Survey Methodology, June 2014
Vol. 40, No. 1, pp. 1-13
Statistics Canada, Catalogue No. 12-001-X

Hierarchical Bayes Modeling of Survey-Weighted Small Area Proportions

Benmei Liu, Partha Lahiri and Graham Kalton [1]

Abstract

The paper reports the results of a Monte Carlo simulation study that was conducted to compare the effectiveness of four different hierarchical Bayes small area models for producing state estimates of proportions based on data from stratified simple random samples from a fixed finite population. Two of the models adopted the commonly made assumptions that the survey weighted proportion for each sampled small area has a normal distribution and that the sampling variance of this proportion is known. One of these models used a linear linking model and the other used a logistic linking model. The other two models both employed logistic linking models and assumed that the sampling variance was unknown. One of these models assumed a normal distribution for the sampling model while the other assumed a beta distribution. The study found that for all four models the credible interval design-based coverage of the finite population state proportions deviated markedly from the 95 percent nominal level used in constructing the intervals.

Key Words: Weighted proportions; Hierarchical Bayes modeling; Beta distribution; credible interval.

1 Introduction

Small area estimation methods are often used to estimate the proportions of units with a given characteristic for small areas. For example, small area estimation methods are used: in the Census Bureau's Small Area Income and Poverty Estimates (SAIPE) program to estimate poverty rates for states, counties, and school districts (Citro and Kalton 2000; Maples and Bell 2005); with data from the National Survey on Drug Use and Health (NSDUH) to estimate substance rates for states (Wright, Sathe and Spagnola 2007); and with data from the National Assessment of Adult Literacy (NAAL) to estimate proportions at the lowest level of literacy for states and counties (Mohadjer, Rao, Liu, Krenzke and Van De Kerckhove 2012). In each case, the survey's sample sizes in the small areas are not large enough to support direct estimates of adequate precision. A wide variety of methods have been developed to address such small area estimation problems. See Rao (2003) and Jiang and Lahiri (2006a) for reviews, and Chattopadhyay, Lahiri, Larsen and Reimnitz (1999), Farrell, MacGibbon and Tomberlin (1997), Malec, Sedransk, Moriarity and LeClere (1997) and Malec, Davis and Cao (1999) for methods specifically for estimating small area proportions. The range of methods includes both empirical best prediction (EBP) and hierarchical Bayes (HB) approaches and models developed at both the area and unit levels. We focus on HB area level models in this paper.

When an HB area level model is used to produce estimates of proportions of units with a given characteristic for small areas, it is commonly assumed that the survey-weighted proportion for each sampled small area has a normal sampling distribution and that the sampling variance of this proportion is known. However, these assumptions are problematic when the small area sample size is small or when the true proportion is near 0 or 1. Reliance on the central limit theorem for approximate normality of the sampling distribution of a proportion requires reasonably large samples, particularly when the population

1. Benmei Liu, Division of Cancer Control and Population Sciences, National Cancer Institute, 9609 Medical Center Drive Room 4E524, Bethesda, Maryland 20892; E-mail: liub2@mail.nih.gov; Partha Lahiri, JPSM, University of Maryland, 1218 Lefrak Hall, College Park, Maryland 20742; Graham Kalton, Westat, 1600 Research Boulevard, Rockville, Maryland 20850. The majority of this research took place while the first author was a graduate student in the Joint Program in Survey Methodology at the University of Maryland.

proportion is very small or very large (e.g., under 0.1 or over 0.9). Moreover, with very small or very large proportions, the sampling variance of a sample proportion is highly sensitive to the actual value of the proportion, thus making it difficult to establish a suitable value for the sampling variance. In an effort to overcome these problems, we propose two alternative models for small area proportions and compare them with two commonly used models. The models are described in Section 3. The four models are compared by means of a Monte Carlo simulation study in which stratified simple random samples are generated from a fixed finite population. The simulation study is described in Section 4 and the results are presented in Section 5. The paper finishes with some concluding remarks in Section 6. First, however, we introduce the notation for a stratified simple random sample design in Section 2.

2 Notation

Let N_{ih} denote the population size in stratum h in area i of a finite population $(i=1,...,m; h=1,...,H_i)$. Let y_{ihk} be the binary response for the characteristic of interest for unit k in stratum h in area i $(k=1,...,N_{ih})$. The parameters to be estimated are the small area proportions $P_i = \sum_h \sum_k y_{ihk} / N_{ih}$.

With the stratified simple random sample design under study, n_{ih} units are selected from the N_{ih} units in stratum (ih). The standard direct survey estimator for P_i is:

$$p_{iw} = \frac{\sum_h^{H_i} \sum_k^{n_{ih}} w_{ih} y_{ihk}}{\sum_h^{H_i} \sum_k^{n_{ih}} w_{ih}},\ i=1,...,m; \tag{2.1}$$

where w_{ih} denotes the sampling weight given by $w_{ih} = N_{ih} / n_{ih}$.

The variance of p_{iw} can be expressed as

$$VAR_{st}(p_{iw}) = \frac{P_i(1-P_i)}{n_i} DEFF_i, \tag{2.2}$$

where $DEFF_i$ is the design effect reflecting the effect of the complex sample design (Kish 1965). For a stratified simple random sample with negligible sampling fractions in all strata, the design effect is given approximately by:

$$DEFF_i = \frac{\sum_h W_{ih}^2 P_{ih}\left(1-P_{ih}\right)/n_{ih}}{P\left(1-P_i\right)/n_i}, \tag{2.3}$$

where $W_{ih} = N_{ih} / N_i$, $N_i = \sum_h N_{ih}$, $n_i = \sum_h n_{ih}$ and P_{ih} is the population proportion in stratum h in area i.

The design effect $DEFF_i$ is a function of the P_{ih}, which are unknown. If $P_{ih}(1-P_{ih}) \approx P_i(1-P_i)$, $DEFF_i$ can be approximated by $deff_{iw} = n_i \sum_h W_{ih}^2 / n_{ih}$. The value of $deff_{iw}$ can be readily computed since it does not depend on any unknown parameters.

Small area estimation procedures can be used to address the problem that p_{iw} is very imprecise when the sample size n_i is small. Section 3 describes the HB area level models investigated in this study.

3 Models Studied

A general area-level small area model has two components. One—the sampling model—is a model for the sampling error of the direct survey estimates. The other—the linking model—relates the population value for an area to area-specific auxiliary variables $x_i = (x_{i1}, \ldots, x_{ip})'$.

Section 3.1 describes two area models that are often used for estimating small area proportions and Section 3.2 outlines some problems associated with these models. Section 3.3 describes two alternative models that may serve to address these problems.

3.1 Two Commonly Used Models

We study two commonly used models for comparison with the new models described in Section 3.4. The first is the Fay-Herriot model (Fay and Herriot 1979), which assumes known sampling variances and normal distributions for both the sampling and the linking models. The second is the normal-logistic model, which differs from the Fay-Herriot model only by the replacement of a logit-normal distribution for the normal distribution in the linking model.

Model 1: (Fay-Herriot normal-normal model)

Sampling model:

$$p_{iw} \mid P_i \overset{ind}{\sim} N(P_i,\ \psi_i) \tag{3.1}$$

Linking model:

$$P_i \mid \beta, \sigma_v^2 \overset{ind}{\sim} N(x_i'\beta, \sigma_v^2) \tag{3.2}$$

Model 2: (normal-logistic model)

Sampling model:

$$p_{iw} \mid P_i \overset{ind}{\sim} N(P_i,\ \psi_i) \tag{3.3}$$

Linking model:

$$g(P_i) \mid \beta, \sigma_v^2 \overset{ind}{\sim} N(x_i'\beta, \sigma_v^2) \tag{3.4}$$

In both models the sampling variance ψ_i is assumed to be known. Model 1 is referred as a matched model because the sampling and linking models can be combined to produce a relatively simple linear mixed model. However, a nonlinear linking model is often preferred for modeling proportions, leading to unmatched sampling and linking models, as in Model 2 (see, for example, You and Rao 2002). The link function $g(\cdot)$ can be empirically determined by checking the model fit. The *log* and *logit* link functions have been used. The $\text{logit}(P_i)$ linking model is chosen here in order to guarantee that the estimate of P_i always falls within the allowable range of (0,1).

3.2 Issues with Models 1 and 2

There are two main issues associated with Models 1 and 2. The first is that both models assume known sampling variances ψ_i, whereas in practice they have to be estimated. A simple approach is to use the direct variance estimate but that estimate is very imprecise when P_i is either very small or very large and when the sample size n_i is small. An alternative, more complex, approach is to develop an approximate estimate of P_i, say p_{isyn}, from a simple model such as a logistic model for p_{iw} in terms of the auxiliary variables, and then use that estimate in the following synthetic variance estimator:

$$\text{var}_{stsyn} = \frac{p_{isyn}(1 - p_{isyn})}{n_i} deff_{iw}. \tag{3.5}$$

When there are no auxiliary variables available, the overall sample proportion may be used for p_{isyn} in the computation of the synthetic variance estimator.

The second issue concerns the normality assumption in the sampling model, which is based on a large sample approximation. As noted in Section 1, when the sample size n_i is small and P_i is near 0 or 1, as is often the case with small area estimation, that assumption is problematic.

3.3 Two Alternative Models

Under Models 1 and 2, the unknown sampling variances ψ_i are estimated in some way, and then the resultant estimates are treated as if they were the known true values. A possible alternative approach is to treat the ψ_i as unknown parameters in the HB model, as has been done in a number of studies. For example, Arora and Lahiri (1997) applied an HB model to model the design-based variances for the sample estimates. Singh, Folsom and Vaish (2005) proposed the use of a generalized design effect model to smooth the sampling covariance matrix in small area modeling with survey data. Recently, You (2008) proposed the use of equal design effects over time to model the sampling variances in estimating small area unemployment rates using a cross-sectional and time series log-linear model. The approach of treating the sampling variances ψ_i as unknown is adopted in Model 3, as a variant of Model 2. One approach for addressing the non-normality of the sampling distributions of the survey-weighted small area proportions is to replace the normal distribution assumption by an alternative distribution. That approach is applied in Model 4 with the assumption of a beta sampling distribution, a distribution that has the desirable property of having a (0,1) range. In other regards Model 4 is the same as Model 3, including treating the ψ_i, $i = 1,...,m$ as unknown parameters. Model 4 was previously considered by Jiang and Lahiri (2006b) in an illustrative example to estimate finite population domain means using an EBP approach.

Model 3: (normal-logistic model with unknown sampling variance)

Sampling model:

$$p_{iw} \mid P_i \overset{ind}{\sim} N(P_i,\ \psi_i) \tag{3.6}$$

Linking model:

$$logit(P_i) \mid \beta, \sigma_v^2 \overset{ind}{\sim} N(x'\beta, \sigma_v^2) \tag{3.7}$$

Model 4: (beta-logistic model with unknown sampling variance)

Sampling model:

$$p_{iw} \mid P_i \overset{ind}{\sim} beta(a_i, b_i) \tag{3.8}$$

Linking model:

$$logit(P_i) \mid \beta, \sigma_v^2 \overset{ind}{\sim} N(x_i'\beta, \sigma_v^2) \tag{3.9}$$

For both Model 3 and Model 4, the approximate variance function $\psi_i = [P_i(1 - P_i)/n_i] deff_{iw}$ is used. The parameters a_i and b_i in Model 4 are given by:

$$a_i = P_i\left(\frac{n_i}{deff_{iw}} - 1\right), \text{ and } b_i = (1 - P_i)\left(\frac{n_i}{deff_{iw}} - 1\right).$$

HB small area estimates can be computed from all four models using the Metropolis-Hastings algorithm within the Gibbs sampler. Details of the algorithm, which draws random samples based on the full conditional distributions of the unknown parameters starting with one or multiple sets of initial values, are given by Robert and Casella (1999) and Chen, Shao, and Ibraham (2000). You and Rao (2002) also describe in detail how the Metropolis-Hastings algorithm works within the Gibbs sampler for models similar to Models 1 and 2. The algorithm works for Models 3 and 4 in the same way as for Model 2. The full conditional distributions under each model are provided in Appendix A.

4 Simulation Study

4.1 The Study Population and the Sample Design

This section describes the simulation study that was conducted to compare the efficiency of the small area estimates produced by the four HB models. The simulation study was based on the 2002 Natality public-use data file that covered all births occurring within the United States in that calendar year. The file contained data obtained from the certificates filed for births occurring in each state and territory (for details see U.S. National Center for Health Statistics 2009).

The finite population studied was restricted to the 4,024,378 records of live births that occurred in 2002 in the 50 states of U.S. and the District of Columbia (DC) and that had birth weights reported. The parameter of interest is the state level low birthweight rate P_i, $i = 1, ..., 51$, where low birthweight is defined as less than 2,500 grams. The value of P_i varied from 5 percent to 11 percent across the states.

Within each state, a stratified SRS design was used to draw samples from the birth records. Mother's race (White, Black, and Other) was used as the stratification variable. The national sample size was set to be about 1,500 birth records for each race group. A uniform sampling fraction was used across the states for each race group, subjecting to the condition that at least two birth records were sampled within each race group in each state. The resultant national sample size turned out to be $n = 4,526$ birth records. The state sample sizes n_i ranged from 7 (for small states such as Vermont) to 690 (for California), with a median sample size of 61. This sampling procedure was repeated $R = 1,000$ times, creating 1,000 independent sample data sets. The sampling weights remained the same over different simulation runs.

4.2 Computation of the HB Estimates

For simplicity, the following assumptions were made for the HB models:

1. No auxiliary variables were used, so that $x_i' \beta = \mu$.
2. For Models 1 and 2, the sampling variances were taken to be given by $\psi_i = [p_w(1 - p_w)/n_i] deff_{iw}$, where $p_w = \sum\sum\sum w_{ih} y_{ihk} / \sum\sum n_i w_{ih}$ is the national estimate of the proportion of low birthweight live births. (A check on the use of $deff_{iw}$ as an approximation for $DEFF_i$ showed that the approximation was reasonable: the two quantities were close, with a product moment correlation of 0.96 and an average ratio of 1.08 between $deff_{iw}$ and $DEFF_i$.)
3. Flat prior for μ, i.e., $f(\mu) \propto 1$, and inverse gamma for σ_v^2, i.e., $\sigma_v^2 \sim IG(0.001,\ 0.001)$.

For each sample data set, the first step in the computations was to calculate the state direct sample estimates. The estimates for each sample data set were then used in turn as input to the WinBUGS software (Lunn, Thomas, Best and Spiegelhalter 2000), which was used to produce the HB estimates for all four models.

In a sizable number of the states with small n_i, the direct estimates were zero in some of the sample data sets. Since WinBUGS can handle direct estimates of zero only for Model 1, the zero direct estimates were perturbed to very small positive numbers for the other models.

For each WinBUGS run, three independent chains were used. For each chain, burn-ins of 10,000 samples were produced, with 10,000 samples after burn-in. The samples after burn-in were thinned by a factor of two to reduce auto-correlation of the MCMC samples. The resultant 15,000 MCMC samples from the three chains after burn-in were then used to compute the posterior mean and percentiles for each HB model based on each sample data set. The potential scale reduction factor $\hat{R}$ was used as the primary measure for convergence (see Gelman and Rubin 1992). The WinBUGS code is given in Appendix B.

5. Simulation Results

In Section 5.1 we report our main results for the credible intervals obtained for the state proportions of low birthweight live births from the application of each of the four models. Section 5.2 then examines the biases and root mean square errors of these estimates.

5.1 Model estimates and credible intervals

Let P_i^{HB} denote an HB estimator of P_i, the percentage of low birthweight live births in state i, and let $P_{i,q}^{HB}$ denote the q^{th} percentile of the posterior distribution of P_i. Based on the results from the 1,000 simulation data sets, Table 5.1 presents the following for each model: the noncoverage probability for the 95 percent credible intervals of P_i , i.e., the probability that the interval from $P_{i,.025}^{HB}$ to $P_{i,.975}^{HB}$ fails to cover P_i and the mean width of the credible intervals $P_{i,.975}^{HB}$ - $P_{i,.025}^{HB}$. The corresponding Monte Carlo simulation standard errors are also reported in the table in parentheses.

To examine the effect of state sample size on the simulation results, the 50 states and the District of Columbia are divided into three groups according to their sample sizes: the 15 states with small sample sizes $(n_i \leq 30)$; the 24 states with medium sample sizes $(30 < n_i \leq 100)$; and the 12 states with large sample sizes $(n_i > 100)$. The results presented in Table 5.1 are overall averages across all states and averages for the three groups separately.

It can be seen from the upper half of Table 5.1 that the Fay-Herriot model (M1) credible intervals are very conservative, giving nearly zero noncoverage. The lower half of the table shows that this result is obtained at the cost of the largest average credible interval width among the four models. The M1 credible interval widths are very stable. A small proportion of the M1 credible intervals had negative lower bounds.

A possible explanation for the low level of noncoverage with M1 is that the sampling variances were overestimated, perhaps because $deff_{iw}$ was used in place of $DEFF_i$. To examine this possibility, we used $DEFF_i$ in computing the sampling variance and found virtually no difference in the noncoverage rate. We also ran the model with the true variance as defined in (2.2) and again found no appreciable difference in the noncoverage rates. The non-normality of the sampling distribution of p_{iw} could also be a source of this problem.

Table 5.1
Percentage of times that the 95 percent credible intervals fail to cover P_i, mean 95 percent credible interval width, along with the Monte Carlo simulation standard errors based on 1,000 simulations (in percentages)

State sample size n_i	M1*	M2	M3	M4
	Noncoverage percentage (Monte Carlo simulation standard error)			
Overall	0.40 (0.028)	8.24 (0.109)	6.52 (0.101)	4.36 (0.088)
$n_i \leq 30$ (15 states)	0.05 (0.019)	11.39 (0.239)	8.45 (0.216)	6.21 (0.190)
$30 < n_i \leq 100$ (24 states)	0.46 (0.043)	9.44 (0.167)	7.61 (0.156)	4.52 (0.132)
$n_i > 100$ (12 states)	0.70 (0.076)	1.91 (0.122)	1.94 (0.124)	1.74 (0.119)
	Mean width of the 95% credible interval (Monte Carlo simulation standard error)			
Overall	9.05 (0.004)	5.52 (0.009)	6.20 (0.009)	8.45 (0.014)
$n_i \leq 30$ (15 states)	10.27 (0.009)	5.94 (0.020)	6.78 (0.021)	9.30 (0.034)
$30 < n_i \leq 100$ (24 states)	9.16 (0.005)	5.60 (0.013)	6.28 (0.013)	8.71 (0.021)
$n_i > 100$ (12 states)	7.29 (0.004)	4.84 (0.012)	5.30 (0.013)	6.88 (0.017)

*Note: For Model 1, a small proportion of the credible intervals had negative lower bounds.

At 8.2 percent, the overall noncoverage rate of the credible intervals for the normal-logistic model (M2) is appreciably above the nominal rate of 5 percent. This model has the smallest average interval width. The noncoverage rate for the normal-logistic model with unknown variance (M3) is closer to the nominal rate, with an overall interval width that is somewhat larger than that for M2.

The noncoverage rate for the beta-logistic model (M4) of 4.4 percent overall is closest to the nominal noncoverage rate. However, the average width of the credible intervals is larger than those for M2 and M3 and the Monte Carlo standard error of the interval width is larger than that of the other three models. This instability may be due to the complexity of the full conditional distribution for the beta model. The large proportion of the 1,000 direct estimates that were 0 for some of the states with small sample sizes may also have caused significant problems in fitting the beta distribution.

As is to be expected, for all four models the mean width of the credible intervals declines with increasing state sample size and the variation in the widths also declines with increased sample size. Even with these declines, however, the noncoverage rates also decline with increasing sample size for Models 2, 3, and 4. The noncoverage rates are in fact very small for the states with large n_i, suggesting that the credible intervals are not adequately reflecting the effect of the greater precision of the direct estimates in the states with large sample sizes.

5.2 Biases and RMSEs of the model-based estimates

For further investigation of these results, we examined the bias and the root mean square errors (RMSEs) of the estimates P_i^{HB} for each model. The results are presented in Table 5.2 in the same format as Table 5.1. The biases for the estimates under M1, M2, and M3 exhibit a similar pattern: the biases are large and positive for the small states, and offset to some extent by relatively small negative biases for the medium and large states. The biases for the estimates for M4 have a very different pattern: they are almost zero for the small states and have large negative values for the medium and large states. This indicates that M4 would perform better than the other three models in terms of bias when the small area sample sizes are small.

Table 5.2
The biases and the root mean square errors of the estimates of P_i based on the four models (in percentages)

State sample size n_i	M1		M2		M3		M4	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
Overall	0.165	1.518	0.071	1.346	-0.009	1.411	-0.214	1.712
$n_i \le 30$ (15 states)	0.621	1.651	0.572	1.630	0.466	1.652	0.009	1.922
$30 < n_i \le 100$ (24 states)	-0.006	1.547	-0.123	1.386	-0.201	1.452	-0.319	1.775
$n_i > 100$ (12 states)	-0.063	1.294	-0.167	0.911	-0.219	1.026	-0.283	1.323

6. Discussion

In this paper, we report the results of a simulation study from a real finite population to evaluate the credible intervals obtained from four different hierarchical models in terms of their interval lengths and their design-based coverage properties. To the best of our knowledge, such a design-based evaluation of small area credible (or confidence) intervals has not previously been performed in the evaluation of small area estimates.

In the simulation study, we have compared the design-based coverage properties of credible intervals resulting from different hierarchical Bayes models for estimating small area proportions from a stratified simple random sample design. Overall, none of the models emerges as a clear winner and so we are not in a position to recommend any of the models studied.

The hierarchical Bayes version of the well-known Fay-Herriot model appears to produce overly conservative credible intervals. The non-normality of both the sampling and the linking models is a possible source of this problem. The credible intervals for the beta-logistic hierarchical model achieve almost the nominal coverage for the finite population proportions and the bias property for this model is the best among the four models being compared when the sample sizes are small. However, since one of the full conditionals for the beta-logistic model involves the survey-weighted proportions, there is a problem with the MCMC whenever the survey-weighted proportion is zero. The credible intervals for this model are also wider than those for the other two models with a logistic linking model. It may be possible to reduce the width of the credible interval for the beta-logistic model by modifying the model in some way, such as by employing a suitable two-part mixture random effect model that will avoid the problem of survey-weighted proportions of zero. Further investigation is needed. Also consideration could usefully be given to other possible models, possibly a discrete probability model for Level 1, to improve on interval estimation of small proportions for small areas.

The simulation study found that the coverage of the Bayesian credible intervals for the finite population proportions was far from the nominal 95 percent level for all four models, and a similar finding was also obtained for the design-based coverage of the widely-used Fay-Herriot model. In the light of these findings we carried out a number of further analyses in a search for an explanation. These analyses included: adding predictor variables to the models; using a uniform prior distribution for σ_v^2 (based on arguments made by Gelman 2006); the use of empirical best prediction approach for the M1 model; increasing the sample size in states with few births to a minimum of 50; and applying the methods to estimate the proportion of births in each state below the national median birthweight. Although there are some differences in the coverage properties for the state finite population proportions, none of these analyses produced coverage rates close to the nominal rates. The only case where the nominal rates coincided with the actual coverage rates was for a simulated dataset constructed under model M1 for the state proportions below the national median birthweight; the average coverage rates were 5.1 and 5.2 percent for the EBP and HB approaches, respectively.

This simulation study was restricted to a single stage sample design. In addition, for simplicity no auxiliary variables were included in the linking models in our main analyses, whereas in practice the inclusion of such variables is routine and almost essential. Further simulation studies are needed to cover different sample designs, different sample sizes, and to incorporate some auxiliary variables in the linking models. We hope that our study will encourage others to conduct similar design-based simulations to evaluate small area estimation methods. Based on our limited results, users of small area estimates need to be cautioned about the interpretation of the credible intervals associated with the estimates.

Acknowledgements

The authors wish to thank the associate editor and two referees for a number of constructive suggestions which led to substantial improvement of the original manuscript. The second author's research was supported by the National Science Foundation SES-085100.

Appendix A

A1. Full conditional distributions for the parameters of each model

Let $\vec{p} = (p_{1w}, \ldots, p_{mw})^t$ and $r_i = \dfrac{\psi_i}{\psi_i + \sigma_v^2}$.

The full conditional distributions for the Fay-Herriot model (M1) are given as follows:

i) $\theta_i \mid \mu, \sigma_v^2, \vec{p} \sim N((1-r_i)p_{iw} + r_i\mu, \quad \psi_i(1-r_i))$;

ii) $\mu \mid \theta_i, \sigma_v^2, p \sim N\left(\dfrac{1}{m}\displaystyle\sum_{i=1}^{m}\theta_i, \dfrac{\sigma_v^2}{m}\right)$;

iii) $\sigma_v^2 \mid \mu, \theta_i, \vec{p} \sim ING\left(a + \dfrac{1}{2}m, b + \dfrac{1}{2}\displaystyle\sum_{i=1}^{m}(\theta_i - \mu)^2\right)$.

The full conditional distributions for the Normal-Logistic model (M2) are given as follows:

i) $\theta_i \mid \mu, \sigma_v^2, \vec{p} \propto \dfrac{1}{\theta_i(1-\theta_i)\sigma_v\sqrt{\psi_i}} \exp\left(-\dfrac{(p_{iw} - \theta_i)^2}{2\psi_i} - \dfrac{(\text{logit}(\theta_i) - \mu)^2}{2\sigma_v^2}\right)$;

ii) $\mu \mid \theta_i, \sigma_v^2, p \sim N\left(\dfrac{1}{m}\displaystyle\sum_{i=1}^{m}\text{logit}(\theta_i), \ \dfrac{\sigma_v^2}{m}\right)$;

iii) $\sigma_v^2 \mid \mu, \theta_i, \vec{p} \sim ING\left(a + \dfrac{1}{2}m, b + \dfrac{1}{2}\displaystyle\sum_{i=1}^{m}(\text{logit}(\theta_i) - \mu)^2\right)$.

The full conditional distributions for the Normal-Logistic model with unknown variance (M3) are the same as those of M2 except that replacing ψ_i by $\theta_i(1-\theta_i)deff_{iw} / n_i$ for the distribution of θ_i given other parameters.

Let $\delta_{iw} = \dfrac{n_i}{deff_{iw}} - 1$. The full conditional distributions for the Beta-Logistic model (M4) are given as follows:

i) $\theta_i \mid \mu, \sigma_v^2, \vec{p} \propto \dfrac{1}{\theta_i(1-\theta_i)\sigma_v} \dfrac{p_{iw}^{\theta_i\delta_{iw}-1}(1-p_{iw})^{(1-\theta_i)\delta_{iw}-1}}{\Gamma(\theta_i\delta_{iw})\Gamma((1-\theta_i)\delta_{iw})} \exp\left(-\dfrac{(\text{logit}(\theta_i) - \mu)^2}{2\sigma_v^2}\right)$;

ii) $\mu \mid \theta_i, \sigma_v^2, p \sim N\left(\frac{1}{m}\sum_{i=1}^{m}\text{logit}(\theta_i),\ \frac{\sigma_v^2}{m}\right);$

iii) $\sigma_v^2 \mid \mu, \theta_i, \vec{p} \sim ING\left(a + \frac{1}{2}m, b + \frac{1}{2}\sum_{i=1}^{m}(\text{logit}(\theta_i) - \mu)^2\right).$

Appendix B

WinBUGS code for Model 1:

```
model {
      for ( i in 1:N)  {
         pobs[i] ~ dnorm(theta[i], D[i])
         D[i] <- 1/varhat[i]
         theta[i]<-u+v[i]
         v[i]~dnorm(0, tau)
                           }
        u~dflat()
       tau~dgamma(0.001, 0.001)
       sigma_v2<-1/tau
         }
```

WinBUGS code for Model 2:

```
model {
      for ( i in 1:N)  {
         pobs[i] ~ dnorm(theta[i], D[i])
         D[i] <- 1/varhat[i]
         logit(theta[i])<-u+v[i]
         v[i]~dnorm(0, tau)
                            }
       u~dflat()
       tau~dgamma(0.001, 0.001)
       sigma_v2<-1/tau
         }
```

WinBUGS code for Model 3:

```
model {
      for ( i in 1:N)  {
         pobs[i] ~ dnorm(theta[i], E[i])
         E[i] <- SAMPn[i]/(theta[i]*(1-theta[i])*DEFF_kish[i])
         logit(theta[i])<-u+v[i]
         v[i]~dnorm(0, tau)
```

```
    D[i]<-1/E[i]
                    }
  u~dflat()
  tau~dgamma(0.001, 0.001)
  sigma_v2<-1/tau
     }
```

WinBUGS code for Model 4:

```
model {
     for ( i in 1:N)  {
        pobs[i] ~ dbeta(a[i], b[i])
        a[i] <- theta[i]*(theta[i]*(1-theta[i])/D[i]-1)
        b[i] <- (1-theta[i])*(theta[i]*(1-theta[i])/D[i]-1)
        logit(theta[i])<-u+v[i]
        v[i]~dnorm(0, tau)
        D[i]<-theta[i]*(1-theta[i])*DEFF_kish[i]/SAMPn[i]
                    }
  u~dflat()
  tau~dgamma(0.001, 0.001)
  sigma_v2<-1/tau
     }
```

References

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

Chattopadhyay, M., Lahiri, P., Larsen, M., and Reimnitz, J. (1999). Composite estimation of drug prevalence for sub-state areas. *Survey Methodology*, 25, 81-86.

Chen, M., Shao, Q., and Ibraham, J.G. (2000). *Monte Carlo Methods in Bayesian Computation.* New York: Springer-Verlag.

Citro, C., and Kalton, G. (Eds.) (2000). *Small-Area Income and Poverty Estimates: Priorities for 2000 and Beyond*. Washington, DC: National Academy Press.

Farrell, P.J., MacGibbon, B., and Tomberlin, T.J. (1997). Empirical Bayes estimators of small area proportions in multistage designs. *Statistical Sinica*, 7, 1065-1083.

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

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1, 515-533.

Gelman, A., and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequence. *Statistical Science*, 7, 457-472.

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

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

Kish, L. (1965). *Survey sampling*. New York: John Wiley.

Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS – A Bayesian modeling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10, 325-337.

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

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

Maples, J., and Bell, W.R. (2005). Evaluation of school district poverty estimates: Predictive models using IRS income tax data. *Proceedings of the Survey Research Methods Section, American Statistical Association,* 1322-1329.

Mohadjer, L., Rao, J.N.K., Liu, B., Krenzke, T., and Van De Kerckhove, W. (2012). Hierarchical Bayes small area estimates of adult literacy using unmatched sampling and linking models. *Journal of the Indian Society of Agricultural Statistics,* 66 (1), 55-63.

Rao, J.N.K. (2003). *Small area estimation*. New York: John Wiley and Sons.

Robert, C.P., and Casella, G. (1999). *Monte Carlo Statistical Methods*. New York: Springer-Verlag.

Singh, A.C., Folsom, R.E. JR., and Vaish, A.K. (2005). Small area modeling for survey data with smoothed error covariance structure via generalized design effects. *Federal Committee on Statistical Methods Statistical Working Paper*, No. 39 (http://www.fcsm.gov/05papers/Singh_etal_IIIC.pdf).

U.S. National Center for Health Statistics (2009). *National Vital Statistics System*. *Birth Data*. (http://www.cdc.gov/nchs/births.htm)

Wright, D., Sathe, N., and Spagnola, K. (2007). *State Estimates of Substance Use from the 2004-2005 National Surveys on Drug Use and Health.* (DHHS Publication No. SMA 07-4235, NSDUH Series H-31). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

You, Y. (2008). An integrated modeling approach to unemployment rate estimation for subprovincial areas of Canada. *Survey Methodology*, 34, 19-27.

You, Y., and Rao, J.N.K. (2002). Small area estimation using unmatched sampling and linking models. *Canadian Journal of Statistics*, 30, 3-15.