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 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 approach may be assessed from an explosive growth of research publications. For early frequentist applications of small area estimation, we refer to Fay and Herriot (1979), Battese, Harter and Fuller (1988), and Prasad and Rao (1990). For a recent comprehensive review of this literature, one may refer to Rao (2003), Jiang and Lahiri (2006), and Pfeffermann (2013).

The landmark article 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. These authors 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 \( i \)th 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 \( \theta_i \) is a summary measure of the characteristic to be estimated for the \( i \)th 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 \( \theta_i \) from its linear regression on \( x_i, x_i^T \beta \). We assume that \( e_1, \ldots, e_m \) are independent and normally distributed with \( e_i \sim N(0, D_i) \), and are independent of \( v_1, \ldots, v_m \), which are i.i.d. \( N(0, \sigma^2_v) \). The sampling variances \( D_i \)'s are treated as known, but the model parameters \( \beta \) and \( \sigma^2_v \) 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 \( \theta_i \) 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 \( \theta_i \), Prasad and Rao (1990) and Lahiri and Rao (1995) considered an empirical best linear unbiased predictor (EBLUP) of \( \theta_i \). For the normal mixed linear model in (1), EB predictor and EBLUP predictor of \( \theta_i \) are identical. Prasad and Rao (1990), Lahiri and Rao (1995), Datta and Lahiri (2000), and Datta, Rao, and Smith (2005) accurately approximated the mean squared error (MSE) of EBLUP of \( \theta_i \) by ignoring all terms of order \( o(m^{-1}) \). These authors also provided second-order unbiased estimators of the MSE with an approximate bias to the order \( o(m^{-1}) \). In the Bayesian approach to predict \( \theta_i \) based on model in (1), a hierarchical Bayes (HB) model, given below, is completed by assigning a prior distribution on the model parameters \( \psi = (\beta^T, \sigma^2)^T \).

Model \( H_1 \):

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

The aforementioned HB model has been considered, for example, by Berger (1985), Ghosh (1992), Datta, Rao, and Smith (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 \( \theta_i \). 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, Hall, and Mandal (2011) demonstrated that in the presence of good covariates \( x_i \), the variability of the small area means \( \theta_i \) 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 uncertainty 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 \( \theta_i \) on the covariates \( x_i \) as a test statistic. The discrepancy function is constructed by combining certain deviance measures for all the small areas. The deviance measures 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 \( \theta_i \), the method of Datta, Hall, and Mandal (2011) advocates excluding the small area effects for all the small areas when a nonsignificant discrepancy statistic is realized. On the other hand, when a significant discrepancy statistic is obtained, 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 between the Datta, Hall, and Mandal (2011) method and the 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 (i.e., a degenerate distribution at zero) for any small area, and with probability \( p \) the random effects has a nondegenerate distribution. In Section 2, we introduce this new HB model and compare it with the Fay–Herriot model. The new model provides a more flexible data-dependent shrinkage of the direct estimator to the synthetic estimator. To our knowledge, this type of model has not yet been explored in small area estimation. In Section 3, we explore the propriety of posterior distribution of our HB model that results from an improper prior distribution on the model parameters. As HB predictors of small area means for our proposed model do not admit closed-form expressions, we obtain them 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. Lahiri and Rao (1995) relaxed the normality of the random effects in the Fay–Herriot model by assuming finite higher order moments to estimate the MSE of the EBLUPs of small area means. For the 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 MSE of EBLUP, proposed by Lahiri and Rao (1995) for the nonnormal random effects Fay–Herriot model. We included a simulation study as supplementary materials 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 the proposed model. The article concludes with a brief summary in Section 6.

2. A MODEL WITH UNCERTAIN RANDOM EFFECTS

As in the Fay–Herriot model, in the proposed model we assume that the direct estimator \( Y_i \) is an unbiased estimator of the small area mean \( \theta_i \) admitting a normal distribution with variance \( D_i \), for \( i = 1, \ldots, m \). The HB representation of our model is given as follows:

Model \( H_M \):

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

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

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 $\sigma^2_{vM}$, 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. A priori the hyperparameters $\beta_M, \sigma^2_{vM}$ and $p$ are independently distributed with prior density

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

where we assign a standard improper uniform prior distribution on the bounded parameter $p$.

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

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

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

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 $\beta_M$ and $\sigma^2_{vM}$ 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 $\sigma^2_{vM}$. This necessitates a proper prior distribution for $\sigma^2_{vM}$ 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 $\sigma^2_{vM}$. Ishwaran and Rao (2005) used this type of prior distribution, which is known in the literature as a “spike and slab” 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 nondegenerate random effects to account for a lack of fit of the regression of $\theta_i$ on $x_i$, it is unlikely that all small areas will need nondegenerate 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 $\theta_i$ based on the HB model $H_M$ by posterior mean of $\theta_i, \hat{\theta}_M$ (say), and measure the accuracy of the estimator by the posterior variance of $\theta_i, V_M$ (say). While the numerical computation of these estimators will be performed by Gibbs sampling, to compare the estimators $\hat{\theta}_M$ with the estimator of $\theta_i$ under the Fay–Herriot model, we first consider $\hat{\theta}_M(\beta_M, \sigma^2_{vM}, p, y_i)$, the conditional mean of $\theta_i$, conditional on the model parameters $\beta_M, \sigma^2_{vM}, p$ and data $y_1, \ldots, y_m$. By direct calculation,

$$\hat{\theta}_M(\beta_M, \sigma^2_{vM}, p, y_i) = x_i^T \beta_M + \frac{\sigma^2_{vM}}{\sigma^2_{vM} + D_i}(y_i - x_i^T \beta_M) \times \bar{p}_i(\beta_M, \sigma^2_{vM}, p, y_i),$$

$$y_i = B_M(y_i - x_i^T \beta_M),$$

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

$$B_M = 1 - \frac{\sigma^2_{vM}}{\sigma^2_{vM} + D_i} \bar{p}_i(\beta_M, \sigma^2_{vM}, p, y_i)$$

with $\bar{p}_i(\beta_M, \sigma^2_{vM}, p, y_i) = P(\delta_i = 1|\beta_M, \sigma^2_{vM}, p, y_i)$ being given by

$$\bar{p}_i(\beta_M, \sigma^2_{vM}, p, y_i) = \frac{p}{p + (1 - p)\sqrt{\sigma^2_{vM} + D_i} \exp \left\{ -\frac{1}{2} \frac{(y_i - x_i^T \beta_M)^2}{\sigma^2_{vM} + D_i} \right\}}.$$

Note that $\bar{p}_i$ is increasing in $p$ and $|y_i - x_i^T \beta_M|$, and $B_M$ is decreasing in $\sigma^2_{vM}$ for fixed $\bar{p}_i$, and is decreasing in $\bar{p}_i$ for fixed $\sigma^2_{vM}$. Thus, the shrinkage coefficient $B_M$ will be large if either $p, \sigma^2_{vM}$ 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” $|y_i - x_i^T \beta_M|$ (i.e., the fit of the regression of the direct estimator on the covariate $x_i$), this is not the case for the Fay–Herriot model, where the shrinkage coefficient $B_{FH}$ is given by

$$B_{FH} = \frac{D_i}{\sigma^2_{vM} + D_i} = 1 - \frac{\sigma^2_{vM}}{\sigma^2_{vM} + D_i},$$

and for known model parameters $\beta$ and $\sigma^2_{vM}$, the Bayes predictor of $\theta_i$ is given by

$$\hat{\theta}_{FH} = y_i - B_{FH}(y_i - x_i^T \beta).$$

Note that while the shrinkage coefficient in the Fay–Herriot model depends only on the ratio $D_i/\sigma^2_{vM}$, 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, Hall, and Mandal (2011) used an estimated version of the squared standardized residual $(y_i - x_i^T \beta)^2/D_i$ as the deviance of the $i$th small area to compute a discrepancy statistic that they proposed to test for the absence of a random effect.
While a fraction of small areas may contribute large deviations to the discrepancy statistic pushing it to be significant, suggesting a positive $\sigma^2$ and thereby reducing the shrinkage for all small areas, the shrinkage coefficient $B_M$ for the proposed model is more robust to large deviations from other small areas. It is definitely desirable to borrow more from the regression model by shrinking more 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 $\theta_i$, for $i = 1, \ldots, m$, is rather complex. We will implement Gibbs sampling (see, Gelfand and Smith 1990) to compute the posterior mean and posterior variance of $\theta_i$.

We discussed earlier that we need a proper prior distribution on the variance parameter $\sigma^2_{vM}$ 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

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

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 nondegenerate. Our experience with small area estimation applications suggests that in the presence of good covariates nearby 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 $\beta_M, v = (v_1, \ldots, v_m)^T$, $\delta = (\delta_1, \ldots, \delta_m)^T$, $p$, and $\sigma^2$. Based on the hierarchical model $H_M$ and the prior densities for $\sigma^2_{vM}$, $p$, and $\beta_M$, we get

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

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

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

where $\delta \cdot v = (\delta_1 v_1, \ldots, \delta_m v_m)^T$;

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

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

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 $\beta_M$, we investigate next 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 $\beta_M, v_1, \ldots, v_m, \delta_1, \ldots, \delta_m, p$, and $\sigma^2$. Let $\Sigma(\delta, \sigma^2_{vM}) = \text{diag}(D_1 + \delta_1 \sigma^2_{vM}, \ldots, D_m + \delta_m \sigma^2_{vM})$. Integrating the right-hand side of (10) with respect to $v_1, \ldots, v_m$, we get

$$
\pi(\beta_M, \delta_1, \ldots, \delta_m, p, \sigma^2_{vM} | y_1, \ldots, y_m) = K |\Sigma(\delta, \sigma^2_{vM})|^{-\frac{1}{2}} \\
\times \exp \left[ -\frac{1}{2} (y - X \beta_M)^T \Sigma^{-1}(\delta, \sigma^2_{vM})(y - X \beta_M) \right] \\
\times \pi_{\sigma^2}(\sigma^2_{vM}) \pi_p(p) p^{\sum_{i=1}^{m} \delta_i} (1-p)^{m-\sum_{i=1}^{m} \delta_i}. \tag{11}
$$

Here, and elsewhere in the proof, $K$ is a generic positive constant not depending on $\beta_M, \delta, p$, and $\sigma^2$. Since $\Sigma(\delta, \sigma^2_{vM})$ is positive definite and $\text{rank}(X) = q$, $X^T \Sigma^{-1}(\delta, \sigma^2_{vM}) X$ is also positive definite. Noting that $\int_0^1 p^{\sum_{i=1}^{m} \delta_i} (1-p)^{m-\sum_{i=1}^{m} \delta_i} \pi_p(p)dp \leq 1$, integrating out $p$ and $\beta_M$ from (11) we get

$$
\pi(\delta_1, \ldots, \delta_m, \sigma^2_{vM} | y_1, \ldots, y_m) \\
\leq K |\Sigma(\delta, \sigma^2_{vM})|^{-\frac{1}{2}} |X^T \Sigma^{-1}(\delta, \sigma^2_{vM}) X|^{-\frac{1}{2}} \\
\times \exp \left[ -\frac{1}{2} (y, \Sigma(\delta, \sigma^2_{vM})) \pi_{\sigma^2}(\sigma^2_{vM}) \right] \\
\leq K |\Sigma(\delta, \sigma^2_{vM})|^{-\frac{1}{2}} |X^T \Sigma^{-1}(\delta, \sigma^2_{vM}) X|^{-\frac{1}{2}} \pi_{\sigma^2}(\sigma^2_{vM}). \tag{12}
$$

where $R(y, \Sigma(\delta, \sigma^2_{vM})) = y^T [I_m - \Sigma^{-1}(\delta, \sigma^2_{vM}) X (X^T \Sigma^{-1}(\delta, \sigma^2_{vM}) X)^{-1}] (\delta, \sigma^2_{vM})$. We show below that for each $\delta$,

$$
|\Sigma(\delta, \sigma^2_{vM})|^{-\frac{1}{2}} |X^T \Sigma^{-1}(\delta, \sigma^2_{vM}) X|^{-\frac{1}{2}} \leq K \tag{13}
$$

for all $\sigma^2_{vM}$. Using (13) in (12) and integrating $\sigma^2_{vM}$ from $\pi_{\sigma^2}(\sigma^2_{vM})$, we get

$$
\pi(\delta_1, \ldots, \delta_m, \sigma^2_{vM} | y_1, \ldots, y_m) \leq K \pi_{\sigma^2}(\sigma^2_{vM} d\sigma^2_{vM}) = K < \infty.
$$

Since $\pi(\delta_1, \ldots, \delta_m | y_1, \ldots, y_m)$ is finite for all $\delta$, we get

$$
\sum_{\delta} \pi(\delta_1, \ldots, \delta_m | y_1, \ldots, y_m) < \infty,
$$

and the propriety of the posterior density (11) follows.

We now prove (13). For the case $\delta = (0, \ldots, 0)^T$, $\Sigma(\delta, \sigma^2_{vM})$ is free from $\sigma^2_{vM}$ and (13) holds trivially. Now suppose, $\sum_{i=1}^{m} \delta_i = m_1 \geq 1$. Corresponding to the indices $i_1, \ldots, i_{m_1}$, let $\delta_{i_1} = \cdots =
\[ \delta_{m_1} = 1. \] Define an \( m_1 \times q \) matrix \( X_1 = (x_{11}, \ldots, x_{m_1})^T \). We also define the matrix \( X_2 \) which is formed by the remaining rows of \( X \). Let \( \text{rank}(X_1) = r_1 \) and \( \text{rank}(X_2) = r_2 \). Note that \( 0 \leq r_1 \leq m_1, 0 \leq r_2 \leq m - m_1 \), and \( r_1 + r_2 \geq q \geq \max(r_1, r_2) \).

We first consider the case \( r_1 = 0 \). Then, \( X_1 \) is a null matrix and \( X^T \Sigma^{-1}(\delta, \sigma^2_{\epsilon M})X = \sum_{j \in A} D_j^{-1} x_j x_j^T \) is free from \( \sigma^2_{\epsilon M} \), where the index set \( A = \{ 1, \ldots, m \} \setminus \{ i_1, \ldots, i_m \} \). Since \(|\Sigma(\delta, \sigma^2_{\epsilon M})| \geq |D| \), it follows that

\[
|\Sigma(\delta, \sigma^2_{\epsilon M})|^{-\frac{1}{2}} |X^T \Sigma^{-1}(\delta, \sigma^2_{\epsilon M})X|^{-\frac{1}{2}} 
\leq |D|^{-\frac{1}{2}} \sum_{j \in A} D_j^{-1} x_j x_j^T \leq K.
\]

Now suppose \( r_1 \geq 1 \). We note that

\[
X^T \Sigma^{-1}(\delta, \sigma^2_{\epsilon M})X \geq (D_{r_1} + \sigma^2_{\epsilon M})^{-1} X_1^T X_1 + (D_{r_1} + \sigma^2_{\epsilon M})^{-1} X_2^T X_2 = G(y),
\]

(14)

Let \( P \) be an orthogonal matrix such that \( P^T X_{r_1} P = \text{diag}(\lambda_{1}, \ldots, \lambda_{r_1}, 0, \ldots, 0) \), where \( \lambda_{1} \geq \ldots \geq \lambda_{r_1} > 0 \) are the positive eigenvalues of \( X_{r_1}^T X_{r_1} \). Let \( D_{r_1}^{|A|} P^T X_2 P = B \) be partitioned as \(((B_{11}), (B_{12}))\), where \( B_{11} \) is \( r_1 \times r_1 \) and \( B_{22} = (q - r_1) \times (q - r_1) \). Writing \((D_{r_1} + \sigma^2_{\epsilon M})^{-1} = u \), we note that

\[
p^{TGP} = \begin{bmatrix} B_{11} + u^T \text{diag}(\lambda_{1}, \ldots, \lambda_{r_1}) B_{12} \\ B_{21} \\ B_{22} \end{bmatrix}.
\]

Since \( p^{TGP} \) is a positive-definite matrix, it follows that \( B_{22} \) is positive definite. Consequently, \( B_{11,2} = B_{11} - B_{12} B_{22}^{-1} B_{21} \) is defined. Since \( B \) is a nonnegative definite matrix, so is \( B_{11,2} \). Then, \( |p^{TGP}| = |B_{22}| |B_{11,2} + u \text{diag}(\lambda_{1}, \ldots, \lambda_{r_1})| \geq (\lambda_{1} \ldots \lambda_{r_1}) |B_{22}| |u^T| ) \), which, in turn, implies that

\[
|X^T \Sigma^{-1}(\delta, \sigma^2_{\epsilon M})X|^{-\frac{1}{2}} \leq |G|^{-\frac{1}{2}} = |p^{TGP}|^{-\frac{1}{2}} \leq K u^{-\frac{1}{2}} = K (D_{r_1} + \sigma^2_{\epsilon M})^{-\frac{1}{2}}.
\]

Finally, we get

\[
|\Sigma(\delta, \sigma^2_{\epsilon M})|^{-\frac{1}{2}} |X^T \Sigma^{-1}(\delta, \sigma^2_{\epsilon M})X|^{-\frac{1}{2}} 
\leq K \left( D_{r_1} + \sigma^2_{\epsilon M} \right)^{-\frac{1}{2}} \left( D_{r_1} + \sigma^2_{\epsilon M} \right)^{-\frac{1}{2}},
\]

and since \( r_1 \leq m_1 \), the right-hand side of this inequality is a bounded function of \( \sigma^2_{\epsilon M} \). This completes the proof of (13).

Only If Part: Integrability of the posterior density implies that for each \( \delta_1, \ldots, \delta_m \), the function \( \pi(\beta_M, \sigma^2_{\epsilon M}, p, v_1, \ldots, v_m, \delta_1, \ldots, \delta_m | y) \) is integrable. In particular, corresponding to \( \delta_1 = \ldots = \delta_m = 0 \), the function

\[
\prod_{i=1}^{m} \exp \left( -\frac{1}{2} (y_i - x_i^{T} \hat{\beta}_{\text{WLS}})^2 \right) \pi(\sigma^2_{\epsilon M}) \pi_p(p)
\]

needs to be integrable. This implies \( \prod_{i=1}^{m} \exp \left( -\frac{1}{2} (y_i - x_i^{T} \hat{\beta}_{\text{WLS}})^2 \right) \) needs to be integrable with respect to \( \beta_M \). A simple argument will show that the set \( \{x_1, \ldots, x_m\} \) must contain \( q \) linearly independent vectors, that is, \( m \geq q \).

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 \( i \)th 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 nonfiler 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 residuals from fitting a model for the 1989 census poverty data on these covariates for the year 1989 have 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 of Datta, Hall, and Mandal (2011) suggested to test the absence of a random effects term in the Fay–Herriot model in (1) is given by

\[
T = \sum_{i=1}^{m} \frac{(y_i - x_i^{T} \hat{\beta}_{\text{WLS}})^2}{D_i},
\]

where \( \hat{\beta}_{\text{WLS}} = (X^{T} D^{-1} X)^{-1} X^{T} D^{-1} Y \), and \( y_i - x_i^{T} \hat{\beta}_{\text{WLS}} \) is the deviance of the \( i \)th small area that 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.

![Figure 1. Plot of deviances of different states (small areas) that contributed to the discrepancy statistic T of Datta, Hall, and Mandal (2011).](image-url)
excluding Massachusetts. The new discrepancy statistic is quite small with a nonsignificant $p$-value of 0.284, and most of the individual deviances are smaller than 3.

An exploratory analysis of the poverty data presented earlier suggests adaptive shrinkage of the direct estimates to the synthetic regression-based estimates. We estimate the poverty ratio $\theta_i$ by applying the HB Fay–Herriot model and the proposed model. For the Fay–Herriot model, we use an improper uniform prior $\pi(\beta, \sigma^2_v) = 1$ to conduct the Bayesian analysis. It is well-known that the resulting posterior will be proper (see Berger 1985), and we implement Bayesian computing via Gibbs sampling. For the proposed uncertain random effects model (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 $\sigma^2_{vM}$ is necessary to ensure a proper posterior distribution. We consider an inverse gamma distribution with mean $D$ and variance $D^2$, where $D$ is the average sampling variance. The motivation behind this choice is to scale the prior variance in terms of the (average) error variance $D$ which is commonly done for a hypervariance (see, Scott and Berger 2006, p. 2146). We use improper uniform prior for $\beta_M$, 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 nondegenerate (it is zero if $\delta_i = 0$).

For each model, we obtain a point estimate of the small area mean $\hat{\theta}_i$ 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 $\hat{\theta}_i$ for the proposed model is about 60% higher than the posterior variance of $\hat{\theta}_i$ for the Fay–Herriot model. The average of this ratio for all the states is 0.54.

We compare the shrinkage coefficients for all the small areas for the two models. For the Fay–Herriot model, we use $D_i/(D_i + \hat{\sigma}^2_{FH-HB})$ as the estimated shrinkage for the $i$th small area, here $\hat{\sigma}^2_{FH-HB}$ is the posterior mean of $\sigma^2_v$ 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 $E(B_{iM}^2|y)$, where the posterior expectation is computed by averaging the values of $B_{iM}(\beta_M, \sigma^2_{Mv}, p, y_i)$ over the Gibbs samples generated from this model. In Figure 3, we plot these shrinkage coefficients. 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, 22 include the zero value.

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 random effects is suggested by the data only for Massachusetts.

We noticed earlier that the direct CPS estimate for Massachusetts does not fit the Fay–Herriot model well, and our model shrinks the CPS estimate much less to the synthetic

![Figure 2](image1.png)

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.

![Figure 3](image2.png)

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

![Figure 4](image3.png)

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 random effects is suggested by the data only for Massachusetts.
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 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 nonzero 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 toward 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 toward 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 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

![Figure 4. Plot of \( P(\delta_i = 1|\text{data}) \).](image)

![Figure 5. Histogram of the posterior distribution of \( \delta_i v_i \).](image)
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 red, and overestimated by 2% or more are shaded 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. The Fay–Herriot model overestimates for five states, namely, MS, LA, SD, KY, and WV, and underestimates for three states, namely, KS, DE, and MA. If we turn to the CPS estimates, they are much less accurate and overestimate for five states, namely, MS, LA, SD, KY, and WV, and underestimate for three states, namely, KS, DE, and MA. Additionally, smaller posterior variances associated with the estimates from the proposed model compared to the posterior variances associated with the estimates from 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 50 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 $R = \frac{\sum_{i=1}^{51} (\text{pop}_i')}{\sum_{i=1}^{51} (\text{pop}_i)'}$, is the estimated population of the 5 − 17 age group of related children in the $i$th 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 the 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 $\sum_{i=1}^{51} (\text{pop}_i')c_i' = \sum_{i=1}^{51} (\text{pop}_i)''$, 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 averaged 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 following table. 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.
5.1 EBLUP Prediction for Nonnormal 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 nonnormality of random effects.

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

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

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, the random effects $v_i^*$ are not normally distributed but they have all moments of positive order finite. This setup is a special case of the pioneering article by Lahiri and Rao (1995). They estimated $\beta$ by a two-stage weighted least squares method and the model variance $\sigma^2_{v^*}$ by the ANOVA method (see Prasad and Rao 1990).

We omit the details here. Denoting the estimators by $\hat{\beta}_{GLS}$ and $\hat{\sigma}^2_{v^*}$, the EBLUP of $\theta_i$, derived by Lahiri and Rao (1995) is given by

$$\hat{\theta}_i = F_{BH}(\hat{\sigma}^2_{v^*}) = Y_i - B_iF(\hat{\sigma}^2_{v^*})(Y_i - x_i^T \hat{\beta}_{GLS}(\hat{\sigma}^2_{v^*})).$$

Assuming finite $(8 + \epsilon)$-th moment of $\epsilon^*_i$, with $\epsilon > 0$, Lahiri and Rao (1995) obtained a second-order accurate approximation to the MSE of $\hat{\theta}_i$. In a serendipitous discovery, Lahiri and Rao (1995) noted that the second-order unbiased estimator of the MSE, $\text{mse}(\hat{\theta}_i)$, derived in (5.15) of Prasad and Rao (1990) continues to be a second-order unbiased estimator of the MSE of the EBLUP $\hat{\theta}_i$, under the model (15).

The uncertainty of the EBLUP predictors of the small area means $\theta_i$ are given by the estimated MSE. For our poverty ratio data, these estimates are very similar to the posterior variances of the $\theta_i$ in the Fay–Herriot HB model. In Figure 2, we also plot the ratio of the posterior variance of $\theta_i$ for our proposed model to $\text{mse}(\hat{\theta}_i)$. 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.

Additionally, the estimated shrinkage coefficients, $B_iF(\hat{\sigma}^2_{v^*})$, are very similar to the estimated shrinkage coefficients for the HB Fay–Herriot model with a uniform prior distribution on the model parameters. This results in very similar point estimates of the small area means. Various deviation measures of the $\hat{\theta}_i$ 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, $Y_i, \hat{\theta}_{FH}, \hat{\theta}_{FM}, \hat{\theta}_{EBL}$, our proposed predictors $\hat{\theta}_i$ have the smallest deviation measures.

6. 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 that may be suggested by the Datta, Hall, and Mandal (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 $\theta_i$, if necessary. We use a Bayesian approach to fit our model, which was applied to estimate poverty ratios for schoolgoing children for the U.S. states. Our application shows the superiority of the proposed model when comparing 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.

The uncertain random effects model proposed in this article, which assumes an equal probability $p$ of a nondegenerate effect in all small areas, may be generalized to unequal probabilities $p_i$. Moreover, in the presence of suitable auxiliary variables, these probabilities may be modeled through a logistic regression. However, similar to the case of equal $p$ for all small areas, here also we should use a proper prior on the logistic regression coefficients to ensure a proper posterior.

In this article, the proposed methodology has been illustrated with the example of SAPE data. However, there are other applications that can also benefit by using our new method. For example, a four-person state-level median income data for the
U.S. States reported by Fay (1987) or a hospital data reported by Morris and Christiansen (1995) and analyzed by Jiang and Tang (2011) are examples of potential application. Details of these examples can be found in the supplementary materials. In both examples, the proposed method produces smaller posterior variances than those from the Fay–Herriot HB model.

**SUPPLEMENTARY MATERIALS**

Supplementary materials contain the simulation study mentioned in the article, and additional examples.

[Received April 2014. Revised December 2014.]

**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. [1735]

Berger, J. O. (1985), *Statistical Decision Theory and Bayesian Analysis*. (2nd ed.), New York: Springer-Verlag. [1736,1740]

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*, 106, 362–374. [1736,1737,1739,1743]

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. [1736]

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. [1736]

Fay, R. E. (1987), “Application of Multivariate Regression to Small Domain Estimation,” in *Small Area Statistics*, eds. R. Platek, J. N. K. Rao, C. E. Särndal and M. P. Singh, New York: Wiley, pp. 91–102. [1744]

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. [1735]

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. [1738]

Ghosh, M. (1992), “Constrained Bayes Estimation With Applications,” *Journal of the American Statistical Association*, 87, 533–540. [1736]

Ishwaran, H., and Rao, J. S. (2005), “Spike and Slab Variable Selection: Frequentist and Bayesian Strategies,” *The Annals of Statistics*, 33, 730–773. [1737]

Jiang, J., and Lahiri, P. (2006), “Mixed Model Prediction and Small Area Estimation” (with discussions), *Test*, 15, 1–96. [1735]

Jiang, J., and Tang, E-T (2011), “The Best EBLUP in the Fay-Herriot Model,” *Annals of the Institute of Statistical Mathematics*, 63, 1123–1140. [1744]

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. [1736,1740,1743]

Morris, C. N., and Christiansen, C. L. (1995), “Hierarchical Models for Ranking and for Identifying Extremes With Applications.” in *Bayesian Statistics*, 5, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, New York: Oxford University Press. pp. 277–296. [1744]

Pfeffermann, D. (2013), “New Important Developments in Small Area Estimation,” *Statistical Science*, 28, 40–68. [1735]

Pfeffermann, D., and Nathan, G. (1981), “Regression Analysis of Data From a Cluster Sample,” *Journal of the American Statistical Association*, 76, 681–689. [1735]

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. [1736,1743]

Rao, J. N. K. (2003), *Small Area Estimation*. New York: Wiley. [1735]

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. [1737,1740]