Small Area Shrinkage Estimation

G. Datta and M. Ghosh

Abstract. The need for small area estimates is increasingly felt in both the public and private sectors in order to formulate their strategic plans. It is now widely recognized that direct small area survey estimates are highly unreliable owing to large standard errors and coefficients of variation. The reason behind this is that a survey is usually designed to achieve a specified level of accuracy at a higher level of geography than that of small areas. Lack of additional resources makes it almost imperative to use the same data to produce small area estimates. For example, if a survey is designed to estimate per capita income for a state, the same survey data need to be used to produce similar estimates for counties, subcounties and census divisions within that state. Thus, by necessity, small area estimation needs explicit, or at least implicit, use of models to link these areas. Improved small area estimates are found by "borrowing strength" from similar neighboring areas.

The key to small area estimation is shrinkage of direct estimates toward some regression estimates obtained by using in addition administrative records and other available sources of information. These shrinkage estimates can often be motivated from both a Bayesian and a frequentist point of view, and indeed in this particular context, it is possible to obtain at least an operational synthesis between the two paradigms. Thus, on one hand, while small area estimates can be developed using a hierarchical Bayesian or an empirical Bayesian approach, similar estimates are also found using the theory of best linear unbiased prediction (BLUP) or empirical best linear unbiased prediction (EBLUP).

The present article discusses primarily normal theory-based small area estimation techniques, and attempts a synthesis between both the Bayesian and the frequentist points of view. The results are mostly discussed for random effects models and their hierarchical Bayesian counterparts. A few miscellaneous remarks are made at the end describing the current research for more complex models including some nonnormal ones. Also provided are some pointers for future research.

Key words and phrases: Area-level models, BLUP, confidence intervals, EBLUP, empirical Bayes, hierarchical Bayes, mean squared error, multivariate, second-order unbiased, unit-level models.

1. INTRODUCTION

Small area estimation has become a topic of growing importance in recent years. The need for such esti-

mates is increasingly felt in both the public and private sectors in order to formulate their strategic plans. For instance, to address emerging or existing social issues, many national governments have passed laws that require production of reliable and up-to-date small area estimates on a regular basis. As an example, in the early 1990s, the U.S. Congress passed a law requiring the Secretary of Commerce to produce and publish, at least every two years, starting in 1996, current small area estimates related to the incidence of poverty for

G. Datta is Professor, Department of Statistics, University of Georgia, Athens, Georgia 30602-7952, USA (e-mail: gauri@uga.edu). M. Ghosh is Distinguished Professor, Department of Statistics, University of Florida, Gainesville, Florida 32611-8545, USA (e-mail: ghoshm@stat.ufl.edu).

states, counties, local jurisdictions of governments and school districts. In the private sector, businesses, especially the smaller ones, make decisions based on local income, population and environmental data to evaluate markets for new products and to determine areas for the location, expansion and contraction of their activities.

Small areas may refer to small geographical areas such as counties, subcounties, census tracts, etc. Alternately, they may also refer to small domains cross-classified by age, sex and other demographic characteristics. Other than "small areas" and "small domains," often the terms "local areas," "subdomains" and "substates" are used interchangeably. Throughout this article, we will use the term "small area," possibly the most popular usage of the term, especially in survey sampling.

Shrinkage estimators have even a longer history than small area estimators. An exact definition of these estimators is hard to come by. Lemmer (1988) in his *Encyclopedia of Statistical Sciences* article characterized shrinkage estimators as ones obtained through modification of some standard estimators, for example, maximum likelihood estimator (MLE), uniformly minimum variance unbiased estimator (UMVUE), least squares estimator, etc., in order to minimize some desirable criterion such as mean squared error (MSE), quadratic risk, bias, etc. With these objectives in mind, shrinkage estimators can be interpreted in a very broad sense. In particular, the best linear unbiased predictors (BLUP's), empirical best linear unbiased predictors (EBLUP's), empirical Bayes (EB), hierarchical Bayes (HB), and possibly a host of other estimators fall within this general category. One common feature of all these estimators is that they are usually weighted averages of one of the aforementioned standard estimators and some other estimator reasonable under an appropriate model. Weights to these estimators are determined with the objective of meeting some "optimality" criterion.

Shrinkage estimates have a natural place in small area estimation where direct estimates such as the MLE, UMVUE, etc., are usually unreliable owing to large standard errors and coefficients of variation associated with them. The reason behind this is that the original survey was targeted to achieve accuracy at a higher order of aggregation than that of small areas. Due to limited resources, the same survey data need to be used for producing small area estimates. This necessitates "borrowing strength" from similar other small areas with the objective of "increasing the effective sample size" in order to obtain estimates of increased precision.

The early small area estimators achieved this objective by shrinking the area-specific direct estimators (e.g., county-specific averages) toward some overall estimator (e.g., the state average). Later, with the availability of auxiliary information from administrative records and other sources, the direct estimators are now usually shrunk toward some estimated regression surface. This shrinking process needs explicit (or at least implicit) use of models.

Bayesian estimators have been in existence for more than two centuries. Very often, they can be regarded as shrinkage estimators, shrinking, for example, the sample mean toward the prior mean. The BLUP and EBLUP estimators developed by Henderson (1953) for mixed linear models are also genuine shrinkage estimators, shrinking the direct estimators toward some regression estimators. However, as the title of this special issue suggests, the name "shrinkage" possibly was coined with the seminal paper of Stein (1956). Stein introduced shrinkage estimators to estimate a multivariate normal mean vector and proved under the sum of squared error loss their domination over the sample mean vector in three or higher dimensions. He gave a purely decision-theoretic motivation of his result, and was implicitly considering a balanced one-way ANOVA model for random effects. The original result of Stein involved shrinking the sample mean toward some guessed value of the population mean. Later extensions of his ideas due to Lindley (1962) and Stein (1962) led to shrinkage toward an overall average, and more generally to a regression surface, still with balanced data. Stein's estimators gained immense popularity in the 1970s when Efron and Morris, in a series of articles, gave interesting EB interpretation of these estimators (see, e.g., Efron and Morris, 1973). A pioneering extension of Stein's ideas in the small area estimation context is due to Fay and Herriot (1979) in their highly referred article. The paper showed how Stein-type results (without necessarily the *exact* dominance consideration) could be extended to unbalanced random effect regression models with tremendous potential for application.

It is near impossible to cover all aspects of small area estimation in a single review article. Our primary focus will be on one-way random effects regression models, and connecting the ideas of BLUP and EBLUP with HB and EB estimators. These models are usually referred to in the small area literature as "area-level" models where one begins with some small area summary statistics, and tries to improve on these estimators by shrinking them toward some regression surface. This is in contrast to the so-called "unit-level"

models where one has data available for the sampled units within a small area. We will barely touch upon the latter. Another component of research which has received scant attention in the small area literature is the development of EB confidence intervals. We will discuss this topic also at some length. For a detailed exposure to small area estimation, the reader is referred to the recent book of Rao (2003a) and the review articles of Ghosh and Rao (1994), Pfeffermann (2002), Rao (1999, 2003b) and Datta (2009).

The outline of the remaining sections is as follows. In Section 2, we discuss balanced one-way random effects regression models, and discuss the connection between the BLUP's, EBLUP's, HB, EB, and in particular, the Stein-type shrinkage estimators. Section 3 extends these results to unbalanced one-way models, and compares and contrasts both HB and EB estimators in this setup. MSE approximation of small area estimators is also discussed in this section. Section 4 discusses multivariate small area shrinkage estimators, and discusses one particular application related to adjustment of census counts. Section 5 discusses EB confidence intervals for both balanced and unbalanced data. Section 6 gives a brief account of unit-level models for small area estimation. Section 7 contains a few other small area models such as measurement error models and generalized linear models. This section contains also a discussion of balanced loss functions in the context of small area estimation. Section 8 contains a summary of the results presented, and provides a few pointers toward topics for future research.

2. SHRINKAGE ESTIMATORS FOR BALANCED DATA

The primary objective of this section is to introduce shrinkage estimators of small area means under different paradigms, and point out the interrelationship between them. The corresponding uncertainty measures are also compared. We begin with the following model.

Let y_i $(i = 1, \ldots, m)$ denote the area-level survey estimators for the m small areas. Consider the model

$$(2.1) \quad \begin{aligned} y_i | \theta_i &\overset{ind}{\sim} N(\theta_i, V), \quad \text{and} \\ \theta_i | A &\overset{ind}{\sim} N(\mathbf{x}_i^T \boldsymbol{\beta}, A), \quad i = 1, \ldots, m. \end{aligned}$$

In the above $\mathbf{x}_1, \ldots, \mathbf{x}_m$ are p-dimensional design vectors and $\boldsymbol{\beta}$ $(p \times 1)$ is the unknown regression coefficient. Writing $\theta_i = \mathbf{x}_i^T \boldsymbol{\beta} + u_i$ $(i = 1, \ldots, m)$, it is easy to reexpress (2.1) as a random effects model with

$$(2.2) \quad y_i = \mathbf{x}_i^T \boldsymbol{\beta} + u_i + e_i, \quad i = 1, \ldots, m,$$

where the u_i and the e_i are mutually independent with $u_i \overset{i.i.d.}{\sim} N(0, A)$ and the $e_i \overset{ind}{\sim} N(0, V)$. Further, writing $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_m)^T$, $\mathbf{y} = (y_1, \ldots, y_m)^T$, $\mathbf{u} = (u_1, \ldots, u_m)^T$ and $\mathbf{e} = (e_1, \ldots, e_m)^T$, one can rewrite (2.2) in matrix notation as

$$(2.3) \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} + \mathbf{e}.$$

We assume rank$(\mathbf{X}) = p(< m)$. Noting that marginally, $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, (V + A)\mathbf{I}_m)$, where $\mathbf{I}_m$ is the identity matrix of order m, it is clear that we encounter an identifiability problem when both V and A are unknown. The problem does not occur in a unit-level model when one can find a separate estimate of V by utilizing the unit-level data. However, this option is unavailable in an area-level model, where it is customary to assume a known V. In practice, V is a sort of smoothed estimate, for example, using the generalized variance function approach; see, for example, Wolter (1985) or Otto and Bell (1995).

First assume $A(> 0)$ is known. We begin with the HB model with the prior $\pi(\boldsymbol{\beta}) = 1$. Then we have the following theorem.

THEOREM 1. *Under the given model, the posterior distribution of $\boldsymbol{\theta}$ is $N((1 - B)\mathbf{y} + B\mathbf{P_X}\mathbf{y}, V((1 - B)\mathbf{I}_m + B\mathbf{P_X}))$, where $B = V/(V + A)$ and $\mathbf{P_X} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$.*

PROOF. The result follows by noting that $\boldsymbol{\theta} | \boldsymbol{\beta}, \mathbf{y} \sim N((1 - B)\mathbf{y} + B\mathbf{X}\boldsymbol{\beta}, V(1 - B)\mathbf{I}_m)$ and $\boldsymbol{\beta} | \mathbf{y} \sim N((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}, (V + A)(\mathbf{X}^T\mathbf{X})^{-1})$, and then using the formulas for iterated expectation and variance along with normality of the conditionals. $\square$

REMARK 1. It follows from the above theorem that the posterior mean given by

$$(2.4) \quad \hat{\boldsymbol{\theta}}^B = E(\boldsymbol{\theta}|\mathbf{y}) = (1 - B)\mathbf{y} + B\mathbf{P_X}\mathbf{y}$$

is a weighted average of the direct estimator $\mathbf{y}$ and the regression estimator $\mathbf{P_X}\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \cdot \mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$. It is easy to check that the weights are inversely proportional to the sample variance and the prior variance. Thus $\hat{\boldsymbol{\theta}}^B$ shrinks the direct estimator $\mathbf{y}$ of $\boldsymbol{\theta}$ to the regression estimator $\mathbf{P_X}\mathbf{y}$ of θ, where the amount of shrinking depends on the ratio V/A. In the limiting cases when $B \to 0$ (i.e., when $V \ll A$) or $B \to 1$ (i.e., when $V \gg A$), $\hat{\boldsymbol{\theta}}^B$ tends respectively to the direct estimator $\mathbf{y}$ and the regression estimator $\mathbf{P_X}\mathbf{y}$, quite in keeping with one's intuition. Later, in Theorem 2, we will motivate the estimator in (2.4) as the BLUP of $\boldsymbol{\theta}$ without any distributional assumption. We also point out that this is also the best unbiased predictor under normality.

REMARK 2. It is also important to note that if the parameter $\boldsymbol{\beta}$ were also known, the posterior variance of $\boldsymbol{\theta}$ would be $V(1-B)\mathbf{I}_m$. Thus the term $VB\mathbf{P}_{\mathbf{X}}$ in the posterior variance in Theorem 1 can be interpreted as the additional posterior uncertainty due to unknown $\boldsymbol{\beta}$, but known A. We will examine later in this section the effect of an unknown A as well on the posterior variance.

Next we show that the estimator of $\boldsymbol{\theta}$ given in (2.4) can be motivated without any distributional assumption but using only the first two moments. The following theorem proves that this estimator is a BLUP, that is, it has the smallest mean squared error (MSE) within the class of all linear unbiased estimators (predictors) of $\boldsymbol{\theta}$. Also, the MSE equals the posterior variance given in Theorem 1.

THEOREM 2. *The estimator* $\hat{\boldsymbol{\theta}}^B$ *of* $\boldsymbol{\theta}$ *given in* (2.4) *is the BLUP of* $\boldsymbol{\theta}$. *Also,* $E[(\hat{\boldsymbol{\theta}}^B - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}^B - \boldsymbol{\theta})^T] = V\{(1-B)\mathbf{I}_m + B\mathbf{P}_{\mathbf{X}}\}$.

PROOF. Since $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$, any linear unbiased predictor $\mathbf{C}\mathbf{Y} + \mathbf{b}$ of $\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ must satisfy $\mathbf{C}\mathbf{X}\boldsymbol{\beta} + \mathbf{b} = \mathbf{X}\boldsymbol{\beta}$ for all $\boldsymbol{\beta}$. That is, $\mathbf{b} = \mathbf{0}$, and $\mathbf{C}\mathbf{X} = \mathbf{X}$, or equivalently, $\mathbf{C}\mathbf{P}_{\mathbf{X}} = \mathbf{P}_{\mathbf{X}}$. For such a predictor $\mathbf{C}\mathbf{Y}$, since $\mathbf{C}\mathbf{Y} - \boldsymbol{\theta} = (\mathbf{C} - \mathbf{I}_m)\mathbf{u} + \mathbf{C}\mathbf{e}$,

$$
\begin{aligned}
E[(\mathbf{C}\mathbf{Y} &- \boldsymbol{\theta})(\mathbf{C}\mathbf{Y} - \boldsymbol{\theta})^T] \\
&= A(\mathbf{C} - \mathbf{I}_m)(\mathbf{C} - \mathbf{I}_m)^T + V\mathbf{C}\mathbf{C}^T \\
&= (V+A)\mathbf{C}\mathbf{C}^T - A(\mathbf{C} + \mathbf{C}^T) + A\mathbf{I}_m \\
&= V(1-B)\mathbf{I}_m \\
&\quad + (V+A)\{\mathbf{C} - (1-B)\mathbf{I}_m\} \\
&\quad \cdot \{\mathbf{C} - (1-B)\mathbf{I}_m\}^T.
\end{aligned}
$$
(2.5)

Now subject to the condition $\mathbf{C}\mathbf{P}_{\mathbf{X}} = \mathbf{P}_{\mathbf{X}}$, it can be shown that

$$
\begin{aligned}
\{\mathbf{C} - (1&-B)\mathbf{I}_m\}\{\mathbf{C} - (1-B)\mathbf{I}_m\}^T \\
&= \{\mathbf{C} - (1-B)\mathbf{I}_m - B\mathbf{P}_{\mathbf{X}} + B\mathbf{P}_{\mathbf{X}}\} \\
&\quad \cdot \{\mathbf{C} - (1-B)\mathbf{I}_m - B\mathbf{P}_{\mathbf{X}} + B\mathbf{P}_{\mathbf{X}}\}^T \\
&= \{\mathbf{C} - (1-B)\mathbf{I}_m - B\mathbf{P}_{\mathbf{X}}\} \\
&\quad \cdot \{\mathbf{C} - (1-B)\mathbf{I}_m - B\mathbf{P}_{\mathbf{X}}\}^T + B^2\mathbf{P}_{\mathbf{X}}.
\end{aligned}
$$
(2.6)

Note that $\mathbf{C} = (1-B)\mathbf{I}_m + B\mathbf{P}_{\mathbf{X}}$ satisfies the condition $\mathbf{C}\mathbf{P}_{\mathbf{X}} = \mathbf{P}_{\mathbf{X}}$ and this choice minimizes $E[(\mathbf{C}\mathbf{Y} - \boldsymbol{\theta})(\mathbf{C}\mathbf{Y} - \boldsymbol{\theta})^T]$. Thus the BLUP of $\boldsymbol{\theta}$ is given by $\hat{\boldsymbol{\theta}}^B$. Also, from (2.5) and (2.6), it follows that the mean squared and product matrix of prediction error of the BLUP is $V\{(1-B)\mathbf{I}_m + B\mathbf{P}_{\mathbf{X}}\}$. $\square$

REMARK 3. Under normality of $\mathbf{u}$ and $\mathbf{e}$, the BLUP $\hat{\boldsymbol{\theta}}^B$ of $\boldsymbol{\theta}$ is also the best unbiased predictor of $\boldsymbol{\theta}$; that is, among all unbiased predictors of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}^B$ has the least mean squared error.

Theorems 1 and 2 establish the equivalence of the BLUP and the HB predictor and also of the corresponding uncertainty measures for the balanced one-way random effects model when the parameter A is known. Indeed, the result is also true for the general mixed effects model (see, e.g., Datta, 1992). However, this algebraic equality does not quite hold for unknown A, or equivalently unknown B.

To see this, we will consider separately, the EBLUP (or EB) and HB estimators, and point out where the differences occur. For the given random effects model, $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, (V+A)\mathbf{I}_m)$, which in the Bayesian terminology, is the marginal distribution of $\mathbf{y}$ after integrating out $\boldsymbol{\theta}$. Based on this marginal pdf, $(\hat{\boldsymbol{\beta}}, S = \|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2)$ is minimal sufficient for $(\boldsymbol{\beta}, A)$. Noting that $S \sim (V+A)\chi^2_{m-p}$, the UMVUE of $B = V/(V+A)$ is given by $\hat{B}^{EB} = V(m-p-2)/S$ for $m > p+2$. The corresponding EB or EBLUP estimator of $\boldsymbol{\theta}$ is then given by

$$
\begin{aligned}
\hat{\boldsymbol{\theta}}^{EB} &\equiv \hat{\boldsymbol{\theta}}^{EBLUP} \\
&= \left[1 - \frac{V(m-p-2)}{S}\right]\mathbf{y} \\
&\quad + \frac{V(m-p-2)}{S}\mathbf{X}\hat{\boldsymbol{\beta}},
\end{aligned}
$$
(2.7)

the James–Stein estimator (James and Stein, 1961).

One criticism of the above EB or EBLUP estimator is that the estimator $\hat{B}^{EB}$ of B can assume values bigger than 1 with positive probability. The resulting EB or EBLUP estimator then pulls the direct estimator $\mathbf{y}$ toward the opposite direction of the regression estimator $\mathbf{P}_{\mathbf{X}}\mathbf{y}$. Replacing $\hat{B}^{EB}$ by $(\hat{B}^{EB})^+$, where $(\hat{B}^{EB})^+ = \min(\hat{B}^{EB}, 1)$, the positive part Stein estimator rectifies the problem. However, it was shown by Datta et al. (2002) that $P(\hat{B}^{EB} > 1)$ goes to zero at an exponential rate for large m. So, the estimator $\hat{B}^{EB}$ is usually quite adequate even for moderate m.

In contrast, with the alternative fully Bayesian approach (Morris, 1983a), if one assigns the prior $\pi(\boldsymbol{\beta}, A) = 1$ so that $\pi(\boldsymbol{\beta}, B) = B^{-2}$, one gets $\pi(\boldsymbol{\theta}|B, \mathbf{y})$ the same as given in Theorem 1 for a known B, but needs in addition

$$
\begin{aligned}
\pi(B|\mathbf{y}) &\propto B^{(m-p)/2} \exp\left(-\frac{1}{2V}BS\right)B^{-2}I[0 < B < 1] \\
&= B^{(m-p-4)/2} \exp\left(-\frac{1}{2V}BS\right)I[0 < B < 1].
\end{aligned}
$$

Here, for the sake of simplicity and to present Morris's results, we have considered only a uniform prior for A. It is certainly possible to consider other priors, including inverse gamma priors with appropriate shape and scale parameters of the inverse gamma distribution, so long as the resulting posterior is proper. A prior of the form $\pi(\boldsymbol{\beta}, A) = A^{-k}$ will yield a proper posterior provided $k < 1$ and $m > p - 2k + 2$. Thus, while the uniform prior $\pi(\boldsymbol{\beta}, A) = 1$ yields a proper posterior when $m > p + 2$, the priors A^{-1} or A^{-2} will always yield improper posteriors. For the uniform prior, the posterior mean of $\boldsymbol{\theta}$ is now obtained by replacing B in Theorem 1 with $E(B|\mathbf{y})$, while $V(\boldsymbol{\theta}|\mathbf{y}) = V[(1 - E(B|\mathbf{y}))\mathbf{I}_m + E(B|\mathbf{y})P_{\mathbf{X}}] + V(B|\mathbf{y})(\mathbf{y} - P_{\mathbf{X}}\mathbf{y})(\mathbf{y} - P_{\mathbf{X}}\mathbf{y})^T$. Thus, other than the replacement of B by $E(B|\mathbf{y})$ in the variance formula given in Theorem 1, the additional uncertainty due to estimation of B is also incorporated in this variance formula.

Integrating by parts, one can show that for large m, $E(B|\mathbf{y})$ can be approximated by $(m - p - 2)V/S$ (cf. Theorem 1 of Datta and Ghosh, 1991a). Similarly, $V(B|\mathbf{y})$ can be approximated by $2(m - p - 2)V/S^2$. With these approximations, $E(\boldsymbol{\theta}|\mathbf{y})$ is approximated by $\hat{\boldsymbol{\theta}}^{EB}$, while $V(\boldsymbol{\theta}|\mathbf{y})$ can be approximated as $V(1 - \frac{(m-p-2)V}{S})\mathbf{I}_m + \frac{(m-p-2)V^2}{S}P_{\mathbf{X}} + \frac{2(m-p-2)V^2}{S^2}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T$. These results agree with Morris' (1983b) intuitive approximations for $E(\boldsymbol{\theta}|\mathbf{y})$ and $V(\boldsymbol{\theta}|\mathbf{y})$ for the special case of intercept model. In addition, if instead of the posterior mean, one estimates B by its posterior mode, one gets the estimator $\hat{B}^{MO} = \min((m - p - 4)V/S, 1)$, which leads to an estimator of $\boldsymbol{\theta}$ quite akin to the positive part James–Stein estimator, the only difference being that $m - p - 2$ is now replaced by $m - p - 4$.

It is instructive to find the Bayes risk of $\hat{\boldsymbol{\theta}}^{EB}$ under squared error loss $L(\boldsymbol{\theta}, \mathbf{a}) = \|\boldsymbol{\theta} - \mathbf{a}\|^2$. The following theorem is proved.

THEOREM 3. *Let $m > p + 2$. Then writing $h_{ii} = \mathbf{x}_i^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i$ for all i*:

(a)

$$E[(\theta_i - \hat{\theta}_i^{EB})]^2 = V(1 - B) + VBh_{ii} + \frac{2VB(1 - h_{ii})}{m - p};$$

(b)

$$E\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}^{EB}\|^2 = V[m - (m - p - 2)B].$$

PROOF. Let $\hat{B} = V(m - p - 2)/S$. If $\tilde{\theta}_i = E[\theta_i | \boldsymbol{\beta}, A, \mathbf{y}]$, then $\tilde{\theta}_i = y_i - B(y_i - \mathbf{x}_i^T\boldsymbol{\beta})$ and $V[\theta_i|\boldsymbol{\beta}, A, \mathbf{y}] = V(1 - B)$. Using iterated expectation it follows that

$$(2.8) \quad E[(\theta_i - \hat{\theta}_i^{EB})]^2 = V(1 - B) + E[(\tilde{\theta}_i - \hat{\theta}_i^{EB})]^2.$$

Using the expressions of $\tilde{\theta}_i$, $\hat{\theta}_i^{EB}$, and independence of $\hat{\boldsymbol{\beta}}$ and $\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$, it follows that

$$
\begin{aligned}
(2.9) \quad & E[(\tilde{\theta}_i - \hat{\theta}_i^{EB})]^2 \\
& = E[\{B\mathbf{x}_i^T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\}^2] \\
& \quad + E[(\hat{B} - B)^2(y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}})^2] \\
& = VBh_{ii} + E[(\hat{B} - B)^2(y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}})^2],
\end{aligned}
$$

where $h_{ii} = \mathbf{x}_i^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i$. By Basu's theorem, S and $(y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}})^2/S$ are independently distributed (see Ghosh, 1992a). Then

$$
\begin{aligned}
(2.10) \quad & E[(\hat{B} - B)^2(y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}})^2] \\
& = E[S(\hat{B} - B)^2]E[(y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}})^2/S].
\end{aligned}
$$

By a simple calculation $E[S(\hat{B} - B)^2] = 2VB$. Also, by the independence of S and $(y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}})^2/S$,

$$
\begin{aligned}
(2.11) \quad E[(y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}})^2/S] & = \frac{E(y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}})^2}{E(S)} \\
& = \frac{(\sigma^2/B)(1 - h_{ii})}{(\sigma^2/B)(m - p)} \\
& = \frac{1 - h_{ii}}{m - p}.
\end{aligned}
$$

Combining (2.8)–(2.11), one gets (a). Summing both sides of (a) over i, and noting $\sum_{i=1}^m h_{ii} = \text{tr}[(\mathbf{X}^T\mathbf{X})^{-1} \cdot (\mathbf{X}^T\mathbf{X})] = p$, one gets (b). □

REMARK 4. It is interesting to observe that a comparison of Theorem 3 with Theorem 1 (or Theorem 2) reveals that the excess Bayes risk due to estimation of the unknown variance component A is simply $2VB$. It is easy to see from Theorem 1 or 2 that the Bayes risk with known A is $V[m(1 - B) + pB]$.

REMARK 5. Another interesting observation from Theorem 3 is that an unbiased estimator of the MSE is $V[m - \frac{V(m-p-2)^2}{S}]$ which is simply Stein's unbiased estimator. While this is in agreement with equation (1.18) of Morris (1983b), our expression for the component MSE given by part (a) in Theorem 3 agrees with equation (1.16) of Morris (1983b) only in the special case of an intercept model, that is, when $\theta_i = \mu + u_i$ $(i = 1, \ldots, m)$. We believe that this is due to an oversight in Morris (1983b) in the derivation of the component risk for the general regression model.

We will now see how the above results can be generalized with unequal numbers of observations in the different small areas.

3. SHRINKAGE ESTIMATORS FOR UNBALANCED DATA

The equal sampling variance scenario considered in the previous section hardly arises for small area problems, where sampling variances for small areas are almost always unequal. A widely used area-level model first introduced by Fay and Herriot (1979) is given by

$$(3.1) \quad y_i|\theta_i \overset{ind}{\sim} N(\theta_i, V_i), \quad \theta_i \overset{ind}{\sim} N(\mathbf{x}_i^T\boldsymbol{\beta}, A).$$

Clearly the above model can be viewed also as a random effects model as shown in the previous section.

Fay and Herriot used the above model for estimating the per capita income (PCI) for small places in the United States with population less than 1000. In their case, y_i is the logarithm of per capita income for the ith small area. The auxiliary variables considered were logarithms of the PCI for the associated counties, tax return data, data on housing from the previous decennial census. The Fay–Herriot method was adopted by the U. S. Bureau of the Census to provide updated PCI estimates for small areas.

Fay and Herriot adopted an EB approach in their analysis. Write $\mathbf{G} = \mathrm{Diag}(V_1, \ldots, V_m)$, $\mathbf{D} = \mathbf{G} + A\mathbf{I}_m$, $\mathbf{B} = \mathbf{GD}^{-1} = \mathbf{D}^{-1}\mathbf{G} = \mathrm{Diag}(B_1, \ldots, B_m)$, where $B_i = V_i/(V_i + A), i = 1, \ldots, m$. First, assuming $\boldsymbol{\beta}$ and A to be both known, the Bayes estimator of $\boldsymbol{\theta}$ is $\hat{\boldsymbol{\theta}}^B = (\mathbf{I}_m - \mathbf{B})\mathbf{y} + \mathbf{BX}\boldsymbol{\beta}$. In order to estimate $\boldsymbol{\beta}$ and A as needed in an EB approach, first observe that for A known, the generalized least squares estimator of $\boldsymbol{\beta}$ is

$$(3.2) \quad \begin{aligned} \tilde{\boldsymbol{\beta}}(A) &= (\mathbf{X}^T\mathbf{D}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{D}^{-1}\mathbf{y} \\ &= [\mathbf{X}^T(\mathbf{I}_m - \mathbf{B})\mathbf{X}]^{-1}\mathbf{X}^T(\mathbf{I}_m - \mathbf{B})\mathbf{y}, \end{aligned}$$

where we assume, as before, $\mathrm{rank}(\mathbf{X}) = p(< m)$. We may note here that the corresponding BLUP estimator of $\boldsymbol{\theta}$ is $(\mathbf{I}_m - \mathbf{B})\mathbf{y} + \mathbf{BX}\tilde{\boldsymbol{\beta}}(A)$. In order to estimate A as well, Fay and Herriot (1979) and Datta, Rao and Smith (2005) used the moment identity given by $E[\sum_{i=1}^{m}\{y_i - \mathbf{x}_i^T\tilde{\boldsymbol{\beta}}(A)\}^2/(V_i + A)] = m - p$. Dropping the expectation from the left-hand side we get

$$(3.3) \quad \sum_{i=1}^{m}\{y_i - \mathbf{x}_i^T\tilde{\boldsymbol{\beta}}(A)\}^2/(V_i + A) = m - p.$$

Since the expression in the left-hand side of (3.3) is a nonincreasing function of A, if this expression evaluated at $A = 0$ is less than $m - p$, there will be no solution to the above equation. In this case, the estimate

is taken to be zero. In the other case, taking an initial guess at A and solving (3.2) and (3.3) iteratively, one finds the estimators $\hat{A}$ and $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\hat{A})$. The resulting EB or EBLUP estimator of $\boldsymbol{\theta}$ is given by

$$(3.4) \quad \hat{\boldsymbol{\theta}}^{EB} = (\mathbf{I}_m - \hat{\mathbf{B}})\mathbf{y} + \hat{\mathbf{B}}\mathbf{X}\hat{\boldsymbol{\beta}},$$

where $\hat{\mathbf{B}} = \mathrm{Diag}(V_1/(V_1 + \hat{A}), \ldots, V_m/(V_m + \hat{A}))$.

Morris (1983b) provided a general discussion of the EB approach in this case with the same prescription for estimation of $\boldsymbol{\beta}$ and A. An alternative HB formulation analogous to the one in Section 2 is given by Ghosh (1992a) who also explored an interrelationship between the EB and the HB procedures. The HB model is given by

$$y_i|\theta_i, \boldsymbol{\beta}, A \overset{ind}{\sim} N(\theta_i, V_i),$$

$$(3.5) \quad \theta_i|\boldsymbol{\beta}, A \overset{ind}{\sim} N(\mathbf{x}_i^T\boldsymbol{\beta}, A),$$

$$i = 1, \ldots, m, \quad \pi(\boldsymbol{\beta}, A) = 1.$$

Then the joint posterior density is

$$\pi(\boldsymbol{\theta}, \boldsymbol{\beta}, A|\mathbf{y})$$

$$(3.6) \quad \propto A^{-m/2}\exp[-\tfrac{1}{2}\{(\mathbf{y} - \boldsymbol{\theta})^T\mathbf{G}^{-1}(\mathbf{y} - \boldsymbol{\theta})$$

$$+ A^{-1}\|\boldsymbol{\theta} - \mathbf{X}\boldsymbol{\beta}\|^2\}].$$

Then one gets $\boldsymbol{\theta}|\boldsymbol{\beta}, a, \mathbf{y} \sim N[(\mathbf{I}_m - \mathbf{B})\mathbf{y} + \mathbf{BX}\boldsymbol{\beta}, \mathbf{G}(\mathbf{I}_m - \mathbf{B})]$, $\boldsymbol{\beta}|A, \mathbf{y} \sim N[\tilde{\boldsymbol{\beta}}(A), A\{\mathbf{X}^T(\mathbf{I}_m - \mathbf{B})\mathbf{X}\}^{-1}]$, where $\tilde{\boldsymbol{\beta}}(A) = [\mathbf{X}^T(\mathbf{I}_m - \mathbf{B})\mathbf{X}]^{-1}\mathbf{X}^T(\mathbf{I}_m - \mathbf{B})\mathbf{y}$. The marginal posterior of A is

$$\pi(A|\mathbf{y}) \propto A^{-(m-p)/2}\prod_{i=1}^{m}(1 - B_i)^{1/2}$$

$$(3.7)$$

$$\cdot \left|\sum_{i=1}^{m}(1 - B_i)\mathbf{x}_i\mathbf{x}_i^T\right|^{-1/2}\exp\left[-\frac{1}{2}Q(\mathbf{y})\right],$$

where $Q(\mathbf{y}) = A^{-1}[\sum_{i=1}^{m}(1 - B_i)y_i^2 - \{\sum_{i=1}^{m}(1 - B_i)y_i\mathbf{x}_i\}^T\{\sum_{i=1}^{m}(1 - B_i)\mathbf{x}_i\mathbf{x}_i^T\}^{-1}\{\sum_{i=1}^{m}(1 - B_i)y_i\mathbf{x}_i\}]$. It follows now that

$$(3.8) \quad E(\boldsymbol{\theta}|\mathbf{y}) = [\mathbf{I}_m - E(\mathbf{B}|\mathbf{y})]\mathbf{y} + E[\mathbf{B}H_{\mathbf{X}}\mathbf{y}|\mathbf{y}],$$

$$V(\boldsymbol{\theta}|\mathbf{y}) = E[\{\mathbf{I}_m - \mathbf{B}\}\mathbf{G}|\mathbf{y}]$$

$$(3.9) \quad + E[\{\mathbf{B}(\mathbf{I}_m - H_{\mathbf{X}})\}\mathbf{G}|\mathbf{y}]$$

$$+ V[\mathbf{B}\{\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}(A)\}|\mathbf{y}],$$

where $H_{\mathbf{X}} = \mathbf{X}[\mathbf{X}^T(\mathbf{I}_m - \mathbf{B})\mathbf{X}]^{-1}\mathbf{X}^T(\mathbf{I}_m - \mathbf{B})$. Numerical integration involving one-dimensional integrals needs to be carried out for evaluating both $E(\boldsymbol{\theta}|\mathbf{y})$ and $V(\boldsymbol{\theta}|\mathbf{y})$. In the special case, when $V_1 = \cdots = V_m$,

these expressions simplify to the ones obtained in the previous section. This is because in this special case, $\mathbf{I}_m - \mathbf{B} = (1 - B)\mathbf{I}_m$. We may also reemphasize that the first component in the right-hand side of (3.9) is the posterior variance when both $\boldsymbol{\beta}$ and A are known. The second term provides additional uncertainty due to unknown $\boldsymbol{\beta}$ but known A. The third term accounts for additional uncertainty due to unknown A as well.

In the Bayesian framework, posterior variances are the natural uncertainty measures. In the frequentist approach, a naive method is to substitute A by some suitable estimator in the mean squared prediction error formula for the BLUP [cf. (2.5) and (2.6) for the balanced case]. In the unbalanced case, the mean squared and product matrix of prediction error for the BLUP is given by the sum of the first two terms in the right-hand side of (3.9) without the conditional expectation operator. As it appears, this will miss the third component as it will not account for uncertainty due to estimation of A. This results in an underestimation of the true MSE of the EBLUP.

To account for the error in estimating A, following an earlier work of Kackar and Harville (1984), Prasad and Rao (1990) considered the MSE of the EBLUP. Unlike in the balanced case of Section 2, there is no closed-form expression of this MSE. They obtained an asymptotic expression of the MSE which is accurate to the order $o(m^{-1})$. This approximation is based on an orthogonal decomposition of the MSE $E(\hat{\theta}_i^{EB} - \theta_i)^2$. Specifically, they used the decomposition

$$\hat{\theta}_i^{EB} - \theta_i = \{(1 - B_i)y_i + B_i\mathbf{x}_i^T\boldsymbol{\beta} - \theta_i\}$$
$$+ \{\hat{\theta}_i^B - (1 - B_i)y_i - B_i\mathbf{x}_i^T\boldsymbol{\beta}\}$$
$$+ \{\hat{\theta}_i^{EB} - \hat{\theta}_i^B\};$$

the first component is always orthogonal to the second and the third components. The orthogonality of the last two components holds only for certain specific estimators of A. It is necessary that these estimators are translation invariant under the transformation $g_c(\mathbf{y})$ which maps $\mathbf{y}$ to $\mathbf{y} + \mathbf{X}c$ and are even functions of $\mathbf{y}$. In particular, the ANOVA estimator in Prasad and Rao (1990), the ML and the REML estimators considered in Datta and Lahiri (2000), and the method of moment estimator due to Fay and Herriot (1979), Morris (1983b) and Datta, Rao and Smith (2005) all satisfy these conditions. For these estimators of A, it follows that

$$(3.10) \quad \begin{aligned} E(\hat{\theta}_i^{EB} - \theta_i)^2 \\ = g_{1i}(A) + g_{2i}(A) + g_{3i}(A) + o(m^{-1}), \end{aligned}$$

where $g_{1i}(A) = V_i(1 - B_i)$, $g_{2i}(A) = B_i^2 A\mathbf{x}_i^T \cdot \{\sum_{j=1}^m (1 - B_j)\mathbf{x}_j\mathbf{x}_j^T\}^{-1}\mathbf{x}_i$ and $g_{3i}(A) = V_i^2(A + V_i)^{-3}\text{Var}(\hat{A})$. The derivation of the third term is based on a second-order Taylor expansion [i.e., retaining up to the $O(m^{-1})$ term] of $E[B_i(y_i - \mathbf{x}_i^T\tilde{\boldsymbol{\beta}}(A)) - \hat{B}_i(y_i - \mathbf{x}_i^T\tilde{\boldsymbol{\beta}}(\hat{A}))]^2$. This derivation requires also orthogonality of $\boldsymbol{\beta}$ and A in the Fisherian sense, that is, block diagonality of the relevant components of the Fisher information matrix. An intuitive estimator, say, $mse^I(\hat{A})$, of the MSE in (3.10) is given by

$$(3.11) \quad mse^I(\hat{A}) = g_{1i}(\hat{A}) + g_{2i}(\hat{A}) + g_{3i}(\hat{A}).$$

In view of the fact that $E[g_{1i}(\hat{A})] = g_{1i}(A) - g_{3i}(A) + o(m^{-1})$, and $g_{3i}(A)$ is $O(m^{-1})$, the above estimator is not second-order unbiased. Based on the ANOVA estimator of A, say, $\hat{A}_{PR}$, which is second-order unbiased for A, Prasad and Rao (1990) showed that the estimator

$$(3.12) \quad mse^S(\hat{A}_{PR}) = \hat{g}_{1PRi} + \hat{g}_{2PRi} + \hat{g}_{3PRi}$$

is second-order unbiased in the sense that

$$E[mse^S(\hat{A}_{PR})] = MSE(\hat{\theta}_i^{EB}) + o(m^{-1}),$$

where

$$\hat{g}_{1PRi} = g_{1i}(\hat{A}_{PR}) + g_{3i}(\hat{A}_{PR}), \quad \hat{g}_{2PRi} = g_{2i}(\hat{A}_{PR}),$$
$$\hat{g}_{3PRi} = g_{3i}(\hat{A}_{PR}).$$

See Harville (1990) for similar results for mixed linear models. In the small area context Datta and Lahiri (2000) showed that the expression in (3.12) based on the REML estimator of A is also second-order unbiased. Second-order unbiased estimator of the MSE of the EBLUP using the ML estimator and Fay–Herriot estimator of A are given in Datta and Lahiri (2000) and Datta, Rao and Smith (2005), respectively. For further discussion we may refer to Rao (2003a) and Datta (2009).

The posterior variance of θ_i, on the other hand [see (3.9)], is given by

$$(3.13) \quad \begin{aligned} V(\theta_i|\mathbf{y}) &= V_i[1 - E(B_i|\mathbf{y})] \\ &+ E\left[B_i^2 A\mathbf{x}_i^T\left\{\sum_{j=1}^m (1 - B_j)\mathbf{x}_j\mathbf{x}_j^T\right\}^{-1}\mathbf{x}_i\Big|\mathbf{y}\right] \\ &+ V[B_i\{y_i - \mathbf{x}_i^T\tilde{\boldsymbol{\beta}}(A)\}|\mathbf{y}] \\ &= E[g_{1i}(A)|\mathbf{y}] + E[g_{2i}(A)|\mathbf{y}] \\ &+ V[B_i\{y_i - \mathbf{x}_i^T\tilde{\boldsymbol{\beta}}(A)\}|\mathbf{y}] \\ &= g_{1HBi} + g_{2HBi} + g_{3HBi} \quad \text{(say)}. \end{aligned}$$

Morris (1983b) provided an approximation to the HB estimator $E(\theta_i|\mathbf{y})$ and the associated posterior variance. Denoting Morris' point estimator of θ_i by $\hat{\theta}_i{}^M$,

$$(3.14) \qquad \hat{\theta}_i{}^M = (1 - \hat{B}_i{}^M)y_i + \hat{B}_i{}^M(\mathbf{x}_i^T\hat{\boldsymbol{\beta}}),$$

where $\hat{B}_i{}^M = ((m - p - 2)/(m - p))(V_i/(V_i + \hat{A}))$, and $\hat{\boldsymbol{\beta}}$ and $\hat{A}$ are obtained by solving (3.2) and (3.3) iteratively. It can be checked that (3.2) and (3.3) are equivalent to Morris' (1983b) equations (5.2) and (5.4). Morris (1983b) approximated the posterior variance by s_{iM}^2, given by $s_{iM}^2 = e_{iM} + v_{iM}$, where $e_{iM} = g_{1iM} + g_{2iM}$, $v_{iM} = g_{3iM}$ with

$$(3.15) \quad \begin{aligned} g_{1iM} &= V_i[1 - \hat{B}_i{}^M], \quad g_{2iM} = V_i\hat{B}_i{}^M\hat{t}_i, \\ g_{3iM} &= \frac{2(\hat{B}_i{}^M)^2(y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}})^2}{m - p - 2} \cdot \frac{\bar{V} + \hat{A}}{V_i + \hat{A}}, \end{aligned}$$

and $\hat{t}_i = \mathbf{x}_i^T[\mathbf{X}^T(\mathbf{V} + \hat{A}\mathbf{I})^{-1}\mathbf{X}]^{-1}\mathbf{x}_i/(V_i + \hat{A})$, $i = 1, \ldots, m$, $\bar{V} = \sum_{i=1}^m V_i/m$.

From the three measures of uncertainty given by (3.12), (3.13) and (3.15) we see a close correspondence in the respective terms in the expansion of the MSE of the EB estimator, the posterior variance of θ_i and Morris' approximation of the posterior variance. It is clear, though, that while the posterior variance of θ_i accounts for all sources of uncertainty in a straightforward way, the EB or EBLUP method needs careful evaluation of all terms in the MSE expression and construct a second-order unbiased estimator of this quantity. Morris (1983b) provided a clever approximation to the posterior variance. The estimator of the MSE of the EBLUP displays poor performance when A is estimated by zero or severely underestimated (this happens if the true variance parameter A is small). In such case the first term $\hat{g}_{1PR}$ is too small compared to the first term in the posterior variance. This results from the integration of A with respect to its long tail posterior distribution. Use of posterior variance has been found to be attractive in small area application. As an example, the U.S. Bureau of the Census uses this method in producing small area income and poverty estimates based on American Community Survey data. The corresponding term in Morris' approximation is a clever approximation to the posterior expectation. Although not as small as $\hat{g}_{1PR}$, this also tends to be small. The $g_{1i}(A)$ function evaluated at the point estimator of A, via posterior mode or REML, is usually smaller than its integrated value with respect to the posterior of A. The second and the third terms in these measures of uncertainty, being of lower order of magnitude, usually

show a greater degree of agreement. Another attractive feature of posterior variance is that it depends on the individual small area observation y_i [through the last term in (3.13)]. This is not true for the second-order unbiased estimator of the MSE given in (3.12). However, the estimate of conditional frequentist mean squared error of prediction obtained by conditioning on y_i depends on the individual small area observation (see, e.g., Booth and Hobert, 1998, or Datta et al., 2011). For related discussions comparing the Bayesian and the frequentist measures of uncertainty in small area estimation we refer to Singh, Stukel and Pfeffermann (1998) and Datta, Rao and Smith (2005). Morris' approximation, which closely mimics the posterior variance, also enjoys this feature.

We consider an illustration of the Fay–Herriot model. The U.S. Department of Health and Human Services (HHS) needs estimates of four-person family state median income to implement an energy assistance program to low-income families. The Bureau of the Census (BOC) has provided such estimates for nearly thirty years. The BOC now uses the Fay–Herriot model to provide more sophisticated estimates. In this model the direct estimate of the four-person family state median income, to be denoted by y_i, is obtained from the Current Population Survey (CPS). Auxiliary variables for the multiple regression model are obtained from the per capita income information of the Bureau of the Economic Analysis (BEA) survey and the latest census data for the four-person family median income. In our illustration, we will consider only a subset of the U.S. states and use only one covariate. We consider 15 U.S. states belonging to the southeast U.S. geographical region. While there are 17 states in this region, we excluded Texas and Washington, DC, from our analysis as these two small areas have their sampling variances (V_i's) very much different from the remaining 15 states. In the notation of this section, we have $m = 15$, $p = 2$, $\mathbf{x}_i^T = (1, x_i)$, with x_i, the adjusted census median income, given by

$$x_i = \frac{\text{BEA PCI}(c) \text{ for state } i}{\text{BEA PCI}(b) \text{ for state } i}$$
$$\cdot \text{Census median}(b) \text{ for state } i,$$

where c stands for current year (in our application 1979) and b stands for base year (1969), BEA PCI(b) and BEA PCI(c) are obtained from the BEA data for these two years, and Census median(b) is obtained from the 1969 census.

We present the relevant data in Table 1 below. Also included in the table are the EB estimates ($\hat{\theta}^{bEB}$ in the

TABLE 1
Data for estimating 1,979 four-person family median income for the 15 southeastern U.S. states, and different small area estimates

State	y	x	V	$\hat{\theta}^{bHB}$	$\hat{\theta}^{bEB}$	$\hat{\theta}^{bM}$	$\hat{\theta}^{uHB}$	$\hat{\theta}^{uEB}$	$\hat{\theta}^{uM}$
DE	21,860	23,103	$1,900^2$	21,185	21,787	21,031	21,088	21,802	21,025
MD	26,235	27,607	$1,722^2$	25,399	26,145	25,221	25,227	26,134	25,090
VA	24,160	25,514	$1,418^2$	23,418	24,080	23,264	23,403	24,040	23,262
WV	18,274	21,807	$1,380^2$	19,133	18,367	19,330	19,027	18,397	19,160
NC	20,296	21,408	$1,012^2$	19,634	20,223	19,472	19,849	20,133	19,712
SC	19,282	21,706	$1,795^2$	19,448	19,299	19,472	19,452	19,296	19,454
GA	22,687	22,599	$1,196^2$	21,217	22,524	20,842	21,510	22,402	21,199
FL	19,675	23,944	$1,042^2$	20,884	19,807	21,174	20,480	19,941	20,700
AL	17,978	22,233	$1,282^2$	19,273	18,119	19,575	19,047	18,187	19,264
KY	18,657	21,359	$1,285^2$	19,008	18,695	19,087	18,954	18,716	19,017
TN	19,776	21,240	$1,274^2$	19,351	19,729	19,239	19,430	19,707	19,350
MS	19,167	19,887	$1,762^2$	18,360	19,075	18,131	18,371	19,097	18,274
AR	18,917	20,214	$1,507^2$	18,388	18,858	18,250	18,452	18,859	18,383
LA	18,965	22,861	$1,444^2$	19,996	19,078	20,240	19,878	19,096	20,020
OK	19,295	23,668	$1,675^2$	20,578	19,436	20,894	20,535	19,418	20,673

balanced case, and $\hat{\theta}^{uEB}$ in the unbalanced case), the HB estimates ($\hat{\theta}^{bHB}$ in the balanced case, and $\hat{\theta}^{uHB}$ in the unbalanced case) and Morris' approximation to the HB estimates ($\hat{\theta}^{bM}$ in the balanced case, and $\hat{\theta}^{uM}$ in the unbalanced case). As noted before, the sampling variances are different for the states and the resulting Fay–Herriot model is an unbalanced model. To compare the frequentist and the Bayesian approaches for both the balanced and the unbalanced setup, we have illustrated the balanced Fay–Herriot model given by (2.1) by replacing each V_i by their average 2,162,469. From the last six columns of Table 1, we note that the point estimates of the small area means do not differ substantially either over EBLUP, HB or Morris' estimates, or if the setup is a balanced or an unbalanced Fay–Herriot model. It is usually our experience that the model-based small area point estimates are substantially robust over varying sampling variances or over the method of estimation, Bayes or frequentist.

In Table 2 we include various components of the uncertainty measures for the Prasad–Rao estimated MSE, the posterior variance of the HB estimates and Morris' approximation to the HB moments. From these components we can get the relevant overall uncertainty measure for the EBLUPs, the HB estimates and Morris' approximation of EB estimates. We note that in the balanced case the relative reduction in the Prasad–Rao estimated MSE over the sampling variance (the measure of uncertainty for the direct estimates) ranges between 21 and 62 percent. These numbers clearly show substantial gain in the accuracy of the model-based

estimates. In the unbalanced setup, these relative reductions range between −23 and 80 percent; only two states, NC and FL, have negative improvement, which is somewhat surprising. However, these two states being direct-use states in the CPS, perhaps they enjoy large sample size to produce relatively accurate direct estimates. Also, for these states, the g_3 term is relatively big resulting in a large estimated MSE of the EBLUP. The corresponding improvement numbers for the HB estimates are between 26 and 64 percent in the unbalanced case, and 28 and 58 percent in the balanced case. For Morris' approximation, these numbers are between 21 and 69 percent in the unbalanced case, and 30 and 72 percent in the balanced case.

In Table 2 we present the decomposition of the uncertainty corresponding to the three sources: uncertainty due to estimation of unknown small area mean, uncertainty due to estimation of the regression coefficients and uncertainty due to unknown variance components. We consider the mean squared error of an EBLUP (or EB predictor), the posterior variance and its approximation due to Morris (1983b) for both the balanced and an unbalanced setup. From this table we find that for each method of estimation and each setup, all the three components of uncertainty contribute substantially toward the overall measure of uncertainty for most small areas. Thus it is important to account for the uncertainty in estimating the regression coefficients and the variance components in deriving a reliable overall measure of uncertainty associated with the model-based small area estimates.

TABLE 2
Decomposition of various measures of uncertainty of the model-based small area estimates

State	Setup	g_{1HB}	g_{2HB}	g_{3HB}	g_{1M}	g_{2M}	g_{3M}	$\hat{g}_{1PR}$	$\hat{g}_{2PR}$	$\hat{g}_{3PR}$
DE	Balanced	792,210	97,376	44,113	459,930	120,989	124,922	418,657	142,987	268,279
	Unbalanced	1,129,602	121,110	39,030	937,376	116,803	75,548	377,191	139,692	194,412
MD	Balanced	792,210	696,375	66,040	459,930	865,242	187,017	418,657	1,022,559	268,279
	Unbalanced	1,039,678	858,915	103,302	828,078	915,537	168,214	409,823	1,157,267	229,034
VA	Balanced	792,210	295,960	51,495	459,930	367,728	145,828	418,657	434,588	268,279
	Unbalanced	863,313	277,090	77,930	656,777	332,909	142,517	485,759	453,886	310,056
WV	Balanced	792,210	106,701	71,651	459,930	132,575	202,908	418,657	156,680	268,279
	Unbalanced	839,191	84,658	67,832	636,503	100,218	144,901	497,472	135,006	322,621
NC	Balanced	792,210	125,904	43,600	459,930	156,435	123,472	418,657	184,878	268,279
	Unbalanced	583,606	60,730	43,268	447,884	85,908	98,855	639,733	139,644	477,672
SC	Balanced	792,210	110,835	2,312	459,930	137,711	6,548	418,657	162,750	268,279
	Unbalanced	1,077,706	126,061	1,807	872,021	125,390	3,518	395,558	152,055	213,886
GA	Balanced	792,210	91,351	218,486	459,930	113,503	618,727	418,657	134,140	268,279
	Unbalanced	716,079	57,972	235,098	540,917	74,935	508,347	562,280	108,923	392,592
FL	Balanced	792,210	134,749	144,408	459,930	167,426	408,947	418,657	197,866	268,279
	Unbalanced	605,744	73,260	125,413	462,990	103,414	293,078	626,452	166,675	462,923
AL	Balanced	792,210	94,683	163,838	459,930	117,643	463,970	418,657	139,033	268,279
	Unbalanced	774,894	66,728	156,347	585,124	82,616	342,043	530,290	115,370	357,952
KY	Balanced	792,210	128,846	11,885	459,930	160,090	33,657	418,657	189,198	268,279
	Unbalanced	776,906	93,056	13,089	586,680	115,699	26,677	529,228	162,238	356,806
TN	Balanced	792,210	136,323	18,482	459,930	169,381	52,340	418,657	200,177	268,279
	Unbalanced	769,515	97,600	16,381	580,981	122,136	37,950	533,140	172,245	361,029
MS	Balanced	792,210	269,811	68,948	459,930	335,238	195,254	418,657	396,191	268,279
	Unbalanced	1,060,717	318,630	40,995	852,009	327,641	98,442	401,844	405,527	220,560
AR	Balanced	792,210	239,946	28,598	459,930	298,131	80,987	418,657	352,337	268,279
	Unbalanced	918,002	229,443	18,346	705,154	260,944	45,703	460,413	342,328	282,932
LA	Balanced	792,210	92,944	104,354	459,930	115,483	295,519	418,657	136,480	268,279
	Unbalanced	879,552	79,807	86,164	670,775	91,992	191,249	478,058	121,522	301,805
OK	Balanced	792,210	118,713	164,196	459,930	147,500	464,983	418,657	174,318	268,279
	Unbalanced	1,014,326	130,894	113,081	800,405	138,730	255,182	419,730	174,242	239,566

In this example in the balanced Fay–Herriot setup the estimate of A obtained by Prasad–Rao or Morris' method is 16,1617, which is substantially smaller than the HB estimate given by 1,735,616. The latter estimate is more than ten times the frequentist estimate and it results from a very long tail of the posterior distribution of A. This larger HB estimate of A results in a substantially bigger value of the first component (the g_1 term) of the Bayesian measure than the corresponding component in the other measures. In fact the frequentist estimate of A is so small that, contrary to our expectation, for some areas the estimate of the g_1 term is not the dominant term in the estimated mean squared error (see the columns for Morris' approximation and Prasad–Rao estimates).

We notice that the picture does not change substantially when we consider the unbalanced setup. Here again, the posterior density of A has a long tail resulting in a posterior mean of 2,063,419. The Prasad–Rao estimate is again far too small, only 192,527, and Morris's estimate is in between, which is 515,969, much larger than the Prasad–Rao estimate but much smaller than the HB estimate.

We reiterate that all three components contribute substantially toward the overall measure of uncertainty. In particular, from the seventh and the eighth columns of Table 2, we note that the third term (the g_3 term) is bigger than the second term (the g_2 term) in 14 of the 30 rows. From the last two columns of the table, we note that the third term is bigger than the second term in

22 of the 30 rows. All these indicate that ignoring this component in the frequentist estimate of MSE or Morris' estimate will result in a severe underestimation. It is particularly so for the Prasad–Rao frequentist MSE since the first term (g_1 term) is also adjusted for bias by adding the g_3 term. Incidentally, the HB measure of uncertainty automatically accounts for all sources of uncertainty.

We conclude this section noting that here and in the previous section we assumed the sampling variances V_i's are known. This assumption was necessary to avoid the identifiability problem. If additional independent estimates (independent of Y_i's) of V_i's are available, and V_i's depend on a finite number of parameters, then the previous results can be extended to develop model-based small area estimates of the means and their measures of uncertainty. It can be done for both the EBLUP and HB approaches. This is essentially similar to the unit-level model considered briefly in Section 6. However, if V_i's cannot be assumed to depend on a finite number of parameters, then the mean squared approximation results presented here do not hold. In this scenario Wang and Fuller (2003) assumed that independent $\hat{V}_i$, $i = 1, \ldots, m$, are available which are unbiased for V_i. Assuming independent chi-squared distributions of these estimates, they derived MSE approximation of the EBLUP of θ_i. Their approximation is valid provided both m and d, the minimum of the degrees of freedom of the chi-squared distribution, are large. Their approximation to the MSE is accurate only to the order of $d^{-3/2}$. We refer to this article for details. Another related paper in this setup is by Rivest and Vandal (2004).

4. EXTENSIONS

The Fay–Herriot (1979) model discussed in the previous section can be extended in different directions. First, instead of $\mathbf{y}|\boldsymbol{\theta} \sim N(\boldsymbol{\theta}, \mathbf{G})$, where $\mathbf{G} = \mathrm{Diag}(V_1, \ldots, V_m)$, one can begin with $\mathbf{y}|\boldsymbol{\theta} \sim N(\boldsymbol{\theta}, \mathbf{V})$, where $\mathbf{V}$ is a known positive definite matrix which is not necessarily diagonal. The full model is thus

$$(4.1) \qquad \mathbf{y}|\boldsymbol{\theta} \sim N(\boldsymbol{\theta}, \mathbf{V}), \quad \boldsymbol{\theta} \sim N(\mathbf{X}\boldsymbol{\beta}, A\mathbf{I}_m).$$

Datta et al. (1992) considered this model in the context of adjustment of census undercounts. It is easy to check for $A(> 0)$ known, the BLUP (or the HB predictor with a flat prior for $\boldsymbol{\beta}$) is given by

$$(4.2) \qquad \tilde{\boldsymbol{\theta}}^B = (\mathbf{I}_m - \mathbf{B})\mathbf{y} + \mathbf{B}\mathbf{X}\tilde{\boldsymbol{\beta}}(A),$$

where $\mathbf{B} = (\mathbf{V} + A\mathbf{I}_m)^{-1}\mathbf{V}$ and $\tilde{\boldsymbol{\beta}}(A) = [\mathbf{X}^T(\mathbf{V} + A\mathbf{I}_m)^{-1}\mathbf{X}]^{-1}\mathbf{X}^T(\mathbf{V} + A\mathbf{I}_m)^{-1}\mathbf{y}$. With A unknown, one

can opt either for estimation of A from the marginal distribution of $\mathbf{y}$, namely, $N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V} + A\mathbf{I}_m)$ or put a flat prior for A, that is, $\pi(\boldsymbol{\beta}, A) = 1$. Datta et al. (1992) tried both methods in the context of adjustment of census counts based on 1988 Missouri Dress Rehearsal data, but found very little difference in the estimation of $\boldsymbol{\theta}$.

The work of Datta et al. (1992) is based on modeling the adjustment factors related to census counts. To be specific, let T_i denote the true count and the C_i the census count for the ith small area. Then Cressie (1989) and Isaki, Huang and Tsay (1991) proposed modeling $\theta_i = T_i/C_i$ $(i = 1, \ldots, m)$.

Direct estimates of these adjustment factors are usually obtained from a postenumeration survey (PES) conducted by the Bureau of the Census. In 1990, the Bureau of the Census produced PES estimates of the adjustment factors for 1,392 subdivisions (poststrata) of the total population. The PES sample contained approximately 377,000 persons in roughly 5,200 census blocks. However, prior to the 1990 census, the Census Bureau had a trial run for several test sites in Missouri to obtain direct estimates of these adjustment factors based on (purported) complete enumeration and PES. Datta et al. (1992) conducted an evaluation of this so-called Census Dress Rehearsal Data using the method described earlier in this section.

The HB and the EB estimators of $\boldsymbol{\theta}$ based on (4.2) are given respectively by

$$(4.3) \qquad \hat{\boldsymbol{\theta}}^{HB} = [\mathbf{I}_m - E(\mathbf{B}|\mathbf{y})]\mathbf{y} + E[\mathbf{B}\mathbf{X}\tilde{\boldsymbol{\beta}}(A)|\mathbf{y}],$$

$$(4.4) \qquad \hat{\boldsymbol{\theta}}^{EB} = [\mathbf{I}_m - \hat{\mathbf{B}}]\mathbf{y} + \hat{\mathbf{B}}\mathbf{X}\tilde{\boldsymbol{\beta}}(\hat{A}).$$

The posterior variance $V(\boldsymbol{\theta}|\mathbf{y})$, as before, is given by

$$\begin{aligned} V(\boldsymbol{\theta}|\mathbf{y}) = {} & \mathbf{V}\{\mathbf{I}_m - E(\mathbf{B}|\mathbf{y})\} \\ & + E[\mathbf{B}\mathbf{X}\{\mathbf{X}^T(\mathbf{V} + A\mathbf{I}_m)^{-1}\mathbf{X}\}^{-1}\mathbf{X}^T\mathbf{B}^T|\mathbf{y}] \\ & + \mathrm{Var}[\mathbf{B}\{\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}(A)\}|\mathbf{y}]. \end{aligned}$$

This was found numerically very similar to the *plug-in estimate* of the second-order approximate MSE given by

$$\begin{aligned} & E[\{\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}^{EB}\}\{\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}^{EB}\}^T] \\ & \approx \mathbf{V}(\mathbf{I}_m - \mathbf{B}) + \mathbf{B}\mathbf{X}\{\mathbf{X}^T(\mathbf{V} + A\mathbf{I}_m)^{-1}\}\mathbf{X}^T\mathbf{B}^T \\ & \quad + 2\mathbf{V}\mathbf{K}^3\mathbf{V}[\mathrm{tr}(\mathbf{V}^{-2})]^{-1}, \end{aligned}$$

where

$$\begin{aligned} \mathbf{K} = {} & (\mathbf{V} + A\mathbf{I}_m)^{-1} \\ & - (\mathbf{V} + A\mathbf{I}_m)^{-1}\mathbf{X}\{\mathbf{X}^T(\mathbf{V} + A\mathbf{I}_m)^{-1}\mathbf{X}\}^{-1} \\ & \cdot \mathbf{X}^T(\mathbf{V} + A\mathbf{I}_m)^{-1}. \end{aligned}$$

The study of Datta et al. (1992) revealed that for every poststratum, the EB (or EBLUP) and HB estimators of the adjustment factors outperformed the direct estimators.

There is also a multivariate extension of the Fay–Herriot (1979) model considered in Datta, Fay and Ghosh (1991). Now the data consist of $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m$, where each $\mathbf{y}_i$ is r-dimensional. Bivariate and trivariate versions of the model were used in Datta, Fay and Ghosh (1991), and later in Datta et al. (1996) to estimate median income of four-person families for the 50 states and the District of Columbia. They considered the random effects model

$$(4.5) \qquad \mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{u}_i + \mathbf{e}_i, \quad i = 1, \ldots, m,$$

where $\mathbf{u}_i \overset{i.i.d.}{\sim} N(\mathbf{0}, \mathbf{A})$ and $\mathbf{e}_i \overset{ind}{\sim} N(\mathbf{0}, \mathbf{V}_i)$, the $\mathbf{u}_i$ and the $\mathbf{e}_i$ being mutually independent, and the $V_i(>0)$ are known. Alternatively, in a Bayesian framework, writing $\boldsymbol{\theta}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{u}_i$ $(i = 1, \ldots, m)$, $\mathbf{y}_i | \boldsymbol{\theta}_i \overset{ind}{\sim} N(\boldsymbol{\theta}_i, \mathbf{V}_i)$ and $\boldsymbol{\theta}_i \overset{ind}{\sim} N(\mathbf{X}_i \boldsymbol{\beta}, \mathbf{A})$. Both EB (or EBLUP) and HB estimators of the $\boldsymbol{\theta}_i$ were found. These estimators were shown to outperform the direct estimators with respect to their precision measures.

5. CONFIDENCE INTERVALS IN SMALL AREA ESTIMATION

Morris (1983b) noted that although Stein's shrinkage estimators were widely used for point estimation, a lack of the availability of estimated uncertainty with these estimators delayed development of reliable confidence intervals. An early attempt to construct EB confidence intervals is due to Cox (1975). In the small area estimation terminology, he developed approximate confidence intervals that are accurate to the order of $O(m^{-1})$ for an individual small area mean θ_i for the balanced Fay–Herriot model without any covariate. Again in the small area estimation terminology, Morris (1983a, 1983b) was the first to consider confidence intervals for small area means for the Fay–Herriot model with covariates. He considered both the balanced and the unbalanced sampling variance cases. His method consists essentially in finding an HB confidence interval for θ_i, approximating (using Laplace approximations to integrals) this interval with estimates of the hyperparameters only at the last stage. He constructed these intervals using normal percentile points and provided a heuristic justification of these naive EB intervals. Later Laird and Louis (1987) proposed EB bootstrap confidence intervals in the spirit of Morris (1983a, 1983b), while Carlin and Gelfand (1990),

following a suggestion of Efron, proposed calibrating the naive EB confidence intervals. Indeed, in small area estimation setup, both for unit-level and area-level data, Prasad and Rao (1990) also suggested approximate confidence intervals for small area means. They based their intervals on normal percentile points and used their second-order unbiased estimator of the MSE of the EBLUP. As in Morris (1983a, 1983b), Prasad–Rao intervals also have a coverage error to the order of $O(m^{-1})$.

For the case when $V_1 = \ldots = V_m = V$ and A are both known, a flat prior for $\boldsymbol{\beta}$ will result in a $100(1 - \alpha)\%$ confidence interval of the form $(1 - B)y_i + B\mathbf{x}_i^T \hat{\boldsymbol{\beta}} \pm z_{\alpha/2}V^{1/2}(1 - B + Bm^{-1})^{1/2}$, where we may recall that $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1}\mathbf{X}^T \mathbf{y}$, the least squares estimator of $\boldsymbol{\beta}$. This result follows immediately from Lindley and Smith (1972). A naive EB confidence interval is given by $(1 - B)y_i + B\mathbf{x}_i^T \hat{\boldsymbol{\beta}} \pm z_{\alpha/2}V^{1/2}(1 - B)^{1/2}$, which does not take into account uncertainty due to estimation of $\boldsymbol{\beta}$. Accordingly, the coverage probability will fall short of the target under the said hierarchical model. The Type III bootstrap approach of Laird and Louis (1987) provides a confidence interval identical to the hierarchical Bayesian approach, where the bootstrap samples $\mathbf{y}$ are drawn from the $N(\mathbf{X}\hat{\boldsymbol{\beta}}, (V + A)\mathbf{I}_m)$ pdf. The same confidence interval is also arrived at by the conditional approach of Hill (1990). Hill's approach consists of finding the conditional distribution of $\theta_i - ((1 - B)y_i + B\mathbf{x}_i^T \hat{\boldsymbol{\beta}})$ given the ancillary statistic $U_i = y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}$. Also, it is pointed out by Laird and Louis (Theorem 2.1, page 743) that the Type III bootstrap can never match a hyperprior solution when A is unknown.

For the balanced Fay–Herriot model, Datta et al. (2002) developed an expansion for the coverage probability of confidence intervals derived by Morris (1983a, 1983b) and Prasad and Rao (1990). Based on this expansion they perturbed the endpoints of the confidence interval to achieve asymptotic coverage accurate to the order of $o(m^{-1})$. Also, following the framework of Hill (1990), Datta et al. (2002) studied conditional coverage probabilities of such intervals even for unknown A by conditioning on a suitable ancillary statistic. They obtained an expansion of the conditional coverage probability as well and used the expansion to better calibrate the interval. For some $d > p$ let $\hat{B}(S) \equiv \hat{B}_d(S) = (m - d)\min\{V/S, (m - p)^{-1}\}$. Assuming $\max_{1 \le i \le m} h_{ii} = O(m^{-1})$, where h_{ii} is defined in Theorem 3, they had for any fixed $t > 0$ the following expansion.

THEOREM 4.

$$P\big[\theta_1 \in \big(1 - \hat{B}(S)\big)Y_1 + \hat{B}(S)\mathbf{x}_1^T\hat{\boldsymbol{\beta}} \pm tV\big(1 - \hat{B}(S)\big)^{1/2}\big]$$

$$= 2\Phi(t) - 1$$

$$- t\phi(t)\bigg[\frac{(1+t^2)B^2}{2m(1-B)^2} + \frac{B}{1-B}\bigg\{h_{11} + \frac{5-d}{m}\bigg\}\bigg]$$

$$+ O(m^{-3/2}).$$

Let $z_{\alpha/2}$ denote the upper $\alpha/2$ point of $N(0,1)$ distribution. Taking $t = z_{\alpha/2}$ will result in an underestimation in the nominal coverage $1 - \alpha$. If we take

$$t^* = z_{\alpha/2}\bigg[1 + \frac{(1+z_{\alpha/2}^2)\hat{B}^2}{4m(1-\hat{B})^2} + \frac{(5-d+mh_{11})\hat{B}}{2m(1-\hat{B})}\bigg],$$

it follows that the interval $(1 - \hat{B}(S))Y_1 + \hat{B}(S)\mathbf{x}_1^T\hat{\boldsymbol{\beta}} \pm t^*V(1 - \hat{B}(S))^{1/2}$ has coverage probability equal to $1 - \alpha$ up to $O(m^{-3/2})$ error terms. Although this theorem is presented in the context of EB intervals, Datta et al. (2002) also discussed expansion of coverage probabilities of intervals that are created through the HB argument of Morris (1983a).

Extending the argument of Hill (1990), Datta et al. (2002) also obtained an expansion of the coverage probability of an EB confidence interval of θ_1 by conditioning on an ancillary statistic $U = (Y_1 - \mathbf{x}_1^T\hat{\boldsymbol{\beta}})\sqrt{(m-p)}/\sqrt{S}$. They proved the following theorem.

THEOREM 5.

$$P\big[\theta_1 \in \big(1 - \hat{B}(S)\big)Y_1 + \hat{B}(S)\mathbf{x}_1^T\hat{\boldsymbol{\beta}}$$

$$\pm tV\big(1 - \hat{B}(S)\big)^{1/2}\big|U\big]$$

$$= 2\Phi(t) - 1$$

$$- t\phi(t)\bigg[\frac{(1+t^2)B^2}{2m(1-B)^2} + \frac{(2U^2+3-d)B}{m(1-B)}$$

$$+ \frac{Bh_{11}}{1-B}\bigg]$$

$$+ O_p(m^{-3/2}).$$

The bias corrected confidence intervals for θ_1 are obtained as before with appropriate changes.

Datta et al. (2002) performed a simulation study to evaluate the performance of the approximate confidence intervals given in the two theorems above. In these simulations they used a simple setup with $m = 30$ small areas with no covariates. Since the coverage probability does not depend on $\boldsymbol{\beta}$, it was taken as zero in generating the samples. Also, the coverage probability depends only on B, so without any loss of generality V was taken to be 1. These authors considered

various values of B in the range 0.025 to 0.975. They computed both conditional and unconditional coverage probabilities as discussed in the theorems given above. They found little qualitative difference in performance between the unconditional and conditional coverage probabilities. They also noted that while the extent of underestimation of the coverage probabilities with $t = z_{\alpha/2}$ from the nominal level α was small for small B, the underestimation was severe for B in the upper half. On the other hand, the adjusted intervals appeared to be too large resulting in overestimation of the coverage probabilities. This overestimation is due to an overestimation of the mean squared error of the EB estimator of θ_1. Incidentally, Lahiri and Rao (1995) also noted similar overestimation of the MSE when B approaches 1, that is, when A/V approaches 0.

Smith (2001) in his unpublished Ph.D. dissertation developed EB confidence intervals for the ith small area mean θ_i for the more practical case of unbalanced Fay–Herriot model in (3.1). Associated with the EB or EBLUP $\hat{\theta}_i^{EB}$ of θ_i, let s_i^2 denote some estimated measure of uncertainty. Note that s_i^2 could be a second-order unbiased estimator of the MSE of $\hat{\theta}_i^{EB}$ as in (3.12) or something similar. For some estimator $\hat{A}$ of A, Smith (2001) defined $s_i^2 = h_i^2(\hat{A}) + c_i$, where $h_i^2(A) = g_{1i}(A) + g_{2i}(A)$. The term c_i is an $O_p(m^{-1})$ order term, that may depend on $\hat{A}$ and the data $\mathbf{Y}$, and may be related to g_{3i} term in (3.10) and bias term of $\hat{A}$. There are many possible choices corresponding to various MSE estimates. Rao (2001) proposed a number of area-specific estimators of the MSE of the EBLUP, and they can be included by proper choice of c_i. Alternatively, in the HB setup, c_i may include $(y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}})^2 \text{Var}(B_i(A)|\mathbf{y})$, which is an approximation to the last term in the posterior variance in (3.13). This general choice enabled Smith to study approximate coverage probabilities of confidence intervals constructed in Morris (1983a, 1983b) by using EB and HB methods. Corresponding to c_i, let the parametric function $c_i^*(A)$ be such that $c_i - c_i^*(A) = o_p(m^{-1})$. Also, define $q_i(A) = B_i^2(A)b_{\hat{A}}(A) + c_i^*(A) - 2g_{3i}(A)$, where $b_{\hat{A}}(A)$ is the asymptotic bias of $\hat{A}$. With the above notation we now state Theorem 1.7.1 of Smith (2001) below.

THEOREM 6. *For any $z > 0$,*

$$P[\hat{\theta}_i^{EB} - zs_i \leq \theta_i \leq \hat{\theta}_i^{EB} + zs_i]$$

$$= 2\Phi(z) - 1 + z\phi(z)\frac{q_i(A)}{h_i^2(A)}$$

$$- \frac{(z^3 + z)\phi(z)g_{3i}(A)}{4h_i^4(A)} \frac{D_i^2}{(A + D_i)} + o(m^{-1}).$$

Note that the leading term in the above expansion is the nominal coverage probability. The first-order error term in this expansion is of order $O(m^{-1})$. From this expansion it follows that as in Theorem 4 we can perturb the cut-off point z in order to achieve the nominal coverage probability to the order $o(m^{-1})$. Another point to note is that since the $O(m^{-1})$ term c_i (or equivalently, c_i^*) was not completely specified, for any given z we can choose $c_i^*(A)$ (depending on z and A) to make the $O(m^{-1})$ term in the expansion of the coverage probability disappear. In particular, the choice $c_i = c_i^*(\hat{A})$ with

$$c_i^*(A) = 2g_{3i}(A) - B_i(A)^2 b_{\hat{A}}(A) + \frac{(z^2 + 1)D_i}{4A} g_{3i}(A)$$

will give an EB confidence interval that matches the nominal coverage probability to the order of $o(m^{-1})$.

In this section we have considered confidence intervals for individual small area means, which is the current state of the literature. In the early applications of small area estimation, practitioners were only interested in point estimates (see, e.g., Fay and Herriot, 1979). Only in the last twenty years or so, substantial development of the measures of uncertainty of the model-based estimates of small area means has taken place. Construction of appropriate confidence intervals for small area means is still limited and is restricted only to individual means. While in the EB setup confidence sets for several population means have been considered, this problem is not fully addressed yet in small area estimation. In a recent article, Ganesh (2009) has considered simultaneous credible intervals in small area estimation. However, calibrated confidence sets for multiple small area means in EB or EBLUP approach have not been studied yet.

6. OTHER IMPORTANT DEVELOPMENTS IN SMALL AREA ESTIMATION

We mentioned in the Introduction that both area-level and unit-level data are available in small area estimation. In the previous sections we have concentrated mostly on area-level models. In this section we review some of the results for unit-level models. For a unit-level model let y_{ij} denote the value for the jth unit in the ith small area, with $j = 1, \ldots, N_i$, $i = 1, \ldots, m$, where N_i is the size of the finite population corresponding to the ith small area. Let $\gamma_i = N_i^{-1} \sum_{j=1}^{N_i} y_{ij}$ denote the finite population mean for

the ith small area. For notational simplicity let y_{ij}, $j = 1, \ldots, n_i, i = 1, \ldots, m$ denote values of the characteristic of the sampled units from these m small areas. Let the vector $\mathbf{y}(s)$ denote all the sampled values. A direct estimator of γ_i based on the ith area sample mean $\bar{Y}_{is}$ is usually less reliable due to a small sample size n_i. To borrow strength from the neighboring areas through shrinkage estimation the following model, known as the nested-error regression model, has been found very useful for unit-level data. The model is given by

$$Y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + v_i + e_{ij},$$
(6.1)
$$j = 1, \ldots, N_i, i = 1, \ldots, m,$$

where $\mathbf{x}_{ij}$ is a p-component vector of auxiliary variables, v_i and e_{ij} are independently distributed with $v_i \overset{i.i.d.}{\sim} N(0, \sigma_v^2)$ and $e_{ij} \overset{ind}{\sim} N(0, \sigma_e^2)$, $j = 1, \ldots, N_i$, $i = 1, \ldots, m$. We denote the observations for the sampled units in the ith small area by $\mathbf{Y}_i^{(1)} = (Y_{i1}, \ldots, Y_{in_i})^T$. Similarly, $\mathbf{Y}_i^{(2)}$ is used to denote the vector of observations corresponding to the unsampled units in the ith small area. Battese, Harter and Fuller (1988) and Prasad and Rao (1990) used this model to develop EBLUP estimate of finite population mean γ_i. They approximated γ_i for large N_i by $\theta_i = \bar{\mathbf{X}}_i^T \boldsymbol{\beta} + v_i$ and used the predictor of θ_i to estimate γ_i. Here $\bar{\mathbf{X}}_i = N_i^{-1} \sum_{j=1}^{N_i} \mathbf{x}_{ij}$ is the known mean vector of the auxiliary variables.

Let $\mathbf{Y}^{(1)}$ be obtained by stacking the vectors $\mathbf{Y}_i^{(1)}$ for all the m small areas. Similarly, denote by $\mathbf{X}^{(1)}$ the matrix of p columns obtained by stacking the $\mathbf{x}_{ij}$'s corresponding to the sampled units. We also denote the variance of $\mathbf{Y}^{(1)}$ by $\boldsymbol{\Sigma}_{11}$. From Prasad and Rao (1990) the BLUP of θ_i is obtained as

$$(6.2) \quad \tilde{\theta}_i(\boldsymbol{\psi}, \mathbf{Y}^{(1)}) = \bar{\mathbf{X}}_i^T \tilde{\boldsymbol{\beta}} + \delta_i(\bar{Y}_{is} - \bar{\mathbf{x}}_{is}^T \tilde{\boldsymbol{\beta}}),$$

where $\boldsymbol{\psi} = (\sigma_v^2, \sigma_e^2)$, and

$$(6.3) \quad \tilde{\boldsymbol{\beta}} = (\mathbf{X}^{(1)T} \boldsymbol{\Sigma}_{11}^{-1} \mathbf{X}^{(1)})^{-1} \mathbf{X}^{(1)T} \boldsymbol{\Sigma}_{11}^{-1} \mathbf{Y}^{(1)}$$

is the generalized least squares estimator of $\boldsymbol{\beta}$. Here $\delta_i = \sigma_v^2 (\sigma_v^2 + \sigma_e^2 n_i^{-1})^{-1}$ is the shrinkage coefficient which shrinks the direct estimator $\bar{Y}_{is}$ of γ_i (or θ_i) toward a regression surface.

Under the superpopulation model given by (6.1), from Prasad and Rao (1990) and Datta and Ghosh (1991b) one can show that the BLUP of the finite population mean γ_i under the nested error regression model is given by

$$(6.4) \quad \tilde{\gamma}_i(\boldsymbol{\psi}, \mathbf{Y}^{(1)}) = f_i \bar{Y}_{is} + (1 - f_i)\tilde{\theta}_{i(u)}(\boldsymbol{\psi}, \mathbf{Y}^{(1)}),$$

where $f_i = n_i/N_i$, $\tilde{\theta}_{i(u)}(\boldsymbol{\psi}, \mathbf{Y}^{(1)})$ is given by (6.2), with $\bar{\mathbf{X}}_i$ replaced by $\bar{\mathbf{x}}_{i(u)}$, the mean of $\mathbf{x}_{ij}$'s for the $N_i - n_i$ unsampled units from the ith area. The BLUP of the small area mean γ_i usually depends on variance components, which in practice will be unknown. Estimates of variance components $\boldsymbol{\psi}$ are plugged in to the BLUP to obtain EBLUP estimates. The variance components are estimated from the marginal distribution (by integrating out v_i's) of the data, $\mathbf{Y}^{(1)}$.

While Datta and Lahiri (2000) suggested ML and REML estimation of the variance components, Prasad and Rao (1990) used ANOVA methods to obtain unbiased estimators for variance components in the nested error regression model. Prasad and Rao (1990) first obtained $\hat{e}_{ij}$, $\hat{u}_{ij}$, $j = 1, \ldots, n_i$, $i = 1, \ldots, m$, where $\{\hat{e}_{ij}, j = 1, \ldots, n_i, i = 1, \ldots, m\}$ are the residuals from the ordinary least squares regression of $Y_{ij} - \bar{Y}_{is}$ on $\{\mathbf{x}_{ij} - \bar{\mathbf{x}}_{is}\}$ and $\hat{u}_{ij}$ are the residuals from the ordinary least squares regression of Y_{ij} on $\mathbf{x}_{ij}$. Estimators

$$\hat{\sigma}_e^2 = (n - m - p^*)^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} \hat{e}_{ij}^2, \quad \text{and}$$

(6.5)

$$\hat{\sigma}_v^2 = n_*^{-1} \left[\sum_{i=1}^m \sum_{j=1}^{n_i} \hat{u}_{ij}^2 - (n - p)\hat{\sigma}_e^2 \right]$$

are unbiased, where $n_* = n - \text{tr}[(\mathbf{X}^{(1)T}\mathbf{X}^{(1)})^{-1} \cdot \sum_{i=1}^m n_i^2 \bar{\mathbf{x}}_{is}\bar{\mathbf{x}}_{is}^T]$, and p^* is equal to the number of linearly independent vectors in the set $\{\mathbf{x}_{ij} - \bar{\mathbf{x}}_{is}, j = 1, \ldots, n_i, i = 1, \ldots, m\}$.

Second-order accurate approximations to MSE of the EBLUP of θ_i were developed by Prasad and Rao (1990) and Datta and Lahiri (2000). These authors showed for the nested error regression model the three terms in the approximation [cf. (3.10)] are

$$g_{1i}(\boldsymbol{\psi}) = (1 - \delta_i)\sigma_v^2,$$

(6.6) $\quad g_{2i}(\boldsymbol{\psi}) = (\bar{\mathbf{X}}_i - \delta_i\bar{\mathbf{x}}_{is})^T (\mathbf{X}^{(1)T}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{\psi})\mathbf{X}^{(1)})^{-1}$

$$\cdot (\bar{\mathbf{X}}_i - \delta_i\bar{\mathbf{x}}_{is}),$$

$$g_{3i}(\boldsymbol{\psi}) = n_i^{-2}(\sigma_v^2 + \sigma_e^2/n_i)^{-3}$$

(6.7)

$$\cdot \text{var}(\sigma_v^2\hat{\sigma}_e^2 - \sigma_e^2\hat{\sigma}_v^2).$$

For an estimator $\hat{\boldsymbol{\psi}}$ of $\boldsymbol{\psi}$, from Prasad and Rao (1990) and Datta and Lahiri (2000) a second-order unbiased estimator of the MSE of the EBLUP of θ_i is given by

$$mse(\tilde{\theta}_i(\hat{\boldsymbol{\psi}})) = g_{1i}(\hat{\boldsymbol{\psi}}) + g_{2i}(\hat{\boldsymbol{\psi}}) + 2g_{3i}(\hat{\boldsymbol{\psi}})$$

(6.8)

$$- \mathbf{b}^T(\hat{\boldsymbol{\psi}}; \hat{\boldsymbol{\psi}})\nabla g_{1i}(\hat{\boldsymbol{\psi}}),$$

where $\mathbf{b}^T(\hat{\boldsymbol{\psi}}; \boldsymbol{\psi})$ is the asymptotic bias of $\hat{\boldsymbol{\psi}}$, and $\nabla g_{1i}(\boldsymbol{\psi})$ is the gradient vector of $g_{1i}(\boldsymbol{\psi})$. For estimators of variance components with asymptotic bias of $o(m^{-1})$, the last term in (6.8) drops out. This happens for the ANOVA estimators suggested by Prasad and Rao (1990) and the REML estimators considered by Datta and Lahiri (2000).

Estimation of the MSE of EBLUP outlined above and in Section 3 is based on Taylor's expansion. Alternatively, a resampling-based approach may be used to estimate the MSE. Laird and Louis (1987) suggested a bootstrap measure of accuracy of the EB estimator for the Fay–Herriot model. Subsequently, Butar and Lahiri (2003) adopted their approach in small area estimation. Further references to this literature may be found in Pfeffermann and Tiller (2005), Lahiri (2003) and Hall and Maiti (2006). Jiang, Lahiri and Wan (2002) proposed jackknife methods to estimate the MSE of the EBLUP.

Datta and Ghosh (1991b) proposed a general HB model for unit-level data in small area estimation. Some earlier Bayesian analysis for two-stage sampling in a simpler framework is due to Scott and Smith (1969), with subsequent extension to the multistage sampling by Malec and Sedransk (1985). Based on the superpopulation approach to finite population sampling Datta and Ghosh (1991b) developed HB estimates of small area means by deriving certain predictive distributions. To that objective, they considered the following HB model:

(A) Conditional on $\boldsymbol{\beta}, \boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_t)^T$ and r, let

$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, r^{-1}(\boldsymbol{\Psi} + \mathbf{Z}\mathbf{D}(\lambda)\mathbf{Z}^T)),$$

where $\mathbf{Y}$ is $N \times 1$ vector of characteristics of all the N units in the finite population, $\mathbf{X}$ and $\mathbf{Z}$ are $N \times p$ and $N \times q$ matrices, respectively, for appropriate known p and q.

(B) $\boldsymbol{\beta}, r$ and $\boldsymbol{\lambda}$ have a certain joint prior distribution.

Stage (A) of the above model can be identified as a general mixed linear model (cf. Datta and Ghosh, 1991b). To see this, write

(6.9) $\qquad \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v} + \mathbf{e},$

where $\mathbf{e}$ and $\mathbf{v}$ are mutually independent with $\mathbf{e} \sim N(\mathbf{0}, r^{-1}\boldsymbol{\Psi})$, and $\mathbf{v} \sim N(\mathbf{0}, r^{-1}\mathbf{D}(\lambda))$. Here $\mathbf{e}$ is $N \times 1$, and $\mathbf{v}$ is $q \times 1$ vector of random effects, $\boldsymbol{\Psi}$ is a known positive definite matrix and $\mathbf{D}(\lambda)$ is a $q \times q$ p.d. matrix which is known except for $\boldsymbol{\lambda}$.

In the context of finite population $\mathbf{Y}$ is partitioned as $\mathbf{Y}^T = (\mathbf{Y}^{(1)T}, \mathbf{Y}^{(2)T})$, where $\mathbf{Y}^{(1)}$ corresponds to

the sampled units and $\mathbf{Y}^{(2)}$ corresponds to the unsampled units. Similarly, the design matrices $\mathbf{X}$ and $\mathbf{Z}$ are partitioned. To make inference about certain functions of $\mathbf{Y}$, the Bayesian solution is obtained by deriving the predictive distribution of $\mathbf{Y}^{(2)}$ given $\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}$ (which is the posterior distribution of $\mathbf{Y}^{(2)}$). In small area estimation the vector of sampled units $\mathbf{Y}^{(1)}$ is from m small areas. If $\mathbf{Y}_i^{(1)}$ is the $(n_i \times 1)$ vector of sampled units from the ith small area, then $\mathbf{Y}^{(1)T} = (\mathbf{Y}_1^{(1)T}, \ldots, \mathbf{Y}_m^{(1)T})$. Similarly, the vector $\mathbf{Y}^{(2)}$ corresponding to the unsampled units can be partitioned. The finite population mean γ_i from small area i is a linear function of $\mathbf{Y}^{(2)}$, and its predictive distribution may be derived from the distribution of $\mathbf{Y}^{(2)}$. In particular, based on a quadratic loss function, the HB estimator is given by the posterior mean of γ_i, and a measure of uncertainty is given by the posterior variance of γ_i. While the solution for the general HB model is presented in Datta and Ghosh (1991b), we now spell out below some of the details for the nested error regression model.

For the nested error regression model in (6.1), $t = 1$, $r = \sigma_e^{-2}$ and $\lambda_1 = \sigma_e^2/\sigma_v^2$. To complete the HB model, Datta and Ghosh (1991b) assigned independent prior distribution on $\boldsymbol{\beta}$, σ_e^2 and σ_v^2. They put a uniform prior over R^p for $\boldsymbol{\beta}$, and $\sigma_e^2 \sim IG(a_0/2, g_0/2)$ and $\sigma_v^2 \sim IG(a_1/2, g_1/2)$, where $IG(\beta, \alpha)$ is a distribution whose pdf is proportional to $\exp(-\beta/x)x^{-\alpha-1}$. Quantities a_0, g_0, g_1 are nonnegative and a_1 is positive, and are chosen suitably small to reflect diffused prior information on the variance components.

The HB estimates for any reasonably complex model do not admit any closed-form expressions, and they are evaluated by numerical computations. Required posterior moments can be found either by Gibbs sampling (cf. Gelfand and Smith, 1990) or by numerical integration. Using formulas for iterated expectation and variance, Datta and Ghosh (1991b) have shown that the posterior mean and the posterior variance can be computed by evaluating several one-dimensional integrals with respect to the posterior density of λ_1. In particular, the HB estimate of γ_i is

$$\hat{\gamma}_i^{HB} = E[\tilde{\gamma}_i(\lambda_1, \mathbf{y}^{(1)})|\mathbf{y}^{(1)}],$$

where the expectation $E[\cdot|\mathbf{y}^{(1)}]$ is with respect to the posterior density of λ_1, and $\tilde{\gamma}_i(\lambda_1, \mathbf{y}^{(1)})$ (with a slight abuse of notation) is the same as the expression of $\tilde{\gamma}_i(\boldsymbol{\psi}, \mathbf{Y}^{(1)})$ given in (6.4). Note that the above HB estimate of γ_i is obtained by shrinking the direct small area estimator $\bar{Y}_{is}$ to an estimated regression surface. Similarly, the posterior variance of γ_i can also be computed

by numerical integration involving one-dimensional integrals. Alternatively, the Gibbs sampling can also be implemented very easily for the present model. Indeed Datta and Ghosh (1991b) have shown that the set of complete conditional distributions are given by either multivariate normal or inverse gamma distributions.

7. OTHER SMALL AREA ESTIMATORS

7.1 Measurement Error Models

In our presentation of the unit-level model, we have assumed so far that the covariates are measured without error. However, sometimes it is not possible to obtain exact measurements of these covariates. For example, if in prediction of certain crop yield, the nitrogen level in the soil is a covariate, this covariate needs to be determined by analysis of soil sample. This will result in measurement error of the covariate. For the nested error regression model with a single covariate with measurement error Ghosh and Sinha (2007), Ghosh, Sinha and Kim (2006) and Torabi, Datta and Rao (2009) have considered estimation of small area means. While Ghosh and Sinha (2007) used a functional measurement error model, Ghosh, Sinha and Kim (2006) and Torabi, Datta and Rao (2009) considered a structural measurement error model for estimation of small area means. Ghosh, Sinha and Kim (2006) and Torabi, Datta and Rao (2009) used the model given by

$$
\begin{aligned}
& y_{ij} = \beta_0 + \beta_1 x_i + v_i + e_{ij}; \\
& j = 1, \ldots, n_i; i = 1, \ldots, m,
\end{aligned}
$$

(7.1)

where as before y_{ij} is the response variable of the jth unit in the ith area (or stratum), x_i is the unknown true area-specific covariate associated with y_{ij}. Further, $v_i \overset{i.i.d.}{\sim} N(0, \sigma_v^2)$ and independent of $e_{ij} \overset{i.i.d.}{\sim} N(0, \sigma_e^2)$. Under measurement errors, $X_{ij}(= x_i + u_{ij})$ are observed, where $u_{ij} \overset{i.i.d.}{\sim} N(0, \sigma_u^2)$. They also assumed that $x_i \overset{i.i.d.}{\sim} N(\mu_x, \sigma_x^2)$. The vector of model parameters is given by $\theta = (\beta_0, \beta_1, \mu_x, \sigma_x^2, \sigma_u^2, \sigma_v^2, \sigma_e^2)^T$, and x_i, v_i, e_{ij} and u_{ij} are assumed to be mutually independent.

Based on the preceding model Ghosh, Sinha and Kim (2006) obtained the EB predictor of γ_i by replacing the model parameters by their estimates in the Bayes estimator of γ_i based on the conditional distribution of $Y_{ij}, j = n_i + 1, \ldots, N_i$, given θ and $y_{ij}, j = 1, \ldots, n_i$. Since X_{ij}'s are also stochastic Torabi, Datta and Rao (2009) instead first derived the fully efficient Bayes estimator of γ_i based on the conditional distribution of $Y_{ij}, j = n_i + 1, \ldots, N_i$, given $\theta, y_{ij}, j =$

$1, \ldots, n_i$, and x_{ij}, $j = 1, \ldots, n_i$. Finally, they obtained an EB estimate of γ_i by replacing θ, the model parameters by their estimates as given in Ghosh, Sinha and Kim (2006). Torabi, Datta and Rao (2009) employed the jackknife method to obtain an estimate of mean squared prediction error (MSPE) of the EB predictor. For further details we refer to these two papers.

7.2 Generalized Linear Models

Until now we have considered small area estimation problems only for continuous-valued response. However, often in practice, response variables are binary or categorical. For example, in the SAIPE program, U.S. Census Bureau is interested in estimating the poverty rates among school children. The response variable here is binary taking values 1 and 0 depending on whether the child is in poverty or not. More generally, the response variable may take values in multiple categories. Again, in the disease mapping context, the response is typically the number of occurrences of a rare event. Generalized linear models are needed for the analysis of this kind of data.

Both empirical and hierarchical Bayesian approaches have played an important role in developing small area estimates for discrete data. Dempster and Tomberlin (1980), Farrell, MacGibbon and Tomberlin (1997) and MacGibbon and Tomberlin (1989) have obtained small area estimates of proportions based on EB techniques. A general EB formulation for simultaneous estimation of means from the natural exponential family quadratic variance function family of distributions is due to Ghosh and Maiti (2004). They provided also estimated mean squared errors of the small area estimators. Earlier, for the binary case, Jiang (1998) and Jiang and Zhang (2001) obtained such mean squared error estimators based on the jackknife approach. On the other hand, a general hierarchical Bayesian approach based on generalized linear models in the small area estimation context is due to Ghosh et al. (1998).

7.3 Balanced Loss Functions

HB and EB estimators in the small area context are mostly derived under squared error loss. As an alternative, Ghosh, Kim and Kim (2008) considered the balanced loss introduced and made popular by Zellner (1988, 1994). For simplicity, we go back to the framework of Section 2 where we considered small area models with equal number of observations within each area. For an arbitrary estimator $\mathbf{T} = (T_1, \ldots, T_m)^T$ of θ, the balanced loss is given by $L(\theta, \mathbf{T}) = m^{-1}[w\|\mathbf{y} - \mathbf{T}\|^2 + (1-w)\|\mathbf{T} - \theta\|^2]$, where $\|\cdot\|$ is the Euclidean

norm and $w \in [0, 1]$ is the known weight. The choice of w reflects the relative weight which the experimenter wants to assign to goodness of fit and precision of estimation. The extreme cases $w = 0$ and $w = 1$ refer solely to precision of an estimate and goodness of fit, respectively.

Under the balanced loss with a flat prior for β, it follows from Section 2 that the Bayes estimator of θ is $\hat{\theta}_{\text{BAL}}^B = [1 - (1-w)B]\mathbf{y} + (1-w)B\mathbf{P_X}\mathbf{y}$ with corresponding Bayes risk $m^{-1}E\|\hat{\theta}_{\text{BAL}}^B - \theta\|^2 = V[(1 - B) + bw^2(m-p)/m]$. An EB estimator is obtained by substituting the same estimator $\hat{B}^{EB} = V(m - p - 2)/S$ or $(\hat{B}^{EB})^+ = \min(\hat{B}^{EB}, 1)$ of B as given in Section 2, where we may recall that $S = \|\mathbf{y} - \mathbf{P_X}\mathbf{y}\|^2$. The calculation of the Bayes risk of the resulting EB estimator is similar to that in Section 2. The details are omitted. The special case of the intercept model where $\mathbf{x}_i^T\beta = \mu$ for all i was considered in Ghosh, Kim and Kim (2007, 2008). These authors also considered constrained Bayes estimators along the lines of Louis (1984) and Ghosh (1992a).

8. SUMMARY AND FUTURE RESEARCH

The paper reviews several normal theory-based small area estimation techniques. In particular, the role of shrinkage estimation in the small area context is highlighted, and different variants of Stein-type shrinkers are discussed. Both hierarchical and empirical Bayesian methods are presented in the context of mixed linear models for unbalanced data, and are illustrated with specific small area problems. Empirical Bayes confidence intervals based on hierarchical normal models are provided. Extensions of these results to measurement error models and generalized linear models are also touched upon.

There are several promising areas of future research. As mentioned earlier, small area estimation needs explicit, or at least implicit, use of models. These model-based estimates can differ widely from the direct estimates, especially for areas with very low sample sizes. One potential drawback of the model-based estimates is that when aggregated, the overall estimate for a larger geographical area may be quite different from the corresponding direct estimate, the latter being usually believed to be quite reliable. This is because the original survey was designed to achieve specified inferential accuracy at this higher level of aggregation. The problem can become more severe in the event of model failure as often there is no real check for the validity of the assumed model. Moreover, this overall

agreement with the direct estimates may sometimes be politically necessary to convince the legislators of the utility of small area estimates.

One way to avoid this problem is the so-called "benchmarking approach" which amounts to modifying these model-based estimates so that one gets the same aggregate estimate for the larger geographical area. A simple illustration is to modify the model-based county-level estimates so that one matches the state-level direct estimate. Currently the most popular approach is the so-called "raking" method which involves multiplying all the small area estimates by a constant factor so that the weighted total agrees with the direct estimate. Clearly, this is an ad hoc procedure with very little statistical foundation.

It appears that constrained Bayes small area estimates (Louis, 1984; Ghosh, 1992b) will be particularly appropriate to achieve this end. Instead of matching the first two moments from the empirical histogram of Bayes estimates with those from the posterior histogram of the parameters as in Louis (1984) or Ghosh (1992a), one should require that the aggregate or some weighted aggregate of these small area estimates should equal the large area aggregate estimate. This can possibly be achieved even for fairly complex models. See also Shen and Louis (1998).

The other interesting issue is to extend the measurement error model much further so that one can even handle discrete data and also more complex normal theory models.

ACKNOWLEDGMENTS

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. *J. Amer. Statist. Assoc.* **83** 28–36.

BOOTH, J. G. and HOBERT, J. P. (1998). Standard errors of prediction in generalized linear mixed models. *J. Amer. Statist. Assoc.* **93** 262–272. MR1614632

BUTAR, F. B. and LAHIRI, P. (2003). On measures of uncertainty of empirical Bayes small-area estimators. *J. Statist. Plann. Inference* **112** 63–76. MR1961720

CARLIN, B. P. and GELFAND, A. E. (1990). Approaches for empirical Bayes confidence intervals. *J. Amer. Statist. Assoc.* **85** 105–114. MR1137356

COX, D. R. (1975). Prediction intervals and empirical Bayes confidence intervals. In *Perspectives in Probability and Statistics* (*Papers in Honour of M. S. Bartlett on the Occasion of His 65th Birthday*) (J. Gani, ed.) 47–55. Applied Probability Trust, Sheffield, UK. MR0403046

CRESSIE, N. (1989). Empirical Bayes estimation of undercount in the decennial census. *J. Amer. Statist. Assoc.* **84** 1033–1044. MR1134491

DATTA, G. S. (1992). A unified Bayesian prediction theory for mixed linear models with application. *Statist. Decisions* **10** 337–365. MR1200694

DATTA, G. S. (2009). Model-based approach to small area estimation. In *Handbook of Statistics*: *Sample Surveys*: *Inference and Analysis*, *Volume 29B* (D. Pfeffermann and C. R. Rao, eds.) 251–288. North-Holland, Amsterdam. MR0403046

DATTA, G. S., FAY, R. E. and GHOSH, M. (1991). Hierarchical and empirical multivariate Bayes analysis in small area estimation. In *Proceedings of the Seventh Annual Research Conference of the Bureau of the Census* 63–79. U.S. Department of Commerce, Washington, DC.

DATTA, G. S. and GHOSH, M. (1991a). Asymptotic optimality of hierarchical Bayes estimators and predictors. *J. Statist. Plann. Inference* **29** 229–243. MR1144170

DATTA, G. S. and GHOSH, M. (1991b). Bayesian prediction in linear models: Applications to small area estimation. *Ann. Statist.* **19** 1748–1770. MR1135147

DATTA, G. S., GHOSH, M., HUANG, E., ISAKI, C., SCHULTZ, L. and TSAY, J. (1992). Hierarchical and empirical Bayes methods for adjustment of census undercount: The 1988 Missouri dress rehearsal data. *Survey Methodology* **18** 95–108.

DATTA, G. S., GHOSH, M., NANGIA, N. and NATARAJAN, K. (1996). Estimation of median income of four-person families: A Bayesian approach. In *Bayesian Analysis in Statistics and Econometrics*. (D. A. Berry, K. M. Chaloner and J. K. Geweke, eds.) 129–140. Wiley, New York. MR1367306

DATTA, G. S., GHOSH, M., SMITH, D. D. and LAHIRI, P. (2002). On an asymptotic theory of conditional and unconditional coverage probabilities of empirical Bayes confidence intervals. *Scand. J. Stat.* **29** 139–152. MR1894387

DATTA, G. S., KUBOKAWA, T., MOLINA, I. and RAO, J. N. K. (2011). Estimation of mean squared error of model-based small area estimators. *TEST* **20** 367–388.

DATTA, G. S. and LAHIRI, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statist. Sinica* **10** 613–627. MR1769758

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. MR2158619

DEMPSTER, A. P. and TOMBERLIN, T. J. (1980). The analysis of census undercount from a post-enumeration survey. In *Proceedings of the Conference on Census Undercount* 88–94. U.S. Department of Commerce, Washington, DC.

EFRON, B. and MORRIS, C. (1973). Stein's estimation rule and its competitors—an empirical Bayes approach. *J. Amer. Statist. Assoc.* **68** 117–130. MR0388597

FARRELL, P. J., MACGIBBON, B. and TOMBERLIN, T. J. (1997). Empirical Bayes estimators of small area proportions in multistage designs. *Statist. Sinica* **7** 1065–1083.

FAY, R. E. III and HERRIOT, R. A. (1979). Estimates of income for small places: An application of James–Stein procedures to census data. *J. Amer. Statist. Assoc.* **74** 269–277. MR0548019

GANESH, N. (2009). Simultaneous credible intervals for small area estimation problems. *J. Multivariate Anal.* **100** 1610–1621. MR2535373

GELFAND, A. E. and SMITH, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85** 398–409. MR1141740

GHOSH, M. (1992a). Hierarchical and empirical Bayes multivariate estimation. In *Current Issues in Statistical Inference*: *Essays in Honor of D. Basu. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **17** 151–177. IMS, Hayward, CA. MR1194415

GHOSH, M. (1992b). Constrained Bayes estimation with applications. *J. Amer. Statist. Assoc.* **87** 533–540. MR1173817

GHOSH, M., KIM, M. J. and KIM, D. (2007). Constrained Bayes and empirical Bayes estimation with balanced loss functions. *Comm. Statist. Theory Methods* **36** 1527–1542. MR2396441

GHOSH, M., KIM, M. J. and KIM, D. H. (2008). Constrained Bayes and empirical Bayes estimation under random effects normal ANOVA model with balanced loss function. *J. Statist. Plann. Inference* **138** 2017–2028. MR2406422

GHOSH, M. and MAITI, T. (2004). Small-area estimation based on natural exponential family quadratic variance function models and survey weights. *Biometrika* **91** 95–112. MR2050462

GHOSH, M., NATARAJAN, K., STROUD, T. W. F. and CARLIN, B. P. (1998). Generalized linear models for small-area estimation. *J. Amer. Statist. Assoc.* **93** 273–282. MR1614644

GHOSH, M. and RAO, J. N. K. (1994). Small area estimation: An appraisal. *Statist. Sci.* **9** 55–93. MR1278679

GHOSH, M. and SINHA, K. (2007). Empirical Bayes estimation in finite population sampling under functional measurement error models. *J. Statist. Plann. Inference* **137** 2759–2773. MR2323789

GHOSH, M., SINHA, K. and KIM, D. (2006). Empirical and hierarchical Bayesian estimation in finite population sampling under structural measurement error models. *Scand. J. Stat.* **33** 591–608. MR2298067

HALL, P. and MAITI, T. (2006). On parametric bootstrap methods for small area prediction. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 221–238.

HARVILLE, D. A. (1990). BLUP (best linear unbiased estimation) and beyond. In *Advances in Statistical Methods for Genetic Improvement of Livestock* (D. Gianola and K. Hammond, eds.) 239–276. Springer, New York.

HENDERSON, C. R. (1953). Estimation of variance and covariance components. *Biometrics* **9** 226–252. MR0055650

HILL, J. R. (1990). A general framework for model-based statistics. *Biometrika* **77** 115–126. MR1049413

ISAKI, C. T., HUANG, E. T. and TSAY, J. H. (1991). Smoothing adjustment factors from the 1990 post enumeration survey. In *Proceedings of the Social Statistics Section* 338–343. Amer. Statist. Assoc., Alexandria, VA.

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. In *Proc.* 4*th Berkeley Sympos. Math. Statist. and Prob.*, *Vol. I* 361–379. Univ. California Press, Berkeley, CA. MR0133191

JIANG, J. (1998). Consistent estimators in generalized linear mixed models. *J. Amer. Statist. Assoc.* **93** 720–729. MR1631373

JIANG, J., LAHIRI, P. and WAN, S.-M. (2002). A unified jackknife theory for empirical best prediction with *M*-estimation. *Ann. Statist.* **30** 1782–1810. MR1969450

JIANG, J. and ZHANG, W. (2001). Robust estimation in generalised linear mixed models. *Biometrika* **88** 753–765. MR1859407

KACKAR, R. N. and HARVILLE, D. A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *J. Amer. Statist. Assoc.* **79** 853–862. MR0770278

LAHIRI, P. (2003). On the impact of bootstrap in survey sampling and small-area estimation. *Statist. Sci.* **18** 199–210. MR2019788

LAHIRI, P. and RAO, J. N. K. (1995). Robust estimation of mean squared error of small area estimators. *J. Amer. Statist. Assoc.* **90** 758–766. MR1340527

LAIRD, N. M. and LOUIS, T. A. (1987). Empirical Bayes confidence intervals based on bootstrap samples. *J. Amer. Statist. Assoc.* **82** 739–757. MR0909979

LEMMER, H. H. (1988). Shrinkage estimators. In *Encyclopedia of Statistical Sciences*. *Vol.* 8 (S. Kotz, N. L. Johnson and C. B. Read, eds.) 452–456. Wiley, New York. MR0976456

LINDLEY, D. V. (1962). Discussion of Professor Stein's paper 'Confidence sets for the mean of a multivariate normal distribution'. *J. R. Stat. Soc. Ser. B* **24** 285–287.

LINDLEY, D. V. and SMITH, A. F. M. (1972). Bayes estimates for the linear model. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **34** 1–41.

LOUIS, T. A. (1984). Estimating a population of parameter values using Bayes and empirical Bayes methods. *J. Amer. Statist. Assoc.* **79** 393–398. MR0755093

MACGIBBON, B. and TOMBERLIN, T. J. (1989). Small area estimates of proportions via empirical Bayes techniques. *Survey Methodology* **15** 237–252.

MALEC, D. and SEDRANSK, J. (1985). Bayesian inference for finite population parameters in multistage cluster sampling. *J. Amer. Statist. Assoc.* **80** 897–902. MR0819590

MORRIS, C. N. (1983a). Parametric empirical Bayes confidence intervals. In *Scientific Inference*, *Data Analysis*, *and Robustness* (*Madison*, *Wis.*, 1981) (G. E. P Box, T. Leonard and J. Wu, eds.). *Publ. Math. Res. Center Univ. Wisconsin* **48** 25–50. Academic Press, Orlando, FL. MR0772762

MORRIS, C. N. (1983b). Parametric empirical Bayes inference: Theory and applications. *J. Amer. Statist. Assoc.* **78** 47–65. MR0696849

OTTO, M. C. and BELL, W. R. (1995). Sampling error modeling of poverty and income statistics for states. In *Proceedings of the American Statistical Association*, *Government Statistics Section* 160–165. Amer. Statist. Assoc., Alexandria, VA.

PFEFFERMANN, D. (2002). Small area estimation- new developments and directions. *Int. Statist. Rev.* **70** 125–143.

PFEFFERMANN, D. and TILLER, R. (2005). Bootstrap approximation to prediction MSE for state-space models with estimated parameters. *J. Time Series Anal.* **26** 893–916. MR2203516

PRASAD, N. G. N. and RAO, J. N. K. (1990). The estimation of the mean squared error of small-area estimators. *J. Amer. Statist. Assoc.* **85** 163–171. MR1137362

RAO, J. N. K. (1999). Some recent advances in model-based small area estimation. *Survey Methodology* **25** 175–186.

RAO, J. N. K. (2001). EB and EBLUP in small area estimation. In *Empirical Bayes and Likelihood Inference* (*Montreal*, *QC*, 1997) (S. E. Ahmed and N. Reid, eds.). *Lecture Notes in Statist.* **148** 33–43. Springer, New York. MR1855557

RAO, J. N. K. (2003a). *Small Area Estimation*. Wiley-Interscience, Hoboken, NJ. MR1953089

RAO, J. N. K. (2003b). Some new developments in small area estimation. *J. Iran. Stat. Soc.* **2** 145–169.

RIVEST, L. P. and VANDAL, N. (2004). Mean squared error estimation for small areas when the small area variances are estimated. In *Proceedings of the International Conference on Recent Advances in Survey Sampling*, *Laboratory for Research in Statistics and Probability*. Carleton Univ., Ottawa, Canada.

SCOTT, A. J. and SMITH, T. M. F. (1969). Estimation in multistage surveys. *J. Amer. Statist. Assoc.* **64** 830–840.

SHEN, W. and LOUIS, T. A. (1998). Triple-goal estimates in two-stage hierarchical models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **60** 455–471. MR1616061

SINGH, A. C., STUKEL, D. M. and PFEFFERMANN, D. (1998). Bayesian versus frequentist measures of error in small area estimation. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **60** 377–396. MR1616057

SMITH, D. D. (2001). Minimum Hellinger distance estimation for the exponential distribution and hierarchical bayesian approaches in small area estimation. Unpublished Ph.D. dissertation, Dept. Statistics, Univ. Georgia, Athens, GA.

STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955, *Vol. I* 197–206. Univ. California Press, Berkeley. MR0084922

STEIN, C. M. (1962). Confidence sets for the mean of a multivariate normal distribution. *J. R. Stat. Soc. Ser. B* **24** 265–296. MR0148184

TORABI, M., DATTA, G. S. and RAO, J. N. K. (2009). Empirical Bayes estimation of small area means under a nested error linear regression model with measurement errors in the covariates. *Scand. J. Stat.* **36** 355–368. MR2528989

WANG, J. and FULLER, W. A. (2003). The mean squared error of small area predictors constructed with estimated area variances. *J. Amer. Statist. Assoc.* **98** 716–723. MR2011685

WOLTER, K. M. (1985). *Introduction to Variance Estimation*. Springer, New York. MR0799715

ZELLNER, A. (1988). Bayesian analysis in econometrics. *J. Econometrics* **37** 27–50. MR0926622

ZELLNER, A. (1994). Bayesian and non-Bayesian estimation using balanced loss functions. In *Statistical Decision Theory and Related Topics*, V (*West Lafayette*, *IN*, 1992) 377–390. Springer, New York. MR1286314