Small Area Estimation with Uncertain Random Effects

Gauri Sankar Datta [1 2 3] **and Abhyuday Mandal** [4]

Abstract

Random effects models play an important role in model-based small area estimation. Random effects account for any lack of fit of a regression model for the population means of small areas on a set of explanatory variables. In a recent paper, Datta, Hall and Mandal (2011, *J. Amer. Statist. Assoc.*) showed that if the random effects to account for a lack of fit of a regression model can be dispensed with through a statistical test, then the model parameters and the small area means can be estimated with substantially higher accuracy. The work of Datta et al. (2011) is most useful when the number of small areas, m, is moderately large. For large m, the null hypothesis of no random effects will likely be rejected. Rejection of null hypothesis is usually caused by a few large residuals signifying a departure of the direct estimator (Y_i) from the synthetic regression estimator. As a flexible alternative to the Fay-Herriot random effects model and the approach in Datta et al. (2011), in this paper we consider a mixture model for random effects. It is reasonably expected that small areas with population means explained adequately by covariates have little model error, and the other areas with means not adequately explained by covariates will require a random component added to the regression model. This model is a flexible alternative to the usual random effects model and the data determine the extent of lack of fit of the regression model for a particular small area, and include a random effect if needed. Unlike the Datta et al. (2011) approach which recommends excluding random effects from all small areas if a test of null hypothesis of no random effects is not rejected, the present model is more flexible. We used this mixture model to estimate poverty ratios for 5- to 17-year old related children for the 50 U.S. states and Washington, DC. This application is motivated by the SAIPE project of the US Census Bureau. We empirically evaluated the accuracy of the direct estimates and the estimates obtained

[1]Department of Statistics, University of Georgia, Athens, GA 30602, USA. email:gauri@stat.uga.edu

[2]US Census Bureau

[3]Disclaimer: This report is released to inform interested parties of research and to encourage discussion of work in progress. The views expressed are those of the authors and not necessarily those of the Census Bureau

[4]Department of Statistics, University of Georgia, Athens, GA 30602, USA. email:amandal@stat.uga.edu

from our mixture model and the Fay-Herriot random effects model. These empirical evaluations and a simulation study, in conjunction with a lower posterior variance of the new estimates, show that the new estimates are more accurate than both the frequentist and the Bayes estimates resulting from the standard Fay-Herriot model.

Keywords: Empirical best linear unbiased prediction, Fay-Herriot model, finite mixture models, hierarchical Bayes, mixed effects, SAIPE project.

1 Introduction

Sample surveys are indispensable to estimate various characteristics of a population of interest. Reliability of estimates from sample surveys depends on the size of the sample. While government agencies need to estimate, for example, income and unemployment rates for the entire nation, they are also required to estimate these same characteristics for various geographic and demographic subdomains. Even though a survey may select a large sample to produce an estimate with desired accuracy for the whole population, the subsamples it allocates to various subdomains, also known as small areas, may be rather small to produce reliable direct estimates of small area characteristics based on individual subdomain samples.

Small area estimation methodology provides essential tools to statistical agencies for production of reliable small area estimates. Since a design-based direct estimate of a small area characteristic based on the area sample is usually less reliable, model-based approach to small area estimation has become very popular to produce indirect estimates. Importance and usefulness of model-based small area estimation may be assessed from an explosive growth of research publications. For a recent comprehensive review of the small area estimation literature, one may refer to Rao (2003), Jiang and Lahiri (2006), and Pfeffermann (2013).

The landmark paper by Fay and Herriot (1979) has popularized the model-based small area estimation in government statistics. Model-based small area estimation methods rely on appropriate regression models connecting direct estimates to suitable auxiliary variables that are available from other surveys and administrative records. Fay and Herriot (1979) introduced an area-level model to develop model-based estimates of per capita income for small places in the United States with population less than 1000. Denoting the design-based direct survey estimator of the ith small area characteristic by Y_i and the corresponding area-level auxiliary variable by x_i, a $q \times 1$ vector, Fay and Herriot (1979) introduced the model

$$Y_i = \theta_i + e_i, \quad \theta_i = x_i^T \beta + v_i, \quad i = 1, \ldots, m, \tag{1}$$

where θ_i is a summary measure of the characteristic to be estimated for the ith small area, e_i is the sampling error with the estimator Y_i, and the random effects v_i denotes the model error measuring the departure of θ_i from its multiple linear regression on x_i, $x_i^T\beta$. We assume that $e_1, \ldots, e_m$ are independent and normally distributed with with $e_i \sim N(0, D_i)$, and are independent of $v_1, \ldots, v_m$, which are i.i.d. $N(0, \sigma_v^2)$. The sampling variances D_i's are treated as known, but the model parameters β and σ_v^2 are unknown. Random effects v_i's are also known as small area effects. The model (1) was also introduced independently by Pfeffermann and Nathan (1981).

In model-based small area estimation, the goal is to develop an optimal predictor of θ_i and a measure of error associated with this predictor. In small area estimation, both the frequentist and Bayesian approaches have been extensively used. In the frequentist approach based on the model given by (1), while Fay and Herriot (1979) developed an empirical Bayes (EB) predictor for θ_i, Prasad and Rao (1990) and Lahiri and Rao (1995) considered an empirical best linear unbiased predictor (EBLUP) of θ_i. For the normal mixed linear model in (1), EB predictor and EBLUP predictor of θ_i are identical. Prasad and Rao (1990), Lahiri and Rao (1995), Datta and Lahiri (2000) and Datta et al. (2005) accurately approximated the mean squared error (MSE) of EBLUP of θ_i by ignoring all terms of order $o(m^{-1})$. These authors also provided estimators of the MSE that are approximately unbiased to the order $o(m^{-1})$.

In the Bayesian approach to predict θ_i based on model in (1), a hierarchical Bayes (HB) model, given below, is completed by assigning a prior distribution on the model parameters $\psi = (\beta^T, \sigma_v^2)^T$.

Model H_1:

1. Conditional on $\theta_1, \ldots, \theta_m$ and ψ, direct estimators $Y_i \sim N(\theta_i, D_i)$ for $i = 1, \ldots, m$, independently.
2. Conditional on model parameters ψ, small area means $\theta_i \sim N(x_i^T\beta, \sigma_v^2)$ for $i = 1, \ldots, m$, independently.
3. Model parameters ψ are given a prior distribution with density $\pi(\psi)$.

The above HB model has been considered, for example, by Berger (1985), Ghosh (1992), Datta et al. (2005), among others.

The success of developing reliable model-based small area estimates depends very much on the availability of good covariates that relate strongly with the small area means θ_i. In the presence of good covariates it is expected that random small area effects v_i in (1) will have small variation leading to significant shrinkage of the direct estimator Y_i to the synthetic regression estimator.

In a recent article, Datta et al. (2011) demonstrated that in the presence of good covariates x_i, the variability of the small area means θ_i may be accounted well by x_i, and the inclusion of a random effects term v_i in the model (1) may be unnecessary. They argued that while the random effects may improve the adaptivity and flexibility of the Fay-Herriot model, it also increases the variability of both point and interval estimators of small area means. In an effort to increase the accuracy of the small area estimators, these authors tested a hypothesis of no random effects in the small area model. If the null hypothesis is not rejected, these authors proposed more accurate synthetic estimators to estimate the small area means. They considered a discrepancy statistic measuring the lack of fit of the proposed multiple linear regression model of θ_i on the covariates x_i as a test statistic. The discrepancy function is constructed by combining certain deviance measures for all the small areas measuring the extent of lack of fit of the proposed model for each small area.

In contrast with the Fay-Herriot model which includes for each small area a random effects term to the regression function $x_i^T\beta$ of θ_i, the approach of Datta et al. (2011) advocates excluding the small area effects for all the small areas in the event of a non-significant discrepancy statistic. On the other hand, when a significant discrepancy statistic is realized, it is often the case that the bigger value is due to large deviance from a small number of domains for which the regression model fails to be adequate. To develop a model reflecting this scenario, and as a middle ground to the Datta et al. (2011) approach and Fay-Herriot model, we propose a "spike and slab" distribution for the random small area effects. In this formulation, with certain positive probability $(1-p)$, the random effects is assumed to be absent (that is, a degenerate distribution at zero) for any small area, and with probability p the random effects has a non-degenerate distribution. In Section 2 we introduce this new HB model and compare this model with the Fay-Herriot model. This new model allows more flexible data-dependent shrinkage of the direct estimator to the synthetic estimator. To our knowledge, this type of model has not been explored in small area estimation yet. In Section 3 we explore the propriety of posterior distribution of our HB model that results from the improper prior distribution on the model parameters. As Bayes predictors of small area means for our proposed model do not possess closed-form expressions, we perform computing by Gibbs sampling. We provide the set of required full conditional distributions.

In Section 4, we consider an application of our model to the estimation of poverty ratios of school-age children for the U. S. states. We create a map of the poverty ratios based on the direct estimates and the HB estimates from the proposed model. Our proposed model uses a mixture of a degenerate distribution and a normal distribution for the random effects, admitting positive moments of all order. Lahiri and Rao (1995) relaxed the normality of the random effects in the Fay-Herriot model by assuming existence of higher order moments to estimate the mean squared error of the EBLUPs of small area means. For our application in Section 4, we present in Section 5 an external evaluation of the proposed estimator based on 2000 census data. We also compare the posterior variances of our estimators with the second-order unbiased estimators of the mean squared error of EBLUP, proposed by Lahiri and Rao (1995) for the non-normal random effects Fay-Herriot model. In Section 6, we conducted a simulation study to compare our estimators with the other standard estimators in terms of their measures of uncertainty and empirical measurers of accuracy. The simulation results show superiority of our proposed model. The paper concludes with a brief summary in Section 7.

2 A model with uncertain random effects

As in the Fay-Herriot model, in our proposed model we assume that the direct estimator Y_i is an unbiased estimator of the small area mean θ_i admitting a normal distribution with variance D_i, for $i = 1, \ldots, m$. The HB representation of our model is given below.

Model H_M:

I. Conditional on $\theta_1, \ldots, \theta_m$, $\delta_1, \ldots, \delta_m$, $v_1, \ldots, v_m$, p, β_M and σ^2_{vM}, direct estimators $Y_i \sim N(\theta_i, D_i)$ for $i = 1, \ldots, m$, independently.

II. Conditional on $\delta_1, \ldots, \delta_m$, $v_1, \ldots, v_m$, p, β_M and σ^2_{vM}, θ_i is given by

$$\theta_i = x_i^T \beta_M + \delta_i v_i, \quad i = 1, \ldots, m, \tag{2}$$

where $\delta_1, \ldots, \delta_m$ are independent and identically distributed with

$$P(\delta_i = 1) = p = 1 - P(\delta_i = 0).$$

Conditional on $\delta_1, \ldots, \delta_m$ and σ^2_{vM}, random effects $v_1, \ldots, v_m$ are independently distributed with $v_i = 0$ when $\delta_i = 0$, and conditional on $\delta_i = 1$, $v_i \sim N(0, \sigma^2_{vM})$ for $i = 1, \ldots, m$, independently.

III. Apriori the hyperparameters β_M, σ^2_{vM} and p are independently distributed with prior density

$$\pi(\beta_M, \sigma^2_{vM}, p) = \pi_{\sigma^2_{vM}}(\sigma^2_{vM})\pi_p(p),$$

where we assign a standard improper uniform prior distribution on the regression parameter β_M, and proper densities $\pi_{\sigma^2_{vM}}(\sigma^2_{vM})$ and $\pi_p(p)$ on σ^2_{vM} and p.

While a proper prior distribution on a bounded parameter p is usually suggested, a proper prior distribution on the unbounded model variance parameter σ^2_{vM} is mandatory, for reasons explained below.

Based on stages (I) and (II) of the hierarchical model H_M, after integrating out the $v_1, \ldots, v_m$, and $\delta_1, \ldots, \delta_m$ from the joint density of $(Y_i, v_i, \delta_i), i = 1 \ldots, m$, we get, conditional on β_M,

σ^2_{vM} and p, that $Y_1, \ldots, Y_m$ are independently distributed, with Y_i having a two-component mixture of normal distributions given by

$$f(y_i|\beta_M, \sigma^2_{vM}, p) = \frac{1}{\sqrt{2\pi}} \left[\frac{(1-p)}{\sqrt{D_i}} e^{-\frac{(y_i - x_i^T\beta_M)^2}{2D_i}} + \frac{p}{\sqrt{D_i + \sigma^2_{vM}}} e^{-\frac{(y_i - x_i^T\beta_M)^2}{2(D_i + \sigma^2_{vM})}} \right]. \quad (3)$$

For the model H_M, we use the subscript M to emphasize the mixture part of the model. Also, we use the subscript M to the model parameters β_M and σ^2_{vM} to distinguish them from the model parameters in H_1. From (3) we note that the first component of the mixture in the density of Y_i does not include σ^2_{vM}. This necessitates a proper prior distribution for σ^2_{vM} so that the HB model admits a proper posterior distribution. For a discussion on this point we refer the reader to Scott and Berger (2006).

The prior distribution of v_i in (2) of Model H_M assigns a positive mass $(1-p)$ at 0 and spreads the remaining mass p according to a normal distribution with mean zero and variance σ^2_{vM}. Ishwaran and Rao (2005) used this type of prior distribution, which is known in the literature as a "slab and spike" prior, in gene selection based on microarray data. Scott and Berger (2006) also recommended this prior in multiple hypothesis testing to determine the activated (or expressed) genes in microarray experiments. In their setup in the microarray context, since only a small fraction (p) of genes are expected to be activated, they assigned a prior distribution on p which is concentrated near zero. We are adopting this model in the small area estimation context to modify the prior distribution for the random effects v_i in the Fay-Herriot model in (1). While for some small areas there may be a need to add a non-degenerate random effects to account for a lack of fit of the regression of θ_i on x_i, it is unlikely that all small areas will need non-degenerate random effects. In the spirit of Scott and Berger (2006), we generalize the Fay-Herriot model to specify the more flexible hierarchical model H_M. In fact, in the special case of $p = 1$, our model H_M encompasses the Fay-Herriot model.

We estimate the small area mean θ_i based on the HB model H_M by posterior mean of θ_i, $\hat{\theta}_{iM}$ (say), and measure the accuracy of the estimator by the posterior variance of θ_i, V_{iM} (say). While the numerical computation of these estimators will be performed by Gibbs sampling, to compare the estimators $\hat{\theta}_{iM}$ with the estimator of θ_i under the Fay-Herriot model, we first consider $\tilde{\theta}_{iM}(\beta_M, \sigma^2_{vM}, p, y_i)$, the conditional mean of θ_i, conditional on the model parameters β_M, σ^2_{vM}, p and data $y_1, \ldots, y_m$. By direct calculation,

$$\begin{aligned} \tilde{\theta}_{iM}(\beta_M, \sigma^2_{vM}, p, y_i) &= x_i^T\beta_M + \frac{\sigma^2_{vM}}{\sigma^2_{vM} + D_i}(y_i - x_i^T\beta_M)\tilde{p}_i(\beta_M, \sigma^2_{vM}, p, y_i) \\ &= y_i - B_{iM}(y_i - x_i^T\beta_M), \end{aligned} \quad (4)$$

where the direct estimate y_i of θ_i is shrunk to the regression (or synthetic) "estimate" $x_i^T\beta_M$; the extent of the shrinkage depends on the coefficient

$$B_{iM} = 1 - \frac{\sigma_{vM}^2}{\sigma_{vM}^2 + D_i}\tilde{p}_i(\beta_M, \sigma_{vM}^2, p, y_i), \tag{5}$$

with $\tilde{p}_i(\beta_M, \sigma_{vM}^2, p, y_i) = P(\delta_i = 1|\beta_M, \sigma_{vM}^2, p, y_i)$ is given by

$$\tilde{p}_i(\beta_M, \sigma_{vM}^2, p, y_i) = \frac{p}{p + (1-p)\sqrt{\frac{\sigma_{vM}^2 + D_i}{D_i}}\exp\left\{-\frac{1}{2}\frac{(y_i - x_i^T\beta_M)^2\sigma_{vM}^2}{D_i(D_i+\sigma_{vM}^2)}\right\}}. \tag{6}$$

Note that $\tilde{p}_i$ is increasing in p and $|y_i - x_i^T\beta_M|$, and B_{iM} is decreasing in σ_{vM}^2 for fixed $\tilde{p}_i$, and is decreasing in $\tilde{p}_i$ for fixed σ_{vM}^2. Thus the shrinkage coefficient B_{iM} will be large if either p, σ_{vM}^2 or the "residual" $|y_i - x_i^T\beta_M|$ is small, all of which make intuitive sense. While the shrinkage of the direct estimator for the proposed model is influenced by "residual" $|y_i - x_i^T\beta_M|$ (that is, the fit of the regression of the direct estimator on the covariate x_i), this is not the case for the Fay-Herriot model, for which the shrinkage coefficient B_{iFH} is given by

$$B_{iFH} = \frac{D_i}{\sigma_v^2 + D_i} = 1 - \frac{\sigma_v^2}{\sigma_v^2 + D_i}, \tag{7}$$

where, for known model parameters β and σ_v^2, the Bayes predictor of θ_i is given by

$$\tilde{\theta}_{iFH} = y_i - B_{iFH}(y_i - x_i^T\beta). \tag{8}$$

Note that while the shrinkage coefficient in the Fay-Herriot model depends only on the ratio $\frac{D_i}{\sigma_v^2}$, but not on the residual $|y_i - x_i^T\beta|$, for the proposed mixture model the shrinkage depends on both these quantities and in turn it achieves a greater flexibility.

Datta et al. (2011) used an estimated version of the squared standardized residual $(y_i - x_i^T\beta)^2/D_i$ as the deviance of the ith small area in the discrepancy statistic which they proposed to test the absence of a random effect. While a fraction of small areas may contribute large deviances to the discrepancy statistic pushing it to be significant, suggesting a positive σ_v^2 and thereby reducing the shrinkage for all small areas, the shrinkage coefficient B_{iM} for the proposed model is more robust to large deviances from other small areas. It is definitely desirable to borrow more from the regression model by shrinking a less reliable direct estimator to a regression synthetic estimator.

3 Computational issues and the posterior distribution

The HB mixture model H_M proposed to estimate small area means θ_i, for $i = 1, \ldots, m$, is rather complex. We will implement Gibbs sampling (cf. Gelfand and Smith (1990)) to compute the posterior mean and posterior variance of θ_i. Gibbs sampling is an MCMC method that generates samples from a joint distribution by iteratively sampling from the various conditional distributions associated with this joint distribution.

We discussed earlier that we need a proper prior distribution on the variance parameter σ^2_{vM} since it does not appear in all components of the mixture distribution. We will use a proper inverse gamma distribution as the prior with a corresponding density

$$\pi_{\sigma^2_{vM}}(\sigma^2_{vM}) \propto \exp\left\{-\frac{a}{\sigma^2_{vM}}\right\}(\sigma^2_{vM})^{-(b+1)}, \tag{9}$$

where both the rate parameter $a > 0$ and the shape parameter $b > 0$ are assumed to be known. Here the mixing parameter p is the apriori probability that a random effect is non-degenerate. Our experience with small area estimation applications suggests that in the presence of good covariates nearly half of the small areas will not need any random effect term in the Fay-Herriot model. Based on such consideration we propose a beta distribution with parameters c, d, with $0 < c < d$. We use suitable values for a, b, c and d in our data analysis.

To derive the full conditional distributions needed for Gibbs sampling we start with the joint posterior density of β_M, $v = (v_1, \ldots, v_m)^T$, $\delta = (\delta_1, \ldots, \delta_m)^T$, p and σ^2_v. Based on the hierarchical model H_M and the prior densities for σ^2_{vM}, p and β_M, we get

$$\begin{aligned}
&\pi(v_1, \ldots, v_m, \delta_1, \ldots, \delta_m, \beta_M, p, \sigma^2_{vM}|y_1, \ldots, y_m) \\
&\propto \exp\left[-\frac{1}{2}\sum_{i=1}^m \frac{(y_i - x_i^T\beta_M - \delta_i v_i)^2}{D_i}\right]\prod_{i=1}^m\left\{(\sigma^2_{vM})^{-\frac{1}{2}}\exp\left(-\frac{v_i^2}{2\sigma^2_{vM}}\right)\right\}^{\delta_i}\{I(v_i = 0)\}^{1-\delta_i} \\
&\quad p^{\sum_{i=1}^m \delta_i}(1-p)^{m-\sum_{i=1}^m \delta_i}(\sigma^2_{vM})^{-(b+1)}\exp\left\{-\frac{a}{\sigma^2_{vM}}\right\}p^{c-1}(1-p)^{d-1}.
\end{aligned} \tag{10}$$

Let $y = (y_1, \ldots, y_m)^T$, $D = diag(D_1, \ldots, D_m)$ and $X = (x_1, \ldots, x_m)^T$. We assume that $rank(X) = q$. For an m-dimensional vector $\zeta = (\zeta_1, \ldots, \zeta_m)^T$, by $\zeta_{(-i)}$ we denote the $(m-1)$-dimensional vector $(\zeta_1, \ldots, \zeta_{i-1}, \zeta_{i+1}, \ldots, \zeta_m)^T$. From (10) it follows that

1. $\beta_M|v_1, \ldots, v_m, \delta_1, \ldots, \delta_m, p, \sigma^2_{vM}, y \sim N\Big((X^TD^{-1}X)^{-1}X^TD^{-1}(y-\delta\cdot v), (X^TD^{-1}X)^{-1}\Big)$ where $\delta \cdot v = (\delta_1 v_1, \ldots, \delta_m v_m)^T$;

2. $v_i|v_{(-i)}, \delta, \beta_M, p, \sigma^2_{vM}, y$ is degenerate at zero if $\delta_i = 0$, and if $\delta_i = 1$, then

$$v_i|v_{(-i)}, \delta_i = 1, \delta_{(-i)}, \beta_M, p, \sigma^2_{vM}, y \sim N\left(\frac{\sigma^2_{vM}}{\sigma^2_{vM} + D_i}(y_i - x_i^T\beta_M), \frac{\sigma^2_{vM} D_i}{\sigma^2_{vM} + D_i}\right);$$

3. $P[\delta_i = 1|v, \delta_{(-i)}, \beta_M, p, \sigma^2_{vM}, y] = \tilde{p}_i(\beta_M, \sigma^2_{vM}, p, y_i)$ (see equation (6));

4. $p|v, \delta, \beta_M, \sigma^2_{vM}, y \sim Beta(c + \sum_{i=1}^m \delta_i, d + m - \sum_{i=1}^m \delta_i)$;

5. $\sigma^2_{vM}|v, \delta, \beta_M, p, y$ follows inverse gamma distribution with shape parameter $b + \frac{1}{2}\sum_{i=1}^m \delta_i$ and rate parameter $a + \frac{1}{2}\sum_{i=1}^m \delta_i v_i^2$.

Since an improper prior density has been used for β_M, we investigate below the condition under which the posterior density is proper.

Theorem 3.1. *The posterior distribution corresponding to the HB model H_M is proper, that is, the posterior density is integrable, if and only if $m \geq q$.*

Proof: If Part: To prove the propriety of the posterior distribution we need to show the integrability of the right hand side of (10) with respect to β_M, $v_1, \ldots, v_m$, $\delta_1, \ldots, \delta_m$, p and σ^2_{vM}. Integrating the right hand side of (10) with respect to $v_1, \ldots, v_m$, we get

$$\begin{aligned}\pi(\beta_M, \delta_1, \ldots, \delta_m, p, \sigma^2_{vM}|y_1, \ldots, y_m) &= K\prod_{i=1}^m (D_i + \sigma^2_{vM}\delta_i)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\sum_{i=1}^m \frac{(y_i - x_i^T\beta_M)^2}{D_i + \sigma^2_{vM}\delta_i}\right] \\ &\quad \times \pi_{\sigma^2_{vM}}(\sigma^2_{vM})\pi_p(p) \times p^{\sum_{i=1}^m \delta_i}(1-p)^{m - \sum_{i=1}^m \delta_i}. \end{aligned} \tag{11}$$

Here K is a generic positive constant. Let $\Sigma = diag(D_1 + \delta_1\sigma^2_{vM}, \ldots, D_m + \delta_m\sigma^2_{vM})$. Since $rank(X) = q$, $X^T\Sigma^{-1}X$ is positive definite. Integrating out β_M from (11) and using $p^{\sum_{i=1}^m \delta_i}(1-p)^{m-\sum_{i=1}^m \delta_i} \leq 1$, we have

$$\begin{aligned}&\pi(\delta_1, \ldots, \delta_m, p, \sigma^2_{vM}|y_1, \ldots, y_m) \\ &\quad \leq K\prod_{i=1}^m (D_i + \sigma^2_{vM}\delta_i)^{-\frac{1}{2}} |X^T\Sigma^{-1}X|^{-\frac{1}{2}} \pi_{\sigma^2_{vM}}(\sigma^2_{vM})\pi_p(p) \\ &\quad\quad \times \exp\{-\frac{1}{2}y^T(\Sigma^{-1} - \Sigma^{-1}X(X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1})y\} \\ &\quad \leq K(D_{(1)} + \sigma^2_{vM})^{-\frac{1}{2}\sum_{i=1}^m \delta_i} |X^T\Sigma^{-1}X|^{-\frac{1}{2}} \pi_{\sigma^2_{vM}}(\sigma^2_{vM})\pi_p(p), \end{aligned} \tag{12}$$

where $D_{(1)} = \min_{1\leq i\leq m} D_i > 0$. Let $D_{(m)} = \max_{1\leq i\leq m} D_i > 0$ and $\Delta = diag(D_{(m)} + \delta_1\sigma^2_{vM}, \ldots, D_{(m)} + \delta_m\sigma^2_{vM})$. Suppose for $0 \leq m_1 \leq m$, we have $\delta_{i_1} = \ldots = \delta_{i_{m_1}} = 1$. Corresponding to the indices $i_1, \ldots, i_{m_1}$, we define the matrix $X_1 = (x_{i_1}, \ldots, x_{i_{m_1}})^T$. We also define the matrix X_2 which is formed by the remaining rows of X. If $m_1 = 0$, then X_1 is vacuous and X_2 is the same as X. We also define $rank(X_1) = r_1$ and $rank(X_2) = r_2$. Note that $0 \le r_1 \le m_1$, $0 \le r_2 \le m - m_1$, and $r_1 + r_2 \ge q$. With this notation, we find that

$$X^T \Sigma^{-1} X \quad \ge \quad (D_{(m)} + \sigma^2_{vM})^{-1} X_1^T X_1 + D^{-1}_{(m)} X_2^T X_2. \tag{13}$$

If X_1 is vacuous or if X_1 is a null matrix, then $X^T \Sigma^{-1} X \ge D^{-1}_{(m)} X_2^T X_2$ with $X_2^T X_2$ being positive definite. Otherwise, (13) holds. In call cases, consequently $|X^T \Sigma^{-1} X| \ge K$, where K is a positive number. If X_1 is not a null matrix, r_1 will be positive. Let $\lambda_1 \ge \ldots \ge \lambda_{r_1} > 0$ be the non-null eigenvalues of $X_1^T X_1$ and $\mu_1 \ge \ldots \ge \mu_{r_2} > 0$ be the non-null eigenvalues of $X_2^T X_2$. We are considering the case $r_2 > 0$, as the case $r_2 = 0$ is trivial. Let $u = (D_{(m)} + \sigma^2_{vM})^{-1}$, $w = D^{-1}_{(m)}$ and $G = (D_{(m)} + \sigma^2_{vM})^{-1} X_1^T X_1 + D^{-1}_{(m)} X_2^T X_2 = u X_1^T X_1 + w X_2^T X_2$. Since $X_1^T X_1$ has r_1 positive eigenvalues, there is an orthogonal matrix P of dimension $q \times q$ such that $P^T X_1^T X_1 P = diag(\lambda_1, \ldots, \lambda_{r_1}, 0, \ldots, 0)$. If we partition the matrix $w P^T X_2^T X_2 P$ as $((B_{kl}))$ where B_{11} is $r_1 \times r_1$ and B_{22} is $(q - r_1) \times (q - r_1)$, then $P^T G P$ is identical to $w P^T X_2^T X_2 P$ except for the leading $r_1 \times r_1$ submatrix given by $B_{11} + u\, diag(\lambda_1, \ldots, \lambda_{r_1}, 0, \ldots, 0)$. Positive definiteness of $P^T G P$ implies that B_{22} is positive definite and $|G| = |B_{22}||B_{11} + u\, diag(\lambda_1, \ldots, \lambda_{r_1}) - B_{12} B_{22}^{-1} B_{21}|$. From this it follows that the determinant of G is a polynomial in u of degree t, where $t \le r_1$. Note that u lies between 0 and $D^{-1}_{(m)}$. Consequently, by (13) we claim that for u near zero, $|\Sigma|^{-\frac{1}{2}} |X^T \Sigma^{-1} X|^{-\frac{1}{2}} \sim O(u^{\frac{m_1 - t}{2}})$. Since $t \le r_1 \le m_1$, we conclude that $|\Sigma|^{-\frac{1}{2}} |X^T \Sigma^{-1} X|^{-\frac{1}{2}}$ is a bounded function of u near zero, and hence it is bounded function of u for $u \in (0, D^{-1}_{(m)})$. Since u is a one-to-one function of σ^2_{vM} from $(0, D^{-1}_{(m)})$ to $(0, \infty)$, we conclude that $|\Sigma|^{-\frac{1}{2}} |X^T \Sigma^{-1} X|^{-\frac{1}{2}}$ is a bounded function of σ^2_{vM}. Hence, $(D_{(1)} + \sigma^2_{vM})^{-\frac{1}{2} m_1} |X^T \Sigma^{-1} X|^{-\frac{1}{2}}$ is also a bounded function of σ^2_{vM}. Consequently, from (12) we get that

$$\pi(\delta_1, \ldots, \delta_m, p, \sigma^2_{vM} | y_1, \ldots, y_m) \quad \le \quad K \pi_{\sigma^2_{vM}}(\sigma^2_{vM}) \pi_p(p).$$

Integrating out both p and σ^2_{vM} we get from the last inequality that $\pi(\delta_1, \ldots, \delta_m | y_1, \ldots, y_m) \le K$. Finally, $\sum_\delta \pi(\delta_1, \ldots, \delta_m, | y_1, \ldots, y_m) \le 2^m K < \infty$. This completes the proof of the propriety of the posterior density.

Only If Part: Integrability of the posterior density implies that for each $\delta_1, \ldots, \delta_m$, the function $\pi(\beta_M, \sigma^2_{vM}, p, v_1, \ldots, v_m, \delta_1, \ldots, \delta_m | y)$ is integrable. In particular, corresponding to $\delta_1 = \ldots = \delta_m = 0$, the function $\prod_{i=1}^m e^{-\frac{1}{2D_i}(y_i - x_i^T \beta_M)^2}\ \pi(\sigma^2_{vM}) \pi_p(p)$ needs to be integrable. This implies $\prod_{i=1}^m e^{-\frac{1}{2D_i}(y_i - x_i^T \beta_M)^2}$ needs to be integrable with respect to β_M. A simple argument will show that the set $\{x_1, \ldots, x_m\}$ must contain q linearly independent vectors, that is, $m \ge q$. □

4 Estimation of child poverty ratio

The United States Census Bureau as part of their Small Area Income and Poverty Estimation (SAIPE) program annually provides estimates of poverty measure for various age groups at the state and county level. One measure of particular interest is the poverty ratio for the school going related children for the 5− to 17−year old group, used by the United States Department of Education to implement the No Child Left Behind program. To provide an application of our method we consider estimation of state-level poverty ratio for this demographic group based on the Current Population Survey (CPS) for the year 1999. Direct estimate y_i for the ith state computed from the CPS is usually subject to large sampling error due to small sample size. To develop more accurate estimate of this characteristic the Census Bureau suggested combining the direct estimate with other administrative data through a suitable regression model. The Census Bureau explored using the Internal Revenue Service (IRS) data measuring poverty ratio based on the number of child exemptions (covariate 1) and IRS non-filer rate (covariate 2) which are found to be correlated with the direct CPS estimate of poverty rate y. The Census Bureau also found out that the residual by fitting a model for the 1989 census poverty data on these covariates for the year 1989 has good predictive power in predicting the prevailing child poverty ratio in 1999. In our application we consider the year 1999 since the poverty ratio obtained from the 2000 census data which collects poverty information for the income year 1999 can be used as a benchmark to compare accuracy of our proposed estimates and other various estimates available in the literature.

The discrepancy statistic Datta et al. (2011) suggested to test the absence of a random effects term in the Fay-Herriot model in (1) is given by $T = \sum_{i=1}^{m} \frac{(y_i - x_i^T \hat{\beta}_{WLS})^2}{D_i}$ where $\hat{\beta}_{WLS} = (X^T D^{-1} X)^{-1} X^T D^{-1} y$, and $(y_i - x_i^T \hat{\beta}_{WLS})^2 / D_i$ is the deviance of the ith small area contributed to the discrepancy statistic. For our application, the value of T is 67.72 which was significant with a p-value of 0.025, computed based on chi-square distribution with 47 degrees of freedom. From Figure 1 we note that only 9 or 10 (about 20%) of the 51 deviances have values larger than 4, the largest one being 14.2, corresponding to Massachusetts. While this test for the entire data suggests inclusion of the random effects term in the Fay-Herriot model, the result is dramatically different if we refit the model excluding Massachusetts. The new discrepancy statistic is quite small with a non-significant p-value of 0.284, and most of the individual deviances are smaller than 3.

An exploratory analysis of the poverty data presented above suggests adaptive shrinkage of the direct estimates to the synthetic regression-based estimates. We estimate the poverty ratio θ_i by applying the HB Fay-Herriot model and the proposed model. For the Fay-Herriot model we use an improper uniform prior $\pi(\beta, \sigma_v^2) = 1$ to conduct the Bayesian analysis. It is well-known that the resulting posterior will be proper (cf. Berger, 1985), and we implement Bayesian computing via Gibbs sampling. For the proposed uncertain random effects model

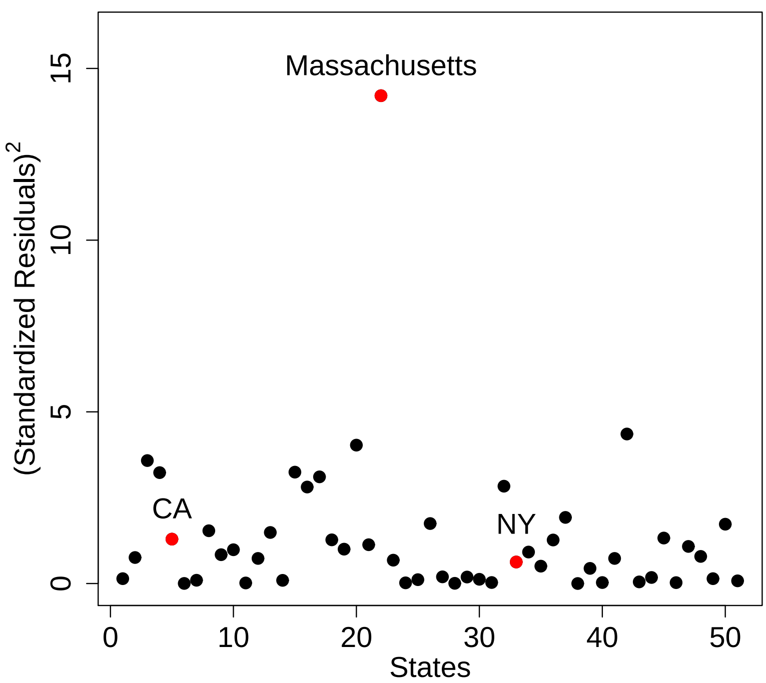

Figure 1: Plot of deviances of different states (small areas) contributed to the discrepancy statistic T of Datta et al. (2011)

(add a random effect if $\delta_i = 1$, or do not add any random effect if $\delta_i = 0$), we use a partially proper prior. It is explained earlier that a proper prior on σ^2_{vM} is necessary to ensure a proper posterior distribution. We consider an inverse gamma distribution with shape parameter 1 with a median $2\bar{D}/\log_e 2$, where $\bar{D}$ is the average sampling variance. Note that this is a flat tail prior distribution with no finite mean. We use improper uniform prior for β_M, and a proper $Beta(1,4)$ prior for p. The beta prior reflects a 20% a priori probability that a random effects to a small area is non-degenerate (it is zero if $\delta_i = 0$).

For each model we obtain a point estimate of the small area mean θ_i by the corresponding posterior mean, and measure the uncertainty of the point estimate by the posterior variance. To demonstrate the effectiveness of the proposed model we have plotted the ratio of the posterior variance of the proposed model to the posterior variance of the Fay-Herriot model in Figure 2. We find that except for Massachusetts, the ratio is less than 80% for all other states, and the ratio is less than 50% for nearly half of the states. Only for Massachusetts the posterior variance of θ_i for the proposed model is about 60% higher than the posterior variance of θ_i for the Fay-Herriot model. The average of this ratio for all the states is 0.54.

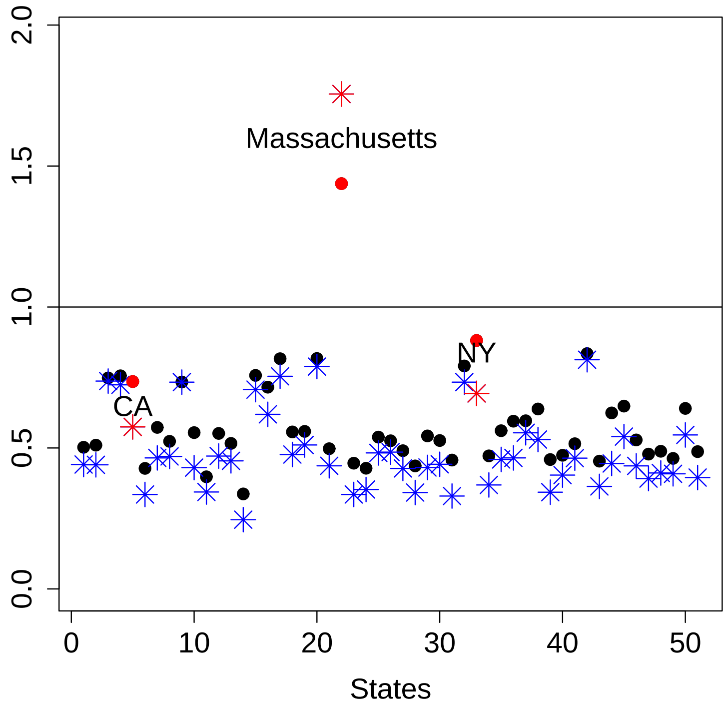

Figure 2: Plot of ratio of uncertainty measures of various model-based estimates of small area means. The solid dot represents the ratio of the posterior variance of the proposed model (2) to the posterior variance of the Fay-Herriot model and the asterisk represents the ratio of posterior variance of the proposed model (2) to the estimated MSE of the Lahiri and Rao (1995) model discussed in Section 5

We compare the shrinkage coefficients for all the small areas for the two models. For the Fay-Herriot model we use $D_i/(D_i + \hat{\sigma}^2_{v,FH\text{-}HB})$ as the estimated shrinkage for the ith small area, here $\hat{\sigma}^2_{v,FH\text{-}HB}$ is the posterior mean of σ^2_v based on the Fay-Herriot HB model (we found that this value is not much different from the average of the Gibbs sample values of $B_{i,FH}(\sigma^2_v)$). For our proposed model we estimated the shrinkage coefficient by $E(B_{iM}|y)$, where the posterior expectation is computed by averaging the values of $B_{iM}(\beta_M, \sigma^2_{vM}, p, y_i)$ over the Gibbs samples generated from this model. In Figure 3, we plot these shrinkage coefficients against the corresponding square-root deviance (which is the same as the absolute standardized residual) of the state. The plot shows that for most states shrinkage coefficients of the proposed model are much larger than the shrinkage coefficients provided by the Fay-Herriot model. This suggests that a random effects term in the Fay-Herriot model may not be needed for many areas. Indeed we found that of the fifty-one 95% prediction intervals for the random effects in the Fay-Herriot model, twenty-two include the zero value.

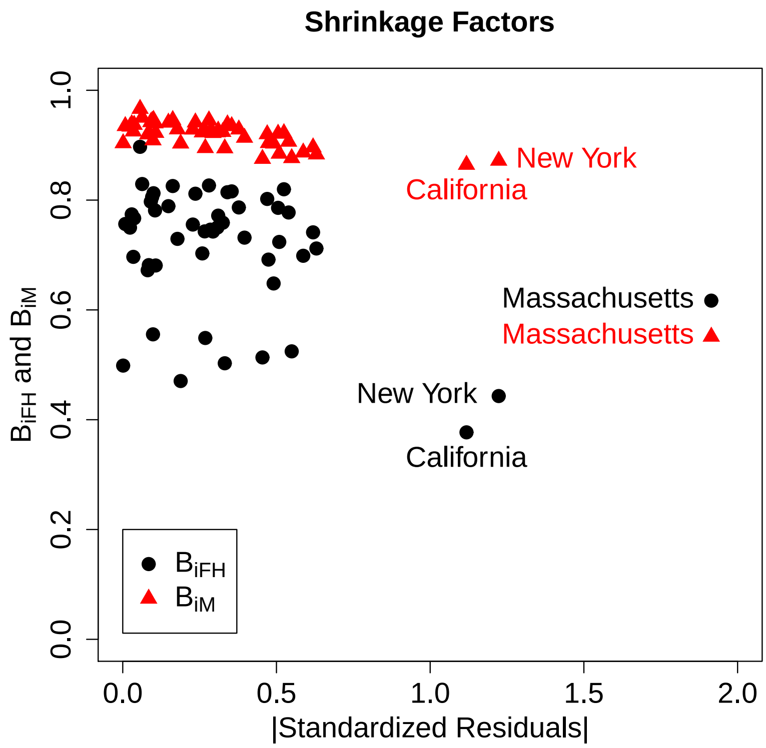

Figure 3: Plot of shrinkage coefficients of different states against the corresponding square-root deviance

For our proposed model we computed the posterior probability of $\{\delta_i = 1\}$ for all the states, which are displayed in Figure 4. Except for Massachusetts, which has this probability 0.83 (nearly 1), all other states have this probability less than 0.30. Compared to a prior odds of 1 to 4, the posterior odds of $\delta = 1$ is nearly 5 to 1 for Massachusetts, and nearly 1 to 2 (or less) for the other states. This indicates that a presence of a random effects is suggested by the data only for Massachusetts.

We noticed earlier that the direct CPS estimate for Massachusetts does not fit well the Fay-Herriot model, and our model shrinks the CPS estimate much less to the synthetic estimate than the Fay-Herriot model shrinks it to the synthetic one. We also note that while our model provides greater shrinkage for CA and NY, the Fay-Herriot model shrinks much less for these two states.

In Figure 5 we provide the posterior density of $\delta_i v_i$ for three states (Massachusetts, California, New York). Each posterior distribution has a large positive mass at $v_i = 0$. The vertical bar

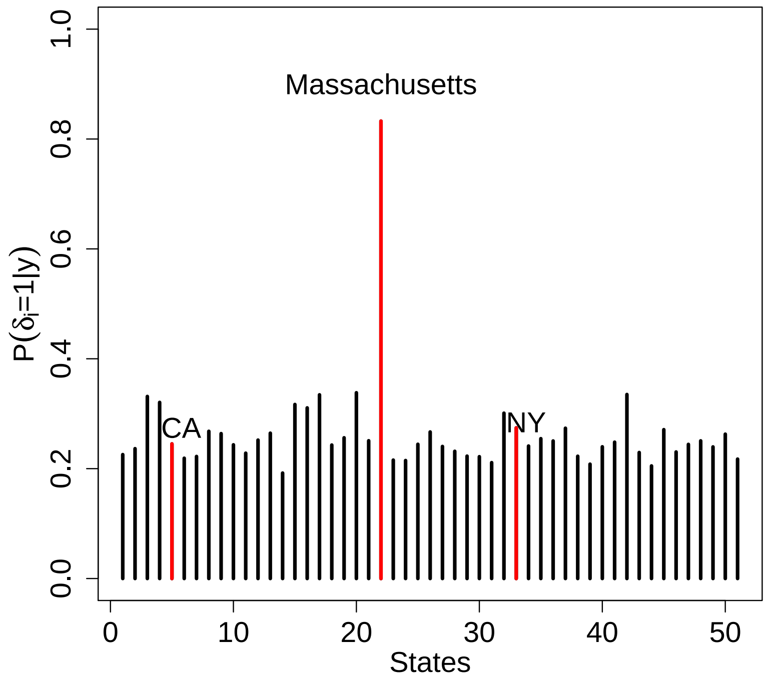

Figure 4: Plot of $P(\delta_i = 1|\text{data})$

at $v_i = 0$ corresponds to the posterior probability of $v_i = 0$ (which is the same as $P(\delta_i = 0|\text{data})$), and the smooth density displays how the remaining probability mass is distributed over non-zero values of v_i. A higher value of the vertical bar and/or a higher concentration of the density near zero will indicate substantial shrinkage of the direct estimator towards the synthetic estimator. From these plots we see that the direct estimate for Massachusetts is subject to moderate shrinkage, since the probability mass of v_i at zero is quite small (0.17) and the remaining mass of 0.83 is spread over 0 to 15 (with practically no mass below zero), indicating that the direct estimator is substantially bigger than the synthetic estimator. This last fact is in agreement with the large positive residual for Massachusetts in Figure 1. On the other hand, direct estimates for CA and NY shrink substantially towards their synthetic values under the proposed model (the respective estimated shrinkage coefficients are 0.86 and 0.87). This large amount of shrinkage is also confirmed from the last two plots in Figure 5. Each of the plots has a large positive mass (bigger than 70%) at $v_i = 0$, and the remaining mass of about 30% is spread over a relatively narrow interval around zero.

Finally, we create a poverty map to compare the CPS estimates with those obtained from

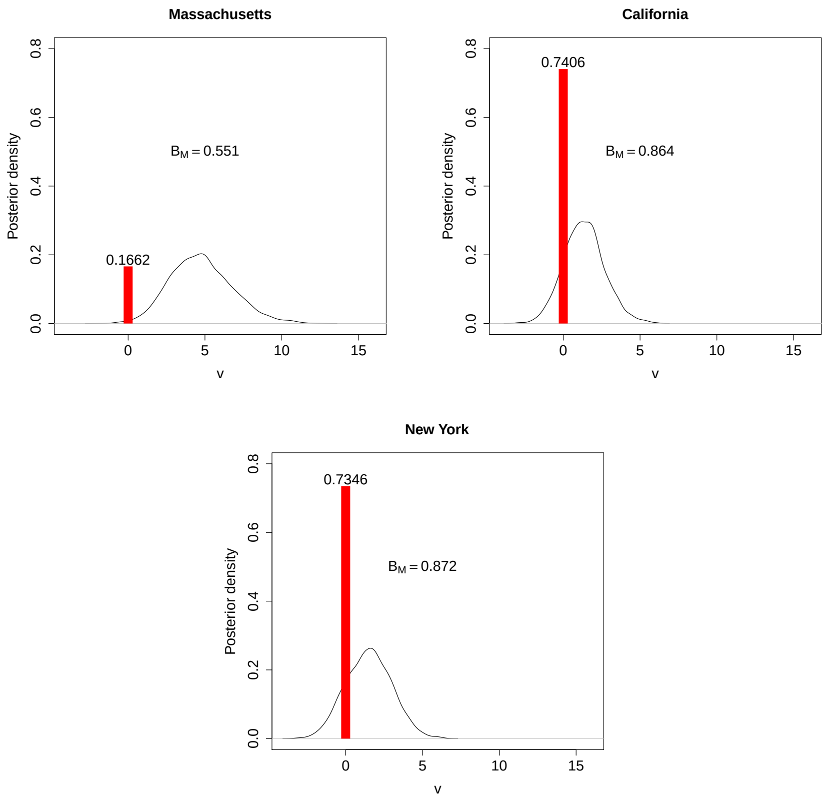

Figure 5: Histogram of the posterior distribution of $\delta_i v_i$

the Fay-Herriot model and the proposed model. These maps, created by subtracting the "true" values (described in Section 5) of the poverty measures from these estimates, show the effectiveness of the proposed model. Panel (d) in Figure 6 provides the state poverty map using the true poverty ratios based on census data. Panels (a)−(c) plot the maps of the departures of the estimates from the true poverty ratios. Panel (a) is for proposed estimates, (b) is for the Fay-Herriot estimates and (c) is for the CPS direct estimates. In panels (a)−(c), states with deviations of the estimated poverty ratios from the true ratios between −2% and 2% are shaded green, states with true poverty ratios underestimated by 2% or more are shaded with red and overestimated by 2% or more are shaded with blue. These maps empirically demonstrate that the proposed model produces more accurate estimates than the Fay-Herriot model. While the proposed model produces estimates that are right on the target for 47 states (with deviations within 2%), it overestimates the poverty ratio by more than 2% for only three states, namely, MS, LA and WV, and underestimates by more than 2% for MA alone. On the other hand, the estimates from the Fay-Herriot model miss the true state poverty ratios by more than 2% for eight states. Fay-Herriot model overestimates for five states, namely, MS, LA, SD, KY and WV, and underestimates for

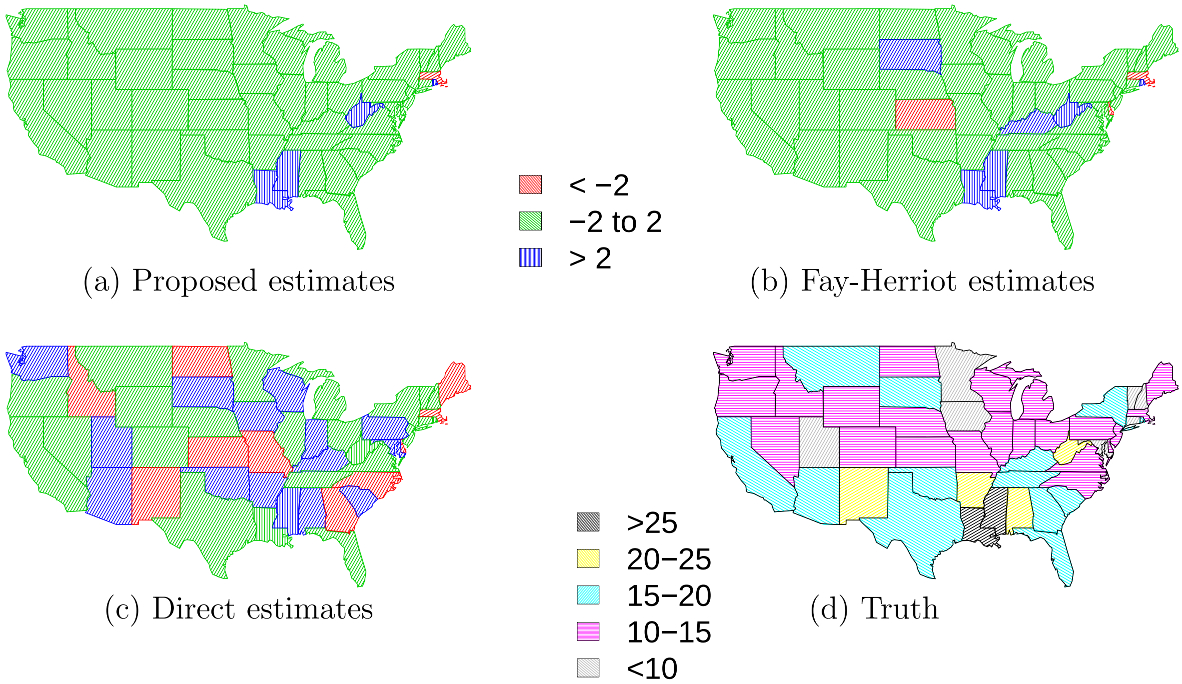

Figure 6: Poverty map: Panel (d) shows the "true poverty map", whereas panels (a), (b) and (c) show the differences of these estimates from the "truth", for the proposed model, FH model and CPS estimates, respectively.

three states, namely, KS, DE and MA. If we turn to the CPS estimates, they are much less accurate and they overestimate for 16 states and underestimate for 10 states. Additionally, smaller posterior variances associated with the estimates of the proposed model compared to the posterior variances associated with the Fay-Herriot model for all states except one justify the superiority of the proposed model.

5 External evaluation and comparison with a robust frequentist method

We evaluate performance of the proposed model by comparing the poverty ratio estimates generated by our model with the corresponding population level poverty ratio we obtained from the 2000 census for the fifty states and Washington, DC. Since the questionnaires measuring poverty in the CPS are usually different from the questionnaires used in the census, there may be a systematic difference between the two sets of numbers. To account for any such difference we compute *ratio benchmarked* census states poverty ratios by multiplying the 51 census poverty ratios by a scale R, where $R = \frac{\sum_{i=1}^{51} (\text{pop})_i y_i}{\sum_{i=1}^{51} (\text{pop})_i c_i}$, $(\text{pop})_i$ is the estimated population of the $5-17$ age group of related children in the ith state, c_i is the poverty ratio for the group obtained from the 2000 census. This adjustment should allow an "apple to apple" comparison of the estimates based on the CPS sample to the "truth" based on the 2000 census.

To compare empirically accuracy of an estimator t, we calculate the set of small area estimates $\{t_i, i = 1, \ldots, 51\}$ and we compute various deviation measures of this set of numbers from the benchmarked census numbers $\{c'_i, i = 1, \ldots, 51\}$, where $c'_i = Rc_i$. Note that $\sum_{i=1}^{51} (\text{pop})_i c'_i = \sum_{i=1}^{51} (\text{pop})_i y_i$, thus the benchmarked "true" census numbers c'_i are calibrated so that national "true" number of poor children agrees with the national CPS number of poor children, the latter is considered quite accurate since it is based on a large national sample. To evaluate effectiveness of an estimator t, we computed average absolute deviation $\text{AAD}(t) = \frac{1}{51} \sum_{i=1}^{51} |t_i - c'_i|$, average squared deviation $\text{ASD}(t) = \frac{1}{51} \sum_{i=1}^{51} (t_i - c'_i)^2$, average absolute relative deviation $\text{ARB}(t) = \frac{1}{51} \sum_{i=1}^{51} \frac{|t_i - c'_i|}{c'_i}$ and average squared relative deviation $\text{ASRB}(t) = \frac{1}{51} \sum_{i=1}^{51} \frac{(t_i - c'_i)^2}{(c'_i)^2}$ of the estimated values $\{t_i, i = 1, \ldots, 51\}$ from their corresponding "true" values $\{c'_i, i = 1, \ldots, 51\}$. We evaluated these four summary performance measures for the direct estimate (y), the FH-HB predictor $\hat{\theta}_{\text{FH-HB}}$ and the proposed HB predictor $\hat{\theta}_{\text{M-HB}}$. We display these summary measures for these three estimates in the table below. The table shows that model-based small area estimates perform much better than the direct estimates. Moreover, the lowest values of these deviation measures are realized by our proposed predictor. Figure 2, which shows that the posterior variance from the proposed model is nearly half of the posterior variance of the Fay-Herriot model, in conjunction with Table 1, establishes the superiority of our proposed model.

Table 1: Effectiveness of the proposed estimator

Estimate	AAD $=\frac{1}{51}\sum \lvert t_i - c'_i \rvert$	ASD $=\frac{1}{51}\sum (t_i - c'_i)^2$	ARB $=\frac{1}{51}\sum \frac{\lvert t_i - c'_i \rvert}{c'_i}$	ASRB $=\frac{1}{51}\sum \frac{(t_i - c'_i)^2}{(c'_i)^2}$
Direct (y)	2.718	12.291	0.196	0.067
$\hat{\theta}_{\text{FH-HB}}$	1.139	2.354	0.076	0.009
$\hat{\theta}_{\text{M-HB}}$	1.007	1.963	0.067	0.007

5.1 EBLUP prediction for non-normal small area means

It is well known that the BLUP of a small area mean based on a mixed linear model depends only on the first two moments of the random effects and the sampling error, and not on their distributional assumptions. In particular, the BLUP continues to remain valid under non-normality of random effects.

Steps I and II of the model H_M can be reexpressed as

$$Y_i = \theta_i + e_i, \quad \theta_i = x_i^T\beta + v_i^*, \quad i = 1, \ldots, m, \tag{14}$$

where $e_i, v_i^*, i = 1, \ldots, m$ are independently distributed with a common mean zero, e_i's are normally distributed with variance D_i, and v_i^*'s are identically distributed as the $v_i\delta_i$'s. Clearly, for any positive integer k, $E[(v_i^*)^{2k}] = pE(v_i^{2k}) < \infty$, that is, all the positive integral moments are finite. The random effects v_i^* are not normally distributed but they have all moments of positive order finite.

Lahiri and Rao (1995) in a pioneering paper investigated the MSE properties of the Prasad-Rao EBLUPs of small area means for a non-normal Fay-Herriot model, which includes the model in (14) as a special case. In particular, they estimated the variance components $V(v_i^*) = \sigma_v^{*2}$ (say), by the ANOVA method. If $\hat{\beta} = (X^TX)^{-1}X^TY$ denotes the ordinary least squares estimator of β, the ANOVA estimator of σ_v^{*2} is given by

$$\hat{\sigma}_v^{*2} = \max\left\{\frac{\sum_{i=1}^m \{(Y_i - x_i^T\hat{\beta})^2 - D_i(1 - h_{ii})\}}{m - q}, 0\right\}, \tag{15}$$

where $h_{ii} = x_i^T(X^TX)^{-1}x_i$.

With this estimator of σ_v^{*2}, the EBLUP of θ_i, derived by Lahiri and Rao (1995) for the model in (14) is given by

$$\hat{\theta}_{i,\text{EBL-LR}} = Y_i - B_{iFH}(\hat{\sigma}_v^{*2})\Big(Y_i - x_i^T\hat{\beta}_{\text{GLS}}(\sigma_v^{*2})\Big), \tag{16}$$

where $\hat{\beta}_{\text{GLS}}(\sigma_v^{*2})$ is the generalized least squares estimator of β, with $\hat{\beta}_{\text{GLS}}(\sigma_v^{*2})$ given by

$$\hat{\beta}_{\text{GLS}}(\sigma_v^{*2}) = \Big(X^T\Phi^{-1}(\sigma_v^{*2})X\Big)^{-1}X^T\Phi^{-1}(\sigma_v^{*2})Y. \tag{17}$$

Here $\Phi(\sigma_v^{*2}) = D + \sigma_v^{*2}I_m$, the marginal variance-covariance matrix of Y under the non-normal Fay-Herriot model given by (14).

Assuming in their non-normal random effects version of the Fay-Herriot model that the random effects v_i^* has a finite $(8+\epsilon)$-th moment, with $\epsilon > 0$, Lahiri and Rao (1995) obtained the second-order accurate approximation to the MSE of the EBLUP $\hat{\theta}_{i,\text{EBL-LR}}$, by ignoring all terms that are of the order $o(m^{-1})$. They showed that this approximate MSE is different from the MSE of the EBLUP of θ_i, derived in Prasad and Rao (1990) under the normal distribution of the random effects in the Fay-Herriot model in (1). The approximate MSE in Lahiri and Rao (1995) depends also on the fourth moment of the non-normal random effects term v_i^*. While the approximate MSEs differ for the two models, to their pleasant surprise Lahiri and Rao (1995) discovered that the second-order unbiased estimator of the MSE, derived in (5.15) of Prasad and Rao (1990) and given by

$$mse(\hat{\theta}_{i,\text{EBL-LR}}) = g_{1i}(\hat{\sigma}_v^{*2}) + g_{2i}(\hat{\sigma}_v^{*2}) + 2g_{3i}(\hat{\sigma}_v^{*2}), \tag{18}$$

continues to be the second-order unbiased estimator of the MSE of the EBLUP $\hat{\theta}_{i,\text{EBL-LR}}$, under the model (14). Thus, while the second-order approximate MSE is not free from the fourth moment of the random effects v_i^*, its second-order unbiased estimator does not require estimation of this unknown fourth moment. From Prasad and Rao (1990) and Lahiri and Rao (1995), expressions of g_{1i}, g_{2i} and g_{3i} are given by

$$\begin{aligned} g_{1i}(\sigma_v^{*2}) &= D_i\Big(1 - B_{iFH}(\sigma_v^{*2})\Big), \\ g_{2i}(\sigma_v^{*2}) &= \Big\{B_{iFH}(\sigma_v^{*2})\Big\}^2 x_i^T\{X^T\Phi^{-1}(\sigma_v^{*2})X\}^{-1}x_i, \\ g_{3i}(\sigma_v^{*2}) &= \frac{2}{m^2 D_i}\Big\{B_{iFH}(\sigma_v^{*2})\Big\}^3 \sum_{j=1}^{m}(\sigma_v^{*2} + D_j)^2. \end{aligned} \tag{19}$$

Expressions (19) may be found, for example, in (7.1.7), (7.1.8) and (7.1.17) of Rao (2003), or in (5.16)−(5.19) of Prasad and Rao (1990).

For the poverty data set the estimated shrinkage coefficients, $B_{iFH}(\hat{\sigma}_v^{*2})$, are very similar to the estimated regression coefficients for the HB Fay-Herriot model with a uniform prior distribution on the model parameters. The similarity of the shrinkage coefficients of these two models also results in very similar point estimates of the small area means that result from them. Various deviation measures of the $\hat{\theta}_{i,\text{EBL-LR}}$ from the "true" census poverty ratios are nearly identical to those for the HB Fay-Herriot model. Among the four sets of predictors of small area means, namely, Y_i, $\hat{\theta}_{i,\text{FHM}}$, $\hat{\theta}_{i,\text{FH-HB}}$ and $\hat{\theta}_{i,\text{EBL-LR}}$, our proposed predictors $\hat{\theta}_{i,\text{FHM}}$ have the smallest deviation measures.

The uncertainty of the EBLUP predictors of the small area means θ_i are given by the estimated MSE, an expression of which is provided in (18). For our poverty ratio data, these estimates are very similar to the posterior variances of the θ_i's in the Fay-Herriot HB model. In Figure 2 we also plot the ratio of the posterior variance of θ_i for our proposed model to the estimated MSE of the EBLUP of θ_i (given by (18)). This plot displays that except for the state MA, for the other 50 states, the proposed model results in predictors of small area means which are more accurate than the predictors, frequentist or Bayes, based on the traditional Fay-Herriot model.

6 Simulation study

To further investigate the effectiveness of the proposed HB mixture model we conduct a simulation study. The setup of our simulation is similar to our application to estimate the state-level poverty ratios of $5-17$ year old related children for 51 small areas. In our simulation study, we take the covariates from this application, and the value of the regression coefficients β is considered as $(2, 0.75, 0.50, 0.25)^T$, motivated by the estimate of the real data considered in Section 4. The data are generated for different values of p and $\sigma^2_{v,M}$, as shown

Table 2: Performance of the proposed methodology − average of ratios of posterior variances and ratio of various deviation measures from the truth and ratios of posterior variances of the Fay-Herriot model to the proposed model

p	$\sigma^2_{v,M}$	AAD	ASD	ARB	ASRB	Post Var
0.1	25	1.198	1.183	1.215	1.254	1.667
	50	1.301	1.309	1.317	1.401	1.751
	75	1.389	1.466	1.401	1.588	1.811
	100	1.474	1.660	1.492	3.121	1.872
0.2	25	1.164	1.076	1.169	1.085	1.677
	50	1.261	1.204	1.261	1.262	1.682
	75	1.339	1.334	1.330	1.532	1.704
	100	1.405	1.431	1.404	2.774	1.732
0.5	25	1.040	0.952	1.029	0.910	1.400
	50	1.090	1.040	1.076	1.087	1.284
	75	1.117	1.096	1.094	1.099	1.263
	100	1.137	1.129	1.104	1.119	1.273
0.75	25	0.979	0.922	0.967	0.876	1.205
	50	1.005	0.974	0.996	0.974	1.117
	75	1.012	0.997	0.998	0.974	1.094
	100	1.015	1.004	1.000	0.981	1.084

in Table 2. We evaluated the four summary performance measures, namely AAD, ASD, ARB and ASRB, mentioned in Section 5, for the FH-HB predictor $\hat{\theta}_{\text{FH-HB}}$ and the proposed HB predictor $\hat{\theta}_{\text{M-HB}}$ via 1000 simulations. The average values of 1000 performance measures, reported in Table 2, demonstrate that except for the case with small $\sigma^2_{v,M}$ and large p, which is close to the "true" random effects model discussed by Fay-Herriot, the proposed method performs better than the FH method. The average ratios of posterior variances of the Fay-Herriot model to the proposed one also demonstrate the superiority of the proposed model. Figure 7 plots the average AAD ratios and the ratios of the posterior variances reported in Table 2.

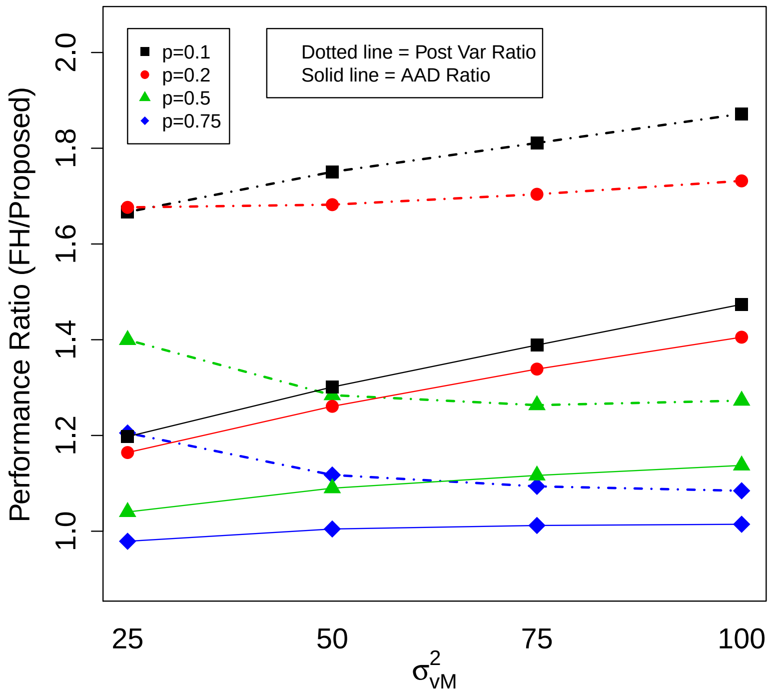

Figure 7: Performance of the proposed methodology as reported in Table 2. The dotted lines correspond to the ratio of posterior variances and the solid lines correspond to the ratio of Average Absolute Differences. Here the solid squares, circles, triangles and diamonds correspond to $p = 0.1, 0.2, 0.5$ and 0.75 respectively.

7 Conclusion

We propose a flexible alternative to the Fay-Herriot model for estimation of small area means and compare performance of the resulting small area predictors with other predictors of small area means. Our model includes as special cases both the Fay-Herriot model and the model without any random effects term suggested by the Datta et al. (2011) method. Our model determines the "goodness-of-fit" of the proposed regression model for the small area mean, and adaptively includes a random effects to the model for small area mean θ_i, if necessary. We use a Bayesian approach to fit our model, which was applied to estimate poverty ratios for school going children for the U.S. states. Our application shows the superiority of the proposed model when compared resulting estimators with the HB and EBLUP predictors of small area means based on the Fay-Herriot model, in terms of posterior variance or estimated mean squared error of the estimates and various deviations of the estimates from the "true" values of the means based on census data. This superior performance continues to carry over for various simulated scenarios of our model.

Acknowledgment

The research of Gauri S. Datta was partially supported through the grants DMS-09-14603 from the National Science Foundation and H98230-11-1-0208 from the National Security Agency. The research of Abhyuday Mandal was partially supported through the grants DMS-09-0573 from the National Science Foundation and H98230-13-1-0251 from the National Security Agency. The authors are thankful to Dr. William R. Bell for providing and explaining the poverty data used in our application. They also thank Jerry Maples of Census Bureau for valuable comments.

References

Battese, G. E., Harter, R. M. and Fuller, W. A. (1988). An error components model for prediction of county crop area using survey and satellite data. *Journal of the American Statistical Association*, 83, 28-36.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. 2nd edn. New York: Springer-Verlag.

Datta, G. S., Hall, P. and Mandal, A. (2011). Model selection by testing for the presence of small-area effects in area-level data. *Journal of the American Statistical Association - Theory and Methods*, **106**, 362-374.

Datta, G. S. and Lahiri. P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica*, **10**, 613-627.

Datta, G. S., Rao, J. N. K. and Smith, D. D. (2005). On measuring the variability of small area estimators under a basic area level model. *Biometrika*, **92**, 183-196.

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

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398-409.

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

Ishwaran, H. and Rao, J. S. (2005). Spike and slab variable selection: frequentist and Bayesian strategies. *The Annals of Statistics*, **Vol. 33**, 730-773.

Jiang, J., and Lahiri, P. (2006). Mixed model prediction and small area estimation (with discussions). *Test*, **15**, 1-96.

Lahiri, P. and Rao, J. N. K. (1995). Robust estimation of mean squared error of small area estimators. *Journal of the American Statistical Association*, **90**, 758-766.

Pfeffermann, D. (2013). New important developments in small area estimation. *Statistical Science*, **28**, 40-68.

Pfeffermann, D. and Nathan, G. (1981). Regression analysis of data from a cluster sample. *Journal of the American Statistical Association*, **76**, 681-9.

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

Rao, J. N. K. (2003). *Small Area Estimation.* New York: John Wiley & Sons Inc.

Scott, J. G. and Berger, J. O. (2006). An exploration of aspects of Bayesian multiple testing. *Journal of Statistical Planning and Inference*, **136**, 2144-2162.