RESEARCH REPORT SERIES
(Statistics #2009-01)

Bayesian Benchmarking with Applications to Small Area Estimation

G.S. Datta[1]
M. Ghosh[2]
R. Steorts[2]
J. Maples

[1] University of Georgia
[2] University of Florida

Statistical Research Division
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: January 29, 2009

Bayesian Benchmarking with Applications to Small Area Estimation

G.S. Datta, M. Ghosh, R. Steorts and J. Maples
University of Georgia, University of Florida and US Bureau of the Census

Abstract

It is well-known that small area estimation needs explicit, or at least implicit use of models. These model-based estimates can differ widely from the direct estimates, especially for areas with very low sample sizes. While model-based small area estimates are very useful, one potential difficulty with such estimates is that when aggregated, the overall estimate for a larger geographical area may be quite different from the corresponding direct estimate, the latter being usually believed to be quite reliable. This is because the original survey was designed to achieve specified inferential accuracy at this higher level of aggregation. The problem can be more severe in the event of model failure as often there is no real check for validity of the assumed model. Moreover, an overall agreement with the direct estimates at an aggregate level may sometimes be necessary for policy reasons to convince the legislators of the utility of small area estimates.

One way to avoid this problem is the so-called "benchmarking approach" which amounts to modifying these model-based estimates so that one gets the same aggregate estimate for the larger geographical area. Currently, the most popular approach is the so-called "raking" or ratio adjustment method which involves multiplying all the small area estimates by a constant data-dependent factor so that the weighted total agrees with the direct estimate. There are alternate proposals, mostly from frequentist considerations, which meet also the aforementioned benchmarking criterion.

We propose in this paper a general class of constrained Bayes estimators which achieve as well the necessary benchmarking. Interestingly enough, many of the frequentist estimators, including the raked estimators, follow as special cases of our general result. In the process, some deficiencies of the raked estimators will be pointed out. Explicit Bayes estimators are derived that benchmark the weighted mean or both the weighted mean and variability. We illustrate our methodology by developing poverty rates in school-aged children at the state level, and then benchmarking these estimates to match at the national level.

Keywords: Area-level, penalty parameter, two-stage, weighted mean, weighted variability.

1 Introduction

Empirical Bayesian (EB) and hierarchical Bayesian (HB) methods are now widely used for simultaneous inference. The biggest advantage of these methods is their ability to enhance the precision of individual estimators by "borrowing strength" from other similar estimators. These methods have been used very successfully in a wide array of disciplines including sociology, epidemiology, business, economics, political science and insurance.

One of the biggest applications of EB and HB methods is in small area estimation. The topic has become a prime area of research globally in many government statistical agencies. As an example, in the United States Bureau of the Census, the Small Area Income and Poverty Estimates (SAIPE) and Small Area Health Insurance Estimates (SAHIE) research groups are actively engaged in various small area projects. A typical small area estimation problem involves simultaneous estimation of quantities of interest for several small geographical areas (for example, counties) or several small domains cross-classified by age, sex, race and other demographic/geographic characteristics. The need for borrowing strength arises in these problems because the original survey was designed to achieve a specific accuracy at a higher level of aggregation than that of small areas or domains. Due to limited resources, the same data needs to be used at lower levels of geography, but individual direct estimates are usually accompanied by large standard errors and coefficients of variation.

It is well-known that small area estimation needs explicit, or at least implicit, use of models. These model-based estimates can differ widely from the direct estimates, especially for areas with very low sample sizes. One potential difficulty with model-based estimates is that when aggregated, the overall estimate for a larger geographical area may be quite different from the corresponding direct estimate, the latter being often believed to be quite reliable. This is because the original survey was designed to achieve specified inferential accuracy at this higher level of aggregation. As an example, the SAIPE county estimates of the United States Bureau of the Census, based on the American Community Survey (ACS) data, are controlled so that the overall weighted estimates agree with the corresponding state estimates which though model-based, are quite close to the direct estimates. The problem can be more severe in the event of model failure as often there is no real check for validity of the assumed model. Pfeffermann and Tiller (2006), in the context of time series models for small area estimation, noted that benchmarked estimates reflect a sudden change in the direct estimates due to some external shock not accounted for in the model much faster than the model-based estimates. Moreover, an overall agreement with the direct estimates at some higher level may sometimes be necessary for policy reasons to convince the legislators of the utility of small area estimates (Fay and Herriot, 1979).

One way to avoid this problem is the so-called "benchmarking approach" which amounts to modifying these model-based estimates so that one gets the same aggregate estimate for the larger geographical area. A simple illustration is to modify the model-based state level estimates so that one matches the national estimates. Currently the most popular approach is the so-called "raking" or ratio adjustment method which involves multiplying all the small area estimates by a constant factor so that the weighted total agrees with the direct estimate. The raking approach is ad-hoc, although, later in this paper, we have given it a constrained Bayes interpretation.

The objective of this paper is to develop a general class of Bayes estimators which achieves the necessary benchmarking. For definiteness, we will concentrate only on area-level models. As we will see later, many of the currently proposed benchmarked estimators including the raked ones belong to the proposed class of Bayes estimators. In particular, some of the estimators proposed in Pfeffermann and Barnard (1991), Isaki, Tsay and Fuller (2000), Wang and Fuller (2002) and You, Rao and Dick (2004) are members of this class.

The proposed Bayesian approach has been motivated from a decision-theoretic framework, and is similar in spirit to one in Louis (1984) and Ghosh (1992) who considered constrained Bayes and empirical Bayes estimators with a slightly different objective. It was pointed out in these papers that the empirical histogram of the posterior means is underdispersed compared to the posterior histogram of the corresponding population parameters. Thus, adjustment of Bayes estimators is needed in order to meet the twin objectives of accuracy and closeness of the histogram of the estimates with the posterior estimate of the parameter histogram. In contrast, the present method achieves matching with some aggregate measure such as a national total. In addition, if necessary, we can also match the empirical variability of the estimates with the posterior variability of the parameters of interest or even some preassigned number.

The organization of the remaining sections is as follows. In Section 2, we develop the constrained Bayes estimators requiring only the matching of a weighted average of small area means with some prespecified estimators. These prespecified estimators can be a weighted average of the direct small area estimators, a situation which will be referred to as *internal benchmarking*. On the other hand, if the prespecified estimator is obtained from some other source, for example, a different survey, census or other administrative records, then it becomes an instance of *external benchmarking*. We will also point out how the proposed benchmarked estimators arise as the limit of a general class of Bayes estimators where one needs only partial benchmarking. The general result is illustrated with several constrained benchmarked estimators of area-level means based on the usual random effects or the Fay-Herriot (1979) (also Pfeffermann and Nathan, 1981) model.

In Section 3, we develop benchmarked Bayes estimators which meet the dual objective of overall matching with the prespecified benchmarks as well as the variability agreement as mentioned in the previous paragraph. Multiparameter extensions of these results will also be given in this section. Section 4 contains an application of the proposed method in a real small area problem. Section 5 contains a summary of the results developed in this paper along with a few suggestions for future research.

Before concluding this section, we discuss some of the existing benchmarking literature (mostly frequentist) for small area estimation. We begin with a simple stratified sampling model with m strata having population sizes $N_1, \cdots, N_m$. You and Rao (2002) required estimates $\hat{\theta}_i$ of the stratum means θ_i such that $\sum_{i=1}^m N_i\hat{\theta}_i$ equals the direct survey regression estimator of the overall total. You and Rao (2002, 2003) considered unit-level small area models with known survey weights attached to the different units. More recently, Fuller (2007) has considered a procedure which allows sampling of small areas from a larger pool of small areas and requires a weighted sum of small area predictors equal to a design consistent estimator of the population total.

Pfeffermann and Tiller (2006) considered benchmarking in small area estimation based on time series data. They took a frequentist approach and used the Kalman filter for time series to obtain first the model-based estimators of the small area means. Then they obtained the benchmarked estimators satisfying certain agreement of these estimators with some direct estimators at a higher level of aggregation. They did not use any cross-sectional model to borrow strength from other areas. This is an example of internal benchmarking. Earlier examples of internal benchmarking include those of Pfeffermann and Barnard (1991), Isaki, Tsay and Fuller (2000) and Wang and Fuller (2002).

You *et al.* (2004) considered benchmarking of HB estimates through ratio adjustment for area-level models. Nandram et al. (2007) suggested a different benchmarked HB estimation of small area means based on unit level models. In this exact benchmarking, they proceeded with the conditional distribution of the unobserved units within a small area given the benchmark constraint on the total of all the units in that area. A disadvantage to such an approach is that results can differ depending on which unit is dropped.

2 Benchmarked Bayes Estimators

Let $\hat{\theta}_1, \cdots, \hat{\theta}_m$ denote the direct estimators of the m small area means $\theta_1, \ldots, \theta_m$. We write $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \cdots, \hat{\theta}_m)^T$ and $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_m)^T$. Initially, we seek the benchmarked Bayes estimator $\hat{\boldsymbol{\theta}}^{BM1} = (\hat{\theta}_1^{BM1}, \cdots, \hat{\theta}_m^{BM1})^T$ of $\boldsymbol{\theta}$ such that $\sum_{i=1}^m w_i\hat{\theta}_i^{BM1} = t$, where either t is prespecified from some other source or $t = \sum_{i=1}^m w_i\hat{\theta}_i$. The w_i are given weights attached to the direct estimators $\hat{\theta}_i$'s, and without any loss of generality, $\sum_{i=1}^m w_i = 1$. These weights may depend on $\hat{\boldsymbol{\theta}}$ (which is most often not the case), but do not depend on $\boldsymbol{\theta}$. For example, one may take $w_i = N_i / \sum_{j=1}^m N_j$, where the N_i are the population sizes for the m small areas.

A Bayesian approach to this end is to minimize the posterior expectation of the weighted squared error loss $\sum_{i=1}^m \phi_i E[(\theta_i - e_i)^2|\hat{\boldsymbol{\theta}}]$ with respect to the e_i's satisfying $\bar{e}_w = \sum_{i=1}^m w_i e_i = t$. These ϕ_i may be the same as the w_i, but that need not always be the case. Also, like w_i, ϕ_i may depend on $\hat{\boldsymbol{\theta}}$, but not on $\boldsymbol{\theta}$. Wang and Fuller (2002) considered the same loss, but minimized instead the MSE (which amounts to conditioning on θ), and came up with a solution different from ours. Moreover, they restricted themselves to linear estimators. One of the advantages of the proposed Bayesian approach is that adjustment is possible for any general Bayes estimator, linear or non-linear.

The ϕ_i can be regarded as weights for a multiple-objective decision process. That is, each specific weight is relevant only to the decision-maker for the corresponding small area, who may not be concerned with the weights related to decision-makers in other small areas. Combining losses in such situations in a linear fashion is discussed for example in Berger (1985, p. 279).

We now prove a theorem which provides a solution to our problem. A few notations are needed before stating the theorem. Let $\hat{\theta}_i^B$ denote the posterior mean of θ_i, $i = 1, \cdots, m$ under a certain prior. The vector of posterior means is denoted by $\hat{\boldsymbol{\theta}}^B = (\hat{\theta}_1^B, \cdots, \hat{\theta}_m^B)^T$ and the weighted average $\bar{\hat{\theta}}_w^B = \sum_{i=1}^m w_i \hat{\theta}_i^B$. Also, let $\boldsymbol{r} = (r_1, \cdots, r_m)^T$, where $r_i = w_i/\phi_i$, $i = 1, \cdots, m$, and $s = \sum_{i=1}^m w_i^2/\phi_i$. Then we have the following theorem.

Theorem 1. The minimizer $\hat{\boldsymbol{\theta}}^{BM1}$ of $\sum_{i=1}^m \phi_i E[(e_i - \theta_i)^2|\hat{\boldsymbol{\theta}}]$ subject to $\bar{e}_w = t$ is given by

$$\hat{\boldsymbol{\theta}}^{BM1} = \hat{\boldsymbol{\theta}}^B + s^{-1}(t - \bar{\hat{\theta}}_w^B)\boldsymbol{r}. \tag{1}$$

Proof. First rewrite $\sum_{i=1}^m \phi_i E[(e_i - \theta_i)^2|\hat{\boldsymbol{\theta}}] = \sum_{i=1}^m \phi_i V(\theta_i|\hat{\boldsymbol{\theta}}) + \sum_{i=1}^m \phi_i (e_i - \hat{\theta}_i^B)^2$. Now the problem reduces to minimization of $\sum_{i=1}^m \phi_i (e_i - \hat{\theta}_i^B)^2$ subject to $\bar{e}_w = t$. A Lagrangian multiplier approach provides the solution. But then one needs to show in addition that the solution provides a minimizer and not a maximizer. Alternately, we can use the identity

$$\sum_{i=1}^m \phi_i (e_i - \hat{\theta}_i^B)^2 = \sum_{i=1}^m \phi_i \{e_i - \hat{\theta}_i^B - s^{-1}(t - \bar{\hat{\theta}}_w^B) r_i\}^2 + s^{-1}(t - \bar{\hat{\theta}}_w^B)^2. \tag{2}$$

The solution is now immediate from (2).

Remark 1. The above constrained Bayes benchmarked estimators can be viewed also as limiting Bayes estimators under the loss

$$L(\boldsymbol{\theta}, \boldsymbol{e}) = \sum_{i=1}^m \phi_i (\theta_i - e_i)^2 + \lambda (t - \bar{e}_w)^2,$$

where $\boldsymbol{e} = (e_1, \cdots, e_m)^T$, and $\lambda(> 0)$ is the penalty parameter. Like the ϕ_i, the penalty parameter λ can differ for different policy makers. The Bayes estimator of $\boldsymbol{\theta}$ under the above loss (after some algebra) is given by

$$\hat{\boldsymbol{\theta}}_\lambda^B = \hat{\boldsymbol{\theta}}^B + (s + \lambda^{-1})^{-1}(t - \bar{\hat{\theta}}_w^B)\boldsymbol{r}. \quad (3)$$

Clearly, when $\lambda \to \infty$, i.e., when one invokes the extreme penalty for not having the exact equality $\bar{e}_w = t$, one gets the estimator given in (1). Otherwise, λ serves as a trade-off between t and $\bar{\hat{\theta}}_w^B$ since

$$\boldsymbol{w}^T \hat{\boldsymbol{\theta}}_\lambda^B = \frac{s\lambda}{s\lambda + 1} t + \frac{1}{s\lambda + 1} \bar{\hat{\theta}}_w^B.$$

Remark 2. The balanced loss of Zellner (1986, 1988, 1994) is not quite the same as the one in Remark 1, and is given by

$$L(\boldsymbol{\theta}, \boldsymbol{e}) = \sum_{i=1}^m \phi_i (\theta_i - e_i)^2 + \lambda \sum_{i=1}^m (\hat{\theta}_i - e_i)^2.$$

This leads to the Bayes estimator $\hat{\boldsymbol{\theta}}^B + \lambda(\lambda \boldsymbol{I} + \boldsymbol{\phi})^{-1}(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^B)$, where $\boldsymbol{I}$ is the identity matrix and $\boldsymbol{\phi} = \text{Diag}(\phi_1, \cdots, \phi_m)$ which is a compromise between the Bayes estimator $\hat{\boldsymbol{\theta}}^B$ and the direct estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$, and converges to the direct estimator as $\lambda \to \infty$ and to the Bayes estimator when $\lambda \to 0$.

Remark 3. It is easy to see why the raked Bayes estimators, considered for example in You and Rao (2004), belong to the general class of estimators proposed in Theorem 1. If one chooses (possibly quite artificially) $\phi_i = w_i / \hat{\theta}_i^B$, $i = 1, \cdots, m$, then $\boldsymbol{r} = \hat{\boldsymbol{\theta}}^B$ and $s = \bar{\hat{\theta}}_w^B$. Consequently, the constrained Bayes estimator proposed in Theorem 1 simplifies to $(t / \bar{\hat{\theta}}_w^B) \hat{\boldsymbol{\theta}}^B$, which is the raked Bayes estimator. In particular, one can take $t = \bar{\hat{\theta}}_w$. We may also note that this choice of the ϕ_i's is different from the one in Wang and Fuller (2002) who considered $\phi_i = w_i / \hat{\theta}_i$. Also, both choices point out the deficiency of the raked estimators. This is because the ϕ_i are usually supposed to be positive and either choice can lead to a negative ϕ_i when $\hat{\theta}_i^B < 0$ or $\hat{\theta}_i < 0$.

Remark 4. Other applications exist. Consider for example the usual random effects model as considered in Fay and Herriot (1979) or Pfeffermann and Nathan (1981). Under this model, $\hat{\theta}_i | \theta_i \overset{ind}{\sim} N(\theta_i, D_i)$ and $\theta_i \overset{ind}{\sim} N(\boldsymbol{x}_i^T \boldsymbol{\beta}, \sigma_u^2)$, the $D_i (> 0)$ being known. For the HB approach, one then uses the the prior $\pi(\boldsymbol{\beta}, \sigma_u^2) = 1$ although other priors are also possible as long as the posteriors are proper. The HB estimators $E(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}})$ cannot be obtained analytically, but it is possible to find them numerically either through Markov chain Monte Carlo (MCMC) or through numerical integration. Denoting the HB estimators by $\hat{\theta}_i^B$, one can obtain the benchmarked Bayes estimators $\hat{\theta}_i^{BM1} (i = 1, \cdots, m)$ by applying Theorem 1.

Remark 5. Wang and Fuller (2002) considered a slightly varied form of the above random effects model, the only change, in our notations, being that the marginal variance of the θ_i are now $z_i^2\sigma_u^2$, where the z_i are known. They did not assume normality, but restricted their attention to the class of linear estimators of $\boldsymbol{\theta}$, and benchmarked the best linear unbiased predictor (BLUP) of $\boldsymbol{\theta}$ when σ_u^2 is known. For this example, the benchmarked estimators given in (6) of Wang and Fuller are also derivable from Theorem 1. First for known σ_u^2, consider the uniform prior for $\boldsymbol{\beta}$. Write $B_i = D_i/(D_i + z_i^2\sigma_u^2)$, $\boldsymbol{B} = \mathrm{Diag}(B_1, \cdots, B_m)$, $\boldsymbol{\Sigma} = \mathrm{Diag}(D_1 + z_1^2\sigma_u^2, \cdots, D_m + z_m^2\sigma_u^2)$, $\boldsymbol{X}^T = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m)$, and $\tilde{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\theta}}$, assuming $\boldsymbol{X}$ to be a full column rank matrix. Then the Bayes estimator of $\boldsymbol{\theta}$ is

$$\tilde{\boldsymbol{\theta}}^B = (\boldsymbol{I}_m - \boldsymbol{B})\hat{\boldsymbol{\theta}} + \boldsymbol{B}\boldsymbol{X}\tilde{\boldsymbol{\beta}}, \tag{4}$$

which is the same as the BLUP of $\boldsymbol{\theta}$ as well. Now identify the r_i/s in this paper with the a_i of Wang and Fuller (2002) to get (26) in their paper.

Remark 6. As an example, the Pfeffermann-Barnard (1991) estimator belongs to this general class of estimators where one chooses $\phi_i = w_i/\mathrm{Cov}(\hat{\theta}_i^B, \bar{\hat{\theta}}^B)$, where the covariance is calculated over the joint distribution of $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}$ treating $\boldsymbol{\beta}$ as an unknown but fixed parameter. Then $\boldsymbol{r}$ contains the elements of $\mathrm{Cov}(\hat{\theta}_i^B, \bar{\hat{\theta}}^B)$ as its components, while $s = V(\bar{\hat{\theta}}_w^B)$. A similar result with different notations appears in Wang and Fuller (2002).

Instead of the constrained Bayes estimators as given in (1), it is possible to obtain constrained empirical Bayes (EB) estimators as well when one estimates the prior parameters from the marginal distribution of $\hat{\boldsymbol{\theta}}$ (after integrating out $\boldsymbol{\theta}$). The resulting EB estimators are given by

$$\hat{\boldsymbol{\theta}}^{EBM1} = \hat{\boldsymbol{\theta}}^{EB} + s^{-1}(t - \bar{\hat{\boldsymbol{\theta}}}_w^{EB})\boldsymbol{r}, \tag{5}$$

where $\hat{\boldsymbol{\theta}}^{EB} = (\hat{\theta}_1^{EB}, \cdots, \hat{\theta}_m^{EB})^T$ is an EB estimator of $\boldsymbol{\theta}$ and $\bar{\hat{\boldsymbol{\theta}}}_w^{EB} = \sum_{i=1}^m w_i\hat{\theta}_i^{EB}$.

Remark 7. Back to the random effects model as considered in Pfeffermann and Barnard (1991), Isaki, Tsay and Fuller (2000) and Wang and Fuller (2002), for unknown σ_u^2, one gets estimators of $\boldsymbol{\beta}$ and σ_u^2 simultaneously from the marginals $\hat{\theta}_i \overset{\text{ind}}{\sim} N(\boldsymbol{x}_i^T\boldsymbol{\beta}, D_i + z_i^2\sigma_u^2)$ (Fay and Herriot, 1979; Prasad and Rao, 1990; Datta and Lahiri, 2000; Datta, Rao and Smith, 2005). Denoting the estimator of σ_u^2 by $\hat{\sigma}_u^2$, one estimates $\boldsymbol{\Sigma}$ by $\hat{\boldsymbol{\Sigma}} = \mathrm{Diag}(D_1 + z_1^2\hat{\sigma}_u^2, \cdots, D_m + z_m^2\hat{\sigma}_u^2)$, $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\theta}}$ and $\boldsymbol{B}$ by $\hat{\boldsymbol{B}} = \boldsymbol{D}\hat{\boldsymbol{\Sigma}}^{-1}$, where $\boldsymbol{d} = \mathrm{Diag}(D_1, \cdots, D_m)$. Denoting the resulting EB estimator of $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}^{EB}$, one gets

$$\hat{\boldsymbol{\theta}}^{EB} = (\boldsymbol{I}_m - \hat{\boldsymbol{B}})\hat{\boldsymbol{\theta}} + \hat{\boldsymbol{B}}\boldsymbol{X}\hat{\boldsymbol{\beta}}. \tag{6}$$

The benchmarked EB estimator is now obtained from (5).

Remark 8. Similar to Wang and Fuller (2002), the benchmarked EB estimators include those proposed in Isaki, Tsay and Fuller (2000), where one takes ϕ_i as the reciprocal of the ith diagonal element of $\hat{\boldsymbol{\Sigma}}$ for all $i = 1, \cdots, m$. Another option is to take ϕ_i as the reciprocal of an estimator of $\mathrm{V}(\hat{\theta}_i^B)$, the variance being computed once again under the joint distribution of $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}$, treating $\boldsymbol{\beta}$ as an unknown but fixed parameter.

Remark 9. As mentioned earlier, the choice of ϕ_i is left to the particular decision-maker depending on the severity of misspecification of true parameters. For the special random effects model, Wang and Fuller have argued in favor of the reciprocal of the ith diagonal element of $\hat{\boldsymbol{\Sigma}}$ as the choice from an interesting frequentist consideration.

Before concluding this section, we prove a generalization of Theorem 1 where one considers multiple instead of one single constraint. As an example, for the SAIPE county level analysis, one may need to control the county estimates in each state so that their weighted total agrees with the corresponding state estimates. Also, one can consider a more general quadratic loss given by

$$L(\boldsymbol{\theta}, \boldsymbol{e}) = (\boldsymbol{e} - \boldsymbol{\theta})^T \boldsymbol{\Omega} (\boldsymbol{e} - \boldsymbol{\theta}),$$

where $\boldsymbol{\Omega}$ is a positive definite matrix. The following theorem provides a Bayesian solution for the minimization of $E[L(\boldsymbol{\theta}, \boldsymbol{e})|\hat{\boldsymbol{\theta}}]$ subject to the constraint $\boldsymbol{W}^T \boldsymbol{e} = \boldsymbol{t}$, where $\boldsymbol{t}$ is a q-component vector, and $\boldsymbol{W}$ is a $m \times q$ matrix of rank $q(< m)$.

Theorem 2. The constrained Bayesian solution in the above set-up is given by
$\hat{\boldsymbol{\theta}}^{MBM} = \hat{\boldsymbol{\theta}}^B + \boldsymbol{\Omega}^{-1} \boldsymbol{W} (\boldsymbol{W}^T \boldsymbol{\Omega}^{-1} \boldsymbol{W})^{-1} (\boldsymbol{t} - \bar{\hat{\boldsymbol{\theta}}}_w^B)$, where $\bar{\hat{\boldsymbol{\theta}}}_w^B = \boldsymbol{W}^T \hat{\boldsymbol{\theta}}^B$.
Proof. First write
$E[(\boldsymbol{e} - \boldsymbol{\theta})^T \boldsymbol{\Omega} (\boldsymbol{e} - \boldsymbol{\theta})|\hat{\boldsymbol{\theta}}] = E[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}^B)^T \boldsymbol{\Omega} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}^B)|\hat{\boldsymbol{\theta}}] + (\boldsymbol{e} - \hat{\boldsymbol{\theta}}^B)^T \boldsymbol{\Omega} (\boldsymbol{e} - \hat{\boldsymbol{\theta}}^B)$. Hence, the problem again reduces to minimization of $(\boldsymbol{e} - \hat{\boldsymbol{\theta}}^B)^T \boldsymbol{\Omega} (\boldsymbol{e} - \hat{\boldsymbol{\theta}}^B)$ with respect to $\boldsymbol{e}$ subject to $\boldsymbol{W}^T \boldsymbol{e} = \boldsymbol{t}$. The result follows from the identity

$$\begin{aligned}(\boldsymbol{e} - \hat{\boldsymbol{\theta}}^B)^T \boldsymbol{\Omega} (\boldsymbol{e} - \hat{\boldsymbol{\theta}}^B) &= [(\boldsymbol{e} - \hat{\boldsymbol{\theta}}^B - \boldsymbol{\Omega}^{-1} \boldsymbol{W} (\boldsymbol{W}^T \boldsymbol{\Omega}^{-1} \boldsymbol{W})^{-1} (\boldsymbol{t} - \hat{\boldsymbol{\theta}}^B)]^T \\ &\times \boldsymbol{\Omega}[(\boldsymbol{e} - \hat{\boldsymbol{\theta}}^B - \boldsymbol{\Omega}^{-1} \boldsymbol{W} (\boldsymbol{W}^T \boldsymbol{\Omega}^{-1} \boldsymbol{W})^{-1} (\boldsymbol{t} - \hat{\boldsymbol{\theta}}^B)] \\ &+ (\boldsymbol{t} - \bar{\hat{\boldsymbol{\theta}}}_w^B)^T (\boldsymbol{W}^T \boldsymbol{\Omega}^{-1} \boldsymbol{W})^{-1} (\boldsymbol{t} - \bar{\hat{\boldsymbol{\theta}}}_w^B).\end{aligned}$$

3 Further Benchmarking Results

There are situations, however, where in addition to benchmarking the first moment, one demands also benchmarking the variability of the Bayes estimators as well. We will address this issue in the special case when $\phi_i = cw_i$ for some $c(> 0)$, $i = 1, \cdots, m$. In this case, $\hat{\theta}_i^{BM1}$ given in (1) simplifies to $\hat{\theta}_i^B + (t - \bar{\hat{\theta}}_w^B)$ for all $i = 1, \cdots, m$. This itself is not a very desirable estimator since then $\sum_{i=1}^m w_i(\hat{\theta}_i^{BM1} - t)^2 = \sum_{i=1}^m w_i(\hat{\theta}_i^B - \bar{\hat{\theta}}_w^B)^2$. It can be shown as in Ghosh (1992) that $\sum_{i=1}^m w_i(\hat{\theta}_i^B - \bar{\hat{\theta}}_w^B)^2 < \sum_{i=1}^m w_i E[(\theta_i - \bar{\theta}_w)^2|\hat{\boldsymbol{\theta}}]$. In other words, the weighted ensemble variability of the estimators $\hat{\theta}_i^{BM1}$ is an underestimate of the posterior expectation of the corresponding weighted ensemble variability of the population parameters. To address this issue, or from other considerations, we will consider estimators $\hat{\theta}_i^{BM2}$, $i = 1, \ldots, m$ which satisfy two constraints, namely, (i) $\sum_{i=1}^m w_i\hat{\theta}_i^{BM2} = t$ and (ii) $\sum_{i=1}^m w_i(\hat{\theta}_i^{BM2} - t)^2 = H$, where H is a preassigned number taken from some other source, for example from census data, or it could be $\sum_{i=1}^m w_i E[(\theta_i - \bar{\theta}_w)^2|\hat{\boldsymbol{\theta}}]$ more in the spirit of Louis (1984) and Ghosh (1992). Subject to these two constraints, one minimizes $\sum_{i=1}^m w_i E[(\theta_i - e_i)^2|\hat{\boldsymbol{\theta}}]$. The following theorem provides the resulting estimator.

Theorem 3. Subject to (i) and (ii), the benchmarked Bayes estimators of θ_i $(i = 1, \cdots, m)$ are given by

$$\hat{\theta}_i^{BM2} = t + a_{CB}(\hat{\theta}_i^B - \bar{\hat{\theta}}_w^B), \tag{7}$$

where $a_{CB}^2 = H / \sum_{i=1}^m w_i(\hat{\theta}_i^B - \bar{\hat{\theta}}_w^B)^2$. Note that $a_{CB} \geq 1$ when $H = \sum_{i=1}^m w_i E[(\theta_i - \bar{\theta}_w)^2|\hat{\boldsymbol{\theta}}]$
.

Proof. As in Theorem 1, the problem reduces to minimization of $\sum_{i=1}^m w_i(e_i - \hat{\theta}_i^B)^2$. We will write

$$\sum_{i=1}^m w_i(e_i - \hat{\theta}_i^B)^2 = \sum_{i=1}^m w_i[(e_i - \bar{e}_w) - (\hat{\theta}_i^B - \bar{\hat{\theta}}_w^B)]^2 + (\bar{e}_w - \bar{\hat{\theta}}_w^B)^2. \tag{8}$$

Now define two discrete random variables Z_1 and Z_2 such that

$$P(Z_1 = e_i - \bar{e}_w, Z_2 = \hat{\theta}_i^B - \bar{\hat{\theta}}_w^B) = w_i,$$

$i = 1, \cdots, m$. Hence,

$$\sum_{i=1}^m w_i[(e_i - \bar{e}_w) - (\hat{\theta}_i^B - \bar{\hat{\theta}}_w^B)]^2 = V(Z_1) + V(Z_2) - 2Cov(Z_1, Z_2)$$

which is minimized when the correlation between Z_1 and Z_2 equals 1, i.e.

$$e_i - \bar{e}_w = a(\hat{\theta}_i^B - \bar{\hat{\theta}}_w^B) + b, \tag{9}$$

$i = 1, \cdots, m$ with $a > 0$. Multiplying both sides of (9) by w_i and summing over $i = 1, \cdots, m$, one gets $b = 0$. Next squaring both sides of (9), then multiplying both sides by w_i and summing over $i = 1, \cdots, m$, one gets $H = a^2 \sum_{i=1}^m w_i(\hat{\theta}_i^B - \bar{\hat{\theta}}_w^B)^2$ due to condition (ii). Finally, by condition (i), the result follows from (9).

Remark 10. As in the case of Theorem 1, it is possible to work with arbitrary ϕ_i rather than $\phi_i = w_i$ for all $i = 1, \cdots, m$. But then one does not get a closed form minimizer, although it can be shown that such a minimizer exists. We can also provide an algorithm for finding this minimizer numerically.

The multiparameter extension of the above result proceeds as follows. Suppose now $\hat{\boldsymbol{\theta}}_1, \cdots, \hat{\boldsymbol{\theta}}_m$ are the q-component direct estimators of the small area means $\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_m$. We generalize the constraints (i) and (ii) as (iM) $\bar{\boldsymbol{e}}_w = \sum_{i=1}^m w_i \boldsymbol{e}_i = \boldsymbol{t}$ for some specified $\boldsymbol{t}$ and (iiM) $\sum_{i=1}^m w_i(\boldsymbol{e}_i - \bar{\boldsymbol{e}}_w)(\boldsymbol{e}_i - \bar{\boldsymbol{e}}_w)^T = \boldsymbol{H}$, where $\boldsymbol{H}$ is a positive definite (possibly data dependent matrix), and is often taken as $\sum_{i=1}^m w_i E[(\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}}_w)(\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}}_w)^T | \hat{\boldsymbol{\theta}}]$. The second condition is equivalent to $\boldsymbol{c}^T\{\sum_{i=1}^m w_i(\boldsymbol{e}_i - \bar{\boldsymbol{e}}_w)(\boldsymbol{e}_i - \bar{\boldsymbol{e}}_w)^T\}\boldsymbol{c} = \boldsymbol{c}^T\boldsymbol{H}\boldsymbol{c}$ for every $\boldsymbol{c} = (c_1, \cdots, c_q)^T \neq \boldsymbol{0}$ which simplifies to $\sum_{i=1}^m w_i\{\boldsymbol{c}^T(\boldsymbol{e}_i - \bar{\boldsymbol{e}}_w)\}^2 = \boldsymbol{c}^T\boldsymbol{H}\boldsymbol{c}$. An argument similar as before now leads to $\boldsymbol{c}^T\hat{\boldsymbol{\theta}}_i^{BM2} = \boldsymbol{c}^T\bar{\hat{\boldsymbol{\theta}}}_w + a_{CB}\boldsymbol{c}^T(\hat{\boldsymbol{\theta}}_i^B - \bar{\hat{\boldsymbol{\theta}}}_w^B)$ for every $\boldsymbol{c} \neq \boldsymbol{0}$, where $\hat{\boldsymbol{\theta}}_i^B$ is the posterior mean of $\boldsymbol{\theta}_i$, $\bar{\hat{\boldsymbol{\theta}}}_w^B = \sum_{i=1}^m w_i\hat{\boldsymbol{\theta}}_i^B$, and $a_{CB}^2 = \boldsymbol{c}^T\boldsymbol{H}\boldsymbol{c}/\sum_{i=1}^m w_i\{\boldsymbol{c}^T(\hat{\boldsymbol{\theta}}_i^B - \bar{\hat{\boldsymbol{\theta}}}_w^B)\}^2$. The coordinatewise benchmarked Bayes estimators are now obtained by putting $\boldsymbol{c} = (1, 0, \cdots, 0)^T, \cdots, (0, 0, \cdots, 1)^T$ in succession.

The proposed approach can be extended also to a two-stage benchmarking somewhat similar to what is considered by Pfeffermann and Tiller (2006). To cite an example, consider the SAIPE scenario where we want to estimate the number of poor school children in different counties within a state, as well as those numbers within the different school districts in all these counties. Let $\hat{\theta}_i$ denote the Current Population Survey (CPS) estimate of θ_i, the true number of poor school children for the ith county and $\hat{\theta}_i^B$ the corresponding Bayes estimate, namely the posterior mean. Subject to the constraints $\bar{e}_w = \sum_{i=1}^m w_i e_i = \sum_{i=1}^m w_i\hat{\theta}_i = \bar{\hat{\theta}}_w$, and $\sum_{i=1}^m w_i(e_i - \bar{e}_w)^2 = H$, the benchmarked Bayes estimate for θ_i in the ith county is $\hat{\theta}_i^{BM2}$ as given in (7). Next, suppose that $\hat{\xi}_{ij}$ is the CPS estimator of ξ_{ij}, the true number of poor school children for the jth school district in the ith county and η_{ij} is the weight attached to the direct CPS estimator of ξ_{ij}, $j = 1, \ldots, n_i$. We seek estimators e_{ij} of ξ_{ij} such that (i) $\bar{e}_{i\eta} = \sum_{j=1}^{n_i} \eta_{ij} e_{ij} = \bar{\hat{\xi}}_{i\eta}^{BENCH}$, the benchmarked estimator of $\bar{\xi}_{i\eta} = \sum_{j=1}^{n_i} \eta_{ij}\xi_{ij}$, and (ii) $\sum_{j=1}^{n_i} \eta_{ij}(e_{ij} - \bar{e}_{i\eta})^2 = H_i^*$ for some preassigned H_i^*, where again H_i^* can be taken as $\sum_{j=1}^{n_i} \eta_{ij} E[(\xi_{ij} - \bar{\xi}_{i\eta})^2 | \hat{\boldsymbol{\xi}}_i]$, $\hat{\boldsymbol{\xi}}_i$ being the vector with elements ξ_{ij}. A benchmarked estimator similar to (7) can now be found for the ξ_{ij} as well.

4 An Example

The Small Area Income and Poverty Estimates (SAIPE) program at the U.S. Bureau of the Census produces model-based estimates of the number of poor school-aged children (5-17 years old) at the state, county and school district levels. The school district estimates are used by the Department of Education to allocate funds under the No Child Left Behind Act of 2001. In the SAIPE program, the model-based state estimates were benchmarked to the national design-based estimate of the number of poor school aged children from the Annual Social and Economic Supplement (ASEC) of the CPS up through 2004, while ACS estimates are used from 2005 onwards. Additionally, the model-based county estimates are benchmarked to the model-based state estimates in a hierarchical fashion using ratio adjustments. In this section we will consider the implications of different benchmarking methods, using the results from Theorems 1 and 3, on the state level estimates.

In the SAIPE program, the state model for poverty rates in school-aged children follows the basic Fay-Herriot framework (see e.g. Bell, 1999):

$$\begin{aligned} \hat{\theta}_i &= \theta_i + e_i && (10) \\ \theta_i &= \boldsymbol{x}_i^T \boldsymbol{\beta} + u_i && (11) \end{aligned}$$

where θ_i is the true state level poverty rate, $\hat{\theta}_i$ is the direct survey estimate (from CPS ASEC), e_i is the sampling error term with assumed known variance D_i, $\boldsymbol{x}_i$ are the predictors, $\boldsymbol{\beta}$ is the vector of regression coefficients and u_i is the model error with constant variance σ_u^2. The explanatory variables in the model are: IRS income tax based pseudo-estimate of the child poverty rate, IRS non-filer rate and the residual term from the regression of the 1990 Census estimated child poverty rate. The parameters $(\boldsymbol{\beta}, \theta_i, \sigma_u^2)$ are estimated using numerical integration for Bayesian inference (Bell, 1999).

The state estimates were benchmarked to the CPS direct estimate of the national school-aged child poverty rate until 2004. The weights, w_i, to calibrate the state's poverty rates to the national poverty rate, are proportional to the population estimates of the number of school-aged children in each state. Three different sets of risk function weights, ϕ_i, will be used to benchmark the estimated state poverty rates based on Theorem 1. The first set of weights will be the weights used in the benchmarking, i.e. $\phi_i = w_i$. The second set of weights creates the ratio adjusted benchmarked estimators, $\phi_i = w_i/\hat{\theta}_i^B$ (Remark 3). The third set of weights uses the results from Pfeffermann and Barnard (1991) where $\phi_i = w_i/Cov(\hat{\theta}_i^B, \bar{\hat{\theta}}_w^B)$ (Remark 6). Let the set of benchmarked estimates be denoted as $\hat{\boldsymbol{\theta}}^{(1)}$, $\hat{\boldsymbol{\theta}}^{(r)}$ and $\hat{\boldsymbol{\theta}}^{(pf)}$ respectively. Finally, we will benchmark the state poverty estimates using the results from Theorem 2 and denote the estimator as $\hat{\boldsymbol{\theta}}^{(2)}$.

To compute $\sum_i w_i E[(\theta_i - \bar{\theta}_w)^2 | \hat{\boldsymbol{\theta}}]$, two methods are given. The first method uses the first two posterior moments of $\boldsymbol{\theta}$.

$$\begin{aligned} H &= \sum_i w_i E[(\theta_i - \bar{\theta}_w)^2 | \hat{\boldsymbol{\theta}}] \\ &= E\left[\boldsymbol{\theta}^T (\boldsymbol{W} - \boldsymbol{w}\boldsymbol{w}^T) \boldsymbol{\theta} | \hat{\boldsymbol{\theta}}\right] \\ &= \text{trace}\left[(\boldsymbol{W} - \boldsymbol{w}\boldsymbol{w}^T)(\text{Var}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) + E(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) E(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})^T)\right], && (12) \end{aligned}$$

where $\boldsymbol{W} = \text{diag}(w_1, \ldots, w_m)$ and $\boldsymbol{w} = (w_1, \ldots, w_m)^T$. The second method is to compute the posterior mean of $\sum_i w_i (\theta_i - \bar{\theta}_w)^2$ from the MCMC output of the Gibbs sampler. While both give equivalent values to benchmark the variability, the second method may

Table 1: Benchmarking Statistics for ASEC CPS

year	t	$\bar{\hat{\theta}}_w^B$	a_{CB}
1995	18.7	17.9	1.09
1997	18.4	17.8	1.07
1998	17.5	16.8	1.14
1999	15.9	14.9	1.14
2000	14.6	15.4	1.11
2001	14.8	15.3	1.10

be more practical as the number of small areas becomes large because it does not require manipulating an $m \times m$ matrix.

For benchmarking, as given by Theorems 1 and 3, the key summary quantities are $t = \sum_i w_i \hat{\theta}_i$, $\bar{\hat{\theta}}_w^B = \sum_i w_i \hat{\theta}_i^B$ and a_{CB}. As noted earlier in Theorem 3, with the choice of H as given in (12), $a_{CB} \geq 1$. Six years of historical data from the CPS and the SAIPE program are analyzed and benchmarked using the four criteria mentioned above. Table 1 gives the key quantities for these six years. The hierarchical Bayes estimates for the years 1995-1999 underestimate the benchmarked poverty rate and overestimate the poverty rate in 2000 and 2001. Even if the estimate $\bar{\hat{\theta}}_w^B$ is close to the benchmarked value t, there is still a strong desire to have exact agreement between the quantities when producing official statistics.

Figure 1 shows the differences of the various benchmarked estimates from the hierarchical Bayes estimate, $\hat{\theta}_i^B$, made for the year 1999 when the overall poverty level had to be raised to agree with the national direct estimate. Figure 2 shows the differences for year 2000 when the overall poverty level had to be lowered to obtain agreement. The differences of the benchmarked estimators $\hat{\boldsymbol{\theta}}^{(1)}$, $\hat{\boldsymbol{\theta}}^{(r)}$ and $\hat{\boldsymbol{\theta}}^{(2)}$ from the HB estimator all fall on straight lines that pass through the same point $(\bar{\hat{\theta}}_w^B, t)$. In fact, these benchmarked estimators can be written in the form:

$$\hat{\theta}_i^{BM} = t + \alpha(\hat{\theta}_i^B - \bar{\hat{\theta}}_w^B)$$

where $\alpha = 1$ for $\hat{\boldsymbol{\theta}}^{(1)}$, $\alpha = t/\bar{\hat{\theta}}_w^B$ for $\hat{\boldsymbol{\theta}}^{(r)}$ and $\alpha = a_{CB}$ for $\hat{\boldsymbol{\theta}}^{(2)}$. The slopes of the lines in Figures 1 and 2 for differences in the benchmarked estimates from $\hat{\theta}_i$ are $\alpha - 1$. Since $a_{CB} \geq 1$, the slope for the difference $\hat{\theta}_i^{(3)} - \hat{\theta}_i^B$ will always be non-negative. The slopes for $\hat{\theta}_i^{(1)}$ and $\hat{\theta}_i^{(r)}$ depend on whether the benchmarked total t is larger or smaller than the model-based estimate $\bar{\hat{\theta}}_w^B$. The Pfeffermann-Barnard benchmarked estimator does not follow this form. However, it does show a trend in a similar direction as the benchmarked estimator $\hat{\theta}_i^{(r)}$ based on ratio adjustment.

Figure 1: Change due to Benchmarking: 1999

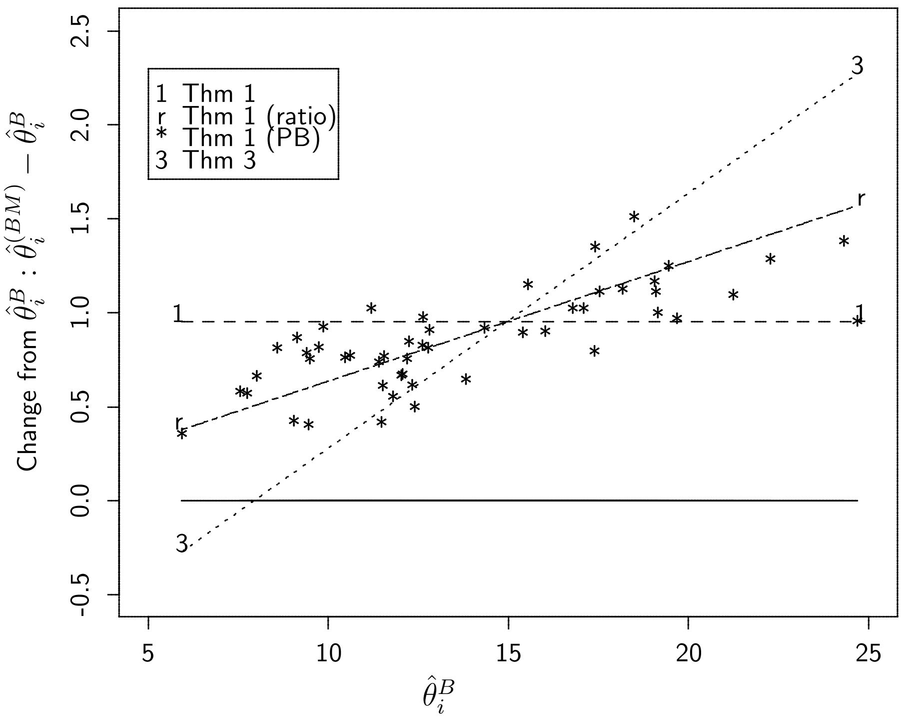
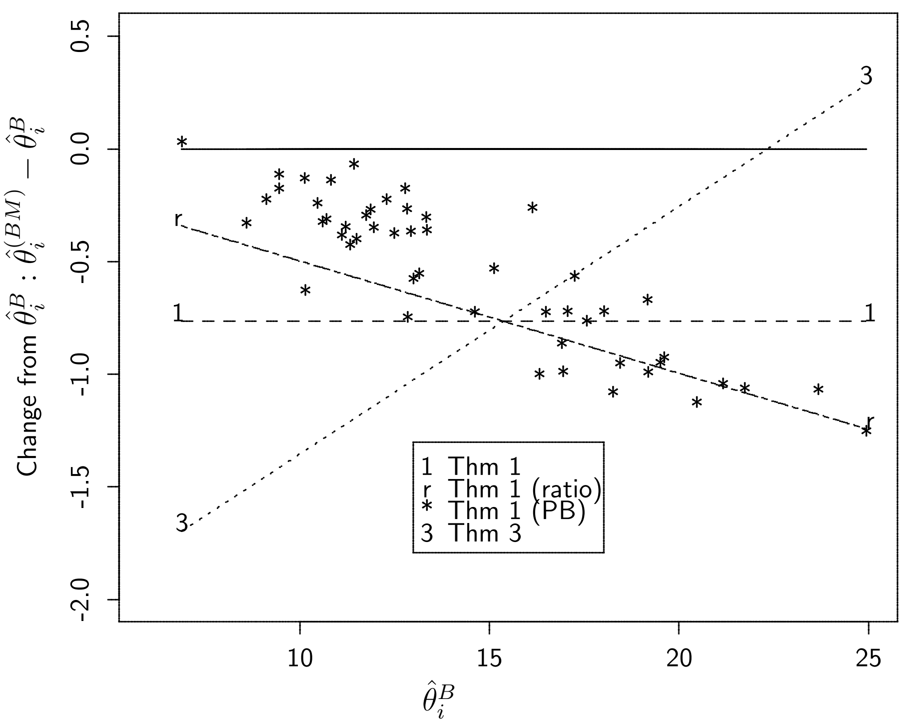


Figure 2: Change due to Benchmarking: 2000



5 Summary and Conclusion

The paper develops some Bayesian benchmarked estimators for area-level models. Benchmarking is achieved either with respect to some weighted mean or with respect to both weighted mean and weighted variability. The proposed benchmarked Bayes estimators include as special cases many benchmarked estimators proposed earlier.

We have mentioned only how to calculate empirical Bayes estimators for area-level models. The next step will be to evaluate and estimate their mean squared errors and also develop empirical Bayes confidence intervals.

Acknowledgments

We thank Don Luery for a comment which led to the general loss as considered in Theorem 1. The paper has also benefitted from the comments of William R. Bell and Donald J. Malec. Most of this work was completed when Ghosh was visiting the United States Bureau of the Census as an ASA/NSF/Census Senior Research Fellow. His research was also partially supported by an NSF Grant SES-0631426 and an NSA Grant MSPF-07G-097. Datta's research was partially supported by an NSA Grant MSPF-07G-082. This paper is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and do not necessarily reflect those of the United States Bureau of the Census.

References

1. Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd Edition. Springer-Verlag, New York.

2. Datta, G.S. and Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica*, **10**, 613-627.

3. Datta, G.S., Rao, J.N.K. and Smith, D.D. (2005). On measuring the variability of small area estimators under a basic area level model. *Biometrika*, **92**, 183-196.

4. Fay, R.E. and Herriot, R.A. (1979). Estimates of income from small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association*, **74**, 269-277.

5. Fuller, W.A. (2007). Small area prediction subject to a restriction. Preprint.

6. Ghosh, M. (1992). Constrained Bayes estimation with applications. *Journal of the American Statistical Association*, **87**, 533-540.

7. Isaki, C.T., Tsay, J.H. and Fuller, W.A. (2000). Estimation of census adjustment factors. *Survey Methodology*, **26**, 31-42.

8. Louis, T.A. (1984). Estimating a population of parameter values using Bayes and empirical Bayes methods. *Journal of the American Statistical Association*, **79**, 393-398.

9. Pfeffermann, D. and Barnard, C.H. (1991). Some new estimators for small area means with application to the assessment of farmland values. *Journal of Business and Economic Statistics*, **9**, 31-42.

10. Pfeffermann, D. and Nathan, G. (1981). Regression analysis of data from a cluster sample. *Journal of the American Statistical Association*, **76**, 681-689.

11. Pfeffermann, D. and Tiller, R. (2006). Small-area estimation with state-space models subject to benchmark constraints. *Journal of the American Statistical Association*, **101**, 1387-1397.

12. Prasad, N.G.N. and Rao, J.N.K. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, **85**, 163-171.

13. Nandram, B., Toto, Ma. C.S. and Choi, J.W. (2007). A Bayesian benchmarking for small areas. Preprint.

14. Rao, J.N.K. (2003). *Small Area Estimation*. Wiley, New York.

15. Wang, J. and Fuller, W.A. (2002). Small area estimation under a restriction. *Proceedings of the Section on Survey Research methods of the American Statistical Association*, 3627-3630.

16. You, Y. and Rao, J.N.K. (2002). A pseudo-empirical best linear unbiased prediction approach to small area estimation using survey weights. *The Canadian Journal of Statistics*, **30**, 431-439.

17. You, Y. and Rao, J.N.K. (2003). Pseudo hierarchical Bayes small area estimation combining unit level models and survey weights. *Journal of Statistical Planning and Inference*, **111**, 197-208.

18. You, Y., Rao, J.N.K. and Dick, P. (2004). Benchmarking hierarchical Bayes small area estimators in the Canadian census undercoverage estimation. *Statistics in Transition*, **6**, 631-640.

19. Zellner, A. (1986). Further results on Bayesian minimum expected loss (MELO) estimates and posterior distributions for structural coefficients. In *Advances in Econometrics.* Ed. D.J. Slottje. JAI Press Inc., pp, 171-182.

20. Zellner, A. (1988). Bayesian analysis in econometrics. *Journal of Econometrics*, **37**, 27-50.

21. Zellner, A. (1994). Bayesian and non-Bayesian estimation using balanced loss functions. In *Statistical Decision Theory and Related Topics V*. Eds. S.S. Gupta and J.O. Berger. Springer-Verlag, New York, pp. 377-390.