Statistica Sinica **7**(1997), 1053-1063

ON THE SUPERIORITY OF THE BAYESIAN METHOD OVER THE BLUP IN SMALL AREA ESTIMATION PROBLEMS

Vipin Arora and P. Lahiri

InfoStat Inc. and University of Nebraska-Lincoln

Abstract: The Bayes estimator of a small area mean is shown to have strictly smaller mean squared error (MSE) than that of the corresponding best linear unbiased predictor (BLUP) for the Kleffe-Rao model, an extended mixed model with random sampling variances. The model is then extended to incorporate sampling weights, covariances and unequal sample sizes. A hierarchical Bayesian procedure which takes into account various sources of variabilities has been proposed. A specific small area estimation problem using data from the U.S. Consumer Expenditure Survey is considered. Based on a robust (i.e., model free) evaluation criterion, the proposed hierarchical Bayes estimator turns out to be superior to both estimated BLUP and the direct survey estimators. The posterior variances which measure the accuracy of the hierarchical Bayes estimates are always smaller than the corresponding variances of the direct survey estimates. The current state of estimated BLUP theory is not rich enough to provide reliable estimates of the MSE of the estimated BLUP for the example considered in this article.

Key words and phrases: Adaptive rejection sampling, consumer expenditure, Gibbs sampling, integrated Bayes risk, mean squared error, sampling weight.

1. Introduction

The sampling design and the sample size of most of the large scale national surveys are usually determined so as to produce a national estimate of a parameter of interest with a desired level of precision. Quite often there is a need to produce similar estimates with the same level of precisions for certain subnational regions (for example, state, counties, etc.). This task cannot be achieved by the regular design-based procedures which use survey data only from the subnational region under consideration simply because of the availability of a smaller sample (relative to the national sample). Similar situations arise when estimates are needed for many domains obtained by classifying the population according to various demographic characteristics (for example, age, sex, race, etc.). Such problems in survey sampling literature are known as small area estimation problems.

Due to budgetary constraints, it is unrealistic to increase sample size for the small areas. Thus, estimates which use implicit or explicit models to borrow strength from the related sources have been proposed. For a review of small area estimation procedures and their applications the reader is referred to Ghosh and Rao (1994) and Rao (1986).

The Bayes and the best linear unbiased prediction (BLUP) methods have been widely used to produce small area statistics. For a very general mixed linear model with *fixed* variance components, Theorem 4 of Datta and Ghosh (1991) shows that the BLUP and the corresponding Bayes approaches produce identical point predictors (estimators) of a small area characteristic. The BLUP method is very popular among classical statisticians and has been extensively used in the mixed linear model and small area literature. To the best of our knowledge there is no situation cited in the literature where the Bayes procedure produces better (in terms of the mean squared error) point estimators than the corresponding BLUP. In Sections 2 of this paper, we cite a situation (an extended mixed model with *random* variance components) where the Bayes estimator of a small area mean has *strictly* smaller MSE than the corresponding BLUP (see Theorem 1). However, unlike the BLUP, in order to calculate the Bayes estimator, one needs to specify the parametric form of the hierarchical Bayes model and the Bayes estimator is not linear.

In Section 3, the well known small area model due to Fay and Herriot (1979) is revisited. All the previous classical and Bayesian estimation procedures, which used the Fay-Herriot model or its extensions, assumed known sampling variances. But, in practice, they are usually not known and are estimated from the generalized variance curves (see, for example, Wolter (1985), chap. 5). The effect of such estimated sampling variances on the estimation of the MSE or posterior variance has not been studied so far. This is a very important issue for the researchers of various federal agencies in the U.S. and other countries. In this section we consider an alternate modeling which addresses this issue. This model can be viewed as an extension of the small area model due to Kleffe and Rao (1992) to incorporate the sampling weights and relevant covariates which are generally available from most sample surveys. Kleffe and Rao (1992) provided a second order approximation to the MSE of the EBLUP (estimated BLUP). Their formula relies on the assumption that n, the sample size in each small area, is bounded but m, the number of small areas, tends to infinity. This assumption may not be satisfied in certain situations (for example, the data set considered in Section 4). The hierarchical Bayesian procedure proposed in Section 3 has a clear edge over the EBLUP procedure in such a situation since it does not depend on any sample size assumption in producing the *exact* measures of accuracy of the hierarchical Bayes estimates.

Generally, improper non-informative priors are put on the hyper parameters in standard hierarchical Bayesian analysis (see Ghosh and Mukerjee (1992)). However, the improper priors could lead to improper posterior distributions and nonexistence of the first two moments of the posterior distributions of the parameters of interest. To circumvent the problem, we assume proper uniform priors on finite subsets of the real lines or real spaces. Such choices of prior distributions ensure that all the posterior distributions as well as their moments exist. In our real data analysis, the model has been found to be non sensitive within the class of uniform prior distributions we have considered. We propose the Gibbs sampler (see Geman and Geman (1984), Gelfand and Smith (1990), Gelman and Rubin (1992)) to carry out the Bayesian computations. The full conditional distributions required to implement the Gibbs sampler are provided in this section. In our case, one of the full conditional distributions is known only up to a multiplicative constant. However, since this full conditional distribution turns out to be log-concave, we have been able to use the adaptive rejection sampling algorithm of Gilks and Wild (1992).

In Section 4, we apply the hierarchical Bayes method described in Section 3 to estimate the average weekly consumer expenditure of an item for 43 publication areas (small areas) throughout the U.S. The U.S. Bureau of Labor Statistics (BLS) needs these estimates to compute the Consumer Price Index (CPI) numbers which are published every month. The CPI is published for various items, goods and services, consumer units and geographical areas. The primary data is collected through the U.S. Consumer Expenditure Survey. In order to compare the hierarchical Bayes estimator with the EBLUP and direct survey estimator of the average weekly consumer expenditure of an item (for example, fresh whole milk) we follow a robust (i.e., model independent) method. Specifically, we view the original sample for the year 1989 as a pseudo-population and compute the direct survey, EBLUP and the hierarchical Bayes estimates based on several samples appropriately constructed from this pseudo-population. The estimates are then compared with the direct survey estimates of the original samples, i.e., the corresponding pseudo-population means. The proposed hierarchical Bayes estimator outperforms the other rival estimators. We also note that the posterior variances, the measures of accuracy of the hierarchical Bayes estimates, are always smaller than the corresponding variances of the direct survey estimates. The EBLUP does not seem to have a natural measure of accuracy in our situation. Here we reiterate that the methods proposed by Prasad and Rao (1990) and Kleffe and Rao (1992) cannot be extended to produce measures of accuracy of the EBLUP'S since the number of samples available from a small area is not small in comparison to the number of small areas, an assumption needed in their paper.

2. The BLUP vs. the Bayes Approach

Let Z_{il} be the value of the characteristic of interest for the lth unit of the ith small area $(i = 1, \ldots, m; l = 1, \ldots, n_i)$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)'$ and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_m^2)'$. We compare the Bayes approach with the BLUP approach to estimate the true small area means θ_i $(i = 1, \ldots, m)$ using the following hierarchical model.

Model 1. (i) $Z_{il} \mid \theta_i, \boldsymbol{\sigma}_i^2 \overset{ind}{\sim} N(\theta_i, \sigma_i^2)$, $i = 1, \ldots, m$; $l = 1, \ldots, n_i$; (ii) $\theta_i \overset{ind}{\sim} N(x_i'\beta, \tau^2)$, $i = 1, \ldots, m$; (iii) $\sigma_i^2 \overset{ind}{\sim} p(\sigma_i^2)$, $i = 1, \ldots, m$.

In the above model, β and τ^2 are known constants and $p(\sigma_i^2)$ represents an arbitrary density function of σ_i^2. Kleffe and Rao (1992) considered a special case of the above model when $x_i'\beta = \mu$ and $n_i = n$ $(i = 1, \ldots, m)$. They argued that it is more appropriate to assume different but random small area variances than a constant variance across the small areas.

If β is known, the BLUP of θ_i is given by $\tilde{\theta}_i^{BLUP} = x_i'\beta + w_i^{BLUP}(Z_{i.} - x_i'\beta)$, where $w_i^{BLUP} = n_i\tau^2/(n_i\tau^2 + \xi)$, $\xi = E(\sigma_i^2)$ and $Z_{i.} = n_i^{-1}\sum_{l=1}^{n_i} Z_{il}$.

Let $Z = (Z_1, \ldots, Z_m)$, where $Z_i = (Z_{i1}, \ldots, Z_{in_i})'$. The Bayes estimator of θ_i, under squared error loss, is given by $\tilde{\theta}_i^B = x_i'\beta + w_i^B(Z_{i.} - x_i'\beta)$, where $w_i^B = E\left[n_i\tau^2(n_i\tau^2 + \sigma_i^2)^{-1} \mid Z\right]$, E being the expectation over the posterior density of σ_i^2, given by,

$$f_i(\sigma_i^2|Z_i) \propto \sigma_i^{-(n_i-1)}(\sigma_i^2 + n_i\tau^2)^{-\frac{1}{2}} \exp\Big[-\frac{1}{2}\Big\{\sigma_i^{-2}S_i^2 + n_i(n_i\tau^2 + \sigma_i^2)^{-1} (Z_{i.} - x_i'\beta)^2\Big\}\Big]p(\sigma_i^2),$$

$\sigma_i^2 > 0$, where $S_i^2 = \sum_{l=1}^{n_i}(Z_{il} - Z_{i.})^2$.

Note that for the Kleffe-Rao model the Bayes estimator of θ_i, unlike the BLUP, assigns different weights to different small area sample means $Z_{i.}$'s.

The integrated Bayes risk of an estimator $\hat{\theta}_i$ of θ_i is defined as $r(\hat{\theta}_i) = E(\hat{\theta}_i - \theta_i)^2$, where E is with respect to Model 1. Note that $r(\hat{\theta}_i)$ is also the MSE of $\hat{\theta}_i$ as defined in Kleffe and Rao (1992). The following theorem demonstrates the superiority of $\tilde{\theta}_i^B$ over $\tilde{\theta}_i^{BLUP}$ in terms of the MSE.

Theorem 1. *Under the Model* 1 *and the condition that the density* $p(\cdot)$ of σ_i^2 $(i = 1, \ldots, m)$ *is non degenerate,*

$$(a)\ E(\tilde{\theta}_i^B) = E(\tilde{\theta}_i^{BLUP}) = E(\theta_i),$$
$$(b)\ r(\tilde{\theta}_i^B) < r(\tilde{\theta}_i^{BLUP}),$$

where E is with respect to Model 1.

Proof. (a) For $i = 1, \ldots, m$ we observe that $E(\tilde{\theta}_i^B) = EE(\theta_i \mid Z) = E(\theta_i)$,

$$E(\tilde{\theta}_i^{BLUP}) = x_i'\beta + w_i^{BLUP} E(Z_{i.} - x_i'\beta) = x_i'\beta = E(\theta_i).$$

(b) Note that

$$\begin{aligned} r(\tilde{\theta}_i^{BLUP}) - r(\tilde{\theta}_i^B) &= E(\tilde{\theta}_i^{BLUP} - \tilde{\theta}_i^B)^2 \\ &= E\{(w_i^{BLUP} - w_i^B)^2 (Z_{i.} - x_i'\beta)^2\} \geq 0. \end{aligned} \tag{1}$$

If possible, assume that the equality sign holds in (1), which in turn will imply $(w_i^{BLUP} - w_i^B)^2 (Z_{i.} - x_i'\beta)^2 = 0$ a.e.
$\Longleftrightarrow$ $(w_i^{BLUP} - w_i^B)^2 = 0$ a.e., (since $Z_{i.} - x_i'\beta$ is continuous)
$\Longleftrightarrow$ $w_i^{BLUP} = w_i^B$ a.e. $\Longleftrightarrow$ $\frac{n_i\tau^2}{n_i\tau^2+\xi} = E(\frac{n_i\tau^2}{n_i\tau^2+\sigma_i^2} \mid Z_i)$ a.e.
$\Rightarrow$ $\frac{n_i\tau^2}{n_i\tau^2+\xi} = E(\frac{n_i\tau^2}{n_i\tau^2+\sigma_i^2})$, (taking another expectation), which is not true, since $p(\cdot)$ is non degenerate and Jensen's inequality gives

$$E(\frac{n_i\tau^2}{n_i\tau^2 + \sigma_i^2}) > \frac{n_i\tau^2}{n_i\tau^2 + E(\sigma_i^2)} = \frac{n_i\tau^2}{n_i\tau^2 + \xi}.$$

3. Hierarchical Bayesian Model and Gibbs Sampler

To estimate per capita income of small places (population less than 1000), Fay and Herriot (1979) considered the empirical Bayes (EB) method to combine information from various administrative records in conjunction with the sample survey data available from the U.S. Current Population Survey. According to the Fay-Herriot model $Y_i \mid \theta_i \overset{ind}{\sim} N(\theta_i, D_i), i = 1, \ldots, m$, where the Y_i's are the survey estimates of the true small area means θ_i, and the sampling variances D_i are assumed to be known. *A priori* $\theta_i \overset{ind}{\sim} N(x_i'b, A)$, where $x_i' = (x_{i1}, \ldots, x_{ip})$ is a vector of known benchmark available for the small areas. Many applications of the Fay-Herriot model, its specific cases or its generalizations can be found in the small area literature. Carter and Rolph (1974) used a special case when $x_i'b = \mu$ to estimate fire alarm probabilities. Also, see Morris (1983), Cressie (1992), Datta, Fay and Ghosh (1991), Ghosh, Nangia and Kim (1996), Prasad and Rao (1990), among others.

In a typical survey situation, the direct survey estimates Y_i are of the form $\sum_{l=1}^{n_i} W_{il}\, Z_{il} / \sum_{l=1}^{n_i} W_{il}$, where n_i is the number of respondents for the ith area and Z_{il} (W_{il}) is the value of the characteristic of interest (sampling weight) for the lth unit of the ith area ($i = 1, \ldots, m; l = 1, \ldots, n_i$). The sampling weights W_{il} are usually determined by the reciprocal of the inclusion probabilities and are adjusted for factors such as nonresponse, post stratification, etc. The sampling weight attached to a respondent represents a certain number of population units.

In practice, the D_i's are unknown and are estimated using a design-based method (e.g., jackknife, balanced repeated replication, etc.) Thus, any procedure which uses these estimates as the true sampling variances D_i will not take into account the variability in estimating D_i. In order to incorporate this additional variability, we consider an alternate modeling. The model can be viewed as a Bayesian extension of the small area model due to Kleffe and Rao (1992). Unlike theirs, this model can handle an unbalanced situation but requires a specific form of density $p(\cdot)$. It also incorporates information on sampling weights and relevant covariates. Define $Y_i = \sum_{l=1}^{n_i} W_{il} Z_{il} / \sum_{l=1}^{n_i} W_{il}$, $S_i^2 = \sum_{l=1}^{n_i} (Z_{il} - Z_{i.})^2$, $K_i = \sum_{l=1}^{n_i} W_{il}^2 / (\sum_{l=1}^{n_i} W_{il})^2$, $(i = 1, \ldots, m)$ and $r = (r_1, \ldots, r_m)'$. Note that Y_i is the direct survey estimator of θ_i. We shall use the following hierarchical model.

Model 2. I Conditional on θ_i and r_i, Y_i and S_i^2 $(i = 1, \ldots, m)$ are independent with $Y_i \mid \theta_i, r_i \overset{ind}{\sim} N(\theta_i, {r_i}^{-1} K_i)$, and ${S_i}^2 \mid r_i, \theta_i \overset{ind}{\sim} {r_i}^{-1} \chi^2_{n_i-1}$, $i = 1, \ldots, m$;

II Conditional on b and v, $\theta_i \overset{ind}{\sim} N(x_i' b, v^{-1})$, $i = 1, \ldots, m$, where x_i is a $p \times 1$ column vector of known constants;

III Conditional on α and $\beta, r_i \overset{ind}{\sim} Gamma(\alpha, \beta)$; i.e. $f(r_i) \propto e^{-\alpha r_i} r_i^{\beta-1}$, $i = 1, \ldots, m$;

IV Marginally, $\alpha \sim U_1^+$, $\beta \sim U_1^+$, $v \sim U_1^+$, and $b \sim U_p$,

where ${U_1}^+$ denotes a uniform distribution over a subset of R^+ with large but finite length and U_p denotes a uniform distribution over a p-dimensional rectangle Q_p whose sides are large but of finite length.

It is important to note that in Step IV of Model 2, we have put proper but vague priors on the hyper parameters α, β, v and b. We have observed that Model 2 is not sensitive towards the choice of length of the uniform proper distributions. With the choice of proper priors on all the hyper parameters, all the posterior distributions are proper. Hence we do not face any problem of some posteriors being improper.

Our objective is to obtain the posterior distributions of the θ_i's, $i = 1, \ldots, m$. Due to the high dimensionality of the problem we recommend Gibbs sampling (see Geman and Geman (1984), and Gelfand and Smith (1990)). We choose the method given in Gelman and Rubin (1992) since it provides a measure, known as potential scale reduction factor, to check the convergence of the Gibbs sampler. Thus, we generate $t = 2d$ sets of random variables in each of l paths. The first d iterations from each path are deleted. We then use the S-program developed by Gelman and Rubin (1992) to obtain the potential scale reduction factors and various posterior densities.

From Model 2 we get the following full conditional distributions for Gibbs sampling, TN_s and TG representing an s-variate truncated normal distribution (truncated outside the s-dimensional rectangle Q_s) and a truncated Gamma distribution:

(i) For $i = 1, \ldots, m$:
$[\theta_i \mid Y, r, S, b, \alpha, \beta, v] \overset{ind}{\sim} N((r_i K_i^{-1} + v)^{-1}\{r_i K_i^{-1} Y_i + v\ x_i' b\}, (r_i K_i^{-1} + v)^{-1})$

(ii) For $i = 1, \ldots, m$:
$[r_i \mid Y, \theta, S, b, \alpha, \beta, v] \overset{ind}{\sim} Gamma(\frac{1}{2}\{K_i^{-1}(Y_i - \theta_i)^2 + {S_i}^2\} + \alpha,\ \ \frac{n_i}{2} + \beta)$

(iii) For $i = 1, \ldots, m$:
$[b \mid Y, \theta, r, S, \alpha, \beta, v] \sim TN_p([\sum_{i=1}^m x_i x_i']^{-1} \sum_{i=1}^m x_i \theta_i, v^{-1}[\sum_{i=1}^m x_i x_i']^{-1})$

(iv) For $n_T = \sum_{i=1}^m n_i$:
$[\alpha \mid Y, \theta, r, S, b, \beta, v] \sim TG(\sum_{i=1}^m r_i,\ \ m\beta + 1)$

(v) $[v \mid Y, \theta, S, b, \alpha, \beta] \sim TG(\frac{1}{2}\sum_{i=1}^m (\theta_i - {x_i}'b)^2,\ \frac{1}{2}n_T + 1)$

(vi) $[\beta \mid Y, \theta, S, b, \alpha, v] \propto \{\Gamma(\beta)\}^{-m} \alpha^{m\beta} \prod_{i=1}^m {r_i}^{\beta}$.

Using Gibbs sampling, the joint posterior pdf of $\theta = (\theta_1, \ldots, \theta_m)'$ is approximated by

$$E[\theta \mid Y, r, S, b, \alpha, \beta, v] \approx (l\ d)^{-1} \sum_{s=1}^{l} \sum_{j=d+1}^{2d} E[\theta \mid Y, r_{(js)}, S, b_{(js)}, \alpha_{(js)}, \beta_{(js)}, v_{(js)}].$$

To estimate the posterior mean and variance, we use Rao-Blackwellized estimates as in Gelfand and Smith (1990).

For implementing the Gibbs sampler, we need to draw samples from the full conditional densities (i) - (vi). Simulations from the full conditional densities (i) - (v) can be done by using standard methods. However, the full conditional density of $[\beta \mid Y, \theta, S, b, \alpha, v]$ is known only up to a multiplicative constant. In order to draw samples from this density a general approach is to use the Metropolis-Hastings accept-reject algorithm. Fortunately, the task becomes simpler since log $[\beta \mid Y, \theta, S, b, \alpha, v]$ is a concave function of β (see Arora (1994)).

4. An Example

The U.S. Bureau of Labor Statistics needs estimates of the true average weekly consumer expenditures of various items, goods and services, for $m = 43$ publication areas (small areas) throughout the U.S. We concentrate on estimating the true average expenditure of the item fresh whole milk for the year 1989 for the ith publication area (i.e., $\theta_i,\ i = 1, \ldots, 43$) and use data from the Diary Survey component of the Consumer Expenditure Survey conducted by the U.S. Bureau of the Census for the BLS. Samples are drawn independently for each quarter. Each respondent of the sample receives a sampling weight (i.e., W_{il})

which is determined by the reciprocal of the inclusion probability of the respondent and adjusted for various factors such as post stratification, nonresponse, etc. The sampling weight for a respondent represents a number of population units and the sum of the sampling weights for all the respondents in the sample is approximately equal to the total number of households in the U.S. Each respondent keeps a record of expenditures on various items for two consecutive weeks. Thus, average weekly expenditure on fresh whole milk (Z_{il}) is available for each respondent. Using W_{il} and Z_{il}, we produce the survey estimates Y_i and their variance $(n_i-1)^{-1}S_i^2K_i$ for the ith ($i=1,\ldots,43$) publication area. The sampling variance of Y_i is obtained under the assumption that Z_{il}'s are i.i.d. and can be improved if certain information regarding the sampling design is available.

In Model 2, we used $x_i'b=b_j$ if $i\in j$th major area, a collection of similar publication areas. There are eight major areas in the U.S. For the Gibbs sampler we used $2d=1000$ iterations and $l=8$ independent paths to draw samples from the full conditional densities (i)-(vi) given in Section 3. In each path, the first $d=500$ generated values were deleted. Starting with initial values of the parameters, we draw samples from the full conditional densities of r, b, α, β, v and $\boldsymbol{\theta}$. For drawing samples from the full conditional density of β, which is known only upto a multiplicative constant, we used the adaptive rejection sampling technique of Gilks and Wild (1992). We have observed that the model is not sensitive towards the initial values of the parameters. To study the convergence of the Gibbs sampler we used the S-program written by Gelmen and Rubin (1992). This program computes a *potential scale reduction factor* R which provides a way to quantitatively monitor the convergence of the Gibbs sampler. For all the situations, the measure R seems to converge to unity after the first 500 iterations.

Table 1 exhibits the direct survey estimates and the hierarchical Bayes estimates along with their measures of accuracy for all the 43 publication areas of the U.S. for the year 1989. We note that the the posterior variances are always smaller than the corresponding variances of the direct survey estimates. In order to derive the BLUP of θ_i, we use steps (I)-(III) of Model 2. The BLUP turns out to be $a_iY_i+(1-a_i)x_i'\tilde{b}$, where $a_i=v^{-1}/(v^{-1}+\xi K_i)$, $\xi=E(r_i^{-1})$ and $\tilde{b}=\left[\sum_{i=1}^m(v^{-1}+\xi K_i)^{-1}x_ix_i'\right]^{-1}\sum_{i=1}^m(v^{-1}+\xi K_i)^{-1}x_iY_i$. We then use an ANOVA method to estimate ξ and v^{-1} for obtaining the EBLUP from the BLUP. The EBLUP for all the 43 publication areas of the U.S. is also reported in Table 1. The EBLUP does not seem to have a natural measure of accuracy in this situation. Note that Kleffe-Rao technique cannot be extended to this case since the sample sizes available from the small areas are much larger than the number of small areas.

Table 1. Direct survey estimates, EBLUP and the proposed HB estimates for consumer expenditure on item *fresh whole milk* : Year 1989. The number in the paranthesis represents the corresponding standard error.

Pub. Area	n	Survey Est.	EBLUP	HB Est.	Pub. Area	n	Survey Est.	E BLUP	HB Est.
1	191	1.099(.163)	1.095	1.093(.118)	23	195	1.044(.140)	1.108	1.102(.117)
2	633	1.075(.080)	1.079	1.079(.073)	24	187	1.267(.171)	1.212	1.200(.131)
3	597	1.105(.083)	1.101	1.100(.075)	25	497	1.193(.106)	1.088	1.054(.095)
4	221	.628(.109)	.822	.763(.096)	26	230	.791(.121)	.822	.810(.095)
5	195	.753(.119)	.891	.844(.097)	27	186	.795(.121)	.828	.813(.096)
6	191	.981(.141)	1.000	.967(.104)	28	199	.759(.259)	.809	.816(.132)
7	183	1.257(.202)	1.127	1.055(.127)	29	238	.796(.106)	.825	.811(.088)
8	188	1.095(.127)	1.053	1.028(.100)	30	207	.565(.089)	.714	.657(.085)
9	204	1.405(.168)	1.204	1.136(.126)	31	165	.886(.225)	.869	.849(.126)
10	188	1.356(.178)	1.178	1.108(.125)	32	153	.952(.205)	.897	.872(.123)
11	149	.615(.100)	.860	.755(.097)	33	210	.807(.119)	.743	.749(.101)
12	290	1.460(.201)	1.365	1.280(.142)	34	383	.582(.067)	.617	.603(.065)
13	250	1.338(.148)	1.291	1.258(.120)	35	255	.684(.106)	.682	.681(.091)
14	194	.854(.142)	1.053	1.032(.121)	36	226	.787(.126)	.757	.756(.097)
15	184	1.176(.149)	1.203	1.193(.108)	37	224	.440(.092)	.579	.537(.086)
16	193	1.111(.145)	1.170	1.162(.106)	38	212	.759(.132)	.743	.741(.099)
17	218	1.257(.135)	1.242	1.232(.102)	39	211	.770(.100)	.748	.751(.085)
18	266	1.430(.172)	1.339	1.296(.118)	40	179	.800(.113)	.759	.765(.093)
19	214	1.278(.137)	1.252	1.242(.103)	41	312	.756(.083)	.744	.746(.074)
20	213	1.292(.163)	1.260	1.243(.112)	42	241	.865(.121)	.799	.800(.096)
21	196	1.002(.125)	1.118	1.098(.100)	43	205	.640(.129)	.685	.679(.098)
22	95	1.183(.247)	1.169	1.163(.151)					

Next we compare the HB (Hierarchal Bayes) estimates with the direct survey estimates and the EBLUP using a robust procedure. We view the data for the year 1989 as a pseudo-population and consider eight 12.5% samples from this pseudo-population. These subsamples are available in the data set and were originally constructed by the U.S. Census Bureau in order to provide variance estimates of the survey estimates at the national level. Thus, this evaluation criterion is quite objective and is not model dependent. In Table 2, we report averaged squared relative deviation ($ASRD$) defined as follows:

$$ASRD = \frac{1}{43}\sum_{i=1}^{43}\frac{(e_i - \theta_i)^2}{\theta_i^2},$$

where e_i is an estimator of the pseudo true small area mean θ_i (i.e., the direct survey estimate based on the entire 1989 data). The HB estimator is better than the direct survey estimator for all the eight samples considered (improvement ranges from 46% to 80%). The HB estimator is better than the EBLUP for six out of eight samples.

Table 2. Comparison of direct survey estimates, EBLUP and the proposed HB for consumer expenditure on item *fresh whole milk*: 1989. The number in the parenthesis represents percent improvement over the survey estimator.

Sample	Average Squared Relative Deviation		
	Survey Est.	EBLUP	HB
1	.3416	.1920(44%)	.1579(54%)
2	.2902	.1474(49%)	.0963(67%)
3	.2515	.0641(75%)	.0719(71%)
4	.1591	.0815(49%)	.0798(50%)
5	.3012	.1333(56%)	.1174(61%)
6	.3144	.1602(49%)	.1188(62%)
7	.2473	.0518(79%)	.1333(46%)
8	.1683	.0617(63%)	.0339(80%)

Acknowledgement

The second author's research was supported in part by National Science Foundation Grant SES-9206326 and SES-9511202 and a research contract from the U.S. Bureau of Labor Statistics. The authors are grateful to the referees and an Associate Editor for many helpful suggestions and comments.

References

Arora, V. (1994). Empirical and Hierarchical Bayes estimation of small area characteristic. Ph.D. dissertation. Univ. of Nebraska-Lincoln.

Carter, G. M. and Rolph, J. E. (1974). Empirical Bayes methods applied to estimating fire alarm probabilities. *J. Amer. Statist. Assoc.* **69**, 880-885.

Cressie, N. (1992). REML estimation in empirical Bayes smoothening of census undercount. *Survey Methodology.* **18**, 75-94.

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

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

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

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

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

Ghosh, J. K. and Mukerjee, R. (1992). Non-informative priors (with discussion). In *Bayesian Statistics* **4** (Edited by J. M. Bernado et al.), 195-210. Oxford University Press.

Ghosh, M., Nangia, N. and Kim, D. (1996). Estimation of median income of four-preson families: a Bayesian time series approach. *J. Amer. Statist. Assoc.* **91**, 1423-1431.

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

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Appl. Statist.* **41**, 337-348.

Kleffe, J. and Rao, J. N. K. (1992). Estimation of mean square error of empirical best linear unbiased predictors under a random error variance linear models. *J. Multivariate Anal.* **43**, 1-15.

Morris, C. N. (1983). Parametric empirical Bayes inference: theory and applications. *J. Amer. Statist. Assoc.* **78**, 47-65.

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.

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

Wolter, K. M. (1985). *Introduction to Variance Estimation.* Springer-Verlag. New York.

3400 Richmond Pkwy # 3214, Richmond CA 94806, U.S.A.

E-mail: varora@unlinfo.unl.edu

Division of Statistics, Department of Mathematics and Statistics, 810 oldfather Hall, University of Nebraska, Lincoln, NE 68588-0323, U.S.A.

E-mail: plahiri@unlinfo.unl.edu

(Received November 1995; accepted February 1997)