Sankhyā : The Indian Journal of Statistics
Special Issue on Sample Surveys
1999, Volume 61, Series B, Pt. 1, pp. 71-90

ADJUSTED BAYES ESTIMATORS WITH APPLICATIONS TO SMALL AREA ESTIMATION

By MALAY GHOSH
University of Florida, Gainesville
and
TAPABRATA MAITI
University of Nebraska-Lincoln

SUMMARY. Much of the recent research on small area estimation considers estimation of parameters of interest simultaneously for several small or local areas. However, often the objective is to classify these areas into multiple subgroups according to some characteristic of interest, and identify those that are above or below certain threshold values. The usual Bayes estimators, namely the posterior means are often inadequate for such purposes, and need adjustment. In this article we review mainly some of the continuing work on adjusted Bayes estimators so that one can match the histogram of the posterior means with the histogram of the population parameters. The resulting estimators need further adjustment if one is interested also in the posterior means of the ranks. Some applications of the general methods will be given.

1. Introduction

Recent years have witnessed a phenomenal growth in research on small area estimation. For most purposes, the main objective here is simultaneous estimation of parameters in local areas like counties, subcounties, census tracts etc. Ghosh and Rao (1994) provide a review of techniques used for small area estimation based on normal models. The paper also contains examples of small area estimation where these techniques are used. More recently, Ghosh, Natarajan, Stroud and Carlin (1998) address small area estimation problems based on generalized linear models.

On occasions, however, the prime objective is to produce an ensemble of parameter estimates with a histogram very close to the histogram of the unknown parameters. This is important, for example, in comparing the variability of surgical rates across small geographic areas for different operations (Louis and

AMS (1991) *subject classification.* 62D05, 62C10
Key words and phrases. Constrained Bayes; exponential family; hierarchical Bayes; empirical Bayes; Gibbs sampling; rank; regression; small area estimation.

DerSimonian, 1982), in communicating heterogeneity when combining evidence from similar clinical trials or repeated carcinogen bioassays of a single chemical agent (Gilbert, McPeek and Mosteller, 1977). Other examples include classifying local areas into multiple categories based on the proportion of persons not in the labour force (Spojtvell and Thomsen, 1987), adjustment of census counts (Cressie, 1989), correcting the bias in the range of state means (Judkins and Liu, 1997), construction of disease maps, especially for identifying percentage of areas with true incidence greater than some fixed surveillance level (Devine and Louis, 1994; Devine, Louis and Halloran, 1994a,b).

A third problem of interest is to rank the local areas according to certain characteristics, for example according to per capita income or growth in population. Ranking on the basis of "optimal" estimates for these local areas need not necessarily produce optimal "ranking".

The primary objective of this paper is to focus attention on ensemble parameter estimation, and review some of the existing Bayesian methods. We discuss briefly the Bayesian approach to ranking as well. Laird and Louis (1989) have pointed out that the "intuitive" approach of ranking the posterior means can have very poor performance, especially if posterior variances vary considerably among the units.

If the only objective is simultaneous estimation of parameters, assuming squared error loss, the optimal Bayes estimate is the vector of posterior means. Louis (1984) showed that such an estimate may not be appropriate for matching the histogram of the parameters with the histogram of the estimates due to overshrinkage towards the prior expectation. In particular, he showed that under normal likelihood with normal prior, the sampling variability of a collection of Bayes estimates is smaller than the posterior expectation of the corresponding population variability. Ghosh (1992) showed that this result indeed holds irrespective of any distributional assumptions. This implies in particular that the expected prior variation among the parameters always exceeds the expected (over both samples and parameters) variation among the corresponding sample estimators.

It is possible, however, to modify the Bayes procedure to make the empirical distribution function of the Bayes estimates close to the empirical distribution function of the unknown parameters. Louis (1984) formalized this idea by matching the first two moments from the histogram of estimates with the corresponding moments from the posterior histogram of m normal means. In this way, the sampling variability of the collection of estimates is a better estimate of the underlying variability among the population parameters. Ghosh (1992) extended this result for an arbitrary distribution, and found adjusted Bayes estimators for this problem. These estimators, usually referred to as "constrained Bayes estimators", have subsequently been used in a wide range of problems, mostly for finding adjusted empirical Bayes estimators in disease mapping or environmental risk assessment, and using them to identify areas with elevated disease incidence rates (Devine and Louis, 1994; Devine, Louis and Halloran, 1994a,b).

The outline of the remaining sections in this paper is as follows. Section 2 contains the main theoretical results of Louis (1984) and Ghosh (1992) on constrained Bayes estimators followed by several examples. We consider also an alternative formulation due to Spojtvell and Thomsen (1987), and compare their results with the ones due to Louis and Ghosh. Section 3 considers the normal regression model in some details, both for the balanced and unbalanced cases. Constrained hierarchical Bayes estimators are given, and are compared against the constrained empirical Bayes estimators of Louis (1984). Section 4 carries the ideas of Sections 2 and 3 to prediction problems. Section 5 considers an illustrative example, where the underlying theory is applied to small area estimation. Section 6 contains multivariate extensions of the results of Section 2. Finally, Section 7 contains some results on ranking local areas.

It is recently shown in Shen and Louis (1998) that if one wants to have a single set of estimates that produce good ranks, good parameter ensemble and good coordinate-specific estimates, then although no set of estimates can optimize all three inferential goals, it is still possible to strike an effective trade-off. The present article, however, does not discuss this approach.

2. The Main Result and Examples

We first show that finite sample variance of a collection of Bayes estimates is smaller than the posterior expectation of the corresponding population variability. To see this, suppose $\theta_1, \cdots, \theta_m$ are the m parameters of interest, and $e_1^B(\mathbf{x}), \cdots, e_m^B(\mathbf{x})$ are the corresponding Bayes estimates based on the data $\mathbf{x}$ under any quadratic loss. Write $\theta = (\theta_1, \cdots, \theta_m)^T$. Assume that

(A) not all of $\theta_1 - \bar{\theta}, \cdots, \theta_m - \bar{\theta}$ have degenerate posteriors.

Note that the assumption (A) is much weaker than the assumption that $V(\theta|\mathbf{x})$ is positive definite.

Noting now that $e_i^B(\mathbf{x}) = E(\theta_i|\mathbf{x})$, $i = 1, \cdots, m$ using the standard notations $\mathbf{I}_m$ for the identity matrix of order m, $\mathbf{1}_m$ for the m-component column vector with each element equal to 1, and $\mathbf{J}_m = \mathbf{1}_m\mathbf{1}_m^T$, one gets

$$\begin{aligned} E[\sum_{i=1}^m(\theta_i - \bar{\theta})^2|\mathbf{x}] &= tr[(\mathbf{I}_m - m^{-1}\mathbf{J}_m)E(\theta\theta^T|\mathbf{x})] \\ &= tr[(\mathbf{I}_m - m^{-1}\mathbf{J}_m)\{V(\theta|\mathbf{x}) + E(\theta|\mathbf{x})(E(\theta|\mathbf{x}))^T\}] \\ &= tr[(\mathbf{I}_m - m^{-1}\mathbf{J}_m)V(\theta|\mathbf{x})] + \sum_{i=1}^m(e_i^B(\mathbf{x}) - \bar{e}^B(\mathbf{x}))^2 \end{aligned}$$

$$\begin{aligned} &= tr[V(\theta - \bar{\theta}\mathbf{1}_m)|\mathbf{x}] + \sum_{i=1}^{m}(e_i^B(\mathbf{x}) - \bar{e}^B(\mathbf{x}))^2 \\ &> \sum_{i=1}^{m}(e_i^B(\mathbf{x}) - \bar{e}^B(\mathbf{x}))^2. \end{aligned} \qquad \ldots(1)$$

The last equality in (1) follows by noting that $\mathbf{I}_m - m^{-1}\mathbf{J}_m$ is symmetric idempotent and the fact that $tr(\mathbf{AB}) = tr(\mathbf{BA})$. The inequality is a consequence of Assumption (A).

We shall use the notations

$$H_1(\mathbf{x}) = tr[V(\theta - \bar{\theta}\mathbf{1}_m)|\mathbf{x}]; \qquad \ldots(2)$$

$$H_2(\mathbf{x}) = \sum_{i=1}^{m}(e_i^B(\mathbf{x}) - \bar{e}^B(\mathbf{x}))^2. \qquad \ldots(3)$$

Next in this section we derive the constrained Bayes estimate

$$\mathbf{e}^{CB}(\mathbf{x}) = (e_1^{CB}(\mathbf{x}), \cdots, e_m^{CB}(\mathbf{x}))^T$$

of θ, where $\mathbf{e}^{CB}(\mathbf{x})$ minimizes

$$E[\sum_{i=1}^{m}(\theta_i - t_i)^2|\mathbf{x}] \qquad \ldots(4)$$

within the class of all estimates $\mathbf{t}(\mathbf{x}) = \mathbf{t} = (t_1, \cdots, t_m)^T$ of θ that satisfy

$$(a)\ E(\bar{\theta}|\mathbf{x}) = m^{-1}\sum_{i=1}^{m} t_i(\mathbf{x}) = \bar{t}(\mathbf{x}), \qquad \ldots(5)$$

and

$$(b)\ E[\sum_{i=1}^{m}(\theta_i - \bar{\theta})^2|\mathbf{x}] = \sum_{i=1}^{m}(t_i(\mathbf{x}) - \bar{t}(\mathbf{x}))^2. \qquad \ldots(6)$$

The usual Bayes estimate $\mathbf{e}^B(\mathbf{x})$ of θ satisfies (a) but not (b). The following theorem shows how a simple modification of $\mathbf{e}^B$ provides the desired solution.

THEOREM 1. *Let $\boldsymbol{e}^B(\boldsymbol{x}) = (e_1^B(\boldsymbol{x}), \cdots, e_m^B(\boldsymbol{x}))^T$ denote the Bayes estimate of $\theta = (\theta_1, \cdots, \theta_m)^T$ under any quadratic loss based on the data $\boldsymbol{x}$. Let $\mathcal{X}_0 = \{\boldsymbol{x} : H_2(\boldsymbol{x}) > 0\}$. Then for $\boldsymbol{x} \in \mathcal{X}_0$, the solution $\boldsymbol{t} = (t_1, \cdots, t_m)^T$ of (4) subject to (5) to (6) is given by $\boldsymbol{e}^{CB}(\boldsymbol{x}) = (e_1^{CB}(\boldsymbol{x}), \cdots, e_m^{CB}(\boldsymbol{X}))^T$, where*

$$e_i^{CB}(\boldsymbol{x}) = a e_i^B(\boldsymbol{x}) + (1-a)\bar{e}^B(\boldsymbol{x}),\ \ i = 1, \cdots, m; \qquad \ldots(7)$$

$$a \equiv a(\boldsymbol{x}) = [1 + \{H_1(\boldsymbol{x})/H_2(\boldsymbol{x})\}]^{1/2}, \qquad \ldots(8)$$

$H_1(\boldsymbol{x})$ and $H_2(\boldsymbol{x})$ being defined in (3).

REMARK 1. The proof of the theorem is given in Ghosh (1992), and is omitted. We shall refer to e^{CB} as the constrained Bayes (CB) estimate of θ. Equation (7) has the deceptive appearance of expressing the components of e^{CB} as convex combinations of the Bayes estimates of the e_i^B and their averages. This is not so, because $a > 1$.

REMARK 2. In many situations-especially in discrete cases-there is a positive probability that $H_2(\boldsymbol{X})$ is 0; that is, $e_1(\boldsymbol{X}) = \cdots = e_m(\boldsymbol{X})$. Although $e^{CB}(\boldsymbol{X})$ remains undefined with positive probability in such instances, asymptotic (as $m \to \infty$) versions of such estimators still may be meaningful. We shall see this in the important binomial and Poisson examples.

REMARK 3. The constrained Bayes estimators, by construction, satisfy the relation $E[(m-1)^{-1}\sum_{i=1}^m (e_i^{CB}(\boldsymbol{X}) - \bar{e}^{CB}(X))^2] = E[(m-1)^{-1}\sum_{i=1}^m(\theta_i - \bar{\theta})^2]$, where expectation is taken over the joint distribution of $\boldsymbol{X}$ and $\boldsymbol{\theta}$. Consider for example the situation where $X_i|\theta_i$ are independent $N(\theta_i, 18)$ and the θ_i are iid $N(0,9)$. Then the Bayes estimator of θ is $e^B(\boldsymbol{X}) = (\frac{1}{3}X_1, \cdots, \frac{1}{3}X_m)^T$, and $E[(m-1)^{-1}\sum_{i=1}^m (e_i^B(\boldsymbol{X}) - \bar{e}^B(X))^2] = \frac{1}{9}(18+9) = 3$, while $E[(m-1)^{-1}\sum_{i=1}^m (e_i^{CB}(\boldsymbol{X}) - \bar{e}^{CB}(X))^2] = 9$, so that the variation of the Bayes estimators underestimates the variation among the θ_i by a factor of 1/3.

We now apply the preceding theorem to find the constrained Bayes estimates in several examples. The first example is taken from Ghosh (1992).

EXAMPLE 1. (One-parameter exponential family). Suppose that $X_1, \cdots, X_m$ are m independent random variables, wherein X_i has pdf (with respect to some σ-finite measure η) given by

$$f_{\phi_i}(x_i) = \exp(n\phi_i x_i - n\psi(\phi_i)), \quad i = 1, \cdots, m. \qquad \ldots (9)$$

Each X_i can be viewed as the average of n iid random variables, each having a pdf belonging to a one-parameter exponential family. It is assumed that $\psi(\cdot)$ is twice differentiable in its argument. Our objective is to estimate $\theta_i = E_{\phi_i}(X_i) = \psi'(\phi_i), \ \ i = 1, \cdots, m$. Assume independent conjugate priors

$$g(\phi_i) = \exp(\nu\phi_i\mu - \nu\psi(\phi_i)) \qquad \ldots (10)$$

for the ϕ_i's. Then under quadratic loss, the Bayes estimates of the θ_i's are given by

$$\begin{aligned} e_i^B(\boldsymbol{x}) &= E(\theta_i|\boldsymbol{x}) = E[\psi'(\phi_i)|\boldsymbol{x}] \\ &= (1-B)x_i + B\mu, \end{aligned} \qquad \ldots (11)$$

where $B = \nu/(n+\nu)$. Also, from the posterior distribution of θ_i, integration by parts gives

$$\begin{aligned} V(\theta_i|\boldsymbol{x}) &= V[\psi'(\phi_i)|x_i] \\ &= (n+\nu)^{-1}E[\psi''(\phi_i)|x_i] = q_i \quad \text{(say)}. \end{aligned} \qquad \ldots (12)$$

It follows from (11) that $H_2(\boldsymbol{x}) = (1-B)^2 \sum_{i=1}^m (x_i - \overline{x})^2$, whereas from (12) one gets $H_1(\boldsymbol{x}) = (1-m^{-1}) \sum_{i=1}^m q_i$. Then the quantity "$a$" is determined from (8).

Further simplification in calculating H_1 is possible when the X_i's are generated from a quadratic variance function (QVF) subfamily of the natural exponential family (NEF) (cf. Morris 1983a). Then,

$$\begin{aligned} \psi''(\phi_i) &= v_0 + v_1\psi'(\phi_i) + v_2(\psi'(\phi_i))^2 \\ &= v_0 + v_1\theta_i + v_2\theta_i^2, \;\; 1 \le i \le m, \end{aligned} \qquad \ldots (13)$$

where v_0, v_1, and v_2 are not simultaneously 0's and $v_2 < n+\nu$. Then, from (12) and (13),

$$q_i = (n+\nu-v_2)^{-1}[v_0 + v_1 e_i^B(\boldsymbol{x}) + v_2(e_i^B(\boldsymbol{x}))^2]$$

so that

$$\begin{aligned} &H_1(\boldsymbol{x}) = (m-1)(n+\nu-v_2)^{-1} \\ &\times[v_0 + v_1\overline{e}^B(\boldsymbol{x}) + v_2\{(\overline{e}^B(\boldsymbol{x}))^2 + m^{-1}H_2(\boldsymbol{x})\}]. \end{aligned} \qquad \ldots (14)$$

Consequently, for $\boldsymbol{x} \in \mathcal{X}_0$

$$\begin{aligned} &a^2(\boldsymbol{x}) = [1 + v_2(n+\nu-v_2)^{-1}(1-m^{-1})] \\ &+(m-1)(n+\nu-v_2)^{-1} \\ &\times[v_0 + v_1\overline{e}^B(\boldsymbol{x}) + v_2(\overline{e}^B(\boldsymbol{x}))^2]/H_2(\boldsymbol{x}). \end{aligned} \qquad \ldots (15)$$

When the X_i's are averages of iid Bernoulli random variables, $v_0 = 0$, $v_1 = 1$, and $v_2 = -1$. Then when the x_i's are not equal,

$$\begin{aligned} &a^2(\boldsymbol{x}) = [1 - (1-m^{-1})(n+\nu+1)^{-1}] \\ &+(m-1)(n+\nu+1)^{-1}\overline{e}^B(\boldsymbol{x})(1-\overline{e}^B(\boldsymbol{x})) \\ &\times(1-B)^{-2} \Big/ \sum_{i=1}^m (x_i - \overline{x})^2. \end{aligned} \qquad \ldots (16)$$

In the Poisson case, $v_0 = v_2 = 0$ and $v_1 = 1$. Then

$$a^2(\boldsymbol{x}) = 1 + (m-1)(n+\nu+1)^{-1}\overline{e}^B(\boldsymbol{x})(1-B)^{-2} \div \sum_{i=1}^m (x_i - \overline{x})^2, \qquad \ldots (17)$$

when the x_i's are not all equal. Finally, in the normal example $v_0 = \sigma^2$ (known), whereas $v_1 = v_2 = 0$. Then

$$\begin{aligned} a^2(\boldsymbol{x}) &= 1 + (m-1)(n+\nu)^{-1}(1-B)^{-2}\sigma^2 \div \sum_{i=1}^m (x_i - \overline{x})^2 \\ &= 1 + (m-1)n^{-1}\sigma^2 \Big/ \left[(1-B)\sum_{i=1}^m (x_i - \overline{x})^2\right]. \end{aligned} \qquad \ldots (18)$$

In the normal case the probability that all the X_i's are equal is 0

Note that marginally $X_1, \cdots, X_m$ are iid with mean μ and variance $C = (\nu^{-1} + n^{-1})E[\psi''(\phi_1)] = (v_0 + v_1\mu + v_2\mu^2)/[nB - v_2(1-B)]$. Hence as $m \to \infty$, $\overline{\boldsymbol{X}}$ and *a fortiori* $\overline{e}^B(\boldsymbol{X})$ converges a.s. to μ and $\sum_{i=1}^m (X_i - \overline{X})^2/(m-1)$ converges a.s. to C. Now after some algebraic simplifications, it follows that $a^2(\boldsymbol{X})$ converges a.s. to $(1-B)^{-1}$. Thus for large m, an approximate CB estimator of θ is given by $e_0^{CB}(\boldsymbol{X}) = (e_{01}^{CB}(\boldsymbol{X}), \cdots, e_{0m}^{CB}(\boldsymbol{X}))$, where

$$e_{0i}^{CB}(\boldsymbol{X}) = (1-B)^{-1/2}[(1-B)X_i + B\mu] + [1-(1-B)^{-1/2}]\{(1-B)\overline{X} + B\mu\}. \quad \ldots (19)$$

In the special Bernoulli example, Spjϕtvell and Thomsen (1987) used a different formulation to arrive at the estimator e^{ST} of θ with the ith component of e^{ST} given by

$$e_i^{ST}(\boldsymbol{X}) = (1-B)^{1/2}X_i + (1-(1-B)^{1/2})\mu, \ \ 1 \le i \le m. \qquad \ldots (20)$$

This estimator is different from $\boldsymbol{e}^{CB}$ or even $\boldsymbol{e}_0^{CB}$ but the estimator works in the same spirit, as there is less shrinking towards the prior mean μ.

It is possible to generalize the results of Spjϕtvell and Thomsen (1987). Suppose one assumes that (a) $X_i|\theta_i$ are independent with means θ_i and variances $\mu_2(\theta_i)$, (b) θ_i are independent with common mean μ and variance τ^2, and (c) $\mu_2(\theta_i)$ have common mean σ^2, $i = 1, \cdots, m$. Then for linear estimators $aX_i + b$ of the θ_i, assuming squared error loss, the Bayes risk is given by $m[a^2\sigma^2 + (1-a)^2\tau^2 + (b-(1-a)\mu)^2]$. Then the unrestricted linear Bayes estimator of $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_m)^T$ is given by $\boldsymbol{e}^B(\boldsymbol{X}) = [(1-B)X_1 + B\mu, \cdots, (1-B)X_m + B\mu]^T$, where $B = \sigma^2/(\sigma^2+\tau^2)$. However, subject to the constraints (i) $E[m^{-1}\sum_{i=1}^m t_i] = E[m^{-1}\sum_{i=1}^m \theta_i]$ and (ii) $E[m^{-1}\sum_{i=1}^m (t_i - \bar{t})^2] = E[m^{-1}\sum_{i=1}^m (\theta_i - \bar{\theta}^2]$, one gets the restricted Bayes estimator $[(1-B)^{1/2}X_1 + (1-(1-B)^{1/2})\mu, \cdots, (1-B)^{1/2}X_m + (1-(1-B)^{1/2})\mu]^T$ of $\boldsymbol{\theta}$. For the one-parameter exponential family with conjugate priors, the linear Bayes estimator, indeed, equals the unrestricted Bayes estimator, but the restricted Bayes estimator of Spjϕtvell and Thomsen does not agree with the Louis-Ghosh estimator even in the limiting case ($m \to \infty$) as noted earlier.

It is possible to use different n_i's instead of the same n in (14), change B into $B_i = \nu/(n_i+\nu)$'s, and obtain resulting CB estimates after suitable modifications of H_1 and H_2. We see this in Section 3 for the important normal example in a more general regression context.

One major difference of the CB estimates compared to the usual Bayes estimates is that unlike the latter, the former change if one uses a weighted distance in (4) instead of the Euclidean distance. This fact was pointed out by Louis (1984) in the normal example. Because in practice specification of the exact loss is difficult if not impossible, we consider only the Euclidean distance in the remainder of this article. Also, closed form estimates are not possible to obtain without the Euclidean distance. For a discussion of weighted squared error loss, see also Cressie (1989).

Finally, in this section we point out that it is possible to avoid any distributional assumptions and work instead with the assumption of posterior linearity, that is,

$$E(\theta_i|\boldsymbol{x}) = c_i x_i + d_i, \;\; 1 \leq i \leq m, \qquad \ldots (21)$$

where X_i is the average of n_i iid rv's. This formulation is essentially the same as the one due to Spjϕtvell and Thomsen (1987). Once again assume the conditions (a)-(c). Then, as Goldstein (1975) has shown, $E(\theta_i|\boldsymbol{x}) = (1 - B_i)x_i + B_i\mu$ and $V(\theta_i|\boldsymbol{x}) = \sigma^2 n_i^{-1}(1 - B_i)$, where $B_i = M/(M + n_i)$ and $M = \sigma^2/\tau^2$. The CB estimator of θ now can be obtained by applying Theorem 1. The idea of posterior linearity appeared in Diaconis and Ylvisaker (1979), Ericson (1969), Hartigan (1969), and Goldstein (1975), among others. Lahiri (1990) used this idea to find constrained CB predictors of finite population means under the slightly more restrictive assumption that $\mu_2(\theta_i)$ is a quadratic in θ_i. A careful analysis of Lahiri's proof reveals that this assumption is not needed, however.

3. **Estimation of Normal Means**

In the event of unknown prior parameters, it is possible to find constrained empirical Bayes (CEB) estimators of θ by substituting estimators of these hyperparameters in $H_1(\boldsymbol{X})$ and $H_2(\boldsymbol{X})$ as given in (3) and thus $a(\boldsymbol{X})$ as given in (8). For example, in the NEF-QVF setup Morris (1988) estimates μ by $\overline{X}$ and B by $\tilde{B}$ ($\hat{B}$ in Morris's notation), where

$$\begin{aligned}\tilde{B} &= (v_2/(n + v_2))(m - 1)m^{-1} + (n/(n + v_2)) \\ &\times [n^{-1}(v_0 + v_1\overline{X} + v_2\overline{X}^2)](m - 3) \div \sum_{i=1}^{m}(X_i - \overline{X})^2.\end{aligned}$$

These substitutions in the $a(\boldsymbol{X})$ expression as given in (20) will enable construction of CEB estimators of θ by applying (7), where the Bayes estimator $\boldsymbol{e}^B$ and accordingly $\overline{e}^B$ also are replaced by their EB versions $\boldsymbol{e}^{EB}$ and $\overline{e}^{EB}$.

The preceding approach can be contrasted with a HB approach that assigns distributions (often improper) to these hyperparameters to reflect their uncertainties. For example, in the NEF-QVF setup one can adopt Stroud's (1991) approach to assign distributions on the hyperparameters μ and ν and compare the resulting constrained HB (CHB) estimators with the CEB's. The CHB estimators generally are fairly complicated even for the NEF-QVF subfamily, due to the lack of closed form expressions for $E(\theta|\boldsymbol{x})$ and $V(\theta|\boldsymbol{x})$.

For simplicity, in this section, we shall consider a hierarchical normal regression model for balanced data. The CHB's will be computed and in a special case will be compared with the CEB's. Also, important special cases of this model will be highlighted.

EXAMPLE 2. Consider the following hierarchical model:

(a) Conditional on θ, β, and $\tau^2(>0)$ let $\boldsymbol{X} \sim N(\theta, \sigma^2 n^{-1}\boldsymbol{I}_m)$, where $\sigma^2(>0)$ is known;

(b) Conditional on β and τ^2, let $\theta \sim N(\boldsymbol{Z}\beta, \tau^2\boldsymbol{I}_m)$, where $\boldsymbol{Z}(m \times r)$ is a known design matrix of rank $r(<m)$;

(c) β and τ^2 are marginally independent with $\beta \sim$ uniform (R^r) and $\tau^2 \sim$ uniform $(0, \infty)$. This particular model was considered in Ghosh (1991) and Morris (1983b). A special case of the model appeared in Morris (1983c).

Write $B = n^{-1}\sigma^2/(n^{-1}\sigma^2 + \tau^2)$, $\hat{B} = E(B|\boldsymbol{X})$, and $\boldsymbol{P}_{\boldsymbol{Z}} = \boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T$. Then it can be shown that (Ghosh 1991)

$$\begin{aligned} \boldsymbol{e}^{HB}(\boldsymbol{X}) &= E(\theta|\boldsymbol{X}) = (1-\hat{B})\boldsymbol{X} + \hat{B}\boldsymbol{P}_{\boldsymbol{Z}}\boldsymbol{X}; \\ V(\theta|\boldsymbol{X}) &= V(B|\boldsymbol{X})(\boldsymbol{I}_m - \boldsymbol{P}_z)\boldsymbol{X}\boldsymbol{X}^T(\boldsymbol{I}_m - \boldsymbol{P}_z) + n^{-1}\sigma^2[(1-\hat{B})\boldsymbol{I}_m + \hat{B}\boldsymbol{P}_{\boldsymbol{Z}}]; \\ H_1(\boldsymbol{X}) &= V(B|\boldsymbol{X})[\boldsymbol{X}^T(\boldsymbol{I}_m - \boldsymbol{P}_{\boldsymbol{Z}})\boldsymbol{X} - m^{-1}\{\boldsymbol{X}^T(\boldsymbol{I}_m - \boldsymbol{P}_{\boldsymbol{Z}})\boldsymbol{1}_m\}^2] \\ &\quad + n^{-1}\sigma^2[(m-1)(1-\hat{B}) + \hat{B}tr\{(\boldsymbol{I}_m - m^{-1}\boldsymbol{J}_m)\boldsymbol{P}_{\boldsymbol{Z}}\}]. \end{aligned} \quad \ldots (22)$$

The HB estimator $\boldsymbol{e}^{HB}$ is a weighted average of the sample estimate $\boldsymbol{X}$ and its projection on the column space of $\boldsymbol{Z}$. The constrained Bayes estimator e^{CHB} usually puts greater weight on $\boldsymbol{X}$ than $\boldsymbol{e}^{HB}$, thereby preventing overshrinking to the regression surface.

REMARK 4. In many important special cases (the usual regression models that include the intercept terms), $\boldsymbol{1}_m \in$ the column space of $\boldsymbol{Z}$. Then $\boldsymbol{P}_{\boldsymbol{Z}}\boldsymbol{1}_m = \boldsymbol{1}_m$. Consequently, $\bar{e}^{HB}(\boldsymbol{X}) = m^{-1}\boldsymbol{1}_m^T\boldsymbol{e}^{HB}(\boldsymbol{X}) = (1-\hat{B})\overline{\boldsymbol{X}} + \hat{B}\overline{\boldsymbol{X}} = \bar{\boldsymbol{X}}$ and then $H_2(\boldsymbol{X}) = \sum_{i=1}^m[(1-\hat{B})(X_i - \overline{X}) + \hat{B}(\boldsymbol{Z}_i^T\hat{\boldsymbol{\beta}} - \overline{X})]^2$. Also, from (22),

$$H_1(\boldsymbol{X}) = V(B|\boldsymbol{X})\boldsymbol{X}^T(\boldsymbol{I}_m - \boldsymbol{P}_{\boldsymbol{Z}})\boldsymbol{X} + n^{-1}\sigma^2[m - 1 - (m-r)\hat{B}]. \quad \ldots (23)$$

REMARK 5. Morris (1983b) dealt with further special case where $\boldsymbol{Z} = \boldsymbol{1}_m$. Then $\boldsymbol{P}_{\boldsymbol{Z}} = m^{-1}\boldsymbol{J}_m$, $\hat{B} = \overline{X}$, $H_2(\boldsymbol{X}) = (1-\hat{B})^2\sum_{i=1}^m(X_i - \overline{X})^2$, and $H_1(\boldsymbol{X}) = V(B|\boldsymbol{X})\sum_{i=1}^m(X_i - \overline{X})^2 + n^{-1}\sigma^2(m-1)(1-\hat{B})$. Consequently, from (18),

$$a_{HB}^2(\boldsymbol{X}) = 1 + V(B|\boldsymbol{X})/(1-\hat{B})^2 + n^{-1}\sigma^2(m-1) \div \left\{(1-\hat{B})\sum_{i=1}^m(X_i - \overline{X})^2\right\}. \quad \ldots (24)$$

It is instructive to compare the CHB estimator with the corresponding CEB estimator in Morris's (1983c) framework. To find the CEB estimator first find the CB estimator of θ under the hierarchical model of Example 2 without the hyperorior given in (c). This is obtained by using the $a(\boldsymbol{X})$ given in (18). In this expression substitute $\overline{\boldsymbol{X}}$ for μ and $\hat{B}$ for B. Writing the corresponding a as a_{EB}, one gets

$$a_{EB}^2(\boldsymbol{X}) = 1 + n^{-1}\sigma^2(m-1) \Big/ \left\{(1-\hat{B})\sum_{i=1}^m(X_i - \overline{X})^2\right\}. \quad \ldots (25)$$

The expression $a_{EB}(\boldsymbol{X})$ essentially appeared in Louis (1984), who also suggested estimating B by $\hat{B}_{JS} = (m-3)/\sum_{i=1}^{m}(X_i-\overline{X})^2$, $(m \geq 4)$, the James-Stein shrinker as a possibility. A comparison of a_{HB} and a_{EB}, however, reveals that unlike the former, the latter fails to incorporate the uncertainty due to estimation of B. The same phenomenon appears when one wants to estimate the variances associated with EB estimators as opposed to HB estimators. In subgroup analyses, where one requires less shrinking of the individual estimates towards an overall average, it often is preferable to use a_{HB} rather than a_{EB}, because $a_{HB} > a_{EB}$. This is especially so when the contribution from $V(B|\boldsymbol{X})(1-\hat{B})^{-2}$ cannot be neglected in comparison with the other terms.

As proved in Datta and Ghosh (1991a), however, asymptotically (as $m \to \infty$), $\hat{B} \to B$ and $V(B|\boldsymbol{X}) \to 0$ a.s. Also, as discussed earlier $\sum_{i=1}^{m}(X_i - \overline{X})^2/(m-1) \to \sigma^2 n^{-1} B^{-1}$ a.s. Then both $a_{HB}(\boldsymbol{X})$ and $a_{EB}(\boldsymbol{X})$ converge to $(1-B)^{-1/2}$ a.s., as does $a(\boldsymbol{X})$ of (18). Thus for large m, one may recommend use of the estimator $e_0^{CHB}(\boldsymbol{X})$ for θ with its ith component given by

$$\begin{aligned} e_{0i}^{CHB}(\boldsymbol{X}) &= (1-\hat{B})^{-1/2}[(1-\hat{B})X_i + \hat{B}\overline{X}] + (1-(1-\hat{B})^{-1/2})\overline{X} \\ &= (1-\hat{B})^{1/2}X_i + (1-(1-\hat{B})^{1/2})\overline{X}. \end{aligned} \quad \ldots (26)$$

This estimator is similar to the one in Spjϕtvell and Thomsen (1987) with μ replaced by $\bar{X}$.

4. Prediction Problems

The ideas in the preceding sections can easily be extended to prediction problems. Suppose $\gamma(m \times 1)$ is a random vector that needs to be predicted on the basis of the data $\boldsymbol{x}$. In a Bayesian framework the problem is the same as the one given in Section 1. First find $\boldsymbol{e}^{BP}(\boldsymbol{x}) = E(\gamma|\boldsymbol{x})$, as well as $V(\gamma|\boldsymbol{x})$. Then one computes $H_1^*(\boldsymbol{x})= \mathrm{tr}[(\boldsymbol{I}_m - m^{-1}\boldsymbol{J}_m)V(\gamma|\boldsymbol{x})]$ and $H_2^*(\boldsymbol{x}) = \sum_{i=1}^{m}(e_i^{BP}(\boldsymbol{x}) - \overline{e}^{BP}(\boldsymbol{x}))^2$, and for $\boldsymbol{x} \in \mathcal{X}^* = \{\boldsymbol{x} : H_2^*(\boldsymbol{x}) > 0\}$, $a^*(\boldsymbol{x}) = [1 + H_1^*(\boldsymbol{x})/H_2^*(\boldsymbol{x})]^{1/2}$. The constrained Bayes predictor (CBP) of γ is denoted by $\boldsymbol{e}^{CBP}(\boldsymbol{x})$ and its ith component is given by

$$\boldsymbol{e}_i^{CHP}(\boldsymbol{x}) = a^*(\boldsymbol{x})e_i^{BP}(\boldsymbol{x}) + (1-a^*(\boldsymbol{x}))\overline{\boldsymbol{e}}^{BP}(\boldsymbol{x}), 1 \leq i \leq m. \quad \ldots (27)$$

One simple illustration of such a prediction procedure is as follows. Suppose there are m strata with the ith stratum containing N_i population units. Denote by $x_{i1}, \cdots, x_{iN_i}$ the measurements of a certain characteristic associated with the N_i population units in the ith stratum. Without loss of generality denote by $x_{i1}, \cdots, x_{in_i}$ the corresponding measurements associated with the n_i sampled units. The objective is to predict $\gamma = (\gamma_1, \cdots, \gamma_m)^T$, where $\gamma_i = N_i^{-1}\sum_{j=1}^{N_i} x_{ij} (i = 1, \cdots, m)$ on the basis of $\boldsymbol{x}_i = (x_{i1}, \cdots, x_{in_i})^T$, $i = 1, \cdots, m$. Write $\boldsymbol{x}_i^* = (x_{\overline{in_i+1}}, \cdots, x_{iN_i})^T$ and $\boldsymbol{x}^T = (\boldsymbol{x}_1^T, \cdots, \boldsymbol{x}_m^T)$.

Consider the following Bayesian model:

(a) conditional on $\theta_1, \cdots, \theta_m$, $\binom{x_i}{x_i^*} \overset{\text{ind}}{\sim} N(\theta_i \mathbf{1}_{N_i}, \sigma^2 \boldsymbol{I}_{N_i})$:

(b) θ_i's are iid $N(\mu, \tau^2)$.

It follows then, from Ghosh and Meeden (1986) or Ghosh and Lahiri (1987), that

$$E(\gamma_i|\boldsymbol{x}) = (1 - f_i B_i)\overline{x}_i + f_i B_i \mu, \;\; 1 \le i \le m; \qquad \ldots (28)$$

$$V(\gamma_i|\boldsymbol{x}) = f_i \sigma^2 [N_i^{-1} + f_i n_i^{-1}(1 - B_i)], \;\; 1 \le i \le m; \qquad \ldots (29)$$

$$\text{Cov}(\gamma_i, \gamma_k|\boldsymbol{x}) = 0, \;\; 1 \le i \ne k \le m, \qquad \ldots (30)$$

where $\overline{x}_i = n_i^{-1} \sum_{j=1}^{n_i} x_{ij}$, $B_i = \sigma^2 n_i^{-1}/(\sigma^2 n_i^{-1} + \tau^2)$, and $f_i = (N_i - n_i)/N_i$. Accordingly one finds the CBP's by computing

$$H_1^*(\boldsymbol{x}) = tr[(\boldsymbol{I}_m - m^{-1}\boldsymbol{J}_m)V(\gamma|\boldsymbol{x})] = (1 - m^{-1}) \sum_{i=1}^{m} V(\gamma_i|\boldsymbol{x}),$$

and

$$H_2^*(\boldsymbol{x}) = \sum_{i=1}^{m} [E(\gamma_i|\boldsymbol{x}) - E(\overline{\gamma}|\boldsymbol{x})]^2,$$

where $\overline{\gamma} = m^{-1} \sum_{i=1}^{m} \gamma_i$. Lahiri (1990) obtained CEBP's in this case by first finding estimates of μ, σ^2, and τ^2 and then substituting them in $E(\gamma_i|\boldsymbol{x})$'s and $V(\gamma_i|\boldsymbol{x})$'s. He also replaced the normality assumption by the weaker assumption of "posterior linearity."

The approach of Lahiri ignores any contribution due to possible posterior dependence between the γ_i's in the construction of $H_i^*(\boldsymbol{x})$. Such a dependence is genuinely introduced if, for example, one introduces prior distributions on μ, σ^2, and τ^2. For instance, one may use a diffuse prior on (μ, σ^2, τ^2). Indeed HB predictors of finite population means in a more general framework that includes convariates were given in Datta and Ghosh (1991b). Constrained HB predictors can be obtained from these HB predictors and the associated variance-covariance matrix by using (3), (7), and (8). We see an application of the Datta and Ghosh (1991b) method for a specific model in the next section.

5. An Application

The present application concerns estimating average wages and salaries of workers in a certain industry consisting of 114 units and spread across 16 small areas. We also want to identify areas with very low or very high average wages. A simple random sample of size 38 is taken, and the sampled units are poststratified into these 16 areas. In the process it turns out that three areas (to be labeled 14, 15, and 16) have no sample representation.

Let x_{ij} denote the average wages and salaries and k_{ij} denote the gross business income for unit j in area $i(j = 1, \cdots, N_i;\ i = 1, \cdots, 16)$. Write $\boldsymbol{x}_{is} = (x_{i1}, \cdots, x_{in_i})^T$, the vector corresponding to the $n_i(\geq 0)$ sampled units in are i. The corresponding vector for unsampled units is denoted by $\boldsymbol{x}_{iu} = (x_{\overline{in_{i+1}}}, \cdots, x_{iN_i})^T$,. Let $\boldsymbol{x}_i^T = (\boldsymbol{x}_{is}^T, \boldsymbol{x}_{iu}^T), i = 1, \cdots, 16$, and $\boldsymbol{x}_s^T = (\boldsymbol{x}_{1s}^T, \cdots, \boldsymbol{x}_{16s}^T)$. Similarly, let $\boldsymbol{k}_{is} = (k_{i1}, \cdots, k_{in_i})^T$, and $\boldsymbol{k}_i^T = (\boldsymbol{k}_{is}^T\ \boldsymbol{k}_{iu}^T)$, $i = 1, \cdots, 16$. The objective is to estimate $\gamma = (\gamma_1, \cdots, \gamma_{16})^T$, where $\gamma_i = N_i^{-1} \sum_{j=1}^{N_i} x_{ij}$, as well as to identify the areas with very low or very high average wage.

The following HB model is used for prediction :

(a) conditional on $\boldsymbol{B} = \boldsymbol{b}, \boldsymbol{R} = r$ and $\Lambda = \lambda$,

$$\boldsymbol{x}_i \overset{\text{ind}}{\sim} N(b_0 \mathbf{1}_{N_i} + b_1 \boldsymbol{k}_i, r^{-1}(\lambda^{-1} \boldsymbol{J}_{N_i} + \boldsymbol{D}_i)), \qquad \ldots (31)$$

where $\boldsymbol{D}_i = \text{diag}(k_{i1}, \cdots, k_{iN_i})$;

(b) $\boldsymbol{B}, R$, and ΛR are marginally independently distributed with $\boldsymbol{B} \sim$ uniform (E^2), E^2 being the two-dimensional Euclidean space, $\boldsymbol{R} \sim \text{Gamma}(\frac{1}{2}a_0, \frac{1}{2}g_0)$ and $\Lambda R \sim \text{Gamma}(\frac{1}{2}a_1, \frac{1}{2}g_1)$. For the sake of illustration we have used diffuse gamma priors on R and λR with $a_0 = g_0 = g_1 = 0$ and $a_1 = .00005$. Putting $a_1 = 0$ can lead to an improper posterior, which we want to avoid.

State (a) of the model can be identified as a mixed effects model with

$$x_{ij} = b_0 + b_1 k_{ij} + v_i + e_{ij}^* k_{ij}^{1/2}, \qquad \ldots (32)$$

where v_i's and e_{ij}^*'s are mutually independent with v_i's iid $N(0, (\lambda r)^{-1})$ and e_{ij}^*'s iid $N(0, r^{-1})$.

The model given in (a) and (b) is a special case of the one given in Datta and Ghosh (1991b). One can use their general formulas and much cumbersome algebra to arrive at the expression for $e_i^{HBP}(\boldsymbol{x}_s) = E(\gamma_i|\boldsymbol{x}_s)$, $V(\gamma_i|\boldsymbol{x}_s)$, $H_1^*(\boldsymbol{x}_s)$, and $H_2^*(\boldsymbol{x}_s)$.

But before providing these expressions, we need to introduce a few notations. Write $\overline{k}_{is} = n_i^{-1} \sum_{j=1}^{n_i} k_{ij}, h_i = \sum_{j=1}^{n_i} k_{ij}^{-1}$,

$$\begin{aligned}
\overline{k}_{iu} &= (N_i - n_i)^{-1} \sum_{j=n_i+1}^{N_i} k_{ij}, \quad f_i = (N_i - n_i)/N_i \\
\boldsymbol{c}_i^T &= [\lambda(\lambda + h_i)^{-1}, \overline{k}_{iu} - n_i(\lambda + h_i)^{-1}] \\
\boldsymbol{t}^T &= \left[\sum_{i=1}^{16} \lambda(\lambda + h_i)^{-1} \left(\sum_{j=1}^{n_i} x_{ij} k_{ij}^{-1} \right),\ \sum_{i=1}^{16} n_i \left\{ \overline{x}_{is} - (\lambda + h_i)^{-1} \sum_{j=1}^{n_i} x_{ij} k_{ij}^{-1} \right\} \right] \\
\boldsymbol{D} &= \begin{bmatrix} \sum_{i=1}^{16} \lambda h_i (\lambda + h_i)^{-1} & \sum_{i=1}^{16} \lambda n_i (\lambda + h_i)^{-1} \\ \sum_{i=1}^{16} \lambda n_i (\lambda + h_i)^{-1} & \sum_{i=1}^{16} n_i (\overline{k}_{is} - n_i (\lambda + h_i)^{-1}) \end{bmatrix}
\end{aligned}$$

$$Q = \sum_{i=1}^{16}\sum_{j=1}^{16} x_{ij}^2 k_{ij}^{-1} - \sum_{i=1}^{16}\left(\sum_{j=1}^{n_i} x_{ij}k_{ij}^{-1}\right)^2 (\lambda+h_i)^{-1} - \boldsymbol{t}^T\boldsymbol{D}^{-1}\boldsymbol{t}.$$

Then writing $n_T = \sum_{i=1}^{16} n_i$,

$$E[\gamma_i|\lambda,\boldsymbol{x}_s] = (1-f_i)\overline{x}_{is} + f_i\left[\left(\sum_{j=1}^{n_i} x_{ij}k_{ij}^{-1}\right)(\lambda+h_i)^{-1} + \boldsymbol{c}_i^T\boldsymbol{D}^{-1}\boldsymbol{c}_i,\right]; \quad \ldots(33)$$

$$V[\gamma_i|\lambda,\boldsymbol{x}_s] = \{(.00005\lambda + Q)/(n_T-2)\} \times\{f_i^2(\lambda+h_i)^{-1} + N_i^{-1}f_i\overline{k}_{iu} + f_i^2\boldsymbol{c}_i^T\boldsymbol{D}^{-1}\boldsymbol{c}_i\}, \quad \ldots(34)$$

and

$$\begin{aligned} H_1^*(\boldsymbol{x}_s) &= \sum_{i=1}^{16} V(\gamma_i|\boldsymbol{x}_s) \\ &\quad -\tfrac{1}{16}E\left[\sum_{i=1}^{16} f_i^2(\lambda+h_i)^{-1} + \sum_{i=1}^{16} N_i f_i\overline{k}_{iu} \right.\\ &\qquad \left. + \left(\sum_{i=1}^{16} f_i\boldsymbol{c}_i\right)^T D^{-1}\left(\sum_{i=1}^{16} f_i\boldsymbol{c}_i\right)\middle|\boldsymbol{x}_s\right] \\ &\quad -\tfrac{1}{16}V\left[\sum_{i=1}^{16}(1-f_i)\overline{x}_{is} + \sum_{i=1}^{16} f_i\left\{\left(\sum_{j=1}^{n_i} x_{ij}k_{ij}^{-1}\right)(\lambda+h_i)^{-1}\right.\right.\\ &\qquad \left.\left. + \boldsymbol{c}_i^T\boldsymbol{D}^{-1}\boldsymbol{c}_i\right\}\middle|\boldsymbol{x}_s\right]. \end{aligned} \quad \ldots(35)$$

Table 1. The true means (M_i), sample means ($\overline{X}_{is}$), HB Predictors (e_i^{HBP}), Associated standard errors (S_i^{HBP}), constrained HB predictors (e_i^{CHBP}) associated standard errors (S_i^{HBP}), sample sizes (n_i) and population sizes (N_i)

i	γ_i	$\overline{X}_{1s}$	e_i^{HBP}	S_i^{HBP}	e_i^{CHBP}	s_i^{CHBP}	n_i	N_i
1	15.1200	6.5833	17.4274	3.563	17.427	3.563	3	6
2	6.8745	3.2250	6.7589	3.183	5.295	3.504	1	4
3	17.4140	2.1710	19.6852	3.994	19.995	4.006	1	8
4	11.5698	18.3890	12.5387	2.608	11.868	2.693	2	6
5	32.4575	32.3325	30.1349	2.726	31.878	3.236	4	6
6	23.3928	19.5270	18.6833	2.793	18.855	2.799	3	6
7	16.5509	22.1341	18.4179	1.672	18.553	1.678	10	27
8	10.8578	5.1900	12.7396	3.748	12.096	3.803	1	5
9	12.1669	10.7585	13.9667	2.895	13.491	2.934	2	12
10	20.3914	10.6780	15.6535	3.573	15.410	3.581	1	7
11	37.5058	6.7797	29.5763	5.239	31.243	5.498	3	6
12	18.3070	12.4655	14.5781	2.033	14.187	2.071	6	13
13	6.4550	4.0950	4.3120	3.425	2.512	3.879	1	2
14	19.4550	–	26.0043	11.221	27.181	11.283	0	1
15	7.4230	–	6.8753	6.697	5.427	6.852	0	1
16	46.8227	–	31.5270	7.006	33.461	7.268	0	4

Table 1 presents our analysis of the data. In the following,

$$\begin{aligned}(S_i^{CHBP})^2 &= E[\theta_i - e_i^{CHBP}(\boldsymbol{x}_s))^2|\boldsymbol{x}_s] \\ &= (S_i^{HBP})^2 + [e_i^{HBP}(\boldsymbol{x}_s) - e_i^{CHBP}(\boldsymbol{x}_s)]^2.\end{aligned}$$

From Table 1 it is clear that the sample averages in most situations are far from the true averages and also that three of the 16 areas have no sample representation. Indeed the average of the squared deviations of the sample means from the true means γ_i - namely $\sum_{i=1}^{13}(\gamma_i - \overline{x}_{is})^2/13$, is 117.1865, whereas the average bias, $(13)^{-1}\sum_{i=1}^{13}(\overline{x}_{is} - \gamma_i)$, is -5.749. Also, the average relative error for the sample means, that is $(13)^{-1}\sum_{i=1}^{13}|\overline{x}_{is} - \gamma_i|/\gamma_i$, is .4373. In contrast, $\sum_{i=1}^{16}(\gamma_i - e_i^{HBP})^2/16$ is 26.8905, $(16)^{-1}\sum_{i=1}^{16}\times(e_i^{HBP} - \gamma_i)$ is -1.493, and $(16)^{-1}\sum_{i=1}^{16}|e_i^{HBP} - \gamma_i|/\gamma_i$ is .1754. Thus on average the HB predictors result in a 77.05% reduction in the average squared deviations and a 59.89% reduction in the average relative error compared to the sample means. Also, $\sum_{i=1}^{16}(e_i^{CHBP} - \gamma_i)^2/16 = 23.8471$, $\sum_{i=1}^{16}(e_i^{CHBP} - \gamma_i)/16 = -1.493$, and $(16)^{-1}\sum_{i=1}^{16}|e_i^{CHBP} - \gamma_i|/\gamma_i = .2061$. Thus the constrained HB predictors effect a 79.65% reduction in the average squared deviations and a 52.87% reduction in the average relative error compared to the sample means.

More important in this context is to identify the areas with low or high wages and salaries. Specifically, we want to identify the areas where average wages and salaries are below \$12,000 or above \$30,000. From the table we find that actually average wages and salaries are below \$12,000 in areas 2, 4, 8, 13, and 15 and above \$30,000 in areas 5, 11, and 16. The usual HB predictors identify areas 2, 13, and 15 with average wages and salaries below \$12,000 and areas 5 and 16 with average wages and salaries above \$30,000. In contrast, the constrained HB predictors identify areas 2, 4, 13, and 15 with average wages and salaries below \$12,000 and areas 5, 11, and 16 with average wages and salaries above \$30,000. It is unfortunate that we have no sample representation from area 16, which has the highest average wages and salaries. Nevertheless, the HB predictors substantially amend the lack of sample representation, and the constrained HB predictors get even closer to the truth. In conclusion, this example suggests very emphatically that when faced with the dual problem of estimation (or prediction) and subgroup identification, it may be wiser to use the constrained Bayes estimator (or predictors) rather than the usual Bayes estimators (or predictors).

6. Multivariate Extensions

Suppose now $\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_m$ are vector-valued parameters. Let $E(\boldsymbol{\theta}_i|\boldsymbol{x}) = \boldsymbol{e}_i^B(\boldsymbol{x}), i =$

$1, \cdots, m$. Then writing $\bar{\boldsymbol{\theta}} = m^{-1}\sum_{i=1}^m \boldsymbol{\theta}_i$, one gets

$$E[\sum_{i=1}^m (\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}})(\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}})^T|\boldsymbol{x}] = \boldsymbol{H}_1(\boldsymbol{x}) + \boldsymbol{H}_2(\boldsymbol{x}), \qquad \ldots (36)$$

where $\boldsymbol{H}_1(\boldsymbol{x}) = \sum_{i=1}^m V(\boldsymbol{\theta}_i|\boldsymbol{x}) - mV(\bar{\boldsymbol{\theta}}|\boldsymbol{x})$, $\boldsymbol{H}_2(\boldsymbol{x}) = \sum_{i=1}^m \left[\boldsymbol{e}_i^B(\boldsymbol{x}) - \bar{\boldsymbol{e}}^B(\boldsymbol{x})\right] \left[\boldsymbol{e}_i^B(\boldsymbol{x}) - \bar{\boldsymbol{e}}^B(\boldsymbol{x})\right]^T$, and $\bar{\boldsymbol{e}}^B(\boldsymbol{x}) = m^{-1}\sum_{i=1}^m \boldsymbol{e}_i^B(\boldsymbol{x})$.

Generalizing the formulation of Section 2, our objective is to find $\boldsymbol{t}_1, \cdots, \boldsymbol{t}_m$ which minimize $E\left[\sum_{i=1}^m (\boldsymbol{\theta}_i - \boldsymbol{t}_i)(\boldsymbol{\theta}_i - \boldsymbol{t}_i)^T|\boldsymbol{x}\right]$ subject to

$$E(\bar{\boldsymbol{\theta}}|\boldsymbol{x}) = m^{-1}\sum_{i=1}^m \boldsymbol{t}_i(\boldsymbol{x}) = \bar{\boldsymbol{t}}(\boldsymbol{x}); \qquad \ldots (37)$$

$$E[\sum_{i=1}^m (\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}})(\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}})^T|\boldsymbol{x}] = \sum_{i=1}^m [\boldsymbol{t}_i(\boldsymbol{x}) - \bar{\boldsymbol{t}}(\boldsymbol{x})][\boldsymbol{t}_i(\boldsymbol{x}) - \bar{\boldsymbol{t}}(\boldsymbol{x})]^T. \qquad \ldots (39)$$

Write

$$\begin{aligned} E[\sum_{i=1}^m (\boldsymbol{\theta}_i - \boldsymbol{t}_i)(\boldsymbol{\theta}_i - \boldsymbol{t}_i)^T|\boldsymbol{x}] &= \sum_{i=1}^m E[(\boldsymbol{\theta}_i - \boldsymbol{e}_i^B(\boldsymbol{x}))(\boldsymbol{\theta}_i - \boldsymbol{e}_i^B(\boldsymbol{x}))^T|\boldsymbol{x}] \\ &+ \sum_{i=1}^m (\boldsymbol{e}_i^B(\boldsymbol{x}) - \boldsymbol{t}_i)(\boldsymbol{e}_i^B(\boldsymbol{x}) - \boldsymbol{t}_i)^T. \end{aligned} \qquad \ldots (39)$$

Thus we need to minimize $\sum_{i=1}^m \left[\boldsymbol{e}_i^B(\boldsymbol{x}) - \boldsymbol{t}_i\right]\left[\boldsymbol{e}_i^B(\boldsymbol{x}) - \boldsymbol{t}_i\right]^T$ with respect to $\boldsymbol{t}_1, \cdots, \boldsymbol{t}_m$ subject to (35) and (36). Since (35) forces $\bar{\boldsymbol{e}}^B(\boldsymbol{x}) = \bar{\boldsymbol{t}}$, one gets

$$\begin{aligned} &\sum_{i=1}^m (\boldsymbol{e}_i^B(\boldsymbol{x}) - \boldsymbol{t}_i)(\boldsymbol{e}_i^B(\boldsymbol{x}) - \boldsymbol{t}_i)^T \\ &= \sum_{i=1}^m \left[(\boldsymbol{e}_i^B(\boldsymbol{x}) - \bar{e}^B(\boldsymbol{x})) - (\boldsymbol{t}_i - \bar{\boldsymbol{t}})\right]\left[(\boldsymbol{e}_i^B(\boldsymbol{x}) - \bar{e}^B(\boldsymbol{x})) - (\boldsymbol{t}_i - \bar{\boldsymbol{t}})\right]^T \\ &= \sum_{i=1}^m \left[\boldsymbol{e}_i^B(\boldsymbol{x}) - \bar{\boldsymbol{e}}^B(\boldsymbol{x})\right]\left[\boldsymbol{e}_i^B(\boldsymbol{x}) - \bar{\boldsymbol{e}}^B(\boldsymbol{x})\right]^T + \sum_{i=1}^m (\boldsymbol{t}_i - \bar{\boldsymbol{t}})(\boldsymbol{t}_i - \bar{\boldsymbol{t}})^T \\ &- \sum_{i=1}^m \left[\boldsymbol{e}_i^B(\boldsymbol{x}) - \bar{\boldsymbol{e}}^B(\boldsymbol{x})\right](\boldsymbol{t}_i - \bar{t})^T - \sum_{i=1}^m (\boldsymbol{t}_i - \bar{\boldsymbol{t}})\left[\boldsymbol{e}_i^B(\boldsymbol{x}) - \bar{\boldsymbol{e}}^B(\boldsymbol{x})\right]^T \\ &= m\left[\boldsymbol{V}(\boldsymbol{Z}_1) + \boldsymbol{V}(\boldsymbol{Z}_2) - Cov(\boldsymbol{Z}_1, \boldsymbol{Z}_2) - Cov(\boldsymbol{Z}_2, \boldsymbol{Z}_1)\right], \end{aligned} \qquad \ldots (40)$$

where $P(\boldsymbol{Z}_1 = \boldsymbol{e}_i^B(\boldsymbol{x}), \boldsymbol{Z}_2 = \boldsymbol{t}_i) = \frac{1}{m}$ for all $i = 1, \cdots, m$. Hence, the expression given in the right hand side of (37) is minimized if $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are linearly related, that is, $\boldsymbol{Z}_2 = \boldsymbol{A}\boldsymbol{Z}_1 + \boldsymbol{b}$. Now, $\bar{\boldsymbol{e}}^B(\boldsymbol{x}) = E(\boldsymbol{Z}_1), \bar{\boldsymbol{t}} = E(\boldsymbol{Z}_2)$ and $\bar{\boldsymbol{e}}^B(\boldsymbol{x}) = \bar{\boldsymbol{t}}$. So, $\boldsymbol{b} = (\boldsymbol{I} - \boldsymbol{A})\bar{\boldsymbol{t}}$ and hence, $\boldsymbol{Z}_2 = \boldsymbol{A}\boldsymbol{Z}_1 + (\boldsymbol{I} - \boldsymbol{A})\bar{\boldsymbol{t}}$, that is $\boldsymbol{t}_i - \bar{\boldsymbol{t}} = \boldsymbol{A}(\boldsymbol{e}_i^B(\boldsymbol{x}) - \bar{\boldsymbol{e}}^B(\boldsymbol{x}))$.

Hence from (34),

$$\begin{aligned}\boldsymbol{H}_1(\boldsymbol{x})+\boldsymbol{H}_2(\boldsymbol{x}) &= \sum_{i=1}^{m}(\boldsymbol{t}_i-\bar{\boldsymbol{t}})(\boldsymbol{t}_i-\bar{\boldsymbol{t}})^T \\ &= \boldsymbol{A}\sum_{i=1}^{m}(\boldsymbol{e}_i^B(\boldsymbol{x})-\bar{\boldsymbol{e}}^B(\boldsymbol{x}))(\boldsymbol{e}_i^B(\boldsymbol{x})-\bar{\boldsymbol{e}}^B(\boldsymbol{x}))^T\boldsymbol{A}^T \\ &= \boldsymbol{A}\boldsymbol{H}_2(\boldsymbol{x})\boldsymbol{A}^T.\end{aligned} \quad \ldots(41)$$

Now observe that $\boldsymbol{A}=(\boldsymbol{H}_1(\boldsymbol{x})+\boldsymbol{H}_2(\boldsymbol{x}))^{\frac{1}{2}}\boldsymbol{H}_2^{-\frac{1}{2}}(\boldsymbol{x})$ is a solution of (39). Hence, one gets

$$\boldsymbol{t}_i=(\boldsymbol{H}_1+\boldsymbol{H}_2)^{\frac{1}{2}}\boldsymbol{H}_2^{-\frac{1}{2}}\boldsymbol{e}_i^B+(\boldsymbol{I}-(\boldsymbol{H}_1+\boldsymbol{H}_2)^{\frac{1}{2}}\boldsymbol{H}_2^{-\frac{1}{2}})\bar{\boldsymbol{e}}^B. \quad \ldots(42)$$

To illustrate the above method, suppose $\boldsymbol{X}_i|\boldsymbol{\theta}_i$ are independent $N(\boldsymbol{\theta}_i,\boldsymbol{\Sigma})$, and $\boldsymbol{\theta}_i$ are iid $N(\boldsymbol{\mu},\boldsymbol{A})$, where $\boldsymbol{\Sigma}$ and $\boldsymbol{A}$ are known. Then the $\boldsymbol{\theta}_i$ are a posteriori independent normal with means $(\boldsymbol{I}-\boldsymbol{B})\boldsymbol{X}_i+\boldsymbol{B}\boldsymbol{\mu}$, where $\boldsymbol{B}=\boldsymbol{\Sigma}(\boldsymbol{\Sigma}+\boldsymbol{A})^{-1}$, and variance $(\boldsymbol{\Sigma}^{-1}+\boldsymbol{A}^{-1})^{-1}$. Assuming squared error loss, the Bayes estimator of the $\boldsymbol{\theta}_i$ are $(\boldsymbol{I}-\boldsymbol{B})\boldsymbol{X}_i+\boldsymbol{B}\boldsymbol{\mu}$. Now, using the formula given earlier, one gets $\boldsymbol{H}_1(\boldsymbol{x})=(m-1)(\boldsymbol{\Sigma}^{-1}+\boldsymbol{A}^{-1})^{-1}$; $\boldsymbol{H}_2(\boldsymbol{x})=(\boldsymbol{I}-\boldsymbol{B})\sum_{i=1}^m(\boldsymbol{X}_i-\bar{\boldsymbol{X}})(\boldsymbol{X}_i-\bar{\boldsymbol{X}})^T(\boldsymbol{I}-\boldsymbol{B})^T$. Now from (40), one gets the constrained Bayes estimators of the $\boldsymbol{\theta}_i$. An interesting application occurs in Devine, Louis and Halloran (1994a) for finding EB estimators of spatially correlated disease incidence rates. They have $m=1$, $\boldsymbol{\Sigma}=\sigma^2\boldsymbol{P}$, where $\boldsymbol{P}$ is a diagonal matrix, and $\boldsymbol{A}=\tau^2 exp(-\gamma\boldsymbol{D})$, where $\boldsymbol{D}$ is a symmetric distance matrix with ijth element equal to d_{ij}, and γ is a parameter reflecting the intensity of the spatial dependence.

7. Rank Estimators

As mentioned in the introduction, ranking based on the posterior means need not necessarily be optimal especially when there is much variability in the posterior variances. Ranking based on the constrained Bayes estimates will not rectify the defect. It is more appropriate to find $E[R(\theta_i)|x]$, where $R(\theta_i)$ is the rank of θ_i among $\theta_1,\cdots,\theta_m$. To find this, we first write $R(\theta_i)$ as

$$R(\theta_i)=1+\sum_{k=1(\neq i)}^{m} I_{[\theta_i\leq\theta_k]} \quad \ldots(43)$$

so that

$$E[R(\theta_i)|\boldsymbol{x}]=1+\sum_{j=1(\neq i)}^{m} P(\theta_i\leq\theta_j|\boldsymbol{x}). \quad \ldots(44)$$

Laird and Louis (1989) used (42) for ranking several schools. If one wants only integer ranks, one can rerank $E[R(\theta_i)|\boldsymbol{x}], (i = 1, \cdots, m)$.

To have a specific example, consider the following hierarchical model:

(i)$X_i|\boldsymbol{\theta}, A, \boldsymbol{b} \overset{ind}{\sim} N(\theta_i, V_i), V_i(known), i = 1, \cdots, m$;

(ii) $\theta_i|A, \boldsymbol{b} \overset{ind}{\sim} N(\boldsymbol{z}_i^T\boldsymbol{b}, A), i = 1, \cdots, m$;

(iii) Marginally $\boldsymbol{b}$ and A are independent with $\boldsymbol{b}$ uniform (R^p), $A \sim$ Uniform $(0, \infty)$.

In the above, $p < m$. Also, we assume $r(\boldsymbol{Z}) = p$, where $\boldsymbol{Z}^T = (\boldsymbol{z}_1, \cdots, \boldsymbol{z}_m)$. Then, writing $\boldsymbol{V} = diag(V_1, \cdots, V_m)$, $\boldsymbol{P_Z} = \boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T$,

$$E(\boldsymbol{\theta}|A, \boldsymbol{x}) = \left[\boldsymbol{V}^{-1} + A^{-1}(\boldsymbol{I} - \boldsymbol{P_Z})\right]^{-1}\boldsymbol{V}^{-1}\boldsymbol{x}; \qquad \ldots(45)$$

$$V(\boldsymbol{\theta}|A, \boldsymbol{x}) = [V^{-1} + A^{-1}(\boldsymbol{I} - \boldsymbol{P_Z})]^{-1}. \qquad \ldots(46)$$

Also, the posterior distribution of A is

$$\Pi(A|\boldsymbol{x}) \propto a^{\frac{1}{2}b}\prod_{i=1}^{m}(a + V_i)^{-\frac{1}{2}}|\sum_{i=1}^{m}(1 - B_i)\boldsymbol{z}_i\boldsymbol{z}_i^T|^{-\frac{1}{2}}\exp(-\frac{1}{2}Q_A(\boldsymbol{x})), \qquad \ldots(47)$$

where $B_i = \frac{V_i}{A+V_i}$, and

$$Q_A(\boldsymbol{x}) = A^{-1}\left[\sum_{i=1}^{m}(1 - B_i)x_i^2 - \{\sum_{i=1}^{m}(1 - B_i)x_i\boldsymbol{z}_i\}^T \times \{\sum_{i=1}^{m}(1 - B_i)\boldsymbol{z}_i\boldsymbol{z}_i^T\}^{-1}\{\sum_{i=1}^{m}(1 - B_i)\boldsymbol{z}_i x_i\}\right].$$

Now writing $\hat{\theta}^A(\boldsymbol{x})$ as the ith component of $E(\boldsymbol{\theta}|A, \boldsymbol{x})$, and $v_{ij}^A(\boldsymbol{x})$ as the (i, j)th component of $V(\boldsymbol{\theta}|A, \boldsymbol{x})$, it follows after some manipulations that

$$\begin{aligned} E[R(\theta_i)|\boldsymbol{x}] &= E\{E[R[\theta_i)|A, \boldsymbol{x}]|\boldsymbol{x}\} \\ &= 1 + \sum_{j=1(\neq i)}^{m} E\left[\Phi\left(\frac{\hat{\theta}_i^A - \hat{\theta}_j^A}{\sqrt{v_{ii}^A + v_{jj}^A - 2v_{ij}^A}}\right)|\boldsymbol{y}\right], \end{aligned}$$

where Φ denotes the standard normal distribution function.

We see now an application of the above result in ranking several states according to the median income of four-person families. Such estimates are needed at the national, state, county and local area levels for a variety of governmental decisions. The U.S. Department of Health and Human Services has a direct need for such data at the state level (the 50 states and the District of Columbia) for formulating its energy assistance program to low income families. Such estimates are provided to HHS annually by the Bureau of Census.

There are three sources of data. The basic source is the annual demographic supplement to the March sample of the current population survey (CPS) which provides annually median income by states for families of different sizes. Second, once every ten years, similar figures are obtained from the decinnial census for the year preceding the census year, for example 1969, 1979, 1989 and so on. Third, the Bureau of the Census also uses annual estimates of per capita income obtained by the Bureau of Economic Analysis (BEA) of the U.S. Department of Commerce.

Fay (1987), Datta, Fay and Ghosh (1991), Datta, Ghosh, Nangia and Natarajan (1996), Ghosh, Nangia and Kim (1996), and Datta, Lahiri and Maiti (1997) have obtained estimates of median income of four-person families for all the 51 states by employing different empirical and hierarchical Bayes methods. This section considers instead ranking 15 out of these 51 states based on the Bayes estimators, constrained Bayes estimators, and reranking the posterior means of the ranks. This analysis is primarily for the sake of illustration - pointing out that ranking based on the three different criteria can be quite distinct.

To this end, let X_i denote the CPS estimator of the median income of four-person families for the ith state ($i = 1, \cdots, 15$) in 1989. Also, let z_i denote the adjusted census median for the ith state. The adjusted census median is found by multiplying the census median for the ith state for the recentmost decennial census, in this case 1980, and then multiplying the same by the ratio of the current per capita income of the ith state as obtained by the Bureau of economic analysis to the per capita income of the ith state at the recentmost decennial census year. The hierarchical model considered is the same as the one in (i)-(iii) with $\boldsymbol{z}_i^T \boldsymbol{b} = b_0 + b_1 z_i$ ($i = 1, \cdots, 15$). The ranking results based on the different estimates mentioned earlier is given in the following table:

Table 2. The ranks based on different methods

Sate	CPS	HB	CB	Est. Rank
ME	1	1	1	1
NH	12	9	9	9
VT	4	4	4	3
MA	15	15	15	15
RI	11	6	6	6
CT	14	3	3	4
NY	8	14	14	14
NJ	13	13	13	11
PA	3	8	8	8
OH	9	10	10	10
IN	2	2	2	2
IL	6	11	11	12
MI	7	12	12	13
WI	5	5	5	5
MN	10	7	7	7

It is clear from the above table that ranking based on the three methods are most often the same, but in some cases they can provide distinct results. In this case, the Bayes and constrained Bayes estimators provide perfect agreement in respective ranking, but reranking the posterior means of ranks can provide sometimes a different answer. For example, Connecticut and Vermont trade places depending on whether ranking is based on Bayes (constrained Bayes) estimators or the posterior means of ranks.

References

Cressie, N. (1989), Empirical Bayes estimation of undercount in the Decennial Census, *Jour. Amer. Statist. Assoc.*, **84**, 1033-1044.

Datta, G.S., and Ghosh, M. (1991a), Asymptotic optimality of hierarchical bayes estimators and predictors, *Jour. Statistical Planning and Inference*, **29**, 229-243.

− − −− (1991b), Bayesian prediction in mixed linear models: Applications to small area estimation, *Ann. Statist.*, **19**, 1748-1770.

Datta, G.S., Fay, R.E., and Ghosh, M. (1991). Hierarchical and empirical multivariate analysis in small area estimation , in *Proceeding of the Bureau of the Census Annual Research Conference,* 63-79. Bureau of the Census, Washington, D.C.

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., Lahiri, P., and Maiti, T. (1997). Empirical Bayes estimation of median income of four-person families by state using Time Series and Cross Sectional Data , unpublished manuscript.

Devine, O. and Louis, T.A. (1994). A constrained empirical Bayes estimator for incidence rates in areas with small populations. *Statistics in Medicine*, **13**, 1119-1133.

Devine, O., Louis, T.A., and Halloran, M.E. (1994a). Empirical Bayes estimators for spatially correlated incidence rates. *Environmetrics*, **5**, 381-398.

− − −− (1994b). Empirical Bayes estimators for stabilizing incidence rates before mapping. *Epidemiology*, **5**, 622-630.

Diaconis, P., and Ylvisakar, D. (1979), Conjugate priors for exponential families, *Ann. Statist.*, **7**, 269-281.

Ericson, W.A. (1969), A note on the posterior mean, *Jour. Royal Statist. Soc.*, Ser. B, **31**, 332-334.

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

Ghosh, M. (1992), Constrained Bayes estimation with applications , *Jour. Amer. Statist. Assoc.*, **87**, 533-540.

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

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

Ghosh, M., Nangia, N. and Kim, Dal Ho (1996). Estimation of median income of four-person families: A Bayesian time series approach , *Jour. Amer. Statist. Assoc.*, **91**, 1423-1431.

Ghosh, M., Natarajan, K., Stroud, T.W.F., and Carlin, B.P. (1998). Generalized linear models for small area estimation, *Jour. Amer. Statist. Assoc.*, **93**, 273-282.

Ghosh, M. and Rao, J.N.K. (1994). Small area estimation: An appraisal , *Statistical Science*, **9**, 55-93.

Gilbert, J.P., McPeek, B., and Mosteller, F. (1977). Statistics and ethics in surgery and anesthesia , *Science*, **198**, 684-689.

GOLDSTEIN, M. (1975), A note on some Bayesian nonparameteric estimates, *Ann. Statist.*, **3**, 736-740.

HARTIGAN, J.A. (1969), Linear Bayes methods, *Jour. Royal Statist. Soc.*, Ser. B, **31**, 446-454.

JUDKINS, D.R. AND LIU, J. (1998). Correcting the bias in the range of a statistic across small areas. Preprint

LAHIRI, P. (1990), Adjusted Bayes and empirical Bayes estimation in finite population sampling, *Sankhyā* B, **42**, 50-66.

LAIRD, N., AND LOUIS, T.A. (1989), Empirical Bayes ranking methods, *Journal of Educational Statistics*, **14**, 29-46.

LINDLEY, D.V., AND SMITH, A.F.M. (1972), Bayes estimates for the linear model (with discussion), *Jour. Royal Statist. Soc.*, Ser. B, **34**, 1-41.

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

LOUIS, T.A. (1991), Using empirical Bayes methods in biopharmaceutical research (with discussion), *Statistics and Medicine*, **10**, 811-829.

LOUIS, T.A., AND DERSIMONIAN, R. (1982). Health statistics based on discrete population groups. In *Regional Variations in Hospital Use*, ed. D.L. Rothberg. Heath.

MORRIS, C. (1983a), Natural exponential families with quadratic variance functions: Statistical theory, *Ann. Statist.*, **11**, 515-529.

MORRIS, C. (1983b), Parametric empirical Bayes inference: Theory and applications (with discussion), *Jour. Amer. Statist. Assoc.*, **78**, 47-65.

– – –– (1983c), Parametric empirical Bayes confidence intervals, in *Scientific Inference, Data Analysis and Robustness*, eds. G.E.P. Box, T. Leonard, and C.F.J. Wu, New York: Academic Press, pp. 25-50.

– – –– (1988), Determining the accuracy of Bayesian empirical Bayes estimates in familiar exponential families, in *Statistical Decision Theory and Related Topics IV* (Vol. 1), eds. S.S. Gupta, and J.O. Berger, New York: Springer-Verlag, pp. 251-263.

SHEN, W. AND LOUIS, T.A. (1998). Triple-threat estimates in exchangeable models. *Jour. Royal Statist. Soc.* B, **60**, 455-471.

SPJϕTVELL, E., AND THOMSEN, I. (1987), Application of some empirical Bayes methods to small area statistics, in *Proceedings of the International Statistical Institute*, (Vol. 2), pp. 435-449.

STROUD, T.W.F. (1991), Hierarchical Bayes predictive means and variances with application to sample survey inference, *Communications in Statistics. Theory and Methods*, **20**, 13-36.

MALAY GHOSH
UNIVERSITY OF FLORIDA
103 GRIFFIN-FLOYD HALL
P.O. BOX 118545
GAINESVILLE, FL 32611-8545
e-mail: ghoshm@stat.ufe.edu

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