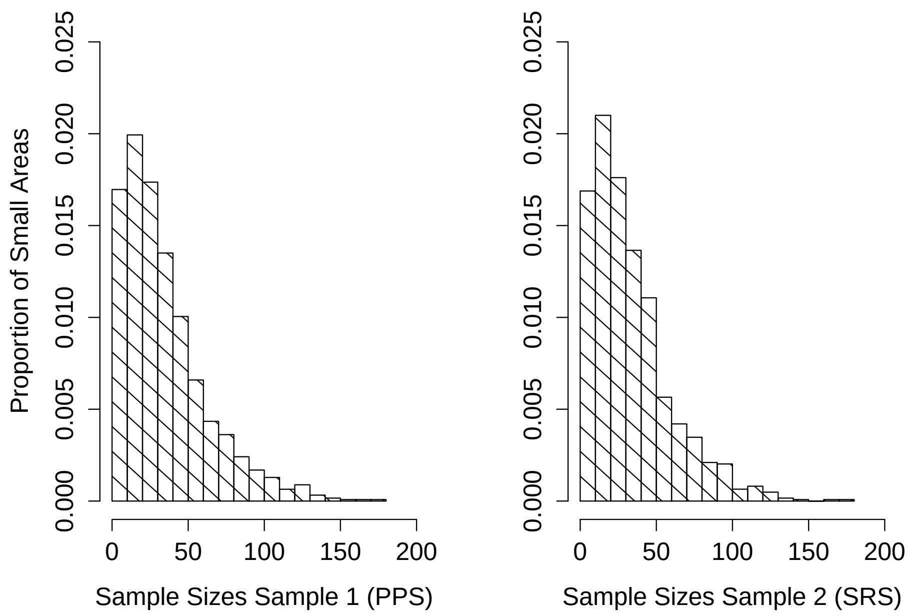

Figure 2.1: Sample Size by Small Area, Independent Samples Taken from the 2007 Census of Governments

Chapter 3: The Models

The expected improvement in model accuracy over a univariate model is supported by the discussion on multivariate estimation versus univariate estimation in [Datta et al., 1998, Sec. 3], as well as an examination of the partial correlations between full- and part-time pay performed by Bac Tran of the Census Bureau (2013: personal communication). Partial correlations between 2007 full- and part-time employment by small area (FT07 and PT07, respectively) conditioned on 2002 full-time employment by small area (FT02) are illustrated in Figure 3. Though the partial correlations are by no means overwhelmingly high for all small areas, 52% (657 of the 1268 valid combinations of state and program function yielding sufficient information) show absolute partial correlations greater than .2, based on the 2002/2007 Censuses of Governments. Partial correlations in Figure 3 are computed within each small area i, as:

$$\rho_{\text{FT07,PT07}|\text{FT02},i} = \frac{\rho_{\text{FT07,PT07},i} - \rho_{\text{FT07,FT02},i} \times \rho_{\text{FT02,PT07},i}}{\sqrt{1-\rho^2_{\text{FT07,FT02},i}}\sqrt{1-\rho^2_{\text{FT02,PT07},i}}}, i = 1, \ldots, m$$

Previously, a univariate linear mixed model was employed on data from the 2002 and 2007 Censuses of Governments, as described in Cheng et al. [2013], which used the strategy of estimating current year full-time employment from the records of the most recent known population data (Census of Governments 2002) and a contemporary sample (a sub-sample of Census of Governments 2007). A univariate model analogous to that model is constructed in Section 3.3, for comparison. Our multivariate model is defined as follows.

3.1 Notation

The following sections contain several lengthy equations, often involving matrix algebra. These guidelines for notation will be observed whenever practical, with exceptions promptly noted.

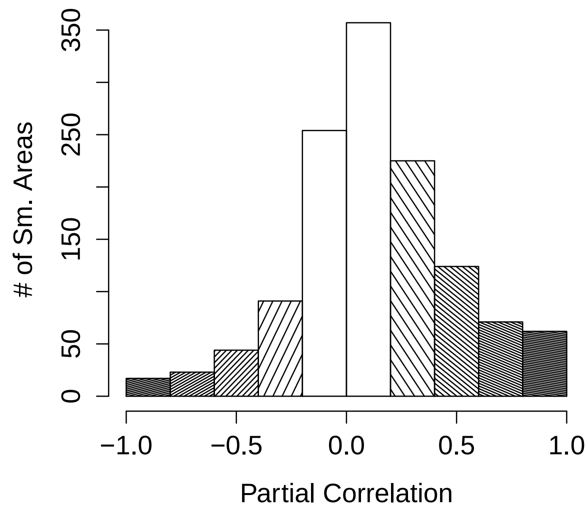

Figure 3.1: 2007 Full- & Part-Time Employment Partial Correlations, by Small Area and 2002 Full-Time Employment

1. Matrices will be represented with boldface capital letters, with dimensions listed in parenthetical subscripts, where number of rows precedes number of columns, e.g. $\mathbf{D}_{(p \times q)}$ denotes a matrix of p rows and q columns.

2. Any single element of a matrix will be represented by the lowercase form of its parent matrix's identifying letter.

3. Column vectors are denoted with arrow accents using lowercase letters. When necessary, their dimension is shown via parenthetical subscript. Vectors denoted by integers have all elements equal to that integer.

4. Transposes of matrices (or vectors) are denoted with the superscript $^\top$ e.g. $\vec{d}^\top_{(a)}$ denotes a row vector of length a.

5. $\mathbf{A} \otimes \mathbf{D}$ denotes the Kronecker product of matrices, such that:

$$\mathbf{A} \otimes \mathbf{D} = \begin{bmatrix} a_{11}d_{11} & a_{11}d_{12} & \dots & a_{12}d_{11} & a_{12}d_{12} & \dots \\ a_{11}d_{21} & a_{11}d_{22} & \dots & a_{12}d_{21} & a_{12}d_{22} & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \\ a_{21}d_{11} & a_{21}d_{12} & \dots & a_{22}d_{11} & a_{22}d_{12} & \dots \\ a_{21}d_{21} & a_{21}d_{22} & \dots & a_{22}d_{21} & a_{22}d_{22} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

3.2 The Multivariate Hierarchical Bayes Model

Let z_{ij1} equal to the number of 2007 full-time employees and z_{ij2} equal to the number of 2007 part-time employees of the j^{th} governmental unit in the i^{th} small area, $i = 1, \dots, m$ and $j = 1, \dots, N_i$ for each of m small areas each containing N_i units. Our model seeks to estimate the total number of full-time employees for each small area, indexed $i = 1, \dots, m$:

$$\theta_i = \sum_{j=1}^{N_i} z_{ij1}. \tag{3.1}$$

Taking $y_{ijh} = \ln(z_{ijh} + 1)$; $h = 1, 2$; x_{ij1} equal to the natural logarithm of the number of 2002 full-time employees plus one, and x_{ij2} equal to the natural logarithm of the number of 2002 part-time employees plus one for the j^{th} governmental unit of the i^{th} small area with $i = 1, \dots, m$ and $j = 1, \dots, N_i$ indexed as previously. The logarithmic transformation is taken to mitigate heteroscedasticity present in the data, as in Cheng et al. [2013].

Define also:

$$\vec{y}_{ij} = \begin{pmatrix} y_{ij1} \\ y_{ij2} \end{pmatrix}, \vec{x}_{ij} = \begin{pmatrix} 1 \\ x_{ij1} \\ x_{ij2} \end{pmatrix}.$$

We then have the following multivariate linear mixed model:

$$\vec{y}_{ij} = \mathbf{B}\vec{x}_{ij} + \vec{v}_i + \vec{e}_{ij}; \tag{3.2}$$

$$\vec{v}_i \overset{i.i.d.}{\sim} \mathcal{N}_2(\vec{0}, \mathbf{\Sigma_v}) \tag{3.3}$$

$$\vec{e}_{ij} \overset{i.i.d.}{\sim} \mathcal{N}_2(\vec{0}, \mathbf{\Sigma_e}); \tag{3.4}$$

$$\text{independently for } i = 1, \ldots, m;\ j = 1, \ldots, N_i; \tag{3.5}$$

where $\mathbf{B} = [(\beta_{hk})]_{(2\times 3)};\ h = 1, 2;\ k = 0, 1, 2$ is a matrix of unknown coefficients, $\vec{v}_i$ are vectors of small area-specific random effects and $\vec{e}_{ij}$ are vectors of sampling errors, such that:

$$\vec{v}_i = \begin{pmatrix} v_{i1} \\ v_{i2} \end{pmatrix}, \mathbf{\Sigma_v} = \begin{pmatrix} \sigma_{v11} & \sigma_{v12} \\ \sigma_{v12} & \sigma_{v22} \end{pmatrix}, \vec{e}_{ij} = \begin{pmatrix} e_{ij1} \\ e_{ij2} \end{pmatrix},\ \text{and } \mathbf{\Sigma_e} = \begin{pmatrix} \sigma_{e11} & \sigma_{e12} \\ \sigma_{e12} & \sigma_{e22} \end{pmatrix} \tag{3.6}$$

for $i = 1, \cdots, m$ and $j = 1, \cdots, N_i$.

Define further:

$$\mathbf{Y}_i = [\vec{y}_{i1} \cdots \vec{y}_{iN_i}], \mathbf{Y} = [\mathbf{Y}_1 \cdots \mathbf{Y}_m], \tag{3.7}$$

$$\mathbf{X}_i = [\vec{x}_{i1} \cdots \vec{x}_{iN_i}], \mathbf{X} = [\mathbf{X}_1 \cdots \mathbf{X}_m], \tag{3.8}$$

$$\mathbf{V} = [\vec{v}_1 \cdots \vec{v}_m], \tag{3.9}$$

$$\mathbf{F} = \bigoplus_{i=1}^{m} \vec{1}_{N_i}^{\top} = \begin{bmatrix} \vec{1}_{(N_1)}^{\top} & \vec{0}_{(N_1)}^{\top} & \cdots & \cdots & \vec{0}_{(N_1)}^{\top} \\ \vec{0}_{(N_2)}^{\top} & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \vec{1}_{(N_i)}^{\top} & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \vec{0}_{(N_{m-1})}^{\top} \\ \vec{0}_{(N_m)}^{\top} & \cdots & \cdots & \vec{0}_{(N_m)}^{\top} & \vec{1}_{(N_m)}^{\top} \end{bmatrix}, \tag{3.10}$$

$$\mathbf{E}_i = [\vec{e}_{i1} \cdots \vec{e}_{iN_i}],\ \text{and } \mathbf{E} = [\mathbf{E}_1 \cdots \mathbf{E}_m]. \tag{3.11}$$

This allows us to specify equation (3.2) more concisely as:

$$\mathbf{Y}{=}\mathbf{BX}{+}\mathbf{VF}{+}\mathbf{E}. \tag{3.12}$$

Following Datta et al. [1999] and Datta et al. [1998], we have the following hierarchical Bayes model with unknown $\mathbf{B}$, $\mathbf{V}$, $\mathbf{\Sigma_v}$ and $\mathbf{\Sigma_e}$:

3.2.1 Model MV

(i) Conditional on $\mathbf{B}$, $\mathbf{V}$, $\mathbf{\Sigma_v}$, and $\mathbf{\Sigma_e}$: $\vec{y}_{ij} \sim \mathcal{N}_2(\mathbf{B}\vec{x}_{ij} + \vec{v}_i, \mathbf{\Sigma_e})$ independently for $i = 1, \ldots, m$ and $j = 1, \ldots, N_i$.

(ii) Conditional on $\mathbf{B}$, $\mathbf{\Sigma_v}$, and $\mathbf{\Sigma_e}$: $\vec{v}_i \sim \mathcal{N}_2(\vec{0}, \mathbf{\Sigma_v})$ independently for $i = 1, \ldots, m$.

(iii) Marginally, $\mathbf{B}$, $\mathbf{\Sigma_v}$, and $\mathbf{\Sigma_e}$ are independently distributed with:
$\mathbf{B} \sim$ *Uniform* on $\mathbb{R}^{2\times 3}$;
${\mathbf{\Sigma_v}}^{-1} \sim \mathcal{W}_a(\mathbf{\Phi}_v)$ for positive-definite $\mathbf{\Sigma_v}$ and positive-semi-definite $\mathbf{\Phi}_v$; and
${\mathbf{\Sigma_e}}^{-1} \sim \mathcal{W}_b(\mathbf{\Phi}_e)$ for positive-definite $\mathbf{\Sigma_e}$ and positive-semi-definite $\mathbf{\Phi}_e$.

$\mathcal{N}_p(\vec{a}, \mathbf{\Sigma})$ denotes the p-variate Normal distribution with mean $\vec{a}_{(p)}$ and variance-covariance matrix $\mathbf{\Sigma}_{(p\times p)}$. $\mathcal{W}_d(\mathbf{\Phi})$ denotes a Wishart distribution with d degrees of freedom and positive-definite scale matrix $\mathbf{\Phi}_{(q\times q)}$. The p.d.f of a Wishart distribution $\mathcal{W}_d(\mathbf{\Phi})$ is specified generally as:

$$p(\mathbf{W}|d, \mathbf{\Phi}) \propto \exp(-\frac{1}{2}\text{trace}(\mathbf{W}\mathbf{\Phi}^{-1}))|\mathbf{W}|^{\frac{d-q-1}{2}},$$

for some positive-definite matrix $\mathbf{W}_{(q\times q)}$, constant q. Taking d equal to $q+1$ and Φ equal to $\mathbf{0}$ the p.d.f. collapses to $p(\mathbf{W}|d, \mathbf{\Phi}) \propto 1$, comparable to a multivariate Uniform distribution for $(q \times q)$ positive-definite matrices. After rearranging the columns of $\mathbf{Y}$, $\mathbf{X}$, $\mathbf{F}$, and $\mathbf{E}$ from equation (3.12) into sampled and non-sampled units, $j = (1, \ldots, n_i$ or $n_i + 1, \ldots, N_i)$, we have:

$$(\mathbf{Y}^{(1)}|\mathbf{Y}^{(2)}) = \mathbf{B}(\mathbf{X}^{(1)}|\mathbf{X}^{(2)}) + \mathbf{V}(\mathbf{F}^{(1)}|\mathbf{F}^{(2)}) + (\mathbf{E}^{(1)}|\mathbf{E}^{(2)}) \tag{3.13}$$

Here $\mathbf{Y}^{(1)}_{(2\times n)}$ is a matrix of sampled units, $\mathbf{Y}^{(2)}_{(2\times(N-n))}$ is a matrix of non-sampled units, $n = \sum_{i=1}^{m} n_i$, and $N = \sum_{i=1}^{m} N_i$. Analogous rearrangements are repeated for matrices $\mathbf{X}$, $\mathbf{F}$, and $\mathbf{E}$. This follows closely the model specification of [Datta et al., 1998, p.4-5]. Under Model MV, defining: $y_{ij1} = \vec{\beta}_{1\cdot}^{\top}\vec{x}_{ij} + v_{i\cdot 1} + e_{ij1}$ for $\vec{\beta}_{1\cdot}^{\top} = (\beta_{10}, \beta_{11}, \beta_{12})$ and $i = 1, \ldots, m$, and $j = 1, \ldots, N_i$, we obtain the estimators:

$$\hat{\theta}_i^{\mathrm{MV}} = \sum_{j=1}^{N_i}[\exp(\vec{\beta}_{1\cdot}^{\top}\mathbf{X}_{ij} + \vec{v}_{i1} + e_{ij1}) - 1];\ i = 1, \ldots, m; \tag{3.14}$$

Equation (3.14) holds true for $\vec{e}_{\cdot\cdot 1}^{\top}$ independent of $\vec{y}_{\cdot\cdot 1}^{\top} = \mathrm{concat}_{1\le i\le m}(\mathrm{concat}_{1\le j\le N_i}(y_{ij1}))$. We will now derive the conditional mean and variance of $\hat{\theta}_i^{\mathrm{MV}}$. Arranging the components of $\mathbf{Y}^{(1)}$ into $\mathbf{Y}_i^{(1)} = \mathrm{col}(\vec{y}_{i1} \ldots \vec{y}_{in_i})_{(2n_i\times 1)}$, for $i = 1, \ldots, m$; and specifying $\mathbf{J}_a$ as square unit matrices and $\mathbf{I}_a$ as identity matrices, each having dimensions of length a, for $\vec{v}_i$ independent of $\vec{e}_{i\cdot\cdot}$,

$$\mathrm{Var}[\mathbf{Y}_i^{(1)}] = \begin{bmatrix} \mathbf{\Sigma_v} + \mathbf{\Sigma_e} & \mathbf{\Sigma_v} & \ldots & \mathbf{\Sigma_v} \\ \mathbf{\Sigma_v} & \mathbf{\Sigma_v} + \mathbf{\Sigma_e} & \ldots & \mathbf{\Sigma_v} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Sigma_v} & \mathbf{\Sigma_v} & \ldots & \mathbf{\Sigma_v} + \mathbf{\Sigma_e} \end{bmatrix} = \mathbf{J}_{n_i} \otimes \mathbf{\Sigma_v} + \mathbf{I}_{n_i} \otimes \mathbf{\Sigma_e}. \tag{3.15}$$

$$\left[\mathrm{Var}[\mathbf{Y}_i^{(1)}]\right]^{-1} = [\mathbf{I}_{n_i} - n_i^{-1}\mathbf{J}_{n_i}] \otimes \mathbf{\Sigma_e}^{-1} + n_i^{-1}\mathbf{J}_{n_i} \otimes [\mathbf{\Sigma_e} + n_i\mathbf{\Sigma_v}]^{-1}. \tag{3.16}$$

$$\mathrm{Cov}[\vec{v}_i, \mathbf{Y}_i^{(1)}] = \mathrm{Cov}\left[\vec{v}_i, \begin{pmatrix} \vec{v}_i \\ \vdots \\ \vec{v}_i \end{pmatrix}\right] = \Big[\mathbf{\Sigma_e}, \ldots, \mathbf{\Sigma_e}\Big]_{2\times 2n_i} = \vec{1}_{n_i}^{\top} \otimes \mathbf{\Sigma_v}. \tag{3.17}$$

Thus,

$$\begin{aligned} \left[\mathrm{Cov}[\vec{v}_i, \mathbf{Y}_i^{(1)}]\right]\left[\mathrm{Var}[\mathbf{Y}_i^{(1)}]\right]^{-1} &= \vec{0}_{n_i} \otimes \mathbf{\Sigma_v}\mathbf{\Sigma_e}^{-1} + \vec{1}_{n_i} \otimes \mathbf{\Sigma_v}(\mathbf{\Sigma_e} + n_i\mathbf{\Sigma_v})^{-1} \\ &= \left[\mathbf{\Sigma_v}[\mathbf{\Sigma_e} + n_i\mathbf{\Sigma_v}]^{-1}, \ldots, \mathbf{\Sigma_v}[\mathbf{\Sigma_e} + n_i\mathbf{\Sigma_v}]^{-1}\right]. \end{aligned} \tag{3.18}$$

Taking $\mathbf{\Sigma_e}^{-1} = \mathbf{\Omega_e}$, $\mathbf{\Sigma_v}^{-1} = \mathbf{\Omega_v}$, $\vec{\bar{y}}_{i\cdot} = n_i^{-1}\sum_{j=1}^{n_i}\vec{y}_{ij}$, and $\vec{\bar{x}}_{i\cdot} = n_i^{-1}\sum_{j=1}^{n_i}\vec{x}_{ij}$,

$$\left[\mathrm{Cov}[\vec{v}_i, \mathbf{Y}_i^{(1)}]\right]\left[\mathrm{Var}[\mathbf{Y}_i^{(1)}]\right]^{-1}\left[\mathbf{Y}_i^{(1)} - \mathbf{B}\begin{pmatrix}\vec{x}_{i1}\\ \vdots \\ \vec{x}_{in_i}\end{pmatrix}\right] = \mathbf{\Sigma_v}[\mathbf{\Sigma_e} + n_i\mathbf{\Sigma_v}]^{-1}\left[\sum_{j=1}^{n_i}(\vec{y}_{ij} - \mathbf{B}\vec{x}_{ij})\right]$$

$$= [\mathbf{\Sigma_e}\mathbf{\Omega_v} + n_i\mathbf{I_2}]^{-1}[n_i(\vec{\bar{y}}_{i\cdot} - \mathbf{B}\vec{\bar{x}}_{i\cdot})] = [\mathbf{\Omega_v} + n_i\mathbf{\Omega_e}]^{-1}\mathbf{\Omega_e}n_i(\vec{\bar{y}}_{i\cdot} - \mathbf{B}\vec{\bar{x}}_{i\cdot}). \quad (3.19)$$

Define also:

$$\mathbf{P}_i = \left[(p_{ihh'})\right]_{(2\times2)} = \mathrm{Var}[\vec{v}_i|\mathbf{B}, \mathbf{\Sigma_v}, \mathbf{\Sigma_e}, \mathbf{Y}^{(1)}]$$

$$= \mathrm{Var}[\vec{v}_i] - \mathrm{Cov}[\vec{v}_i, \mathbf{Y}_i^{(1)}]\left[\mathrm{Var}[\mathbf{Y}_i^{(1)}]\right]^{-1}\left[\mathrm{Cov}[\mathbf{Y}_i^{(1)}, \vec{v}_i]\right] = \mathbf{\Sigma_v} - n_i\mathbf{\Sigma_v}[\mathbf{\Sigma_e} + n_i\mathbf{\Sigma_v}]^{-1}\mathbf{\Sigma_v}$$

$$= \mathbf{\Sigma_v}[\mathbf{\Sigma_e} + n_i\mathbf{\Sigma_v}]^{-1}\mathbf{\Sigma_e} = [\mathbf{\Sigma_v}^{-1} + n_i\mathbf{\Sigma_e}^{-1}]^{-1} = [\mathbf{\Omega_v} + n_i\mathbf{\Omega_e}]^{-1} \quad (3.20)$$

and

$$\vec{\mu}_i = \begin{pmatrix}\mu_{i1}\\ \mu_{i2}\end{pmatrix} = E[\vec{v}_i|\mathbf{B}, \mathbf{\Sigma_v}, \mathbf{\Sigma_e}, \mathbf{Y^{(1)}}](\mathbf{\Omega_v} + n_i\mathbf{\Omega_e})^{-1}\mathbf{\Omega_e}n_i(\vec{\bar{y}}_{i\cdot} - \mathbf{B}\vec{\bar{x}}_{i\cdot}) = \mathbf{P}_i\mathbf{\Omega_e}n_i(\vec{\bar{y}}_{i\cdot} - \mathbf{B}\vec{\bar{x}}_{i\cdot}). \quad (3.21)$$

Making use of the moment-generating function of the log-normal distribution, for $i = 1, \ldots, m$ and $j = 1, \ldots, n_i$:

$$\mathrm{E}[\exp(e_{ij1})|\mathbf{B}, \mathbf{V}, \mathbf{\Sigma_v}, \mathbf{\Sigma_e}, \mathbf{Y}^{(1)}] = \exp(\frac{\sigma_{e11}}{2}) \quad (3.22)$$

$$\mathrm{E}[\exp(v_{i1})|\mathbf{B}, \mathbf{\Sigma_v}, \mathbf{\Sigma_e}, \mathbf{Y^{(1)}}] = \exp(\mu_{1i} + p_{i11}/2). \quad (3.23)$$

Accordingly, for $j = n_i + 1, \ldots, N_i$:

$$\mathrm{E}[z_{ij1}|\mathbf{B}, \mathbf{\Sigma_v}, \mathbf{\Sigma_e}, \mathbf{Y^{(1)}}] = \exp[\vec{\beta}_{1\cdot}^{\top}\vec{x}_{ij} + \mu_{1i} + \frac{1}{2}(p_{i11} + \sigma_{e11})] - 1 = g_{ij}(\vec{\beta}_{1\cdot}^{\top}, \mathbf{\Sigma_v}, \mathbf{\Sigma_e}), \quad (3.24)$$

$$\mathrm{E}[\theta_i^{\mathrm{MV}}|\mathbf{B},\mathbf{\Sigma_v},\mathbf{\Sigma_e},\mathbf{Y}^{(1)}] = \sum_{j=1}^{n_i} z_{ij1} + \sum_{j=n_i+1}^{N_i} g_{ij}(\vec{\beta}_{1\cdot}^{\top},\mathbf{\Sigma_v},\mathbf{\Sigma_e}) = \tilde{\theta}_i^{\mathrm{MV}}(\vec{\beta}_{1\cdot}^{\top},\mathbf{\Sigma_v},\mathbf{\Sigma_e}), \tag{3.25}$$

$$\begin{aligned}
\mathrm{Var}[z_{ij1}|\mathbf{B},\mathbf{\Sigma_v},\mathbf{\Sigma_e},\mathbf{Y}^{(1)}] &= \mathrm{Var}[\exp(\vec{\beta}_{1\cdot}^{\top}\vec{x}_{ij} + \vec{v}_{1i} + e_{ij1}) - 1|\mathbf{B},\mathbf{\Sigma_v},\mathbf{\Sigma_e},\mathbf{Y}^{(1)}] \\
&= \exp(2\vec{\beta}_{1\cdot}^{\top}\vec{x}_{ij})\mathrm{Var}[exp(\vec{v}_{1i} + e_{ij1})|\mathbf{B},\mathbf{\Sigma_v},\mathbf{\Sigma_e},\mathbf{Y}^{(1)}] \\
&= \exp(2\vec{\beta}_{1\cdot}^{\top}\vec{x}_{ij})[\mathrm{E}[exp(2[\vec{v}_{1i} + e_{ij1}])|\mathbf{B},\mathbf{\Sigma_v},\mathbf{\Sigma_e},\mathbf{Y}^{(1)}] - (\mathrm{e}[exp(\vec{v}_{1i} + e_{ij1})|\mathbf{B},\mathbf{\Sigma_v},\mathbf{\Sigma_e},\mathbf{Y}^{(1)}])^2] \\
&= \exp(2\vec{\beta}_{1\cdot}^{\top}\vec{x}_{ij})[\exp(2[\sigma_{e11} + \mu_{1i} + p_{i11}]) - \exp(\sigma_{e11} + 2\mu_{1i} + p_{i11})] \\
&= \exp(2\vec{\beta}_{1\cdot}^{\top}\vec{x}_{ij})\exp(\sigma_{e11} + 2\mu_{1i} + p_{i11})[\exp(\sigma_{e11} + p_{i11}) - 1] \\
&= [g_{ij}(\vec{\beta}_{1\cdot}^{\top},\mathbf{\Sigma_v},\mathbf{\Sigma_e}) + 1]^2[\exp(\sigma_{e11} + p_{i11}) - 1] \\
&= h_{ij}(\vec{\beta}_{1\cdot}^{\top},\mathbf{\Sigma_v},\mathbf{\Sigma_e}).
\end{aligned} \tag{3.26}$$

Proceeding for $(j \neq l) = n_i + 1, \ldots, N_i$:

$$\begin{aligned}
&\mathrm{Cov}[z_{ij1}, z_{il1}|\mathbf{B},\mathbf{\Sigma_v},\mathbf{\Sigma_e},\mathbf{Y}^{(1)}] \\
&= \mathrm{Cov}[\exp(y_{ij1}),\exp(y_{il1})|\mathbf{B},\mathbf{\Sigma_v},\mathbf{\Sigma_e},\mathbf{Y}^{(1)}] \\
&= \exp(\vec{\beta}_{1\cdot}^{\top}\vec{x}_{ij} + \vec{\beta}_{1\cdot}^{\top}\vec{x}_{il})\mathrm{E}[\exp(2\vec{v}_{1i} + e_{ij1} + e_{il1})|\mathbf{B},\mathbf{\Sigma_v},\mathbf{\Sigma_e},\mathbf{Y}^{(1)}] \\
&\quad - [g_{ij}(\vec{\beta}_{1\cdot}^{\top},\mathbf{\Omega_v},\mathbf{\Omega_e}) + 1][g_{il}(\vec{\beta}_{1\cdot}^{\top},\mathbf{\Omega_v},\mathbf{\Omega_e}) + 1] \\
&= \exp(\vec{\beta}_{1\cdot}^{\top}\vec{x}_{ij} + \vec{\beta}_{1\cdot}^{\top}\vec{x}_{il})[\exp([\sigma_{e11} + \sigma_{e11}]/2 + 2[\mu_{1i} + p_{i11}])] \\
&\quad - [g_{ij}(\vec{\beta}_{1\cdot}^{\top},\mathbf{\Omega_v},\mathbf{\Omega_e}) + 1][g_{il}(\vec{\beta}_{1\cdot}^{\top},\mathbf{\Omega_v},\mathbf{\Omega_e}) + 1] \\
&= [g_{ij}(\vec{\beta}_{1\cdot}^{\top},\mathbf{\Omega_v},\mathbf{\Omega_e}) + 1][g_{il}(\vec{\beta}_{1\cdot}^{\top},\mathbf{\Omega_v},\mathbf{\Omega_e}) + 1]e^{p_{i11}} \\
&\quad - [g_{ij}(\vec{\beta}_{1\cdot}^{\top},\mathbf{\Omega_v},\mathbf{\Omega_e}) + 1][g_{il}(\vec{\beta}_{1\cdot}^{\top},\mathbf{\Omega_v},\mathbf{\Omega_e}) + 1] \\
&= [g_{ij}(\vec{\beta}_{1\cdot}^{\top},\mathbf{\Omega_v},\mathbf{\Omega_e}) + 1][g_{il}(\vec{\beta}_{1\cdot}^{\top},\mathbf{\Omega_v},\mathbf{\Omega_e}) + 1][e^{p_{i11}} - 1] \\
&= h_{ijl}(\vec{\beta}_{1\cdot}^{\top},\mathbf{\Omega_v},\mathbf{\Omega_e}).
\end{aligned} \tag{3.27}$$

Estimation of $\theta_i^{\mathrm{MV}}(\vec{\beta}_{1\cdot}^{\top},\mathbf{\Sigma_v},\mathbf{\Sigma_e})$ via the Gibbs sampler is discussed in Section 4.1.

However, as noted in Chandra and Chambers [2011], in the context of log-transformed variables this estimator is inherently biased for skewed data such as those generated by PPS sampling. A secondary multivariate model, Model MVw, incorporates the design weights as covariates in an augmented version of $\mathbf{X}$ and is discussed in Section 3.2.3 for alternative analysis of the PPS sample. Before doing so, we provide the Horvitz-Thompson estimator (HT) in Section 3.2.2.

3.2.2 The Horvitz-Thompson Estimator

The standard method of design based estimation, exclusive of auxiliary data, is to employ an expansion estimator whereby each sampled unit is assigned a weight based on the unit's probability of inclusion in the sample. The intuitive choice of weight is $w_{ij} = \pi_{ij}^{-1}$ for $i = 1, \ldots, m;\ j = 1, \ldots, n_i$. This is a generalization of the Horvitz-Thompson estimator for small-area totals Cochran [1977]. Each total is estimated, per [Rao, 2003, p. 16]:

$$\hat{\theta}_i^{\text{HT}} = \sum_{j=1}^{n_i} \tilde{w}_{ij} z_{ij1};\ i = 1, \ldots, m. \tag{3.28}$$

In our PPS sample, this weight becomes:

$$\tilde{w}_{ij} = \begin{cases} T_h / n_h t_{hj} & \text{for} \quad n_h t_{hj} \leq T_h, j \in \text{“PPS”} \\ 1 & \text{for} \quad n_h t_{hj} > T_h, j \in \text{“PPS”} \\ 1 & \text{for} \quad j \in \text{“Certs”}. \end{cases} \tag{3.29}$$

The set “Certs” designates ‘certainty units’, those governments that were included in the sample irrespective of other criteria, while “PPS” designates those selected through the probability proportional to size methodology. As the weight of all certainty units and PPS units of sufficient size is equal to one, these units become self representing in the analysis, such that the very nature of their selection renders them unable to reliably reflect the traits of the population they are drawn from. The weights $w_{ij} = \ln(\tilde{w}_{ij})$ employed in Models MVw and UVw, specified in Sections 3.2.3 and 3.3.2, respectively, are transformed to maintain their relationship to $\vec{z}_{ij}$.

3.2.3 Model MVw

In an effort to allow the model to correct for the bias imposed by the PPS sample, as noted in Section 3.2.1, Model MVw is an extension of Model MV that includes PPS weights as covariates and is specified as follows. Take

$$\vec{x}^{\mathrm{w}}_{ij} = (col)(1, x_{ij1}, x_{ij2}, w_{ij}), \tag{3.30}$$

$$\mathbf{X}^{\mathrm{w}} = \mathrm{concat}_{1\leq i\leq m}(\mathrm{concat}_{1\leq j\leq N_i}(\vec{x}^{\mathrm{w}}_{ij})), \text{ and} \tag{3.31}$$

$$\mathbf{B}^{\mathrm{w}} = [(\beta^{\mathrm{w}}_h k)] - (2\times 4); h = 1, 2; k = 0, 1, 2, 3. \tag{3.32}$$

Clearly $\vec{e}_{ij}$ and $\vec{v}_i$ are not necessarily equal to their counterparts in Model MV, but as they do not alter the model specification, they will remain undistinguished across models, though contextually free to assume appropriate values. This holds for $\mathbf{\Sigma_v}$, $\mathbf{\Omega_v}$, $\mathbf{\Sigma_e}$, and $\mathbf{\Omega_e}$ as well.

(i) Conditional on $\mathbf{B}^{\mathrm{w}}$, $\mathbf{V}$, $\mathbf{\Sigma_v}$, and $\mathbf{\Sigma_e}$: $\vec{y}_{ij} \sim \mathcal{N}_2(\mathbf{B}^{\mathrm{w}}\vec{x}^{\mathrm{w}}_{ij} + \vec{v}_i, \mathbf{\Sigma_e})$ independently for $i = 1, \ldots, m$; $j = 1, \ldots, N_i$.

(ii) Conditional on $\mathbf{B}^{\mathrm{w}}$, $\mathbf{\Sigma_v}$, and $\mathbf{\Sigma_e}$: $\vec{v}_i \sim \mathcal{N}_2(\vec{0}, \mathbf{\Sigma_v})$ independently for $i = 1, \ldots, m$.

(iii) Marginally, $\mathbf{B}^{\mathrm{w}}$, $\mathbf{\Sigma_v}$ and $\mathbf{\Sigma_e}$ are independently distributed with:
$\mathbf{B}^{w} \sim$ *Uniform* on $\mathbb{R}^{2\times 4}$,
${\mathbf{\Sigma_v}}^{-1} \sim \mathcal{W}_m(\mathbf{\Phi}_v)$ for positive-definite $\mathbf{\Sigma_v}$ and positive-semi-definite scale matrix $\mathbf{\Phi}_v$, and
${\mathbf{\Sigma_e}}^{-1} \sim \mathcal{W}_n(\mathbf{\Phi}_e)$ for positive-definite $\mathbf{\Sigma_e}$ and positive-semi-definite scale matrix $\mathbf{\Phi}_e$.

For known $\vec{w}^{\top} = \mathrm{concat}_{1\leq i\leq m}(\mathrm{concat}_{1\leq j\leq N_i}(w_{ij}))$ by extension of equation (3.29) to the unsampled units (all of which are classified as 'PPS' units), $i = 1, \ldots, m$, and $j = n_i +$

$1, \ldots, N_i$, equation (3.24) becomes:

$$\begin{aligned} \mathrm{E}[z_{ij1}|\mathbf{B}^{w}, \mathbf{V}, \mathbf{\Sigma_v}, \mathbf{\Sigma_e}, \mathbf{Y^{(1)}}] &= \exp[\vec{\beta}_{1\cdot}^{\mathrm{w}\top}\vec{x}_{ij}^{\mathrm{w}} + v_{i1} + \frac{\sigma_{e11}}{2}] - 1 \\ &= g_{ij}^{\mathrm{w}}(\vec{\beta}_{1\cdot}^{\mathrm{w}\top}, (\mathbf{V}), (\mathbf{\Sigma_e})) \end{aligned} \quad (3.33)$$

Similarly,

$$\begin{aligned} &\mathrm{Var}[z_{ij1}|\mathbf{B}^{\mathrm{w}}, \mathbf{V}, \mathbf{\Sigma_v}, \mathbf{\Sigma_e}, \mathbf{Y}^{(1)}] \\ &\quad= \exp(2(\vec{\beta}_{1\cdot}^{\mathrm{w}\top}\vec{x}_{ij}^{\mathrm{w}} + v_{i1}))\exp(\sigma_{e11})[\exp(\sigma_{e11}) - 1] \\ &\quad= [g_{ij}^{\mathrm{w}}(\vec{\beta}_{1\cdot}^{\mathrm{w}\top}, \mathbf{V}, \mathbf{\Sigma_e}) + 1]^2(\exp(\sigma_{e11}) - 1) \\ &\quad= h_{ij}^{\mathrm{w}}(\vec{\beta}_{1\cdot}^{\mathrm{w}\top}, \mathbf{V}, \mathbf{\Sigma_e}), \end{aligned} \quad (3.34)$$

$$\begin{aligned} \mathrm{E}[\theta_i^{\mathrm{MVw}}|\mathbf{B}^{\mathrm{w}}, \mathbf{V}, \mathbf{\Sigma_v}, \mathbf{\Sigma_e}, \mathbf{Y}^{(1)}] &= \sum_{j=1}^{n_i} z_{ij1} + \sum_{j=n_i+1}^{N_i} g_{ij}^{\mathrm{w}}(\vec{\beta}_{1\cdot}^{\mathrm{w}\top}, \mathbf{V}, \mathbf{\Sigma_e}) \\ &= \tilde{\theta}_i^{\mathrm{MVw}}(\vec{\beta}_{1\cdot}^{\mathrm{w}\top}, \mathbf{V}, \mathbf{\Sigma_e}). \end{aligned} \quad (3.35)$$

3.3 The Univariate Hierarchical Bayes Model

Restricting Models MV and MVw to the univariate case, we have the following models: Model UV and Model UVw, described subsequently.

3.3.1 Model UV

Using 2002 full-time employment, x_{ij1}, as the only covariate we have the following model. For $\vec{x}_{ij}^{\mathrm{UV}} = \mathrm{col}(1, x_{ij1})$ and taking $\vec{\beta}^{\mathrm{UV}\top} = (\beta_0^{\mathrm{UV}}, \beta_1^{\mathrm{UV}})$, we specify Model UV as:

(i) Conditional on $\vec{\beta}^{\mathrm{UV}}$, $\vec{v}_\cdot$, σ_v and σ_e: $y_{ij1} \sim \mathcal{N}(\vec{\beta}^{\mathrm{UV}\top}\vec{x}_{ij}^{\mathrm{UV}} + v_i, \sigma_e)$ independently for $i = 1, \ldots, m$; $j = 1, \ldots, N_i$.

(ii) Conditional on $\vec{\beta}^{\mathrm{UV}}$, σ_v, and σ_e: $v_i \sim \mathcal{N}(0, \sigma_v)$ independently for $i = 1, \ldots, m$.

(iii) Marginally, $\vec{\beta}^{\mathrm{UV}}$, σ_v, and σ_e are independently distributed with: $\vec{\beta}^{\mathrm{UV}} \sim$ *Uniform* on $\mathbb{R}^2$, $\sigma_v \sim \mathcal{W}_a^{-1}(\phi_v)$ for $\sigma_v > 0$ and $\phi_v \geq 0$; and $\sigma_e \sim \mathcal{W}_b^{-1}(\phi_e)$ for $\sigma_e > 0$ and $\phi_e \geq 0$.

Under Model UV,

$$\mathrm{E}[z_{ij1}|\vec{\beta}^{\mathrm{UV}}, \sigma_v, \sigma_e, Y_{1\cdot\cdot}^{(1)}] = \exp[\vec{\beta}^{\mathrm{UV}\top}\vec{x}_{ij}^{\mathrm{UV}} + \mu_i + \frac{1}{2}(p_i + \sigma_e)] - 1 = g_{ij}^{UV}(\beta^{\mathrm{UV}}, \sigma_v, \sigma_e), \quad (3.36)$$

where

$$\mu_i = \mathrm{E}[v_i|\beta^{\mathrm{UV}}, \sigma_v, \sigma_e, Y_{1\cdot\cdot}] = (\sigma_v^{-1} + n_i\sigma_e^{-1})^{-1}\sigma_e^{-1}n_i(\bar{y}_{1i\cdot} - \vec{\beta}^{\mathrm{UV}\top}\vec{x}_{i\cdot}^{\mathrm{UV}}) \quad (3.37)$$

and

$$p_i = \mathrm{Var}[v_i|\vec{\beta}^{\mathrm{UV}}, \sigma_v, \sigma_e, Y_{1\cdot\cdot}] = (\sigma_v^{-1} + n_i\sigma_e^{-1})^{-1}. \quad (3.38)$$

Similarly,

$$\begin{aligned}\mathrm{Var}[z_{ij1}|\vec{\beta}^{\mathrm{UV}}, \vec{v}, \sigma_v, \sigma_e, Y_{1\cdot\cdot}^{(1)}] &= \exp(2(\vec{\beta}^{\mathrm{UV}\top}\vec{x}_{ij}^{\mathrm{UV}} + v_i))\exp(\sigma_e)[\exp(\sigma_e) - 1] \\ &= [g_{ij}^{\mathrm{UV}}(\vec{\beta}^{\mathrm{UV}}, \vec{v}, \sigma_e) + 1]^2[\exp(\sigma_e) - 1] = h_{ij}^{\mathrm{UV}}(\vec{\beta}^{\mathrm{UV}}, \vec{v}, \sigma_e) \quad (3.39)\end{aligned}$$

and

$$\mathrm{E}[\theta_i^{\mathrm{UV}}|\vec{\beta}^{\mathrm{UV}}, \vec{v}, \sigma_v, \sigma_e, Y_{1\cdot\cdot}^{(1)}] = \sum_{j=1}^{n_i} z_{ij1} + \sum_{j=n_i+1}^{N_i} g_{ij}^{\mathrm{UV}}(\vec{\beta}^{\mathrm{UV}}, \vec{v}, \sigma_e) = \tilde{\theta}_i^{\mathrm{UV}}(\vec{\beta}^{\mathrm{UV}}, \vec{v}, \sigma_e). \quad (3.40)$$

3.3.2 Model UVw

Similarly, the weighted Univariate Model 'UVw' includes the vector of transformed weights $\vec{w}$ as a component of $\mathbf{X}^{UVw} = \mathrm{concat}_{1\leq i\leq m}(\mathrm{concat}_{1\leq j\leq n_i}(\mathrm{col}(1, x_{ij1}, w_{ij})))$ and $\vec{x}_{ij}^{UVw} = \mathrm{col}(1, x_{1ij}, w_{ij})$, with the results as follows.

Taking $\vec{\beta}^{\mathrm{UVw}\top} = (\beta_0^{\mathrm{UVw}}, \beta_1^{\mathrm{UVw}}, \beta_2^{\mathrm{UVw}})$, we specify Model UVw as:

(i) Conditional on $\vec{\beta}^{\text{UVw}}$, $\vec{v}.$, σ_v and σ_e: $y_{ij1} \sim \mathcal{N}(\vec{\beta}^{\text{UVw}\top}\vec{x}_{ij}^{\text{UVw}} + v_i, \sigma_e)$ independently for $i = 1, \ldots, m$, $j = 1, \ldots, N_i$.

(ii) Conditional on $\vec{\beta}^{\text{UVw}}$, σ_v and σ_e: $v_i \sim \mathcal{N}(0, \sigma_v)$ independently for $i = 1, \ldots, m$.

(iii) Marginally, $\vec{\beta}^{\text{UVw}}$, σ_v and σ_e are independently distributed with: $\vec{\beta}^{\text{UVw}} \sim$ *Uniform* on $\mathbb{R}^{(1\times 2)}$, $\sigma_v^{-1} \sim \mathcal{W}_a(\phi_v)$ for $\sigma_v > 0$ and $\phi_v \geq 0$; and $\sigma_e^{-1} \sim \mathcal{W}_b(\phi_e)$ for $\sigma_e > 0$ and $\phi_e \geq 0$.

Under Model UVw,

$$\text{E}[z_{ij1}|\vec{\beta}^{\text{UVw}}, \vec{v}, \sigma_v, \sigma_e, Y_{1\cdot\cdot}^{(1)}] = \exp[\vec{\beta}^{\text{UVw}\top}\vec{x}_{ij}^{UVw} + \mu_{i1} + \frac{1}{2}(p_i + \sigma_e)] - 1 = g_{ij}^{\text{UVw}}(\vec{\beta}^{\text{UVw}}, \vec{v}, \sigma_e). \tag{3.41}$$

Here, $\mu_i = \text{E}[v_i|\vec{\beta}^{\text{UVw}}, \sigma_v, \sigma_e, Y_{1\cdot\cdot}] = (\sigma_v^{-1} + n_i\sigma_e^{-1})^{-1}\sigma_e^{-1}n_i(\bar{y}_{1i\cdot} - \vec{\beta}^{\text{UVw}\top}\vec{x}_{i\cdot}^{UVw})$ and $p_i = \text{Var}[v_i|\vec{\beta}^{\text{UVw}}, \sigma_v, \sigma_e, Y_{1\cdot\cdot}] = (\sigma_v^{-1} + n_i\sigma_e^{-1})^{-1}$. Similarly,

$$\begin{aligned}\text{Var}[z_{ij1}|\vec{\beta}^{\text{UVw}}, \sigma_v, \sigma_e, Y_{1\cdot\cdot}^{(1)}] &= \exp(2\vec{\beta}^{\text{UVw}\top}\vec{x}_{ij}^{UVw})\exp(2\mu_i + p_i + \sigma_e)[\exp(p_i + \sigma_e) - 1] \\ &= [g_{ij}^{\text{UVw}}(\vec{\beta}^{\text{UVw}}, \sigma_v, \sigma_e) + 1]^2[\exp(\sigma_e + p_i) - 1] = h_{ij}^{\text{UVw}}(\vec{\beta}^{\text{UVw}}, \sigma_v, \sigma_e),\end{aligned} \tag{3.42}$$

$$\begin{aligned}\text{E}[\theta_i^{\text{UVw}}|\vec{\beta}^{\text{UVw}}, \sigma_v, \sigma_e, Y_{1\cdot\cdot}^{(1)}] &= \sum_{j=1}^{n_i} z_{ij1} + \sum_{j=n_i+1}^{N_i} g_{ij}^{\text{UVw}}(\vec{\beta}^{\text{UVw}}, \sigma_v, \sigma_e) \\ &= \tilde{\theta}_i^{\text{UVw}}(\vec{\beta}^{\text{UVw}}, \sigma_v, \sigma_e).\end{aligned} \tag{3.43}$$

Chapter 4: Simulation and Results

4.1 The Gibbs Sampler

Having produced Sample 1, through probability proportional to size sampling, and Sample 2, through simple random sampling, within sampling strata from the combined 2002 and 2007 Census of Governments data we use the univariate and multivariate models specified in chapter 3 to produce estimates of posterior mean and variance of θ_i for each of the m small areas, defined by the intersections of state and program function.

Eschewing higher dimensional numerical integration, we employ the Gibbs Sampler, a specific application of Markov Chain Monte Carlo (MCMC) integration, based on the following specifications.

By iteratively sampling from a complete set of joint conditional distributions, the Gibbs sampler generates a Markov chain which converges to the target distribution as the number of samples, K, increases. A properly specified chain will generate results that may be regarded as dependent samples from the target distribution, in our case $p(\theta)$. Following Rao [2003], Lee [2004], Holmes and Held [2006], Casella and George [1992], and Datta et al. [1998], we have the following joint conditional distributions corresponding to Model MV, as specified in Section 3.2.1:

(i) Conditional on $\mathbf{V}$, $\mathbf{\Sigma_e}$, $\mathbf{\Sigma_v}$, $\mathbf{Y^{(1)}}$:

$$\text{vec}(\mathbf{B}) \sim \mathcal{N}_4\left(\text{vec}(\mathbf{M_B}), \left[\mathbf{X}^{(1)}\mathbf{X}^{(1)^\top}\right]^{-1} \otimes \mathbf{\Sigma_e}\right)$$

with vec($\mathbf{B}$) composed of $\mathbf{B}$ arranged column-wise as a vector, and

$$\text{vec}(\mathbf{M_B}) = \text{vec}\left(\left[[\mathbf{Y}^{(1)} - \mathbf{V}\mathbf{F}^{(1)}]\mathbf{X}^{(1)\top}\right]\left[\mathbf{X}^{(1)}\mathbf{X}^{(1)\top}\right]^{-1}\right)$$

also arranged column-wise as a vector.

(ii) Conditional on $\mathbf{B}, \mathbf{\Sigma_v}, \mathbf{\Sigma_e}, \mathbf{Y}^{(1)}$:

$$\vec{v}_i \sim \mathcal{N}_2 \left(\mathbf{P}_i \mathbf{\Sigma_e} \sum_{j=1}^{n_i} (\vec{y}_{ij} - \mathbf{B}\vec{x}_{ij}), \mathbf{P}_i \right)$$

with $\mathbf{P}_i$ obtained in equation (3.20).

(iii) Conditional on $\mathbf{B}, \mathbf{V}, \mathbf{\Sigma_v}, \mathbf{Y}^{(1)}$:

$$\mathbf{\Sigma_e} \sim \mathcal{W}_{n+a} \left(\left[[\mathbf{Y}^{(1)} - \mathbf{B}\mathbf{X}^{(1)} - \mathbf{V}\mathbf{F}^{(1)}][\mathbf{Y}^{(1)} - \mathbf{B}\mathbf{X}^{(1)} - \mathbf{V}\mathbf{F}^{(1)}]^\top \right]^{-1} \right)$$

with $n = \sum_{i=1}^{m} n_i$, and

(iv) conditional on $\mathbf{B}, \mathbf{V}, \mathbf{\Sigma_e}, \mathbf{Y}^{(1)}$; $\mathbf{\Sigma_v} \sim \mathcal{W}_{m+b} \left([\mathbf{V}\mathbf{V}^\top]^{-1} \right)$.

Here a and b are integers that augment the degrees of freedom of the Wishart distributions such that their choice will render the posterior distribution of θ proper or improper, based on a set of conditions derived in [Datta et al., 1998, appendix]. Values of a and b should be consistent with their values in the specification of the marginal distribution of the model. Thus, per section 3.2.1 we have a and b equal to the dimension of $\vec{y}_{ij}$ plus one or $a = b = 3$ in the multivariate case. The previously specified joint conditionals are readily adaptable to Models MVw, UV, and UVw with the appropriate alterations to the dimensions of the multivariate normal ($\mathcal{N}$) and Wishart ($\mathcal{W}$) distributed variables and appropriate substitutions in the specification of the associated joint conditionals. In all applications a simulation consisted of equal to eight independent chains, generated in parallel, and each with 10,010 iterations. The first 500 iterations of each chain were discarded and the results were further 'thinned' by retaining only every 15$^{\text{th}}$ result. The degree of thinning was determined from analysis of autocorrelations within all chains, choosing a conservative common value for all parameters, chains, and models. The hierarchical Bayes analyses use noninformative priors on the distributions of $\mathbf{B}, \mathbf{\Sigma_v}, \mathbf{\Sigma_e}$ such that:

$$\Pi(\mathbf{B}, \mathbf{\Sigma_v}, \mathbf{\Sigma_e}) = 1.$$

Starting values of $\mathbf{V}$, $\mathbf{\Omega_v}$, and $\mathbf{\Omega_e}$ for each chain, denoted with superscript (0), were generated based on the known values $x_{ijh}; h = 1, 2$ as follows:

$$\mathbf{\Omega_e}^{(0)} \sim \mathcal{W}_a([\mathbf{\Sigma'_e}]^{-1}), \mathbf{\Omega_v}^{(0)} \sim \mathcal{W}_b([\mathbf{\Sigma'_v}]^{-1}), \text{ and} \tag{4.1}$$

$$v_{ih}^{(0)} \sim \text{multivariate Student's } t(\text{d.f.} = b, \text{scale} = \mathbf{\Sigma'_v}, \text{mean} = v'_{ih}). \tag{4.2}$$

Where

$$v'_{ih} = \frac{1}{N_i} \sum_{j=1}^{N_i} (x_{ijh} - \bar{x}_{\cdot\cdot h}), \bar{x}_{\cdot\cdot h} = \frac{1}{N} \sum_{i=1}^{m} \sum_{j=1}^{N_i} xijh, \tag{4.3}$$

$\mathbf{\Sigma'_v}$ is the (2×2) variance-covariance matrix of $\vec{v}_i^{(0)} = ((v_{ih}^{(0)}))^\top$ and $\mathbf{\Sigma'_e}$ is the (2×2) variance-covariance matrix of the residuals $\vec{x}_{ij} - \vec{v}_i^{(0)}$.

4.2 Posterior Mean and Variance

The following Rao-Blackwell estimators of the mean value of θ_i and its posterior variance generated our results for each small area, elaborated fully in the case of Model MV, as follows. Taking $s(K)$ is the set of all K' retained estimates and $\mathbf{B}^{(k)}, \mathbf{\Omega_v}^{(k)}, \mathbf{\Omega_e}^{(k)}$ are the sampled values of $\mathbf{B}, \mathbf{\Omega_v}, \mathbf{\Omega_e}$ from iteration k of the Gibbs sampler; for $j = n_i + 1, \ldots, N_i$ and following equations 3.25 to 3.27,

$$\hat{\theta}_i^{\text{MVHB}} = \text{E}[\theta_i | \mathbf{Y}^{(1)}] = \text{E}[\tilde{\theta}_i^{\text{MV}} | \mathbf{Y}^{(1)}] = (K')^{-1} \sum_{k \in s(K)} \tilde{\theta}_i(\mathbf{B}^{(k)}, \mathbf{\Sigma_v}^{(k)}, \mathbf{\Sigma_e}^{(k)}) \tag{4.4}$$

and

$$
\begin{aligned}
\delta_i^{\text{MVHB}} &= \text{Var}[\theta_i|\mathbf{Y}^{(1)}] \\
&= \text{E}[\text{Var}[\theta_i|\mathbf{B}, \boldsymbol{\Sigma}_{\mathbf{v}}, \boldsymbol{\Sigma}_{\mathbf{e}}, \mathbf{Y}^{(1)}]|\mathbf{Y}^{(1)}] + \text{Var}[\text{E}[\theta_i|\mathbf{B}, \boldsymbol{\Sigma}_{\mathbf{v}}, \boldsymbol{\Sigma}_{\mathbf{e}}, \mathbf{Y}^{(1)}]|\mathbf{Y}^{(1)}] \\
&= \text{E}\left[\sum_{j=n_i+1}^{N_i} h_{ij}(\mathbf{B}, \boldsymbol{\Sigma}_{\mathbf{v}}, \boldsymbol{\Sigma}_{\mathbf{e}}) + \sum_{j=n_i+1}^{N_i}\sum_{l=n_i+1}^{N_i} h_{ijl}|j \neq l\right] + \text{Var}[\tilde{\theta}_i(\mathbf{B}, \boldsymbol{\Sigma}_{\mathbf{v}}, \boldsymbol{\Sigma}_{\mathbf{e}})|\mathbf{Y}^{(1)}] \\
&= (K')^{-1}\sum_{k\in s(K)}\left[\sum_{j=n_i+1}^{N_i} hij(\mathbf{B}^{(k)}, {\boldsymbol{\Sigma}_{\mathbf{v}}}^{(k)}, {\boldsymbol{\Sigma}_{\mathbf{e}}}^{(k)}) + \sum_{j=n_i+1}^{N_i}\sum_{l=n_i+1}^{N_i} hijl(\mathbf{B}^{(k)}, {\boldsymbol{\Sigma}_{\mathbf{v}}}^{(k)}, {\boldsymbol{\Sigma}_{\mathbf{e}}}^{(k)})|j \neq l\right] \\
&\qquad + (K'-1)^{-1}\sum_{k\in s(K)}\left[\tilde{\theta}_i(\mathbf{B}^{(k)}, {\boldsymbol{\Sigma}_{\mathbf{v}}}^{(k)}, {\boldsymbol{\Sigma}_{\mathbf{e}}}^{(k)}) - \hat{\theta}_i^{\text{MVHB}}\right]^2 \qquad (4.5)
\end{aligned}
$$

4.3 Sample 1 (PPS) Results

For the sample drawn under the probability proportional to size methodology we have the following results for 1,244 estimable small areas. In comparison to the HT estimate and models MV, UV, and UVw, the weighted multivariate model produced the estimates with the lowest absolute relative error (ARE) for 565 (45%) of the estimable small areas. Comparatively, the weighted univariate model produced the best estimate in 368 (30%) of the small areas. ARE is taken as the absolute value of an estimate's relative error (RE), calculated for each small area:

$$
\text{RE}_i = \left[\frac{\text{ESTIMATE} - \text{TRUE}}{\text{TRUE}}\right];\ \text{for } i = 1, \ldots, m.
$$

Evaluated by sample size, with 'large' sample sizes referring to those small areas with $n_i > 50$ and 'small' referring to those with $n_i \leq 50$, and $i = 1, \ldots m$, the multivariate weighted model, Model MVw, produced estimates with lowest ARE for 161 (60%) of small areas with large sample sizes and 404 (41%) of small areas with small sample sizes. In the case of the small sample sizes, the weighted univariate estimate produced the estimates with the lowest ARE for a plurality of the estimable small areas.

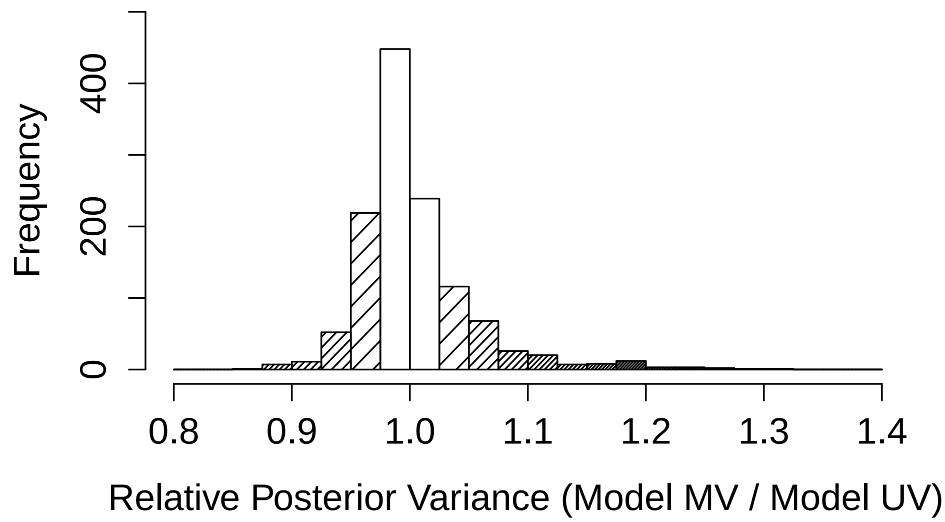

Figure 4.1: PPS Relative Posterior Variance, PPS Sample, Unweighted Models

In terms of posterior variance, calculated via equation 4.5, the weighted multivariate model (MVw) showed lower posterior variance than its univariate counterpart (UVw) in 72% of estimable small areas and the unweighted multivariate model (MV) showed lower posterior variance than its univariate counterpart (UV) in 59% of estimable small areas. Relative posterior variances for the PPS sample and unweighted models (UV and MV) are illustrated in Figure 4.3 with relative posterior variance computed as the posterior variance of Model MV divided by the posterior variance of Model UV. Relative posterior variance for the weighted models (UVw and MVw) is illustrated in Figure 4.3 Evaluated by sample size, the weighted multivariate model (MVw) had lower posterior variance in 55% of small areas with small sample sizes and in 74% of small areas with large sample sizes in comparison to Model UVw. The unweighted multivariate model (MV) had lower posterior variance in 68% of small areas with small sample sizes and 84% of small areas with large sample size, in comparison to Model UV. Relative posterior variances are summarized in Table 4.3.

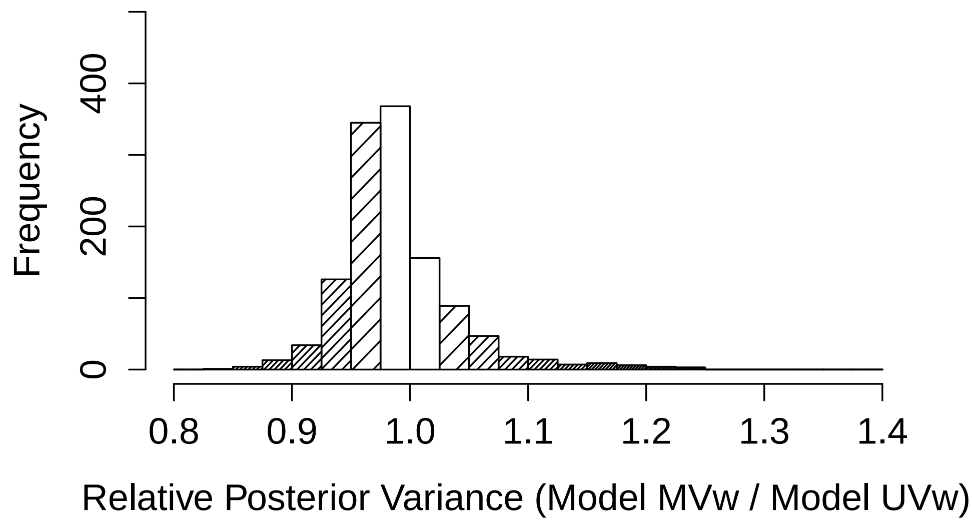

Figure 4.2: PPS Sample Relative Posterior Variance, Weighted Models

Overall the weighted multivariate model (MVw) achieved the lowest mean ARE for the PPS sample, equivalent to 4.13% of the known small area totals. The weighted univariate model (UVw) achieved the lowest median ARE, equivalent to 2.50%. Comparatively, Model MVw achieved the equivalent of 2.51%. The mean and median ARE results are summarised in table 4.3 and illustrated by sample size in Figures 4.3 and 4.3.Figures 4.3 and 4.3 exclude one point, the relative ARE of the total number of full-time transit employees for the state of Louisiana, which had ARE(Model UV) minus ARE(Model MV) equal to 0.2282 for the unweighted models and equal to 0.2327 for the weighted models.

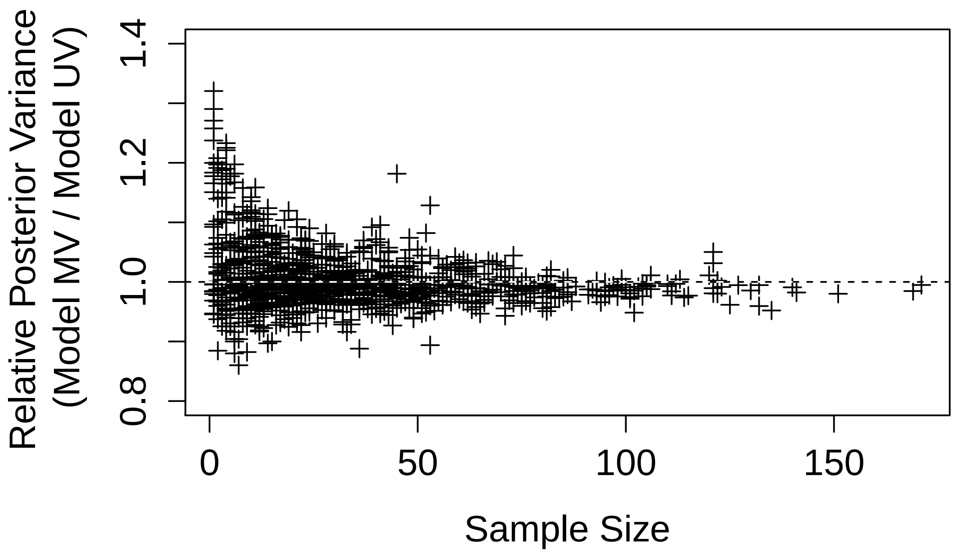

Figure 4.3: PPS Sample Relative Posterior Variance by Sample Size, Unweighted Models

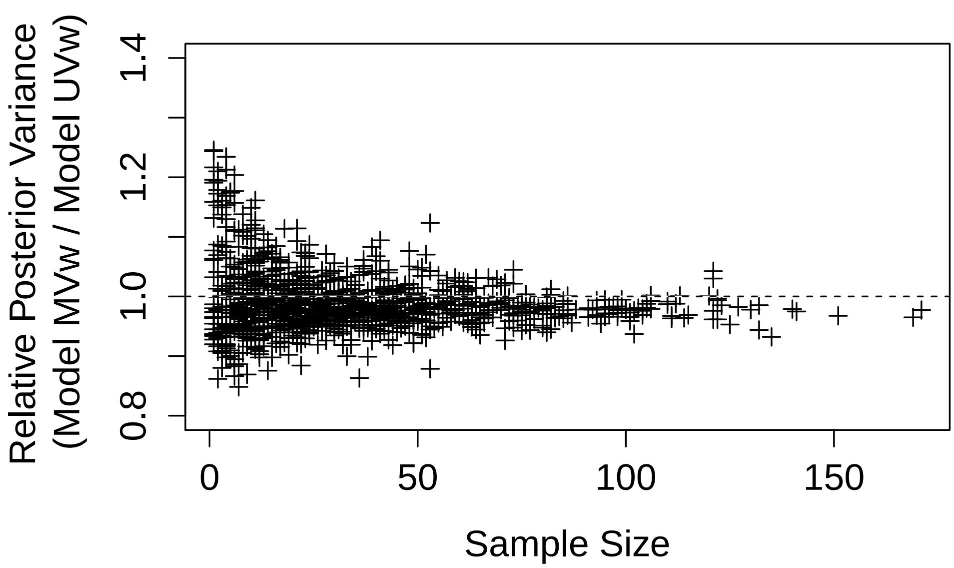

Figure 4.4: PPS Sample Relative Posterior Variance by Sample Size, Weighted Models

Table 4.1: PPS Sample Relative Posterior Variance

Sample Sizes	All		$n_i \leq 50$		$n_i > 50$	
Models	MV/UV	MVw/UVw	MV/UV	MVw/UVw	MV/UV	MVw/UVw
Minimum	0.8601	0.8484	0.8601	0.8484	0.8938	0.8787
1^{st} Quartile	0.9765	0.9616	0.9757	0.9599	0.9778	0.9683
Median	0.9936	0.9805	0.9962	0.9810	0.9896	0.9792
Mean	1.0030	0.9886	1.0060	0.9905	0.9914	0.9815
3^{rd} Quartile	1.0170	1.0040	1.0240	1.0100	1.0010	0.9920
Maximum	1.3200	1.2460	1.3200	1.2460	1.1280	1.1230

Multivariate Posterior Variance Divided by Univariate Posterior Variance

Table 4.2: PPS ARE Results

	Model				
	HT	UV	MV	UVw	MVw
Median	0.9290	0.0344	0.0342	0.0250	0.0251
Mean	0.9155	0.0519	0.0520	0.0418	0.0413
			$n_i \leq 50$		
Median	0.9386	0.0311	0.0313	0.0238	0.0236
Mean	0.9177	0.0529	0.0532	0.0443	0.0439
			$n_i > 50$		
Median	0.9135	0.0440	0.0448	0.0289	0.0288
Mean	0.9074	0.0480	0.0478	0.0327	0.0317

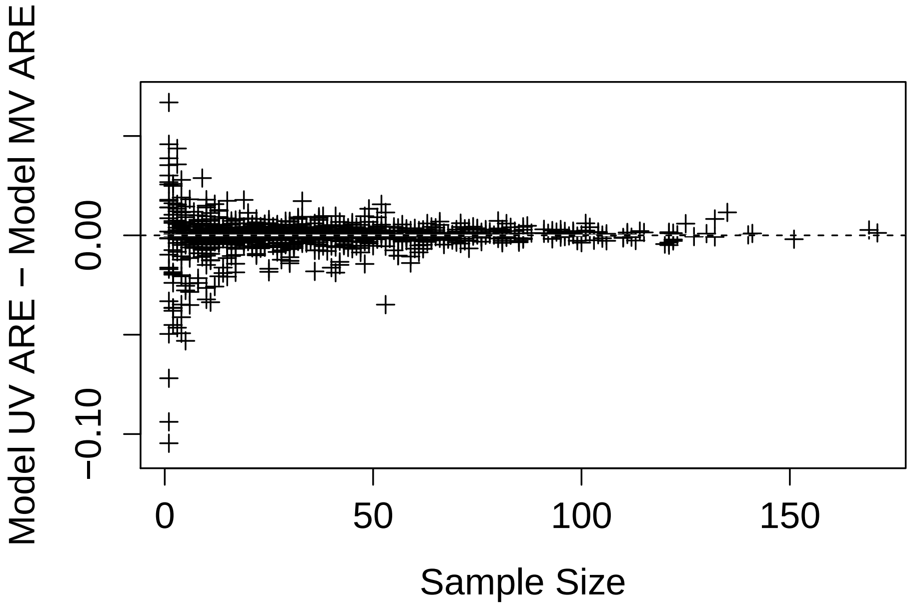

Figure 4.5: PPS Sample Relative ARE by Sample Size, Unweighted Models

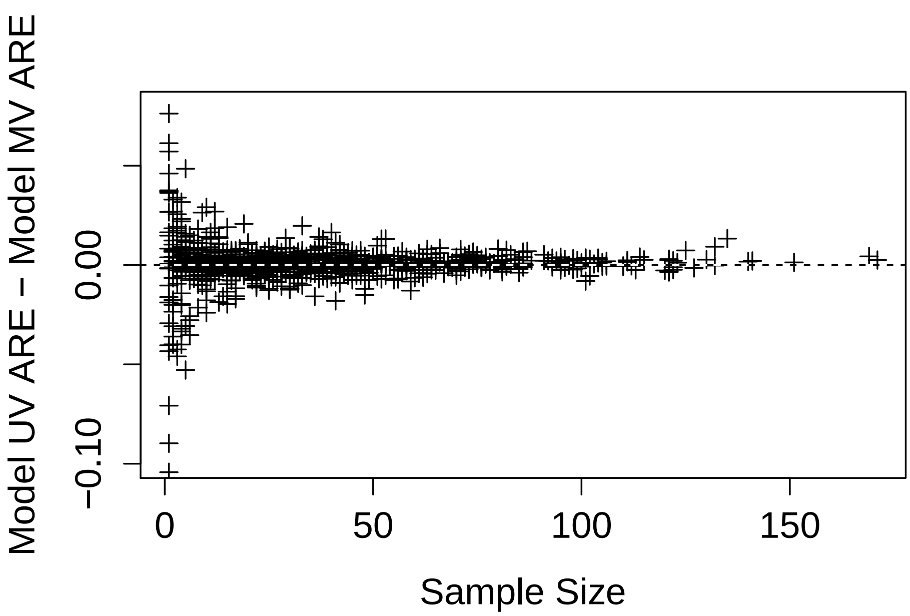

Figure 4.6: PPS Sample Relative ARE by Sample Size, Weighted Models

4.4 Sample 2 (SRS) Results

Under the stratified simple random sampling framework, the expansion estimator w_{ij} becomes $N_i/n_i \ \forall \ j = 1, \ldots, n_i$, thus the inclusion of a weight variable for calibration purposes is unnecessary. The Horwitz-Thompson estimator (HT) for small area totals becomes:

$$\theta_i^{\text{HT}} = \frac{N_i}{n_i} \sum_{j=1}^{n_i} z_{ij1}.$$

Using the unweighted univariate and multivariate models constructed under Sections 3.2.1 and 3.3.1, we have the following results. For 662 (53%) of the 1,238 estimable small areas, the multivariate model (Model MV) produced lower absolute relative error (ARE) than either its univariate counterpart (Model UV) or the Horwitz-Thompson estimator (HT). Evaluated by sample size, Model MV had the lowest ARE in 537 (54%) of the 1002 estimable small areas with 'small' sample sizes. Model MV also had the lowest ARE in 125 (56%) of the 236 estimable small areas with 'large' sample sizes.

Posterior variance of the multivariate model was lower in 65% of all estimable small areas. Relative posterior variances for the SRS sample and unweighted models (UV and MV) are illustrated in Figure 4.4 with relative posterior variance computed as in Section 4.3 using Models MV and UV applied to the SRS sample data. Evaluated by sample size, Model MV had lower posterior variance in 63% of small areas with 'small' sample sizes and 75% of small areas with 'large' sample sizes. Relative posterior variance by sample size is illustrated in Figure 4.4. Relative posterior variances are summarized in Table 4.4.

Overall, the mean ARE of the multivariate model was lower than the univariate model, though the univariate model had a lower median ARE for this analysis. The mean and median ARE of the HT estimate was significantly higher in all cases. ARE results are summarised in table 4.4 and illustrated by sample size in Figure 4.4. Figure 4.4 excludes one point, the relative ARE of the total number of full-time transit employees for the state of Louisiana, which had ARE(Model UV) minus ARE(Model MV) equal to 0.2343.

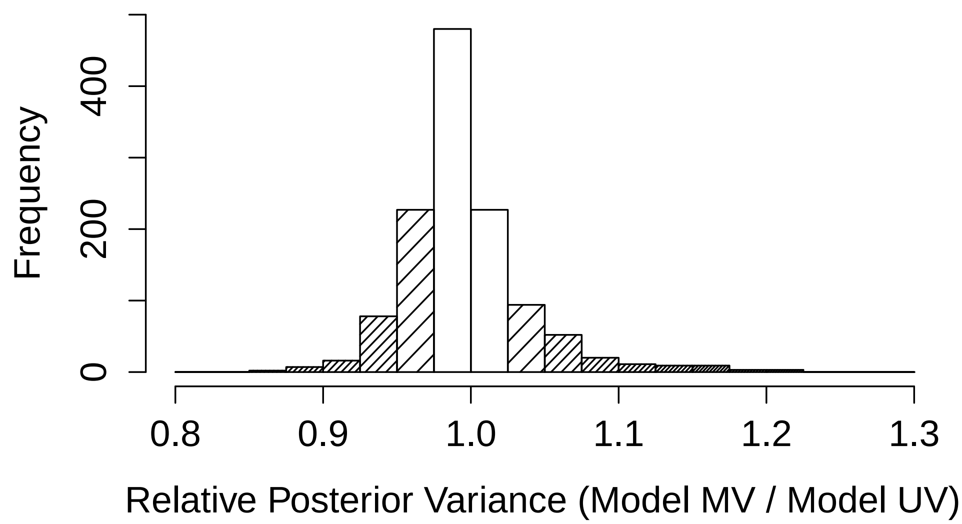

Figure 4.7: SRS Sample Relative Posterior Variance

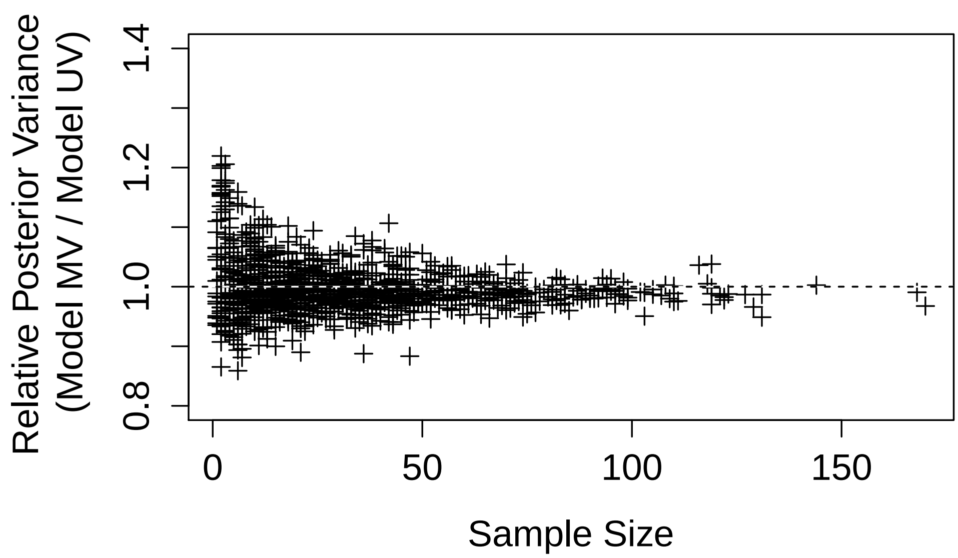

Figure 4.8: SRS Sample Relative Posterior Variance by Sample Size

Table 4.3: SRS ARE Results

	Model		
	HT	UV	MV
Median	0.9168	0.0342	0.0344
Mean	1.2028	0.0495	0.0494
	$n_i \leq 50$		
Median	0.7796	0.0306	0.0301
Mean	1.0820	0.0450	0.0493
	$n_i > 50$		
Median	1.4950	0.0450	0.0459
Mean	1.7140	0.0497	0.0495

Table 4.4: SRS Sample Relative Posterior Variance

Sample Sizes	All	$n_i \leq 50$	$n_i > 50$
Minimum	0.8589	0.8589	0.9457
1st Quartile	0.9734	0.9716	0.9794
Median	0.9888	0.9889	0.9885
Mean	0.9954	0.9967	0.9899
3rd Quartile	1.0100	1.0140	0.9996
Maximum	1.2190	1.2190	1.0420

Multivariate Posterior Variance
Divided by Univariate Posterior Variance

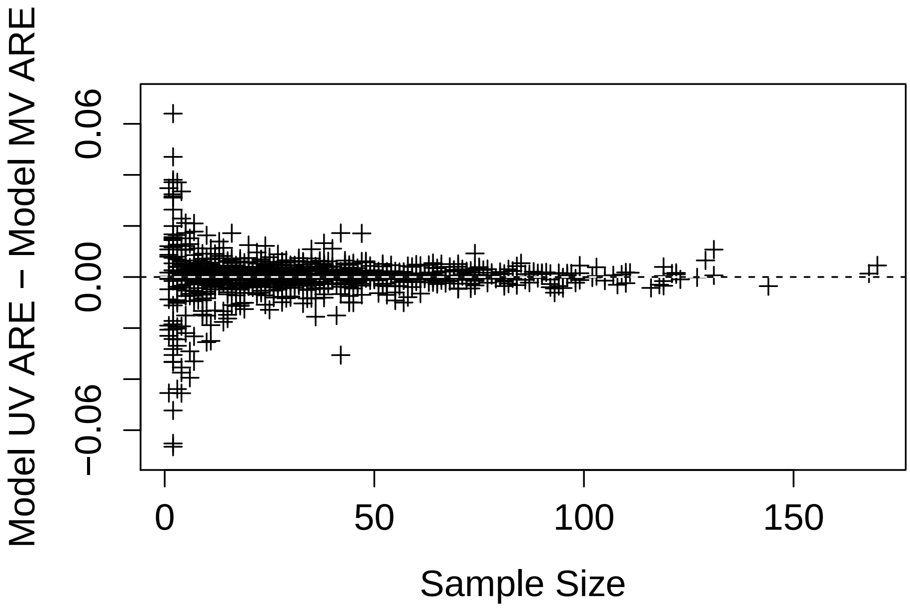

Figure 4.9: SRS Sample Relative ARE by Sample Size

For both small and large sample sizes, as well as overall, the multivariate model produced a higher number of more accurate estimates, in terms of either lower posterior variance or lower ARE, when compared to its univariate counterpart and to the design-based estimator for the sample drawn using stratified random sampling.

Chapter 5: Conclusion

Based on the simulation results, the multivariate hierarchical Bayesian approach to small area estimation shows promise in its apparent ability to produce improved estimates in relation to both design based and univariate approaches, in the case of samples drawn through either the PPS or SRS methodology. The multivariate models produced lower overall posterior variance than their univariate counterparts in both the PPS and SRS cases, while generating estimates that were on average within 5% of the known small area totals. In the analysis of the skewed PPS sampled data, the inclusion of the design weights as covariates provided a further improvement in the overall accuracy of estimates.

The multivariate model could be developed further by including a multivariate analog to the calibration constant introduced in Chandra and Chambers [2011], though this may or may not provide an improvement over the present inclusion of weights as covariates. Additionally, a components of variance error structure may be more appropriate for use in the model if it can be supposed that most or all governments within a state grow (or shrink) at similar rates between ASPEP surveys, reflecting year-to-year increases (or decreases) in state budgets. Alternatively, a components of variance error structure may be appropriate in modeling year-to-year employment trends by program function if they are present.

Appendix A: Example Census of Governments Data

Table A.1: Example Data from the Combined 2002, 2007 Census of Governments

				2002				2007			
				Full-Time		Part-Time		Full-Time		Part-Time	
Government ID	State	Type	Itemcode	Emp.	Pay	Emp.	Pay	Emp.	Pay	Emp.	Pay
11100100100000	11	1	005	5	8159	1	598	31	68634	0	0
11100100100000	11	1	023	9	14657	2	1455	18	38662	2	501
11100100100000	11	1	024	6	11628	2	3806	0	0	0	0
11100300300000	11	1	005	9	11309	3	867	11	14216	0	0
11100300300000	11	1	023	8	16147	0	0	6	13448	1	1050
11100300300000	11	1	025	6	13572	3	3034	7	18457	7	4196
11100400400000	11	1	005	5	6959	0	0	2	5311	2	2657
11100400400000	11	1	023	1	4756	0	0	1	3126	0	0
11100400400000	11	1	025	1	4089	3	1623	3	9188	0	0
11400100300000	11	4	050	5	13683	0	0	5	12691	0	0

Appendix B: Implementing the Gibbs Sampler in R

Implementation of the Gibbs sampler(s) with joint conditional distributions specified in Section 4.1 is carried out using the application "R". "R" as part of the R project for statistical computing as described in and available through:

`http://www.r-project.org/`

The following code is used in R to produce a Gibbs sampler, simulating a number of parallel chains (specified by the 'chains' variable) with the variable 'nmc' specifying the number of iterations per chain and the variable 'nbi' specifying the number of initial 'burn-in' observations to be discarded for each chain.

The program draws samples from the joint conditional distributions of the model (equivalent to equation (3.12)):

$$\mathbf{Y}{=}\mathbf{BX}{+}\mathbf{VF}{+}\mathbf{E},$$

as specified in the text of the Section 4.1.

After partitioning the data into known and unknown values (per equation (3.13), we have $\mathbf{Y}^{(1)}$, denoted Y1, as the $(s \times n)$ sub-matrix of sampled values and $\mathbf{X}^{(1)}$, denoted X1, as the corresponding $(p \times n)$ sub-matrix of covariates. 's' is taken as the number of variables to be modeled per sample (two in Model MV) while 'p' is the number of covariates used in modeling each variable, including the intercept. Here 's' is equal to two and 'p' is equal to three in the unweighted multivariate case.

$\mathbf{F}^{(1)}$, denoted F1, refers to the $(m \times n)$ portion of the design matrix corresponding to Y1 and X1. The final estimation of θ, denoted theta is stored at the end in each pass through an iteration and chain. σ_{e11}, denoted 'sigma_e11' is the first element of the conditional variances of $\mathbf{Y}$, used in calculating theta. Additionally, the matrix of numbers of unsampled units by small area $N_i - n_i$ is denoted 'Nni'. The conditional means and variances of $\vec{v}_i$ are represented by Mu_i and P_i, respectively with first and first-diagonal elements Mu_1i and p11_i, respectively.

```
#initialize output workspace
  Parms.chain=NULL;  theta.chain=NULL; vartheta.chain=NULL
#draw initial values of Omegas, generate V_0
  Omega_e=rWishart(1,s+1,solve(Se_0))[,,1]
  Sigma_e=solve(Omega_e)
  Omega_v=rWishart(1,s+1,solve(Sv_0))[,,1]
  V=matrix(sapply(1:m,function(k){rmt(1,V_0[,k],Sv_0,3)}),nrow=s)
# Begin Chain
  for (j in 1:nmc){
# vec(B) ~ MVN(s*p,Mu_B,Sigma_B)
    B=matrix(rmnorm(1,matrix((Y1-V%*%F1)%*%XtiXXt,s*p),iXXt%x%Sigma_e),s,p)
    YmBX=Y1-B%*%X1
    YmBXF=tcrossprod(YmBX,F1)
# V_i ~ MVN(s,Mu_i,P_i)
    p11=vector("numeric",m)
    Mu1=vector("numeric",m)
    for (k in 1:m){
      P_i=solve(Omega_v+ni[k]*Omega_e)
      p11[k]=P_i[1]
      Mu_i=P_i%*%Omega_e%*%YmBXF[,k]
      Mu1[k]=Mu_i[1]
      V[,k]=t(rmnorm(1,Mu_i,P_i))
    }
# Omega_e ~ Wishart(N,Phi_e)
    Omega_e=rWishart(1,N+s+1,as.matrix(solve(tcrossprod(YmBX-V%*%F1))))[,,1]
```

```
    Sigma_e=solve(Omega_e)
# Omega_v ~ Wishart(m+a,Phi_v,new)
    Omega_v=rWishart(1,m+s+1,solve(tcrossprod(V)))[,,1]
# output parameters only when iterations are past burn-in
#  and iteration modulus thin is equal to zero
    if (j>nbi & j%%thin==0){
      Parms.chain=rbind(Parms.chain,c(
        B,Omega_e[lower.tri(Omega_e,diag=T)],
          Omega_v[lower.tri(Omega_v,diag=T)]))
      gi=lapply(c(1:m),function(k){exp(B[1,]%*%X2[,as.vector(F2[k,]==1)]
          +Mu1[k]+(p11[k]+Sigma_e[1])/2)})
      theta.chain=rbind(theta.chain, as.vector(Z1F1+sapply(gi,sum)-Nni))
      vartheta.chain=rbind(vartheta.chain,sapply(1:m,function(k){
          tcrossprod(gi[[k]]%*%(exp(p11[k]+diag(Sigma_e[1],
            nrow=Nni[k]))-1),gi[[k]])}
        ))
    }
  }
```

Bibliography

Joseph J. Barth, Yang Cheng, and Carma R. Hogue. Reducing the public employment survey sample size. *JSM*, 2009.

George E. Battese, Rachel M. Harter, and Wayne A. Fuller. An error-components model for predictio of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83(401):28–36, March 1988.

George Casella and Edward I. George. Explaining the gibbs sampler. *The American Statistician*, 46(3):167–174, 1992.

Hukum Chandra and Ray Chambers. Small area estimation under transformation to linearity. *Survey Methodology*, 37(1):39–51, June 2011.

Yang Cheng, Casey Corcoran, Joe Barth, and Carma R. Hogue. Estimation procedure for new public employment survey design. In *Section on Survey Research Methods*, Session 479: Analysis of Complex Survey Data, 2009.

Yang Cheng, Bac Tran, Carma R. Hogue, and Partha Lahiri. Small area estimation applications in the us census bureau. *Proceedings 59th International Statistical Institute World Statistics Congress*, 2013.

William G. Cochran. *Sampling Techniques*. John Wiley & Sons, Inc., New York, third edition, 1977.

Gauri S. Datta and Malay Ghosh. Bayesian prediction in linear models: Applications to small area estimation. *The Annals of Statistics*, 19(4):1748–1770, 1991.

Gauri S. Datta, Bannmo Day, and T. Maiti. Multivariate bayesian small area estimation: An application to survey and satellite data. *Sankhyā: The Indian Journal of Statistics*, 60(3):344–362, 1998.

Gauri S. Datta, Bannmo Day, and Ishwar Basawa. Empirical best linear unbiased and empirical bayes prediction in multivariate small area estimation. *Journal of Statistical Planning and Inference*, 75(2):269–279, 1999.

Chris C. Holmes and Leonhard Held. Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis*, 1(1):145–168, 2006.

Peter M. Lee. *Bayesian Statistics: An Introduction*. John Wiley and Sons Ltd., London, 3rd edition, 2004.

J. N. K. Rao. *Small Area Estimation*. Wiley series in survey methodology. John Wiley & Sons, Inc., Hoboken, New Jersey, 2003.

Government Finance and Employment Classification Manual. U.S. Bureau of the Census, 2006 edition, October 2006.

Hui Zheng and Roderick J. A. Little. Inference for the population total from probability-proportional-to-size samples based on predictions from a penalized spline nonparametric model. *Journal of Official Statistics*, 21(1):1–20, 2005.