Sankhyā: The Indian Journal of Statistics
Special Issue on Bayesian Analysis
1998, Volume 60, Series A, Pt. 3, pp. 344-362

MULTIVARIATE BAYESIAN SMALL AREA ESTIMATION: AN APPLICATION TO SURVEY AND SATELLITE DATA*

By G.S. DATTA
B. DAY
University of Georgia, Athens
and
T. MAITI
University of Nebraska-Lincoln

SUMMARY. The importance of small area estimation in survey sampling is increasing, due to the growing demand for reliable small area statistics from both public and private sectors. Appropriate models are used to produce reliable small area estimates by "borrowing strength" from neighboring areas. In many small area problems, data on related multiple characteristics and auxiliary variables are available. In this context, Fay (1987) proposed to borrow strength through multivariate modeling of related characteristics using a multiple regression model. The success of such modeling rests on the strength of dependence among these characteristics. In this article, we consider a superpopulation approach to hierarchical Bayes prediction of small area mean vectors using the multivariate nested error regression model of Fuller and Harter (1987). To compare the performance of the multivariate approach with the usual univariate approach we analyse the survey and satellite data of Battese *et al.* (1988) on crop areas under corn and soybean. Our simulations show that the multivariate approach may result in substantial improvement over its univariate counterpart. Finally, we obtain a set of necessary and sufficient conditions for the propriety of the posterior distributions corresponding to a certain class of improper priors on the components of variance matrices.

1. **Introduction**

During the last two decades many market research companies and various government agencies worldwide have become interested in producing reliable

---

AMS (1991) *subject classifications.* 62D05, 62F15, 62H12, 62J05.
Key words and phrases. Components of variance, Gibbs sampling, multivariate mixed linear model, posterior predictive assessment, superpopulation, inverse Wishart distribution.
* Research of the first author was supported in part by NSF Grant SBR-9705145 and an ASA/NSF/BLS/Census Fellowship. The views expressed here are those of the authors and reflect neither the policy of the Bureau of Labor Statistics nor of the Bureau of the Census.

small area statistics. A small area or a small domain usually refers to a subgroup of a population from which samples are drawn. The subgroup may refer to a geographical region (e.g., state, county, municipality, etc.) or a group obtained by cross-classification of various demographic factors such as age, gender, race, etc. The importance of reliable small area statistics can hardly be over-emphasized, as such statistics are needed in regional planning and resource allocation in many government programs. For a number of important small area estimation problems encountered by various U.S. federal agencies, one may refer to the report of a U.S. federal subcommittee on small area estimation (Schaible 1993).

Most surveys provide little information on individual small areas since they are generally designed to produce accurate estimates at a higher level of aggregation. Small area estimation methods are well suited for settings that involve many domains, with small (or no) samples from individual domains. In this setting, traditional design-based direct survey estimates based only on samples from individual small areas are not reliable. In order to improve on the traditional estimates based on individual area sample, one may "borrow strength" from neighboring or related small areas, or other correlated dependent variables (via multivariate approach) and relevant covariate information available from other sources, such as administrative records, to produce accurate small area estimates.

Borrowing strength to produce reliable small area or small domain estimates is attained through the use of an appropriate model. Various model-based methods in small area estimation include empirical or estimated best linear unbiased prediction (EBLUP) (Battese *et al.* 1988; Prasad and Rao 1990; Lahiri and Rao 1995; Datta and Lahiri 1997; Datta *et al.* 1998), James-Stein shrinkage estimation, or more generally, parametric empirical Bayes (EB) estimation (Carter and Rolph 1974; Fay and Herriot 1979; Ghosh and Meeden 1986; Ghosh and Lahiri 1987a,b, 1988; Arora *et al.* 1997; Butar and Lahiri 1997) and hierarchical Bayes (HB) estimation (Datta and Ghosh 1991; Ghosh and Lahiri 1992; Datta *et al.* 1991, 1992, 1996). For a review of the history of small area estimation, and various small area estimation procedures and their applications, we refer to review articles by Purcell and Kish (1979), Rao (1986), and Ghosh and Rao (1994).

Fay (1987) advocated a multivariate approach to small area estimation to produce reliable estimates of median incomes for four-person families for the fifty states of the U.S. and the District of Columbia by including information from three- and five-person families as well. Fuller and Harter (1987) also considered this approach for estimating a finite population mean vector of multiple characteristics for each small area. Fay (1987), Fuller and Harter (1987) and, more recently, Datta *et al.* (1998) took an EB approach, which is identical to an EBLUP approach, to multivariate small area estimation. Datta *et al.* (1991, 1996) and Ghosh *et al.* (1996) used a multivariate HB approach to the median income estimation problem.

In this article we adopt an HB approach to multivariate small area estimation. In Section 2 we describe our HB formulation for the multivariate nested error regression model of Fuller and Harter (1987) and use that model to develop HB estimates of finite population mean vectors of several small areas. The HB predictors of finite population mean vectors are given by their posterior means and the associated measures of uncertainty are given by their posterior variances. The posterior quantities are obtained by Gibbs sampling. We compare our multivariate HB model with the corresponding univariate HB model considered earlier by Datta and Ghosh (1991). In Subsections 3.1 and 3.2, respectively, we show heuristically and through simulations that the multivariate HB analysis may have smaller posterior variance than the corresponding univariate analysis. Percentage reductions in posterior variance are given in Table 1. Depending on the values of the model parameters, our simulation results show that the improvement due to multivariate modeling may be as high as 84%. In Subsection 3.3 we analyse the corn and soybean data in Battese *et al.* (1988) using the multivariate and the univariate models. We compare these models through Bayesian model checking. Models are compared by computing the divergence measure of Laud and Ibrahim (1995) by using the posterior predictive distribution. Based on this measure the multivariate model emerges as a better model than the corresponding pair of univariate models. Brief concluding remarks are given in Section 4. For HB analysis we use noninformative priors on the model parameters, namely, a uniform distribution on the regression coefficient matrix and inverse Wishart distribution on the components of variance matrices. By choosing the scale matrices of the Wishart distribution as null matrices we get a class of improper priors on the model parameters. We prove a theorem in the Appendix obtaining necessary and sufficient conditions that this class of priors must satisfy to yield proper posterior distributions.

2. The Multivariate Hierarchical Bayes Model

Let $\{\boldsymbol{Y}_{ij}, j = 1, \ldots, N_i, i = 1, \ldots, m\}$ denote a finite population where $\boldsymbol{Y}_{ij}(s \times 1)$ is the response vector associated with the j-th unit in the i-th small area. We assume that $\boldsymbol{c}_{ij}(p \times 1)$ is a vector of covariates associated with $\boldsymbol{Y}_{ij}$. We will consider a superpopulation approach to finite population sampling to predict the finite population mean vectors $\gamma_i = N_i^{-1} \sum_{j=1}^{N_i} \boldsymbol{Y}_{ij}$, $i = 1, ..., m$, of the m small areas based on a multivariate HB model introduced below.

Model M:

(I) Conditional on $\boldsymbol{B}, \boldsymbol{v}_1, ..., \boldsymbol{v}_m$, Σ_e and Σ_v, $\boldsymbol{Y}_{ij} \sim N_s(\boldsymbol{B}\boldsymbol{c}_{ij} + \boldsymbol{v}_i, \Sigma_e)$ independently for $j = 1, ..., N_i$, $i = 1, ..., m$;

(II) Conditional on $\boldsymbol{B}$, Σ_v and Σ_e, $\boldsymbol{v}_i \sim N_s(\boldsymbol{0}, \Sigma_v)$ independently for $i = 1, ..., m$;

(III) Marginally, $\boldsymbol{B}$, Σ_v and Σ_e are independently distributed with $\boldsymbol{B} \sim$ Uniform on $R^{s \times p}$, $\Sigma_v \sim W_a^{-1}(\Phi_v)$, Σ_v p.d., Φ_v n.n.d., and $\Sigma_e \sim W_b^{-1}(\Phi_e)$, Σ_e

p.d., Φ_e n.n.d., where $W_a^{-1}(\cdot)$ denotes an inverse Wishart distribution with a degrees of freedom.

We say that a random p.d. matrix $\boldsymbol{T}(s \times s)$ has an inverse Wishart distribution with p.d. scale matrix $\Delta(s \times s)$ and degrees of freedom d if it has a density of the form

$$p(\boldsymbol{T}) \propto |\boldsymbol{T}|^{-(d+s+1)/2} \exp\{-\frac{1}{2} tr(\boldsymbol{T}^{-1}\Delta)\}, d \geq s,$$

and we denote it by $\boldsymbol{T} \sim W_d^{-1}(\Delta)$. Note that an inverse Wishart is a conjugate prior distribution for a multivariate normal covariance matrix.

In the above model, $\boldsymbol{B}(s \times p)$ is a matrix of regression coefficients, $\Sigma_e(s \times s)$ is a matrix of sampling variance, and $\Sigma_v(s \times s)$ is a matrix of model variance, i.e., the variance-covariance matrix of the small area effect $\boldsymbol{v}_i$. In an HB analysis one assigns a prior distribution on the parameters $\boldsymbol{B}$, Σ_v and Σ_e. We specify this prior distribution in stage (III) of the model where we assume that a, b, Φ_v and Φ_e are known. By appropriately choosing a, b, Φ_v and Φ_e we can make the prior distributions proper or improper.

The first two steps of this model can be identified as a general multivariate mixed linear model. To see this, we introduce some notation. We assume that all matrices that appear below have appropriate dimensions. Let $\oplus_{i=1}^{u} \boldsymbol{B}_i$ denote *block diagonal*$(\boldsymbol{B}_1, \ldots, \boldsymbol{B}_u)$. Also write

$(\boldsymbol{Y}_{11}, \ldots, \boldsymbol{Y}_{1N_1}, \ldots, \boldsymbol{Y}_{m1}, \ldots, \boldsymbol{Y}_{mN_m}) = \boldsymbol{U}$,
$(\boldsymbol{e}_{11}, \ldots, \boldsymbol{e}_{1N_1}, \ldots, \boldsymbol{e}_{m1}, \ldots, \boldsymbol{e}_{mN_m}) = \boldsymbol{E}$, $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m) = \boldsymbol{A}$,
$(\boldsymbol{c}_{11}, \ldots, \boldsymbol{c}_{1N_1}, \ldots, \boldsymbol{c}_{m1}, \ldots, \boldsymbol{c}_{mN_m}) = \boldsymbol{C}^T$, and $\boldsymbol{F}^T = \oplus_{i=1}^{m} \boldsymbol{1}_{N_i}^T$.

With this notation we express (I)-(II) as

$$\boldsymbol{U} = \boldsymbol{B}\boldsymbol{C}^T + \boldsymbol{A}\boldsymbol{F}^T + \boldsymbol{E}, \qquad \ldots(1)$$

where $\boldsymbol{e}_{ij}$'s are i.i.d. $N_s(\boldsymbol{0}, \Sigma_e)$ and independent of $\boldsymbol{v}_1, ..., \boldsymbol{v}_m$. Fuller and Harter (1987) referred to the model given by (1) as a nested error regression model.

Writing $\sum_{i=1}^{m} N_i = N$ and making conformable partitions of $\boldsymbol{U}(s \times N)$, $\boldsymbol{C}^T(p \times N)$, $\boldsymbol{F}^T(m \times N)$ and $\boldsymbol{E}(s \times N)$, after rearranging their columns corresponding to *sampled* and *unsampled* units, we rewrite the model given in (1) as

$$(\boldsymbol{U}^{(1)}|\boldsymbol{U}^{(2)}) = \boldsymbol{B}(\boldsymbol{C}^{(1)T}|\boldsymbol{C}^{(2)T}) + \boldsymbol{A}(\boldsymbol{F}^{(1)T}|\boldsymbol{F}^{(2)T}) + (\boldsymbol{E}^{(1)}|\boldsymbol{E}^{(2)}). \qquad \ldots(2)$$

In (2), $\boldsymbol{U}^{(1)}(s \times n)$ corresponds to the matrix of sampled units from m small areas or domains, while $\boldsymbol{U}^{(2)}(s \times (N-n))$ corresponds to the matrix of unsampled units, where n denotes the total sample from the m small areas. Similarly we describe $\boldsymbol{C}^{(1)T}, \boldsymbol{C}^{(2)T}, \boldsymbol{F}^{(1)T}, \boldsymbol{F}^{(2)T}, \boldsymbol{E}^{(1)}$ and $\boldsymbol{E}^{(2)}$. For m small areas, we can further partition $\boldsymbol{U}^{(\alpha)}$ as $\boldsymbol{U}^{(\alpha)} = (\boldsymbol{U}_1^{(\alpha)}, ..., \boldsymbol{U}_m^{(\alpha)})$ for $\alpha = 1, 2$, where, without loss of generality, $\boldsymbol{U}_i^{(1)} = (\boldsymbol{Y}_{i1}, ..., \boldsymbol{Y}_{in_i})$ is assumed to represent the matrix of n_i sampled observations from the i-th small area, and $\boldsymbol{U}_i^{(2)} = (\boldsymbol{Y}_{i,n_i+1}, ..., \boldsymbol{Y}_{i,N_i})$ is matrix of $N_i - n_i$ unsampled observations from the i-th small area.

We are interested in predicting $\gamma_i = N_i^{-1}\sum_{j=1}^{N_i} \boldsymbol{Y}_{ij}$, the finite population mean for the i-th county, $i = 1, ..., m$. Noting that $\boldsymbol{Y}_{ij} = \boldsymbol{B}\boldsymbol{c}_{ij} + \boldsymbol{v}_i + \boldsymbol{e}_{ij}, j = 1, ..., N_i, i = 1, ..., m$, γ_i can be written as $\gamma_i = \mu_i + \bar{\boldsymbol{e}}_i$, where $\bar{\boldsymbol{e}}_i = N_i^{-1}\sum_{j=1}^{N_i} \boldsymbol{e}_{ij}$, $\mu_i = \boldsymbol{B}\bar{\boldsymbol{c}}_{i(\text{pop})} + \boldsymbol{v}_i$ and $\bar{\boldsymbol{c}}_{i(\text{pop})} = N_i^{-1}\sum_{j=1}^{N_i} \boldsymbol{c}_{ij}$. Under the assumptions of Model **M**, μ_i can be interpreted as the conditional mean (vector) of the i-th small area given the values of covariates $\boldsymbol{c}_{ij}$'s and the realized small area effect $\boldsymbol{v}_i$. Clearly, γ_i is not equivalent to μ_i, because the average of the $\boldsymbol{e}_{ij}$'s over the population units in area i is not identically $\mathbf{0}$. However, if N_i is large and the sampling fraction $f_i = n_i/N_i (i = 1, ..., m)$ are small, then the predictor of the mixed effect (vector) μ_i may be an appropriate predictor of γ_i. In our example these conditions are satisfied. In fact, it was noted for this example in Datta (1990, p. 50 Table 2.3) that there is virtually no difference between a predictor of γ_i and μ_i for an univariate HB analysis of the soybean data. In this article we will thus consider HB prediction of the mixed effect vector $\boldsymbol{B}\boldsymbol{h} + \boldsymbol{A}\lambda$ for known vectors $\boldsymbol{h}(p \times 1)$ and $\lambda(m \times 1)$ based on a model that is obtained from the marginal distribution of the sampled units as implied by the Model **M**.

By rewriting the matrices, $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{U}^{(1)}$ and $\boldsymbol{E}^{(1)}$ in vector notation, we show in Subsection 2.1 that the HB prediction of $\boldsymbol{B}\boldsymbol{h} + \boldsymbol{A}\lambda$ can be viewed as a prediction in linear models (e.g., see Lindley and Smith, 1972; Datta, 1992).

2.1. *Hierarchical Bayes prediction in linear models.* Let $\boldsymbol{Y}^T = (\boldsymbol{Y}_{11}^T, \ldots, \boldsymbol{Y}_{1n_1}^T, \ldots, \boldsymbol{Y}_{m1}^T, \ldots, \boldsymbol{Y}_{mn_m}^T)$, $\boldsymbol{e}^T = (\boldsymbol{e}_{11}^T, \ldots, \boldsymbol{e}_{1n_1}^T, \ldots, \boldsymbol{e}_{m1}^T, \ldots, \boldsymbol{e}_{mn_m}^T)$, $\boldsymbol{v}^T = (\boldsymbol{v}_1^T, \ldots, \boldsymbol{v}_m^T)$, $\beta^T = (b_{11}, \ldots, b_{s1}, \ldots, b_{1p}, \ \ldots, b_{sp})$, $\boldsymbol{X} = \boldsymbol{C}^{(1)} \otimes \boldsymbol{I}_s$, $\boldsymbol{Z} = \boldsymbol{F}^{(1)} \otimes \boldsymbol{I}_s$, $\boldsymbol{F}^{(1)} = \oplus_{i=1}^m \mathbf{1}_{n_i}$, $\boldsymbol{H} = \boldsymbol{h}^T \otimes \boldsymbol{I}_s$, $\Lambda = \lambda^T \otimes \boldsymbol{I}_s$, where $\otimes$ is the symbol for Kronecker product of matrices. With this notation we obtain, from (I)-(II) given above,

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{Z}\boldsymbol{v} + \boldsymbol{e}, \qquad \ldots(3)$$

where $\boldsymbol{e} \sim N_{ns}(\mathbf{0}, \boldsymbol{R}(\psi))$ independently of $\boldsymbol{v} \sim N_{ms}(\mathbf{0}, \boldsymbol{G}(\psi))$. Here, ψ_v and ψ_e are the $\frac{s(s+1)}{2}$-component vectors of $\frac{s(s+1)}{2}$ distinct elements of Σ_v and Σ_e, and $\psi = (\psi_v^T, \psi_e^T)^T$. It can be verified that $\boldsymbol{R}(\psi) = \boldsymbol{I}_n \otimes \Sigma_e$, $\boldsymbol{G}(\psi) = \boldsymbol{I}_m \otimes \Sigma_v$, $\boldsymbol{B}\boldsymbol{h} = \boldsymbol{H}\beta$ and $\boldsymbol{A}\lambda = \Lambda\boldsymbol{v}$. Thus, prediction of $\boldsymbol{B}\boldsymbol{h} + \boldsymbol{A}\lambda$ is equivalent to the prediction of the vector of mixed effects $\boldsymbol{H}\beta + \Lambda\boldsymbol{v}$ under (3).

The HB prediction of a mixed effect vector under a model similar to the one in (3) is given by, e.g., Datta (1992). We assume that $\boldsymbol{X}$ is of full column rank. For known β and ψ, due to the normality assumption in (3), the Bayes predictor of $\boldsymbol{H}\beta + \Lambda\boldsymbol{v}$, under matrix generalization of squared error loss, is given by

$$\begin{aligned} \boldsymbol{e}_{\text{BM}}(\boldsymbol{Y}; \beta, \psi) &= E[\boldsymbol{H}\beta + \Lambda\boldsymbol{v}|\beta, \psi, \boldsymbol{Y}] \\ &= [\boldsymbol{H} - \Lambda\boldsymbol{G}(\psi)\boldsymbol{Z}^T\Sigma^{-1}(\psi)\boldsymbol{X}]\beta + \Lambda\boldsymbol{G}(\psi)\boldsymbol{Z}^T\Sigma^{-1}(\psi)\boldsymbol{Y}, \qquad \ldots(4) \end{aligned}$$

with the associated posterior variance matrix given by

$$\boldsymbol{G}_{1\text{M}}(\psi) = V[\boldsymbol{H}\beta + \Lambda\boldsymbol{v}|\beta, \psi, \boldsymbol{Y}] = \Lambda\boldsymbol{W}(\psi)\Lambda^T, \qquad \ldots(5)$$

where $\Sigma(\psi) = \boldsymbol{R}(\psi) + \boldsymbol{Z}\boldsymbol{G}(\psi)\boldsymbol{Z}^T$ and $\boldsymbol{W}(\psi) = (\boldsymbol{G}^{-1}(\psi) + \boldsymbol{Z}^T\boldsymbol{R}^{-1}(\psi)\boldsymbol{Z})^{-1}$. Since β and ψ are both unknown, $\boldsymbol{e}_{\mathrm{BM}}(\boldsymbol{Y};\beta,\psi)$ can not be used as a predictor. For known ψ, if we integrate β with respect to its posterior distribution conditional on ψ, we obtain the Bayes predictor given by

$$\begin{aligned}\tilde{\boldsymbol{e}}_{\mathrm{BM}}(\boldsymbol{Y};\psi) &= \boldsymbol{e}_{\mathrm{BM}}(\boldsymbol{Y};\tilde{\beta}(\psi),\psi)\\ &= [\boldsymbol{H}-\Lambda\boldsymbol{G}(\psi)\boldsymbol{Z}^T\Sigma^{-1}(\psi)\boldsymbol{X}]\tilde{\beta}(\psi)+\Lambda\boldsymbol{G}(\psi)\boldsymbol{Z}^T\Sigma^{-1}(\psi)\boldsymbol{Y} \qquad \ldots(6)\end{aligned}$$

with the associated posterior variance matrix given by

$$V[\boldsymbol{H}\beta+\Lambda\boldsymbol{v}|\psi,\boldsymbol{Y}] = \boldsymbol{G}_{1\mathrm{M}}(\psi)+\boldsymbol{G}_{2\mathrm{M}}(\psi), \qquad \ldots(7)$$

where $\tilde{\beta}(\psi) = (\boldsymbol{X}^T\Sigma^{-1}(\psi)\boldsymbol{X})^{-1}\boldsymbol{X}^T\Sigma^{-1}(\psi)\boldsymbol{Y}$ and

$$\begin{aligned}\boldsymbol{G}_{2M}(\psi) &= [\boldsymbol{H}-\Lambda\boldsymbol{G}(\psi)\boldsymbol{Z}^T\Sigma^{-1}(\psi)\boldsymbol{X}](\boldsymbol{X}^T\Sigma^{-1}(\psi)\boldsymbol{X})^{-1}\\ &\quad\times[\boldsymbol{H}-\Lambda\boldsymbol{G}(\psi)\boldsymbol{Z}^T\Sigma^{-1}(\psi)\boldsymbol{X}]^T. \qquad \ldots(8)\end{aligned}$$

The predictor in (6) is the BLUP of $\boldsymbol{H}\beta+\Lambda\boldsymbol{v}$. Since ψ is unknown, $\tilde{\boldsymbol{e}}_{\mathrm{BM}}\ (\boldsymbol{Y};\psi)$ can not be used to predict $\boldsymbol{H}\beta+\Lambda\boldsymbol{v}$. Our HB predictor will be obtained by integrating $\tilde{\boldsymbol{e}}_{\mathrm{BM}}(\boldsymbol{Y};\psi)$ with respect to the posterior distribution of ψ. Under the HB model, one can verify that

(A) conditional on $\boldsymbol{B}$, Σ_e, Σ_v and $\boldsymbol{U}^{(1)}$, $\boldsymbol{v}_i$'s are independent s-variate normal with mean vector $\boldsymbol{P}_i\Sigma_e^{-1}(\sum_{j=1}^{n_i}\boldsymbol{y}_{ij}-\sum_{j=1}^{n_i}\boldsymbol{B}\boldsymbol{c}_{ij})$ and variance $\boldsymbol{P}_i = (n_i\Sigma_e^{-1}+\Sigma_v^{-1})^{-1}$;

(B) conditional on Σ_e, Σ_v and $\boldsymbol{U}^{(1)}$, $\beta \sim N_{sp}((\boldsymbol{X}^T\Sigma^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\Sigma^{-1}\boldsymbol{Y}$, $(\boldsymbol{X}^T\Sigma^{-1}$
$\boldsymbol{X})^{-1})$,
where $\boldsymbol{X}^T\Sigma^{-1}\boldsymbol{X}$ is given in Remark 2 below, and

$$\begin{aligned}\boldsymbol{X}^T\Sigma^{-1}\boldsymbol{Y} &= \mathrm{col}_{1\le t\le p}[\sum_{i=1}^{m}\sum_{j=1}^{n_i}\Sigma_e^{-1}\boldsymbol{y}_{ij}(c_{ijt}-\bar{c}_{it(\mathrm{sam})})\\ &\quad+\sum_{i=1}^{m}(n_i\Sigma_v+\Sigma_e)^{-1}n_i\bar{\boldsymbol{y}}_i\bar{c}_{it(\mathrm{sam})}], \qquad \ldots(9)\end{aligned}$$

with $\bar{c}_{it(\mathrm{sam})} = n_i^{-1}\sum_{j=1}^{n_i}c_{ijt}$, $t=1,...,p$, the i-th small area sample mean of the t-th covariate. Then, using (A) and (B), the HB predictor for $\boldsymbol{H}\beta+\Lambda\boldsymbol{v}$ and the posterior variance matrix are given by

$$\boldsymbol{e}_{\mathrm{HBM}}(\boldsymbol{Y}) = E[\boldsymbol{H}\beta+\Lambda\boldsymbol{v}|\boldsymbol{U}^{(1)}] = E[\tilde{\boldsymbol{e}}_{\mathrm{BM}}(\boldsymbol{Y};\psi)|\boldsymbol{Y}] \qquad \ldots(10)$$

and

$$\begin{aligned}V_{\mathrm{HBM}}(\boldsymbol{Y}) &= Var[\boldsymbol{H}\beta+\Lambda\boldsymbol{v}|\boldsymbol{U}^{(1)}]\\ &= E[\boldsymbol{G}_{1\mathrm{M}}(\psi)+\boldsymbol{G}_{2\mathrm{M}}(\psi)|\boldsymbol{Y}]+V(\tilde{\boldsymbol{e}}_{\mathrm{BM}}(\boldsymbol{Y};\psi)|\boldsymbol{Y}). \qquad \ldots(11)\end{aligned}$$

To obtain the HB predictor and posterior variance matrix, the expectation and variance on the right hand sides of (10) and (11) with respect to the high-dimensional posterior distribution of ψ must be computed numerically. Numerical integration in such high dimensions are unreliable so we use the Markov chain Monte Carlo method of integration. In particular, we use Gibbs sampling to generate our Monte Carlo samples (see Geman and Geman 1984; Gelfand and Smith 1990). Simulation from the joint posterior distribution using Gibbs sampling requires sampling from a set of full conditional distributions. The set of full conditional distributions in our case is given by

(i) $vec(\boldsymbol{B})|\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m, \Sigma_e, \Sigma_v, \boldsymbol{Y}$

$$\sim N_{sp}(vec([\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\boldsymbol{y}_{ij}-\boldsymbol{v}_i)\boldsymbol{c}_{ij}^T][\sum_{i=1}^{m}\sum_{j=1}^{n_i}\boldsymbol{c}_{ij}\boldsymbol{c}_{ij}^T]^{-1}), (\sum_{i=1}^{m}\sum_{j=1}^{n_i}\boldsymbol{c}_{ij}\boldsymbol{c}_{ij}^T)^{-1}\otimes\Sigma_e),$$

where $vec(\boldsymbol{B})$ is the usual notation and equal to β;

(ii) conditional on $\boldsymbol{B}$, Σ_e, Σ_v and $\boldsymbol{Y}$, $\boldsymbol{v}_1, ..., \boldsymbol{v}_m$ are independent with $\boldsymbol{v}_i|\boldsymbol{B}, \Sigma_e, \Sigma_v, \boldsymbol{Y} \sim N_s(\boldsymbol{P}_i\Sigma_e^{-1}(\sum_{j=1}^{n_i}(\boldsymbol{y}_{ij}-\boldsymbol{B}\boldsymbol{c}_{ij})), \boldsymbol{P}_i)$, $i = 1, ..., m$;

(iii) conditional on $\boldsymbol{B}$, $\boldsymbol{v}_1, ..., \boldsymbol{v}_m$ and $\boldsymbol{Y}$, Σ_e and Σ_v are independently distributed with

(iiia) $\Sigma_e|\boldsymbol{B}, \boldsymbol{v}_1, ..., \boldsymbol{v}_m, \boldsymbol{Y} \sim W_{n+b}^{-1}(\Phi_{e,\text{new}})$ where $\Phi_{e,\text{new}} = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(\boldsymbol{y}_{ij} - \boldsymbol{B}\boldsymbol{c}_{ij} - \boldsymbol{v}_i)(\boldsymbol{y}_{ij} - \boldsymbol{B}\boldsymbol{c}_{ij} - \boldsymbol{v}_i)^T + \Phi_e$, $n = \sum_{i=1}^{m} n_i$ and

(iiib) $\Sigma_v|\boldsymbol{B}, \boldsymbol{v}_1, ..., \boldsymbol{v}_m, \boldsymbol{Y} \sim W_{m+a}^{-1}(\Phi_{v,\text{new}})$ where $\Phi_{v,\text{new}} = \sum_{i=1}^{m}\boldsymbol{v}_i\boldsymbol{v}_i^T + \Phi_v$.

It was shown by Day (1998) that under some mild conditions on n, m and p the joint posterior distribution will be proper if the prior distribution of Σ_e and Σ_v in step III is proper. However, if we also use an improper prior for Σ_e and Σ_v, we must check whether the resulting posterior is proper or not. For example, choice of Φ_e and Φ_v as null matrices will result in an improper prior of the form

$$\pi(\boldsymbol{B}, \Sigma_e, \Sigma_v) \propto |\Sigma_v|^{-\frac{a_v}{2}}|\Sigma_e|^{-\frac{a_e}{2}}. \quad \ldots(12)$$

Hobert and Casella (1996) showed that Gibbs sampling can be routinely carried out for an improper posterior distribution without encountering computational difficulties. However, it is clearly meaningless to use such Gibbs samples for drawing inference about a distribution which is not proper. Hobert and Casella (1996) obtained necessary and sufficient conditions for a special class of improper priors for the variance components in univariate mixed linear models so that the resulting posterior distribution is proper. Our prior in (12) is a generalization of their priors for two components variance-covariance matrices in the multivariate mixed linear model. For our multivariate model and the prior given by (12), we obtain a set of necessary and sufficient conditions so that the resulting posterior distribution is proper. These conditions are given in Theorem A.1 in the Appendix.

REMARK 1. Instead of integrating ψ in (10), if we replace it by $\hat{\psi}$ (an estimate of ψ obtained from marginal distribution of the data given by (I)-(II)), the resulting estimate $\tilde{\boldsymbol{e}}_{\mathrm{BM}}(\boldsymbol{Y};\hat{\psi})$ is the EB estimate or EBLUP of $\boldsymbol{H}\beta+\Lambda\boldsymbol{v}$ (e.g., see Datta *et al.* 1998). A naive measure of uncertainty for the EB estimate is given by $\boldsymbol{G}_{1\mathrm{M}}(\hat{\psi})+\boldsymbol{G}_{2\mathrm{M}}(\hat{\psi})$, which usually underestimates (11) since it fails to account for the uncertainty in estimating ψ. As in Prasad and Rao (1990) and Datta and Lahiri (1997), a second order accurate approximation of the measure of uncertainty was obtained by Datta *et al.* (1998).

REMARK 2. One can verify that $\boldsymbol{G}_{1\mathrm{M}}(\psi)$ and $\boldsymbol{G}_{2\mathrm{M}}(\psi)$ simplify to

$$\boldsymbol{G}_{1\mathrm{M}}(\psi)=\sum_{i=1}^{m}\lambda_i^2(\Sigma_v^{-1}+n_i\Sigma_e^{-1})^{-1}, \qquad \ldots(13)$$

$$\boldsymbol{G}_{2\mathrm{M}}(\psi)=[\boldsymbol{h}^T\otimes\boldsymbol{I}_s-\textstyle\sum_{i=1}^{m}\lambda_i\bar{\boldsymbol{c}}_{i(\mathrm{sam})}^T\otimes\Sigma_v(\Sigma_v+n_i^{-1}\Sigma_e)^{-1}](\boldsymbol{X}^T\Sigma^{-1}\boldsymbol{X})^{-1}$$
$$\times[\boldsymbol{h}\otimes\boldsymbol{I}_s-\sum_{i=1}^{m}\lambda_i\bar{\boldsymbol{c}}_{i(\mathrm{sam})}\otimes(\Sigma_v+n_i^{-1}\Sigma_e)^{-1}\Sigma_v], \qquad \ldots(14)$$

where $\bar{\boldsymbol{c}}_{i(\mathrm{sam})}=n_i^{-1}\sum_{j=1}^{n_i}\boldsymbol{c}_{ij}$ and $\boldsymbol{X}^T\Sigma^{-1}\boldsymbol{X}$ simplifies to

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\boldsymbol{c}_{ij}-\bar{\boldsymbol{c}}_{i(\mathrm{sam})})(\boldsymbol{c}_{ij}-\bar{\boldsymbol{c}}_{i(\mathrm{sam})})^T\otimes\Sigma_e^{-1}+\sum_{i=1}^{m}\bar{\boldsymbol{c}}_{i(\mathrm{sam})}\bar{\boldsymbol{c}}_{i(\mathrm{sam})}^T\otimes(\Sigma_v+n_i^{-1}\Sigma_e)^{-1}.$$

REMARK 3. If the N_i are not large or the $f_i=n_i/N_i$ are not negligible, one should instead use the HB predictor of γ_i, the i-th finite population mean, given by

$$\hat{\boldsymbol{e}}_{i\mathrm{HBM}}^{\mathrm{F}}(\boldsymbol{Y})=f_i\bar{\boldsymbol{Y}}_i+(1-f_i)\hat{\boldsymbol{e}}_{i\mathrm{HBM}}^*(\boldsymbol{Y}),$$

where $\hat{\boldsymbol{e}}_{i\mathrm{HBM}}^*(\boldsymbol{Y})$ is the HB estimate of $\mu_i^*=\boldsymbol{B}\bar{\boldsymbol{c}}_i^*+\boldsymbol{v}_i$, $\bar{\boldsymbol{Y}}_i=n_i^{-1}\sum_{j=1}^{n_i}\boldsymbol{Y}_{ij}$ and $\bar{\boldsymbol{c}}_i^*=(N_i-n_i)^{-1}\sum_{j=n_i+1}^{N_i}\boldsymbol{c}_{ij}$. The posterior variance associated with $\hat{\boldsymbol{e}}_{i\mathrm{HBM}}^{\mathrm{F}}(\boldsymbol{Y})$ is given by

$$\begin{aligned}V_{i\mathrm{HBM}}^{\mathrm{F}}(\boldsymbol{Y}) &= N_i^{-1}(1-f_i)E[\Sigma_e|\boldsymbol{Y}]+(1-f_i)^2\{E[\boldsymbol{G}_{1i\mathrm{M}}^{\mathrm{F}}(\psi)\\ &+ \boldsymbol{G}_{2i\mathrm{M}}^{\mathrm{F}}(\psi)|\boldsymbol{Y}]+V(\hat{\boldsymbol{e}}_{i\mathrm{BM}}^*(\boldsymbol{Y};\psi)|\boldsymbol{Y})\},\end{aligned}$$

where $\hat{\boldsymbol{e}}_{i\mathrm{BM}}^*(\boldsymbol{Y};\psi)$ is obtained from (6) by replacing $\boldsymbol{H}$ by $\bar{\boldsymbol{c}}_i^*\otimes\boldsymbol{I}_s$, and $\boldsymbol{G}_{1i\mathrm{M}}^{\mathrm{F}}(\psi)$ and $\boldsymbol{G}_{2i\mathrm{M}}^{\mathrm{F}}(\psi)$ are obtained from (13) and (14), taking $\lambda_i=1$ and $\lambda_u=0$ for $u\neq i$, and $\boldsymbol{h}=\bar{\boldsymbol{c}}_i^*$.

3. Multivariate Estimation versus Univariate Estimation

Let $\boldsymbol{b}_\alpha^T$, $\boldsymbol{a}_\alpha^T$ and $\boldsymbol{U}_\alpha^T$ denote, respectively, the αth row of $\boldsymbol{B}$, $\boldsymbol{A}$ and $\boldsymbol{U}^{(1)}$. Denote $\boldsymbol{h}^T\boldsymbol{b}_\alpha + \lambda^T\boldsymbol{a}_\alpha$, the αth component of $\boldsymbol{Bh} + \boldsymbol{A}\lambda$ by c_α. Based on our multivariate model in Section 2, the posterior variance of c_α is (a) $G_{1\mathrm{M}\alpha\alpha}(\psi)$, the αth diagonal element of $\boldsymbol{G}_{1\mathrm{M}}(\psi)$, when both $\boldsymbol{B}$ and ψ are known and (b) $G_{1\mathrm{M}\alpha\alpha}(\psi) + G_{2\mathrm{M}\alpha\alpha}(\psi)$ when $\boldsymbol{B}$ is unknown but ψ is known, where $G_{2\mathrm{M}\alpha\alpha}(\psi)$ is the αth diagonal element of $\boldsymbol{G}_{2\mathrm{M}}(\psi)$.

To predict $c_\alpha = \boldsymbol{h}^T\boldsymbol{b}_\alpha + \lambda^T\boldsymbol{a}_\alpha$, one can also use the corresponding univariate model, ignoring the multivariate nature of the problem. In a univariate context, Datta and Ghosh (1991) used the nested error regression model of Battese *et al.* (1988), based on $\boldsymbol{U}_\alpha^T$ alone. Let $v_{i\alpha}$ and $Y_{ij\alpha}$ denote the αth components of $\boldsymbol{v}_i$ and $\boldsymbol{Y}_{ij}$, respectively, and $\sigma_{e\alpha\alpha}$ and $\sigma_{v\alpha\alpha}$ denote the αth diagonal elements of Σ_e and Σ_v, respectively. The Model $\mathbf{U}_\alpha$ given below, based on the marginal distribution of the αth component, is a special case of a general model of Datta and Ghosh (1991).

Model $\mathbf{U}_\alpha$:

(i) Conditional on $\boldsymbol{b}_\alpha,\ v_{1\alpha},\ldots,v_{m\alpha},\ \sigma_{e\alpha\alpha}$ and $\sigma_{v\alpha\alpha}$, $Y_{ij\alpha} \sim N_1(\boldsymbol{b}_\alpha^T\boldsymbol{c}_{ij} + v_{i\alpha}, \sigma_{e\alpha\alpha})$ independently for $j = 1,\cdots,N_i,\ i = 1,\cdots,m$;

(ii) Conditional on $\boldsymbol{b}_\alpha$, $\sigma_{e\alpha\alpha}$ and $\sigma_{v\alpha\alpha}$, $v_{i\alpha} \sim N_1(0,\sigma_{v\alpha\alpha})$ independently for $i = 1,\ldots,m$;

(iii) Marginally, $\boldsymbol{b}_\alpha$, $\sigma_{v\alpha\alpha}$ and $\sigma_{e\alpha\alpha}$ are independently distributed, with $\boldsymbol{b}_\alpha \sim$ Uniform on R^p, $\sigma_{v\alpha\alpha}^{-1} \sim \mathrm{gamma}(\frac{1}{2}\Phi_{v\alpha\alpha}, \frac{a-(s-1)}{2})$ and $\sigma_{e\alpha\alpha}^{-1} \sim \mathrm{gamma}(\frac{1}{2}\Phi_{e\alpha\alpha}, \frac{b-(s-1)}{2})$, where $\Phi_{v\alpha\alpha}$ and $\Phi_{e\alpha\alpha}$ are the (α,α)th elements of Φ_v and Φ_e, respectively.

A random variable R with gamma(ρ,τ) distribution has pdf $f(r) \propto \exp(-\rho r)\, r^{\tau-1}I_{[r>0]}$.

3.1. *A heuristic comparison.* Based on the above univariate model, let $g_{1\mathrm{U}\alpha\alpha}(\psi_\alpha)$ denote the posterior variance of c_α when both $\boldsymbol{b}_\alpha$ and $\psi_\alpha^T = (\sigma_{e\alpha\alpha}, \sigma_{v\alpha\alpha})$ are known, and let $g_{1\mathrm{U}\alpha\alpha}(\psi_\alpha) + g_{2\mathrm{U}\alpha\alpha}(\psi_\alpha)$ denote the posterior variance when $\boldsymbol{b}_\alpha$ is not known. From (13) and (14), for $s = 1$, we obtain

$$g_{1\mathrm{U}\alpha\alpha}(\psi_\alpha) = \sum_{i=1}^m \lambda_i^2(\sigma_{v\alpha\alpha}^{-1} + n_i\sigma_{e\alpha\alpha}^{-1})^{-1}, \qquad \ldots(15)$$

$$\begin{aligned} g_{2\mathrm{U}\alpha\alpha}(\psi_\alpha) &= [\boldsymbol{h} - \sum_{i=1}^m \lambda_i\sigma_{v\alpha\alpha}(\sigma_{v\alpha\alpha} + n_i^{-1}\sigma_{e\alpha\alpha})^{-1}\bar{\boldsymbol{c}}_{i(\mathrm{sam})}]^T(\boldsymbol{X}^T\Sigma^{-1}\boldsymbol{X})^{-1} \\ &\times [\boldsymbol{h} - \sum_{i=1}^m \lambda_i\sigma_{v\alpha\alpha}(\sigma_{v\alpha\alpha} + n_i^{-1}\sigma_{e\alpha\alpha})^{-1}\bar{\boldsymbol{c}}_{i(\mathrm{sam})}], \qquad \ldots(16) \end{aligned}$$

where

$$\boldsymbol{X}^T\Sigma^{-1}\boldsymbol{X} = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(\boldsymbol{c}_{ij}-\bar{\boldsymbol{c}}_{i(\text{sam})})(\boldsymbol{c}_{ij}-\bar{\boldsymbol{c}}_{i(\text{sam})})^T\sigma_{e\alpha\alpha}^{-1} + \sum_{i=1}^{m}\bar{\boldsymbol{c}}_{i(\text{sam})}\bar{c}_{i(\text{sam})}^T(\sigma_{v\alpha\alpha}+n_i^{-1}\sigma_{e\alpha\alpha})^{-1}.$$

The following lemma was established by Datta *et al.* (1998).

LEMMA 1. *For* $g_{1\text{U}\alpha\alpha}(\psi_\alpha)$, $g_{2\text{U}\alpha\alpha}(\psi_\alpha)$, $G_{1\text{M}\alpha\alpha}(\psi)$, *and* $G_{2\text{M}\alpha\alpha}(\psi)$ *defined above,*

(i) $G_{1\text{M}\alpha\alpha}(\psi) \le g_{1\text{U}\alpha\alpha}(\psi_\alpha)$ *for all* ψ,

(ii) $G_{1\text{M}\alpha\alpha}(\psi) + G_{2\text{M}\alpha\alpha}(\psi) \le g_{1\text{U}\alpha\alpha}(\psi_\alpha) + g_{2\text{U}\alpha\alpha}(\psi_\alpha)$ *for all* ψ.

REMARK 4. To see the usefulness of the above lemma, note that in the posterior variance matrix in (11), the dominating term is $E[\boldsymbol{G}_{1\text{M}}(\psi)|\boldsymbol{Y}]$, which is $O(1)$, whereas the other two terms are both $O(m^{-1})$. In view of this lemma, one may heuristically expect that a multivariate analysis may result in a smaller posterior variance of c_α than the corresponding posterior variance in univariate analysis, at least when m is large. To investigate actual improvement we have done a small simulation study, explained in the next subsection.

3.2. *A simulation study.* For our simulation we consider the LANDSAT data of Battese *et al.* (1988), who originally used this data to produce EBLUP of the crop areas under corn and soybean for each of the 12 (value of m) counties in North Central Iowa, using farm-interview data collected by USDA field reporting staff in conjunction with the satellite data from LANDSAT. Each county was divided into area segments, and areas under corn and soybean were determined for sampled segments by interviewing farm operators. The number of sampled segments in a county range from 1 to 5 (n_i values), and the population segments (N_i values) for the counties range from 394 to 965. Clearly, in this example, the ratios n_i/N_i's are negligible. Auxiliary data in the form of the numbers of pixels (a term used for "picture element" of about 0.45 hectares) classified as corn and soybean are available for all the area segments, including the sample segments, for all 12 counties based on LANDSAT readings. Battese *et al.* (1988) used separate univariate models for corn and soybean. They used the univariate nested error regression model that can be written from (i)-(ii) of Model $\mathbf{U}_\alpha$. They obtained the EBLUPs and the mean squared errors (MSE) approximation of the EBLUPs of average county areas under corn and soybean.

In this simulation, $s = 2,\ p = 3,\ m = 12$ and we use the sample sizes and the auxiliary variables described in detail by Battese *et al.* (1988) as our n_i and $\boldsymbol{c}_{ij}$ values. To generate data based on Model M in our simulation, we use estimates of $\boldsymbol{B}$, $\sigma_{e\alpha\alpha}$, $\sigma_{v\alpha\alpha}$, $\alpha = 1, 2$ obtained by Battese *et al.* (1988) as the true values, but choose different values for ρ_e and ρ_v, where ρ_e and ρ_v are the correlation coefficients for the variance matrices Σ_e and Σ_v, respectively. In our simulations we used 0, ± 0.5, ± 0.95 as values for ρ_e and ρ_v (($\rho_e,\ \rho_v$)=(0, 0) was excluded as it implies independence—multivariate analysis is the same as the univariate analysis). We mention that as long as we use a uniform prior on $\boldsymbol{B}$, the true values of $\boldsymbol{B}$ used in our simulation have no role in the posterior variance comparison. This is because the posterior variance in (11) is influenced by the values of $\boldsymbol{B}$ used in the simulation only through the posterior distribution of Σ_e and Σ_v. It follows from (A.10) that this posterior depends on $\boldsymbol{y}$ only through $\boldsymbol{y}^T\boldsymbol{K}\boldsymbol{y}$, which remains invariant under transformation $\boldsymbol{y}+\boldsymbol{X}\boldsymbol{b}$, where $\boldsymbol{b}$ is an sp-component vector.

For given values of $\sigma_{v\alpha\alpha},\ \sigma_{e\alpha\alpha},\ \alpha=1,2$, and ρ_e and ρ_v, Σ_v and Σ_e were computed. We then use steps (I) and (II) of Model $\mathbf{M}$ in conjunction with $\boldsymbol{c}_{ij}$-values to generate sample $\boldsymbol{y}_{ij}$-values for 200 such data sets. On each generated data we performed both univariate and multivariate HB analyses using Model $\mathbf{U}_\alpha, \alpha=1,2$ and Model $\mathbf{M}$, respectively. We obtained the posterior variance matrix of the vector μ_i in multivariate analysis using the methods of Section 2. We used $\Phi_{e\alpha\alpha}=\Phi_{v\alpha\alpha}=10^4,\ \alpha=1,2,\ \Phi_{e12}=10^4\rho_e,\ \Phi_{v12}=10^4\rho_v$, and $a=b=3$ in (III) of Model $\mathbf{M}$. As noted earlier, the univariate method follows as a special case of the multivariate method by taking $s=1$.

Let $V_{i\mathrm{HBM}}$ denote the posterior variance matrix of μ_i from (11). The αth diagonal element of $V_{i\mathrm{HBM}}$, denoted by $V_{i\mathrm{HBM}_{\alpha\alpha}}$, is the posterior variance of $\mu_{i\alpha}$ based on the multivariate HB analysis. Similarly, we use $V_{i\mathrm{HBU}_{\alpha\alpha}}$ to denote the posterior variance of $\mu_{i\alpha}$ based on the univariate analysis using Model $\mathbf{U}_\alpha$. To obtain our HB estimates for each generated data set, we used the algorithm of Gelman and Rubin (1992) with ten parallel paths and 1000 iterations (where the first 500 iterations are discarded as "burn-in") per path. Based on the average values of $V_{i\mathrm{HBM}_{\alpha\alpha}}$ and $V_{i\mathrm{HBU}_{\alpha\alpha}}$ over the 200 data sets, we computed the percentage reductions in the posterior variance for the multivariate method over the univariate method. Percentage reductions are computed for each of the two (recall $s=2$) components of μ_i and for all 12 counties (recall i ranges from 1 to $m=12$). In Table 1 below, for various choices of ρ_e and ρ_v, we have included the minimum, the average, and the maximum of the 12 percentage reductions only for the first component. Similar conclusion emerges from the simulations for the second component.

Table 1: Percentage improvement in posterior variance for multivariate analysis over univariate analysis

(ρ_e,ρ_v)	Min	Avg	Max	(ρ_e,ρ_v)	Min	Avg	Max
(0, 0.5)	5.7	9.1	13.1	(0,−0.5)	6.0	8.8	13.5
(0, 0.95)	30.3	34.3	36.6	(0,−.95)	31.8	32.8	34.8
(0.5,0.5)	0.4	2.8	6.5	(0.5,−0.5)	16.2	21.0	26.5
(0.5,0.95)	18.0	19.9	22.1	(0.5,−0.95)	49.2	51.7	54.8
(0.95,0.5)	1.8	6.2	12.3	(0.95,−0.5)	29.1	39.8	54.5
(0.95,0.95)	4.4	7.8	11.4	(0.95,−0.95)	67.3	74.9	84.3
(−0.5,0.5)	15.4	20.8	26.9	(−0.5,−0.5)	0.3	3.3	7.0
(−0.5,0.95)	48.2	52.2	55.9	(−0.5,−0.95)	19.2	20.1	21.4
(−0.95,0.5)	30.3	41.1	55.5	(−0.95,−0.5)	1.4	6.1	12.4
(−0.95,0.95)	67.2	75.1	84.3	(−0.95,−0.95)	4.6	8.2	12.6

We see from Table 1 below that the percentage reduction can be as small as 0.3 % and as large as 84.3 %, depending on the values of the various parameters ρ_e and ρ_v. Similar simulation results were obtained by Datta *et al.* (1998) for MSE comparisons of the univariate and bivariate EBLUPs.

REMARK 5. Theoretical considerations suggest that the percentage reduction in the posterior variance for the parameter settings $(\rho_e, \rho_v) = (x, y)$ and $(\rho_e, \rho_v) = (-x, -y)$ for any feasible x and y, and fixing other parameters, should be the same. Our simulation results (Table 1) agree with this. The largest improvement is realized when the difference between ρ_e and ρ_v is maximal, corresponding to one at 0.95 and the other -0.95. We put forward a reason to explain this situation. Writing $r_\alpha = \sigma_{v\alpha\alpha}/\sigma_{e\alpha\alpha},\ \alpha = 1, 2$, it can be shown that

$$Pct_{i1}(\psi) = \frac{100 \times n_i(\rho_v\sqrt{r_2} - \rho_e\sqrt{r_1})^2}{(1 + n_i r_1)[(1 - \rho_e^2) + n_i r_2(1 - \rho_v^2)] + n_i(\rho_v\sqrt{r_2} - \rho_e\sqrt{r_1})^2}$$

where $Pct_{i1}(\psi) = 100 \times (g_{i1\mathrm{U}_{11}}(\psi) - G_{i1\mathrm{M}_{11}}(\psi))/g_{i1\mathrm{U}_{11}}(\psi)$, $G_{i1\mathrm{M}_{11}}(\psi)$, where the first diagonal element of $G_{i1\mathrm{M}}(\psi)$ is obtained from (13) and $g_{i1\mathrm{U}_{11}}(\psi)$ is obtained from (15), taking $\lambda_i = 1$ and the other elements equal to zero. Note that $Pct_{i1}(\psi)$ gets bigger when ρ_e and ρ_v have opposite signs. Ignoring $O(m^{-1})$ terms in the posterior variance in (11), and using Laplace's approximation to the integrals, the percentage reduction in the multivariate HB posterior variance over the univariate posterior variance is approximated by $Pct_{i1}(\hat{\psi})$, where $\hat{\psi}$ is the posterior mode of ψ.

To determine if the choice of the scale matrices Φ_e and Φ_v significantly influences the percentage reduction of the posterior variance, we ran simulations with $\rho_e = -0.95$, $\rho_v = 0.95$ and two choices of Φ_v and Φ_e, namely (i) $\Phi_v = \Phi_e = 10\boldsymbol{I}_2$, and (ii) $\Phi_v = \Phi_e = 0.1\boldsymbol{I}_2$. In both cases we ignored the correlation in the scale matrices Φ_v and Φ_e, and we used $a = b = 3$. While in case (i) the minimum percentage reduction is 59.4%, in case (ii) this value is 67.1%. This shows that, even if the scale matrices do not have the same correlation as do the variance matrices, the percentage reductions still remain significantly large.

3.3. *Data analysis and model comparison.* We now analyse the corn and soybean data in Battese *et al.* (1988) using the multivariate and the univariate models. In our data analysis we use the same prior distribution as the one used in our simulations for Table 1. In Table 2 below we report the posterior means and the posterior standard deviations within parentheses for corn and soybean. Identification of the counties may be found from Battese *et al.* (1988). The univariate results are given in the second line below the multivariate results in each cell.

In order to check if the multivariate model gives a better fit to the given data than the corresponding pair of (marginal) univariate models, we implement Bayesian model checking via posterior predictive assessment approach of Gelman *et al.* (1996) (see also Sinha and Dey, 1997). We assess the predictive power of

Table 2: The HB estimates and posterior SDs (in parentheses) for corn and soybean. multivariate estimates in line 1 and univariate estimates in line 2 in each cell. county sample sizes in parentheses are next to county number.

County	Corn Estimate (SD)	Soybean Estimate (SD)	County	Corn Estimate (SD)	Soybean Estimate (SD)
1 (1)	120.8 (10.0) 121.8 (10.9)	76.4 (10.4) 78.1 (12.4)	7 (3)	111.2 (6.2) 112.2 (7.2)	98.8 (6.4) 97.9 (8.0)
2 (1)	127.7 (9.7) 126.7 (10.8)	95.8 (10.1) 94.7 (12.0)	8 (3)	122.0 (6.2) 122.0 (7.2)	112.8 (6.5) 112.3 (8.1)
3 (1)	101.9 (9.4) 105.0 (10.8)	85.7 (9.9) 87.2 (11.8)	9 (4)	116.7 (5.3) 115.8 (6.4)	109.6 (5.6) 109.8 (7.0)
4 (2)	105.2 (7.9) 107.5 (9.0)	77.8 (8.2) 80.7 (10.4)	10 (5)	124.4 (5.0) 124.4 (5.8)	101.3 (5.1) 100.7 (6.5)
5 (3)	146.2 (6.2) 145.0 (7.3)	64.6 (6.4) 66.1 (8.2)	11 (5)	105.4 (4.9) 106.4 (5.9)	120.1 (5.1) 119.0 (6.6)
6 (3)	113.8 (6.1) 112.7 (7.2)	114.0 (6.4) 113.8 (7.9)	12 (5)	144.4 (5.2) 143.5 (6.3)	75.2 (5.5) 75.2 (6.8)

the multivariate and the univariate models (Model M and Model $\mathbf{U}_\alpha, \alpha = 1, 2$) by computing the divergence measure of Laud and Ibrahim (1995). Let $\boldsymbol{y}_{\text{obs}}$ and $\boldsymbol{y}_{\text{new}}$ be the observed and the generated data respectively, where $\boldsymbol{y}_{\text{new}}$ is generated from the posterior predictive distribution $f(\boldsymbol{y}_{\text{new}}, \theta | \boldsymbol{y}_{\text{obs}})$ and θ denotes the parameter vector for the entertained model. To compare between the two models, we calculate the expected divergence measure of Laud and Ibrahim (1995) $d(\boldsymbol{y}_{\text{new}}, \boldsymbol{y}_{\text{obs}}) = E[n^{-1}||\boldsymbol{y}_{\text{new}} - \boldsymbol{y}_{\text{obs}}||^2]$ where the expectation will be taken w.r.t. the posterior predictive distribution. For neither model this expectation can be obtained in closed form. Using the simulated data vector from the posterior predictive distribution $f(\boldsymbol{y}_{\text{new}}, \theta \mid \boldsymbol{y}_{\text{obs}})$, we estimate the divergence measure by $(nB)^{-1} \sum_{l=1}^{B} || \boldsymbol{y}^{(l)} - \boldsymbol{y}_{\text{obs}} ||^2$ where $\{\boldsymbol{y}^{(l)} : 1 \leq l \leq B\}$ denotes the simulated sample from the posterior predictive distribution. These simulated values are obtained by generating $\boldsymbol{y}$ from $f(\boldsymbol{y}|\theta^{(l)})$ corresponding to a Gibbs replicate of θ. Whereas a small value of this divergence measure is expected under an adequate model, a large value of this measure will indicate a lack of fit. Between the two models, we prefer that model which yields a smaller value of this divergence measure. For the crop data in Battese *et al.* (1988) this measure is 1841.76 for the univariate and 1612.28 for the multivariate model, which is slightly in favor of the multivariate Model M.

4. Concluding Remarks

In this paper, we considered multivariate HB prediction of small area means using the multivariate nested error regression model of Battese *et al.* (1988).

Advantages (efficiency gains) of using a multivariate approach over a univariate approach were demonstrated via heuristic considerations and simulations. Similar efficiency gains were noted for the multivariate model by Datta *et al.* (1998) in their EB analysis of the same data. We also characterized a class of improper priors on the components of variance matrices that results in a proper posterior distribution.

Appendix

THEOREM AND PROOF. We introduce notation to state and prove our result on the propriety of the posterior distribution for a model which uses the prior distribution given by (12) and the first two steps of Model M. Let

$$\begin{aligned} \boldsymbol{L} &= \boldsymbol{I}_n - \boldsymbol{C}^{(1)}(\boldsymbol{C}^{(1)T}\boldsymbol{C}^{(1)})^{-1}\boldsymbol{C}^{(1)T},\ \boldsymbol{P}_2 = \boldsymbol{F}^{(1)T}\boldsymbol{L}\boldsymbol{F}^{(1)},\ \ \mathrm{rank}(\boldsymbol{P}_2) = t, \\ \boldsymbol{S}_0 &= \boldsymbol{U}^{(1)}\boldsymbol{U}^{(1)T},\ \ \boldsymbol{S}_1 = \boldsymbol{U}^{(1)}\{\boldsymbol{L} - \boldsymbol{L}\boldsymbol{F}^{(1)}\boldsymbol{P}_2^{-}\boldsymbol{F}^{(1)T}\boldsymbol{L}\}\boldsymbol{U}^{(1)T}, \end{aligned}$$

and

$$\boldsymbol{K} = \Sigma^{-1} - \Sigma^{-1}\boldsymbol{X}(\boldsymbol{X}^T\Sigma^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\Sigma^{-1}. \qquad \ldots(A.1)$$

THEOREM A.1. *Under the multivariate HB model given by (I)-(II) of Model M and the prior distribution in (13), the posterior distribution is proper if and only if*

(i) $n + a_e + a_v - 3s - p - 1 > 0$

(ii) $-t + 2s < a_v < 2$.

The proof follows the arguments of Mukerjee (1997: personal communication), who proved the above theorem for the univariate case (for $s = 1$) in balanced one-way ANOVA model. To prove the theorem we need the following lemmas.

LEMMA A.2. *Let* $\nu_1 < \cdots < \nu_t$ *denote the* t *positive eigenvalues of* $\boldsymbol{P}_2$. *Then*

$$|\Sigma_e|^{n-p-t}|\Sigma_e + \nu_1\Sigma_v|^t \leq \frac{1}{d}|\Sigma||\boldsymbol{X}^T\Sigma^{-1}\boldsymbol{X}| \leq |\Sigma_e|^{n-p-t}|\Sigma_e + \nu_t\Sigma_v|^t, \quad \ldots(A.2)$$

where $d = |\boldsymbol{C}^{(1)T}\boldsymbol{C}^{(1)}|^s$.

PROOF. Recall $\Sigma = \boldsymbol{R} + \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}^T$, where $\boldsymbol{R} = \boldsymbol{I}_n \otimes \Sigma_e$, $\boldsymbol{G} = \boldsymbol{I}_m \otimes \Sigma_v$, and $\boldsymbol{Z} = \boldsymbol{F}^{(1)} \otimes \boldsymbol{I}_s$. Using Exercise 2.9 and Exercise 2.4 of Rao (1973, pp. 32-33) we get

$$\Sigma^{-1} = \boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{Z}(\boldsymbol{G}^{-1} + \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\boldsymbol{R}^{-1}, \qquad \ldots(A.3)$$

$$|\Sigma| = |\boldsymbol{R}||\boldsymbol{G}||\boldsymbol{G}^{-1} + \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z}|. \qquad \ldots(A.4)$$

Then, since $\boldsymbol{X}^T\Sigma^{-1}\boldsymbol{X} = \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{X} - \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{Z}(\boldsymbol{G}^{-1} + \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{X}$, using Exercise 2.4 of Rao (1973, p. 32) again,

$$\begin{aligned} |\boldsymbol{X}^T\Sigma^{-1}\boldsymbol{X}| &= |\boldsymbol{G}^{-1} + \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z} - \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{Z}| \\ &\times\ (|\boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{X}|/|\boldsymbol{G}^{-1} + \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z}|). \end{aligned} \qquad \ldots(A.5)$$

Then, after simplification, it follows from (A.4) and (A.5) that

$$|\Sigma||\boldsymbol{X}^T\Sigma^{-1}\boldsymbol{X}| = d|\Sigma_e|^{n-p}|\Sigma_v|^m|\boldsymbol{I}_m \otimes \Sigma_v^{-1} + \boldsymbol{P}_2 \otimes \Sigma_e^{-1}|. \qquad \ldots (A.6)$$

Writing $\boldsymbol{N} = Diag(\nu_1, ..., \nu_t, 0, ..., 0)$ and $\boldsymbol{N}_1 = Diag(\nu_1, ..., \nu_t)$, it follows that

$$\begin{aligned} |\boldsymbol{I}_m \otimes \Sigma_v^{-1} + \boldsymbol{P}_2 \otimes \Sigma_e^{-1}| &= |\boldsymbol{I}_m \otimes \Sigma_v^{-1} + \boldsymbol{N} \otimes \Sigma_e^{-1}| \\ &= |\boldsymbol{I}_t \otimes \Sigma_v^{-1} + \boldsymbol{N}_1 \otimes \Sigma_e^{-1}||\boldsymbol{I}_{m-t} \otimes \Sigma_v^{-1}| \end{aligned} \qquad \ldots (A.7)$$

It follows by applying Exercise 5(ii) of Rao (1973, p. 70) that

$$|\boldsymbol{I}_t \otimes (\Sigma_v^{-1} + \nu_1\Sigma_e^{-1})| \le |\boldsymbol{I}_t \otimes \Sigma_v^{-1} + \boldsymbol{N}_1 \otimes \Sigma_e^{-1}| \le |\boldsymbol{I}_t \otimes (\Sigma_v^{-1} + \nu_t\Sigma_e^{-1})|,$$

$$\text{i.}e., \;\; \begin{aligned} |\Sigma_v|^{-t}|\Sigma_e|^{-t}|\Sigma_e + \nu_1\Sigma_v|^t &\le |\boldsymbol{I}_t \otimes \Sigma_v^{-1} + \boldsymbol{N}_1 \otimes \Sigma_e^{-1}| \\ &\le |\Sigma_v|^{-t}|\Sigma_e|^{-t}|\Sigma_e + \nu_t\Sigma_v|^t. \end{aligned} \qquad \ldots (A.8)$$

The proof of the lemma follows from (A.6)–(A.8).

Lemma A.3. *For a p.d. matrix $\boldsymbol{M}(a \times a)$, $a \times b$ matrix $\boldsymbol{F}$, and a positive scalar c,*

$$\lim_{l\to\infty}(\boldsymbol{M} + lc\boldsymbol{F}\boldsymbol{F}^T)^{-1} = \boldsymbol{M}^{-1} - \boldsymbol{M}^{-1}\boldsymbol{F}(\boldsymbol{F}^T\boldsymbol{M}^{-1}\boldsymbol{F})^{-}\boldsymbol{F}^T\boldsymbol{M}^{-1}.$$

A proof of this lemma is found in Sallas and Harville (1981).

Lemma A.4. *For the matrices $\boldsymbol{L}$ and $\boldsymbol{P}_2$ defined earlier, a lower and an upper bound of $\boldsymbol{y}^T\boldsymbol{K}\boldsymbol{y}$ is given, respectively, by*

$$\boldsymbol{y}^T[(\boldsymbol{L} - \boldsymbol{L}\boldsymbol{F}^{(1)}\boldsymbol{P}_2^{-}\boldsymbol{F}^{(1)T}\boldsymbol{L}) \otimes \Sigma_e^{-1}]\boldsymbol{y} \;\text{ and }\; \boldsymbol{y}^T\boldsymbol{R}^{-1}\boldsymbol{y}.$$

Proof. Since $\Sigma = \boldsymbol{R} + \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}^T$, for $l > 0$, $(\Sigma + l\boldsymbol{X}\boldsymbol{X}^T)^{-1} - \boldsymbol{R}^{-1}$ is n.p.d., we have

$$\boldsymbol{y}^T(\Sigma + l\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{y} \le \boldsymbol{y}^T\boldsymbol{R}^{-1}\boldsymbol{y}.$$

Since the right hand side does not contain l, the upper limit follows by taking l in limit to infinity and using Lemma A.3.

For the lower limit note that for any $c > 1$ and $l > 0$

$$\boldsymbol{y}^T(\boldsymbol{R} + l\boldsymbol{X}\boldsymbol{X}^T + c\lambda_{\max}\boldsymbol{Z}\boldsymbol{Z}^T)^{-1}\boldsymbol{y} \le \boldsymbol{y}^T(\Sigma + l\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{y},$$

where $\lambda_{\max}$ is the largest eigenvalue of $\boldsymbol{G}$. Since the right hand side does not contain c, by letting c go to infinity and using Lemma A.3, we obtain

$$\begin{aligned} \boldsymbol{y}^T[(\boldsymbol{R} + l\boldsymbol{X}\boldsymbol{X}^T)^{-1} &- (\boldsymbol{R} + l\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{Z}\{\boldsymbol{Z}^T(\boldsymbol{R} + l\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{Z}\}^{-}\boldsymbol{Z}^T \\ &\times (\boldsymbol{R} + l\boldsymbol{X}\boldsymbol{X}^T)^{-1}]\boldsymbol{y} \le \boldsymbol{y}^T(\Sigma + l\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{y}. \end{aligned}$$

Letting l go to infinity in the above inequality and simplifying we obtain

$$\boldsymbol{y}^T[(\boldsymbol{L} - \boldsymbol{L}\boldsymbol{F}^{(1)}\boldsymbol{P}_2^{-}\boldsymbol{F}^{(1)T}\boldsymbol{L}) \otimes \Sigma_e^{-1}]\boldsymbol{y} \le \boldsymbol{y}^T\boldsymbol{K}\boldsymbol{y}.$$

Additional details are in Day (1998).

We now state the following lemma without a proof.

LEMMA A.5. *For two matrices* $\Gamma(n \times n)$ *and* $\boldsymbol{T}(s \times s)$,

$$\boldsymbol{y}^T(\Gamma \otimes \boldsymbol{T})\boldsymbol{y} = tr[\boldsymbol{U}^{(1)}\Gamma\boldsymbol{U}^{(1)T}\boldsymbol{T}].$$

LEMMA A.6. *For a p.d. matrix* $\boldsymbol{Q}(s \times s)$

$$|\Sigma_v|^{-\frac{a_v}{2}}|\Sigma_e|^{-\frac{n-p-t+a_e}{2}}|\Sigma_e+\Sigma_v|^{-\frac{t}{2}}\exp[-\frac{1}{2}tr(\Sigma_e^{-1}\boldsymbol{Q})] \qquad \ldots (A.9)$$

is integrable if and only if
(a) $n + a_e + a_v - 3s - p - 1 > 0$
(b) $-t + 2s < a_v < 2$.

PROOF. Using the transformation $(\Sigma_v, \Sigma_e) \to (\Omega_1, \Sigma_e)$ with $\Omega_1 = \Sigma_e^{-1/2}\Sigma_v \times \Sigma_e^{-1/2}$ and noting the Jacobian of transformation $|\Sigma_e|^{\frac{s+1}{2}}$, the integrand in (A.9) is integrable if and only if

$$|\Sigma_e|^{-\dfrac{n-p-t+a_e+a_v+t-s-1}{2}}\exp[-\frac{1}{2}tr(\Sigma_e^{-1}\boldsymbol{Q})]|\Omega_1|^{-\frac{a_v}{2}}|\boldsymbol{I}_s+\Omega_1|^{-\frac{t}{2}}$$

is integrable.

The last integrand will be integrable if and only if both I_1 and I_2 are finite, where

$$I_1 = \int |\Sigma_e|^{-\frac{n-p+a_e+a_v-s-1}{2}}\exp[-\frac{1}{2}tr(\Sigma_e^{-1}\boldsymbol{Q})]d\Sigma_e,$$
$$I_2 = \int |\Omega_1|^{-\frac{a_v}{2}}|\boldsymbol{I}_s+\Omega_1|^{-\frac{t}{2}}d\Omega_1.$$

Using a property of the inverse Wishart (see Anderson 1984, p. 268, 251) it follows that I_1 is finite if and only if

$$n + a_e + a_v - 3s - p - 1 > 0.$$

Similarly, by a property of the multivariate beta integral (see Anderson 1984, p. 411), it follows that I_2 is finite if and only if

$$-t + 2s < a_v < 2.$$

PROOF OF THEOREM A.1. The posterior distribution will be proper if and only if the joint pdf of $\boldsymbol{y}, \beta, \Sigma_e$ and Σ_v, given by

$$f(\boldsymbol{y},\beta,\Sigma_e,\Sigma_v) \propto |\Sigma|^{-1/2}|\Sigma_e|^{-\frac{a_e}{2}}|\Sigma_v|^{-\frac{a_v}{2}}\exp\{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{X}\beta)^T\Sigma^{-1}(\boldsymbol{y}-\boldsymbol{X}\beta)\},$$

is integrable with respect to β, Σ_e and Σ_v. Equivalently upon integrating β, the posterior is proper if and only if

$$f(\boldsymbol{y}, \Sigma_e, \Sigma_v) \propto \{|\Sigma||\boldsymbol{X}^T\Sigma^{-1}\boldsymbol{X}|\}^{-1/2}|\Sigma_e|^{-\frac{a_e}{2}}|\Sigma_v|^{-\frac{a_v}{2}}\exp\{-\frac{1}{2}\boldsymbol{y}^T\boldsymbol{K}\boldsymbol{y}\} \qquad \ldots(A.10)$$

is integrable with respect to Σ_e and Σ_v, where $\boldsymbol{K}$ is defined in (A.1).

Using Lemmas A.2, A.4, and A.5, we obtain from (A.10)

$$\begin{aligned} & d^{-1/2}|\Sigma_v|^{-\frac{a_v}{2}}|\Sigma_e|^{-\frac{n-p-t+a_e}{2}}|\Sigma_e+\nu_t\Sigma_v|^{-t/2}\exp[-\tfrac{1}{2}tr(\Sigma_e^{-1}\boldsymbol{S}_0)] \\ & \le \{|\Sigma||\boldsymbol{X}^T\Sigma^{-1}\boldsymbol{X}|\}^{-1/2}|\Sigma_e|^{-\frac{a_e}{2}}|\Sigma_v|^{-\frac{a_v}{2}}\exp\{-\tfrac{1}{2}\boldsymbol{y}^T\boldsymbol{K}\boldsymbol{y}\} \\ & \le d^{-1/2}|\Sigma_v|^{-\frac{a_v}{2}}|\Sigma_e|^{-\frac{n-p-t+a_e}{2}}|\Sigma_e+\nu_1\Sigma_v|^{-t/2}\exp[-\tfrac{1}{2}tr(\Sigma_e^{-1}\boldsymbol{S}_1)]. \end{aligned} \qquad \ldots(A.11)$$

The posterior will be proper if the upper bound in (A.11) is integrable, and if the posterior is proper, then the lower bound in (A.11) will be integrable.

Proof of Theorem A.1 is completed by using Lemma A.6.

Acknowledgement. The authors are grateful to Professor Rahul Mukerjee for many valuable discussions. They are also thankful to Professor J.K. Ghosh, Professor Malay Ghosh and an anonymous referee for many useful suggestions, and to Dr. M. Kramer for his careful reading of the manuscript.

References

ANDERSON, T.W. (1984), *An Introduction to Multivariate Statistical Analysis*, 2nd Ed., Wiley, New York.

ARORA, V., LAHIRI, P., AND MUKHERJEE, K. (1997), Empirical Bayes estimation of finite population means from complex surveys, *J. Amer. Statist. Assoc.*, **92**, 1555-1562.

BATTESE, G.E., HARTER, R.M., AND FULLER, W.A. (1988), An error-components model for prediction of county crop areas using survey and satellite data, *J. Amer. Statist. Assoc.*, **80**, 28-36.

BUTAR, F., AND LAHIRI, P. (1997), On the measures of uncertainty of empirical Bayes small-area estimators, unpublished manuscript.

CARTER, G.M., AND ROLPH, J.E. (1974), Empirical Bayes methods applied to estimating fire alarm probabilities, *J. Amer. Statist. Assoc.*, **74**, 269-277.

DATTA, G.S. (1990), Bayesian prediction in mixed linear models with applications to small area estimation, *Unpublished Ph.D. Dissertation*, University of Florida.

−−−− (1992), A unified Bayesian prediction theory for mixed linear models with application, *Statistics and Decisions*, **10**, 337-365.

DATTA, G.S., DAY, BANNMO, AND BASAWA, I.V. (1998), Empirical best linear unbiased and empirical Bayes prediction in multivariate small area estimation, *Journal of Statistical Planning and Inference*, (to appear).

DATTA, G.S., FAY, R.E., AND GHOSH, M. (1991), Hierarchical and empirical multivariate Bayes analysis in small area estimation, In *Proc. of the Bureau of the Census Annual Research Conference*, 63-79. Bureau of the Census, Washington, D.C.

DATTA, G.S., AND GHOSH, M. (1991), Bayesian prediction in linear models: applications to small area estimation, *Ann. Statist.*, **19**, 1748-1770.

Datta, G.S., Ghosh, M., Huang, E.T., Isaki, C.T., Schultz, L.K., and Tsay, J.H. (1992), Hierarchical Bayes and empirical Bayes method for adjustment of the census undercount: the 1988 Missouri dress rehearsal data, **18**, 75-94.

Datta, G.S., Ghosh, M., Nangia, N., and Natarajan, K. (1996), Estimation of median income of four-person families: a Bayesian approach, *Bayesian Analysis in Statistics and Econometrics,* (Eds.: D.A. Berry, K.M. Chaloner and J.K. Geweke), 129-140, Wiley.

Datta, G.S., and Lahiri, P. (1997), A unified measure of uncertainty of estimated best linear unbiased predictor in small-area estimation problems, *Technical Report* 97-7, Department of Statistics, University of Georgia.

Day, Bannmo (1998), Bayes and empirical Bayes estimation with application to small area estimation, *Unpublished Ph.D. Dissertation*, University of Georgia.

Fay, R. E. (1987), Application of multivariate regression to small domain estimation, *Small Area Statistics*, Eds.: R. Platek, J.N.K. Rao, C.E. Sarndal, and M.P. Singh. Wiley, New York, pp. 91-102.

Fay, R. E., and Herriot, R. A. (1979), Estimates of income for small places: an application of James-Stein procedure to census data, *J. Amer. Statist. Assoc.*, **74**, 269-277

Fuller, W.A., and Harter, R.M. (1987), The multivariate components of variance model for small area estimation, *Small Area Statistics*, Eds.: R. Platek, J.N.K. Rao, C.E. Sarndal, and M.P. Singh. Wiley, New York, pp. 103-123.

Gelfand, A., and Smith, A.F.M. (1990), Sampling based approaches to calculating marginal densities, *J. Amer. Statist. Assoc.*, **85**, 398-409.

Gelman, A., Meng, X.-L., and Stern, H. (1996), Posterior predictive assessment of model fitness via realized discrepancies (with discussions), *Statistica Sinica*, **6**, 733-807.

Gelman, A., and Rubin, D.B. (1992), Inference from iterative simulation using multiple sequences, *Statist. Sci.*, **7**, 457-511, (with discussion).

Geman, S., and Geman, D. (1984), Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Trans. Pattern Analysis and Machine Intelligence*, **6**, 721-741.

Ghosh, M., and Lahiri, P. (1987a), Robust empirical Bayes estimation of means from stratified samples, *J. Amer. Statist. Assoc.*, **82**, 1153-1162.

Ghosh, M., and Lahiri, P. (1987b), Robust empirical Bayes estimation of variance from stratified samples, *Sankhyā*, B, **49**, 78-89.

– – –– (1988), Bayes and empirical Bayes analysis in multistage sampling, *Statist. Dec. Theory and Related Topics IV*, Eds.: S.S. Gupta and J.O. Berger, Springer-Verlag, New York, Vol. 1, pp. 195-212.

Ghosh, M., and Lahiri, P. (1992), A hierarchical Bayes approach to small area estimation with auxiliary information (with discussion), *Bayesian Analysis in Statistics and Econometrics,* Eds.: P.K. Goel and N. Iyenger, Springer, Berlin, 107-125.

Ghosh, M., and Meeden, G. (1986), Empirical Bayes estimation in finite population sampling, *J. Amer. Statist. Assoc.*, **81**, 1058-1062.

Ghosh, M., and Rao, J.N.K. (1994), Small area estimation: an appraisal, *Statist. Sci.*, **9**, 55-93, (with discussion).

Hobert, J., and Casella, G. (1996), The effect of improper priors on Gibbs sampling in hierarchical linear mixed models, *J. Amer. Statist. Assoc.*, **91**, 1461-1473.

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.

Laud, P., and Ibrahim, J. (1995), Predictive model selection, *J. Roy. Statist. Soc., Ser. B*, **57**, 247-262.

Lindley, D.V., and Smith, A.F.M. (1972), Bayes estimates for the linear model, *J. Roy. Statist. Soc., Ser. B*, **34**, 1-41, (with discussion).

Mukerjee, R. (1997), On propriety of posterior distributions of the variance components in one-way balanced ANOVA model, (Personal Communication).

Prasad, N.G.N., and Rao, J.N.K. (1990), The estimation of mean squared errors of small area estimators, *J. Amer. Statist. Assoc.*, **85**, 163-171.

Purcell, N.J., and Kish, L. (1979), Estimation for small domains, *Biometrics*, **35**, 365-384.

RAO, C.R. (1973), *Linear Statistical Inference and Its Applications*, 2nd Ed., Wiley, New York.

RAO, J.N.K. (1986), Synthetic estimators, SPREE and the best model based predictors, *Proceedings of the Conference on Survey Methods in Agriculture 1-16.*, U.S. Dept. of Agriculture. Washington, D.C.

SALLAS, W.M., AND HARVILLE, D.A. (1981), Best linear recursive estimation for linear models, *J. Amer. Statist. Assoc.*, **76**, 860-869.

SCHAIBLE, W.L. (ED.) (1993), *Indirect estimators in federal programs*, Statistical Policy Working Paper 21, Prepared by Subcommittee on Small Area Estimation, Federal Committee on Statistical Methodology, Statistical Policy Office, Office of Information and Regulatory Affairs, Office of Management and Budget.

SINHA, D., AND DEY, D. (1997), Semiparametric Bayesian analysis of survival data, *Jour. Amer. Statist. Assoc.*, **92**, 1195-1212.

G.S. DATTA AND B. DAY
DEPARTMENT OF STATISTICS
UNIVERSITY OF GEORGIA
ATHENS, GA 30602.
e-mail: Gaurisankar@hotmail.com

T. MAITI
DIVISION OF STATISTICS
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF NEBRASKA-LINCOLN
NE 68588-0323.
e-mail: Tmaiti@unlinfo.unl.edu