BAYESIAN VERSUS FREQUENTIST MEASURES OF UNCERTAINTY FOR SMALL AREA ESTIMATORS

A.C. Singh, D.M. Stukel (Statistics Canada) and D. Pfeffermann (Hebrew University)
D.M. Stukel, SSMD, Statistics Canada, Ottawa, Canada, K1A 0T6

KEY WORDS: Empirical Best Linear Unbiased Predictors; Maximum Likelihood; MSE; Small Area Estimation.

1. INTRODUCTION

Statistical bureaus are often required to provide reliable estimators for small area means. The problem with the production of such estimators is that the sample sizes within those areas are usually too small to allow the use of direct survey estimators. Hence, new estimators have been proposed in recent years which combine auxiliary information representing individual or small area characteristics with the data observed for the response variable in all the small areas. Unlike direct survey estimators, these small area estimators "borrow strength" from other small areas. See Ghosh and Rao (1994) for a recent review and discussion of the major developments in small area estimation over the last two decades.

A general class of models giving rise to this kind of estimators are the mixed linear models. A special, yet quite general member of this class, often analyzed in the literature and underlying our study is the nested error regression model defined as

$$y_{ij} = x_{ij}'\beta + \nu_i + e_{ij};\ i = 1, \ldots, k;\ j = 1, \ldots, m_i \tag{1.1}$$

where y_{ij} is the value of the response variable for the j-th unit sampled within the i-th small area, x_{ij}' is the corresponding vector of auxiliary variables values (including a possible intercept term), β is a vector of fixed regression coefficients and ν_i and e_{ij} are independent white noise random variables such that the ν_i have mean 0 and common variance λ_1, and the e_{ij} have mean 0 and common variance λ_2. Here the ν_i represent the joint effects of small area characteristics not included in the measurements x_{ij}. The true small area means are defined under the model as $\bar{Y}_{i(p)} = \bar{X}_{i(p)}'\beta + \nu_i + \bar{e}_{i(p)}$ where $\bar{X}_{i(p)}'$ and $\bar{e}_{i(p)}$ are the population means of the x_{ij} and e_{ij} for small area i. Assuming, however, that the small area population sizes are sufficiently large, then $\bar{e}_{i(p)} \cong 0$ and so we can take the true small area means to be

$$\theta_i = \bar{X}_{i(p)}'\ \beta + \nu_i. \tag{1.2}$$

Further, assuming that the sampling fractions within the small areas are sufficiently small, then any reasonable predictor for θ_i should also be an appropriate predictor for $\bar{Y}_{i(p)}$. The θ_i include fixed and random effects and so both frequentist and Bayesian approaches arise naturally for estimating these quantities. In fact, if the variance components inherent in the models are known, then the frequentist-based best linear unbiased predictor (BLUP) of θ_i and its associated mean squared error (MSE) coincide respectively with the Bayesian posterior mean and variance of θ_i, assuming normality of the ν_i and e_{ij} and assuming a noninformative prior on β.

When the variance components are unknown, it is common practice to substitute suitable consistent estimates for them in the expressions for the BLUP or the posterior mean. Under the frequentist approach, this gives rise to what is known as the "empirical BLUP" (EBLUP), while the use of estimated parameters under the Bayesian framework gives rise to "parametric empirical Bayes" (PEB) estimators. It is interesting to note that under a hierarchical Bayes framework, the PEB estimator can also be obtained as an approximation to the posterior mean, irrespective of the prior distribution on the variance components $\lambda' = (\lambda_1, \lambda_2)$, when the number of small areas is large — for example, see Kass and Steffey (1989).

The use of the EBLUP, or equivalently the PEB estimator, raises the question of how to assess the prediction errors under the two approaches. The work of Kackar and Harville (1984), (henceforth K-H), and Prasad and Rao (1990), (henceforth P-R), under the frequentist framework, and Morris (1983) and Kass and Steffey (1989), (henceforth K-S) under the Bayesian framework indicates that a naive replacement of the unknown variance components by their estimates in the theoretical expressions for the MSE or the posterior variance may result in severe underestimation. This is because the resultant estimators fail to account for the additional uncertainty arising from the estimation of the unknown variance components in the expressions for the small area predictors. In the studies cited above, the authors propose modifications to account for this extra variability.

In the present paper we study the performance of the above modifications and propose some alternative methods, aimed to measure the uncertainty in small area estimation. Specifically, we address the following issues:

A. Following the frequentist approach, P-R correct the bias in the estimator of the MSE of the EBLUP proposed by K-H, so that the neglected terms are of order $o(k^{-1})$. K-S develop two analagous approximations to the posterior variance associated with the PEB estimator under an asymptotic hierarchical Bayesian framework; the first, denoted KS-I, has neglected terms of order $O(k^{-1})$, while the second, denoted KS-II, has neglected terms of order $O(k^{-2})$. We offer an alternative to KS-II (denoted KS-II*), which is simpler than KS-II and allows for a term by term comparison with P-R, unlike KS-II.

B. The modification proposed by P-R is based on the δ (linearization) method, which requires first order partial derivatives of the EBLUP with respect to the variance components. These derivatives can be cumbersome to derive under more general models and hence we explore the use of Monte Carlo Integration (MCI) approximations to the P-R approximation (which are of the same order as the P-R approximations).

C. Hamilton (1986) proposes an MCI procedure for assessing the posterior variance of unobservable components in state-space time series models. We borrow his idea and apply it in a different context, that is, as an alternative method for approximating the posterior variance associated with the PEB estimator. We also propose a modified version of Hamilton's approximation whose neglected terms are of order $o(k^{-1})$, and which can be considered an MCI analogue to KS-II*.

D. The aforementioned bias modifications are asymptotically correct; that is, they are based on the number of small areas increasing to infinity. We are interested in modifications to the approximations to the MSE or posterior variance from two separate standpoints: i) We prefer small overestimation to any underestimation which may arise in the approximations to the MSE or posterior variance, since conservative approximations are always desirable. That is, we prefer a small positive bias to a negative one, and ii) We are interested in a reduction in the absolute bias of the approximations insomuch as it does not cause a corresponding increase in the MSE (of the approximations to the MSE or posterior variance). We perform a Monte Carlo simulation study which explores these two issues as they relate to the various approximations discussed in A-C above.

The Monte Carlo study explores the frequentist properties of all of the approximations even though the approximations to the posterior variance are motivated by Bayesian theory considerations. As illustrated by Hulting and Harville (1991), there are some theoretical results suggesting that the use of Bayesian predictors and their associated posterior variances may be appropriate for use in a frequentist context. For example, a $100(1-\alpha)\%$ posterior credible interval based on a Jeffrey's prior may have a frequentist coverage of approximately $100(1-\alpha)\%$. Similarly, K-S argue that the variance approximations derived in their study can "also be justified as variance estimates in non-Bayesian theory".

The organization of this paper is as follows. In section 2 we review the frequentist approximations to measures of uncertainty as proposed by K-H and P-R, and consider an alternative method to the latter, based on MCI methods (see B above). In section 3 we discuss Bayesian approximations emerging from the work of K-S and Hamilton (1986). These approximations have a bias of order $O(k^{-1})$ and we propose bias corrections which reduce the order of the bias to $o(k^{-1})$ (see A and C above). The empirical results obtained from the simulation study (see D above) are summarized in section 4. In section 5, we close with some concluding remarks.

2. APPROXIMATIONS UNDER THE FREQUENTIST APPROACH: EXISTING AND PROPOSED METHODS

In the following discussion, we assume the underlying model to be defined by (1.1), and the parameter of interest, θ_i, to be defined by (1.2). To simplify the discussion, we assume that $\lambda_2 = Var(e_{ij})$ is known, but in the empirical study we consider the more general case where λ_2 is unknown as well. Let $\lambda' = (\lambda_1, \lambda_2)$ and $y = (y_{ij})$.

2.1 Existing Methods

Case 1: β and λ are Known

The BLUP of θ_i is given by

$$\hat{\theta}_i(y, \lambda, \beta) = \bar{X}'_{i(p)}\beta + \gamma_i(\bar{y}_i - \bar{x}'_{i(s)}\beta) \tag{2.1}$$

where $(\bar{y}_i, \bar{x}'_{i(s)}) = \sum_{j=1}^{m_i}(y_{ij}, x'_{ij})/m_i$ are the sample means of the y_{ij} and x_{ij} over small area i and $\gamma_i = \lambda_1(\lambda_1 + \lambda_2 m_i^{-1})^{-1}$ is the "shrinkage factor". Notice that $E[\hat{\theta}_i(y, \lambda, \beta) - \theta_i] = 0$. The corresponding MSE is given by

$$\begin{aligned} \text{MSE}\left[\hat{\theta}_i(y, \lambda, \beta)\right] &= E\left[\hat{\theta}_i(y, \lambda, \beta) - \theta_i\right]^2 \\ &= \lambda_1\lambda_2 m_i^{-1}(\lambda_1 + \lambda_2 m_i^{-1})^{-1} \equiv g_{1i}(\lambda). \end{aligned} \tag{2.2}$$

Case 2: λ Known and β Unknown

The BLUP of θ_i has the same form as in (2.1), except that β is now replaced by the weighted (Aitken) least squares estimator

$$\hat{\beta} = (X'V^{-1}X)X'V^{-1}y \tag{2.3}$$

where X and V are defined in the usual manner. The BLUP of θ_i and its corresponding MSE are therefore given by:

$$\hat{\theta}_i(y, \lambda) = \bar{X}'_{i(p)}\hat{\beta} + \gamma_i(\bar{y}_i - \bar{x}'_{i(s)}\hat{\beta}) \tag{2.4}$$

and

$$\begin{aligned} &\text{MSE}\left[\hat{\theta}_i(y, \lambda)\right] \\ &= g_{1i}(\lambda) + (\bar{X}_{i(p)} - \gamma_i\bar{x}_{i(s)})'(X'V^{-1}X)^{-1}(\bar{X}_{i(p)} - \gamma_i\bar{x}_{i(s)}) \\ &\equiv g_{1i}(\lambda) + g_{2i}(\lambda) \end{aligned} \tag{2.5}$$

where $g_{1i}(\lambda)$ is defined by (2.2).

Case 3: λ and β Both Unknown

This case arises most often in practice. As discussed in the introduction, it is common to replace the unknown elements of λ by consistent sample estimates $\hat{\lambda}$ in the expression $\hat{\theta}_i(y, \lambda)$. The resulting predictor, $\hat{\theta}_i(y, \hat{\lambda})$, is referred to in the literature as the "empirical BLUP" (EBLUP). It is interesting to note that the EBLUP remains unbiased provided that $\hat{\lambda}$ is an even, translation invariant function of y, and the distributions of ν_i and e_{ij} are both symmetric.

Now an exact expression for the MSE of the EBLUP is intractable for most models. A first attempt at an estimate may lead one to simply subsitute $\hat{\lambda}$ for λ into (2.5), but this estimator seriously underestimates the MSE of the EBLUP; the underestimation arises due to the fact that this estimator, commonly known as the "Naive Estimator", ignores the uncertainty in estimating λ.

K-H show that when $\hat{\lambda}$ is translation invariant and the errors ν_i and e_{ij} are normal,

$$\begin{aligned} \text{MSE}[\hat{\theta}_i(y, \hat{\lambda})] &= \text{MSE}[\hat{\theta}_i(y, \lambda)] \\ &+ E[\hat{\theta}_i(y, \hat{\lambda}) - \hat{\theta}_i(y, \lambda)]^2. \end{aligned} \tag{2.6}$$

Using the δ-method, K-H approximate the second term on the right hand side of (2.6) correct to order of $o(k^{-1})$ as $E[d(\lambda)'(\hat{\lambda} - \lambda)]^2$ where $d(\lambda) = \partial\hat{\theta}_i(y, \lambda)/\partial\lambda$. They further simplify the approximation as $tr\{A(\lambda)E[(\hat{\lambda} - \lambda)(\hat{\lambda} - \lambda)']\}$ where $A(\lambda)$ is the covariance matrix of $d(\lambda)$, using heuristic arguments for approximate independence of $\hat{\lambda}$ and $d(\lambda)$. For the special case where only λ_1 is unknown, the K-H approximation to the MSE of the EBLUP simplifies to

$$\begin{aligned} \text{MSE}_{KH}[\hat{\theta}_i(y, \hat{\lambda}_1, \lambda_2)] &= g_{1i}(\lambda) \\ &+ g_{2i}(\lambda) + E[d^2(\lambda_1)]\ Var(\hat{\lambda}_1) + o(k^{-1}). \end{aligned} \tag{2.7}$$

P-R simplify $E[d^2(\lambda_1)]$ by noting (implicitly) that in calculating $d(\lambda_1)$, $\hat{\beta}$ can be regarded as fixed for the order of approximation under consideration. Thus, the P-R approximation to the last term of (2.7) is given by

$$\begin{aligned} &E\{\partial[\hat{\gamma}_i(\bar{y}_i - \bar{x}'_{i(s)}\beta)]/\partial\hat{\lambda}_1|_{\hat{\lambda}_1=\lambda_1}\}^2 Var(\hat{\lambda}_1) \\ &= m_i^{-2}\lambda_2^2(\lambda_1 + \lambda_2 m_i^{-1})^{-3}Var(\hat{\lambda}_1) \end{aligned}$$

so that

$$\begin{aligned} &\text{MSE}_{PR}[\hat{\theta}_i(y, \hat{\lambda}_1, \lambda_2)] \\ &= g_{1i}(\lambda) + g_{2i}(\lambda) + g_{3i}(\lambda) + o(k^{-1}) \end{aligned} \tag{2.8}$$

where $g_{3i}(\lambda) = m_i^{-2}\lambda_2^2(\lambda_1 + \lambda_2 m_i^{-1})^{-3}Var(\hat{\lambda}_1) + o(k^{-1})$. An expression for $Var(\hat{\lambda}_1)$ can easily be derived assuming normality of the random errors, when $\hat{\lambda}_1$ is the "Method of Fitting Constants" or MFC estimator of λ_1; this expression is again a function of λ_1.

Now the approximation (2.8) is still a function of the unknown variance component λ_1 and thus can be estimated by substituting $\hat{\lambda}_1$ for λ_1. However the resulting estimator, $g_{1i}(\hat{\lambda}_1, \lambda_2) + g_{2i}(\hat{\lambda}_1, \lambda_2) + g_{3i}(\hat{\lambda}_1, \lambda_2)$ has a bias of order

$O(k^{-1})$ since $E[g_{1i}(\hat{\lambda}_1, \lambda_2) - g_{1i}(\lambda_1, \lambda_2)] = O(k^{-1})$. In order to reduce the order of the bias, P-R proceed as follows: they consider the Taylor expansion of $g_{1i}(\hat{\lambda}_1, \lambda_2)$ about λ_1:

$$g_{1i}(\hat{\lambda}_1, \lambda_2) = g_{1i}(\lambda) + \lambda_2^2 m_i^{-2} (\lambda_1 + \lambda_2 m_i^{-1})^{-2} (\hat{\lambda}_1 - \lambda_1) - \lambda_2^2 m_i^{-2} (\lambda_1 + \lambda_2 m_i^{-1})^{-3} (\hat{\lambda}_1 - \lambda_1)^2 + o_p(k^{-1}). \tag{2.9}$$

Applying the expectation operator to both sides and assuming that $E(\hat{\lambda}_1 - \lambda_1) = o(k^{-1})$ and that $E(o_p(k^{-1})) = o(k^{-1})$ (both conditions hold when λ_1 is estimated by the MFC), then

$$E[g_{1i}(\hat{\lambda}_1, \lambda_2)] = g_{1i}(\lambda) - g_{3i}(\lambda) + o(k^{-1}). \tag{2.10}$$

The P-R bias-corrected estimator of $\mathrm{MSE}_{PR}[\hat{\theta}_i(y, \hat{\lambda}_1, \lambda_2)]$ is therefore

$$\hat{\mathrm{MSE}}_{PR}[\hat{\theta}_i(y, \hat{\lambda}_1, \lambda_2)] = g_{1i}(\hat{\lambda}_1, \lambda_2) + g_{2i}(\hat{\lambda}_1, \lambda_2) + 2g_{3i}(\hat{\lambda}_1, \lambda_2) \tag{2.11}$$

noting that $E\{\hat{\mathrm{MSE}}_{PR}[\hat{\theta}_i(y, \hat{\lambda}_1, \lambda_2)] - \mathrm{MSE}_{PR}[\hat{\theta}_i(y, \hat{\lambda}_1, \lambda_2)]\} = o(k^{-1})$ since $g_{2i}(\hat{\lambda}_1, \lambda_2)$ and $g_{3i}(\hat{\lambda}_1, \lambda_2)$ have biases of order $o(k^{-1})$. See P-R for details.

2.2 Proposed Alternative to the P-R Estimator Based on MCI

We conclude this section by outlining an MCI procedure for approximating $\hat{MSE}_{PR}[\hat{\theta}_i(y, \hat{\lambda}_1, \lambda_2)]$. The use of this procedure is general and avoids the computation of the derivatives $d(\lambda)$ which can be cumbersome under more complicated models.

To motivate the idea, note that the second term on the right hand side of (2.6) can be written in general as

$$\begin{aligned} &E[\hat{\theta}_i(y, \hat{\lambda}_1, \lambda_2) - \hat{\theta}_i(y, \lambda)]^2 \\ &= E\left\{(\bar{X}_{i(p)} - \hat{\gamma}_i \bar{x}_{i(s)})' [\hat{\beta}(\hat{\lambda}_1) - \hat{\beta}] + (\hat{\gamma}_i - \gamma_i)(\bar{y}_i - \bar{x}_{i(s)}' \beta)\right\}^2 \\ &= E[(\hat{\gamma}_i - \gamma_i)(\bar{y}_i - \bar{x}_{i(s)}' \beta)]^2 + o(k^{-1}) \\ &\approx E(\hat{\gamma}_i - \gamma_i)^2 E(\bar{y}_i - \bar{x}_{i(s)}' \beta)^2 + o(k^{-1}) \end{aligned} \tag{2.12}$$

where $\hat{\beta}(\hat{\lambda}_1)$ is defined as in (2.3) but with $\hat{\lambda}_1$ substituted for λ_1. The first equality follows directly from (2.4) whereas the second equality follows from the results of P-R, where they note implicitly that $\hat{\beta}$ (and hence $\hat{\beta}(\hat{\lambda}_1)$) can be regarded as fixed for the order of approximation under consideration. The third equality assumes an approximate independence between $\hat{\gamma}_i$ and $\bar{y}_i$, which can be justified heuristically by noting that $\hat{\gamma}_i = \hat{\lambda}_1(\hat{\lambda}_1 + \lambda_2 m_i^{-1})^{-1}$ is based on the data in all the small areas whereas $\bar{y}_i$ depends only on the data from a single small area. Also see K-H and P-R. Now the term $E(\hat{\gamma}_i - \gamma_i)^2$ in (2.12) is difficult to evaluate since the expectation is taken with respect to y, and since $\hat{\gamma}_i$ is nonlinear in y. Note that we may alternatively consider the expectation as taken with respect to $\hat{\lambda}_1$ since $\hat{\gamma}_i$ depends on y only through $\hat{\lambda}_1$.

Thus consider a Taylor expansion of $\hat{\gamma}_i = \gamma_i(\hat{\lambda}_1, \lambda_2)$ about λ_1:

$$\gamma_i(\hat{\lambda}_1, \lambda_2) = \gamma_i(\lambda_1, \lambda_2) + \lambda_2 m_i^{-1} (\lambda_1 + \lambda_2 m_i^{-1})^{-2} (\hat{\lambda}_1 - \lambda_1) + O_p(k^{-1}). \tag{2.13}$$

To obtain a Taylor approximation to the MSE of $\gamma_i(\hat{\lambda}_1, \lambda_2)$ as an estimator of $\gamma_i(\lambda_1, \lambda_2)$, one can simply rearrange (2.13), square both sides, apply the operator $E_{\hat{\lambda}_1}$, and then substitute $\hat{\lambda}_1$ for the unknown λ_1, obtaining:

$$\hat{E}_{\hat{\lambda}_1}[\gamma_i(\hat{\lambda}_1, \lambda_2) - \gamma_i(\lambda_1, \lambda_2)]^2 = \lambda_2^2 m_i^{-2} (\hat{\lambda}_1 + \lambda_2 m_i^{-1})^{-4} \hat{E}_{\hat{\lambda}_1}(\hat{\lambda}_1 - \lambda_1)^2 + o_p(k^{-1}). \tag{2.14}$$

It can be shown that this is mathematically equivalent to reversing the roles of λ_1 and $\hat{\lambda}_1$, and taking a Taylor expansion of γ_i about $\hat{\lambda}_1$, proceeding as before, except applying the operator E_{λ_1} (defined below) instead. Thus, we obtain:

$$E_{\lambda_1}[\gamma_i(\lambda_1, \lambda_2) - \gamma_i(\hat{\lambda}_1, \lambda_2)]^2 = \lambda_2^2 m_i^{-2} (\hat{\lambda}_1 + \lambda_2 m_i^{-1})^{-4} E_{\lambda_1}(\lambda_1 - \hat{\lambda}_1)^2 + o_p(k^{-1}). \tag{2.15}$$

To define E_{λ_1}, it is sufficient as well as convenient to endow λ_1 with a working distribution so that it is normal having mean $\hat{\lambda}_1$ and variance $\hat{Var}(\hat{\lambda}_1)$, and we can evaluate the expectation on the left hand side of (2.15) by MCI. That is, we can draw a large number of realizations (say L) of $\lambda_{1l} \sim N(\hat{\lambda}_1, \hat{Var}(\hat{\lambda}_1))$ and then compute $MS_{\gamma_i} = \sum_{l=1}^{L} [\gamma_i(\lambda_{1l}, \lambda_2) - \gamma_i(\hat{\lambda}_1, \lambda_2)]^2 / L$. Thus, the MCI approximation to the MSE of $\gamma_i(\hat{\lambda}_1, \lambda_2)$ is computed as

$$E_{\lambda_1}[\gamma_i(\hat{\lambda}_1, \lambda_2) - \gamma_i(\lambda_1, \lambda_2)]^2 \approx MS_{\gamma_i}. \tag{2.16}$$

Multiplying MS_{γ_i} by $E(\bar{y}_i - \bar{x}_{i(s)}' \beta)^2 = (\lambda_1 + \lambda_2 m_i^{-1})$ yields an MCI approximation to $g_{3i}(\lambda)$ (see last expression in (2.12)). Note that switching the roles of λ_1 and $\hat{\lambda}_1$ and proceeding as above does not alter the order of the aproximation.

Finally to estimate the MSE of the EBLUP correct to order $o(k^{-1})$, one needs to adjust for the bias $E[g_{1i}(\hat{\lambda}_1, \lambda_2) - g_{1i}(\lambda_1, \lambda_2)]$ in a similar vein to P-R (see equation 2.10). This can again be implemented by MCI, that is, by computing

$$E_{\lambda_1}[g_{1i}(\hat{\lambda}_1, \lambda_2) - g_{1i}(\lambda_1, \lambda_2)] \approx \sum_{l=1}^{L} [g_{1i}(\hat{\lambda}_1, \lambda_2) - g_{1i}(\lambda_{1l}, \lambda_2)] / L \equiv M_{g_{1i}}. \tag{2.17}$$

Thus, an MCI approximation to the P-R estimator of the MSE is given by (compare with (2.11)):

$$\hat{MSE}_{MCI(I)}[\hat{\theta}_i(y, \hat{\lambda}_1, \lambda_2)] = g_{1i}(\hat{\lambda}_1, \lambda_2) + g_{2i}(\hat{\lambda}_1, \lambda_2) + M_{g_{1i}} + (\hat{\lambda}_1 + \lambda_2 m_i^{-1}) MS_{\gamma_i} \tag{2.18}$$

where $(\hat{\lambda}_1 + \lambda_2 m_i^{-1}) = \hat{E}(\bar{y}_i - \bar{x}_{i(s)}' \beta)^2$ (see last expression in (2.12)). Notice that $E[\hat{MSE}_{MCI(I)} - MSE_{PR}] = o(k^{-1})$ provided that $E(\hat{\lambda}_1 - \lambda) = o(k^{-1})$; the latter condition is satisfied when $\hat{\lambda}_1$ is estimated by MFC.

A possible disadvantage to using (2.18) instead of (2.11) is that the bias correction $M_{g_{1i}}$ is not necessarily positive, unlike $g_{3i}(\hat{\lambda}_1, \lambda_2)$ used in (2.11). An alternative MCI approximation to $\hat{MSE}_{PR}$ (having the same order) which ensures a positive bias correction is obtained by replacing $g_{3i}(\hat{\lambda}_1, \lambda_2)$ in (2.11) by $(\hat{\lambda}_1 + \lambda_2 m_i^{-1}) MS_{\gamma_i}$, i.e.

$$\hat{MSE}_{MCI(II)}[\hat{\theta}_i(y, \hat{\lambda}_1, \lambda_2)] = g_{1i}(\hat{\lambda}_1, \lambda_2) + g_{2i}(\hat{\lambda}_1, \lambda_2) + 2(\hat{\lambda}_1 + \lambda_2 m_i^{-1}) MS_{\gamma_i}. \tag{2.19}$$

Note that the last term of (2.19) mimicks the $2g_{3i}(\hat{\lambda}_1, \lambda_2)$ term of (2.11).

3. ASYMPTOTIC BAYESIAN APPROXIMATIONS: EXISTING METHODS AND MODIFICATIONS

In this section we consider the PEB approach for the small area estimation problem. As in section 2 we concentrate primarily on measures of uncertainty (prediction errors). To motivate the reasons behind considering the Bayesian approach, we begin by discussing commonalities between the two approaches (Bayesian and frequentist).

Case 1: λ and β are Known

In this case the posterior mean of θ_i which is the Bayesian predictor under a quadratic loss function coincides with the BLUP under frequentist theory, defined by (2.1). Similarly the posterior variance coincides with the MSE of the BLUP, defined by (2.2).

Case 2: λ Known and β Unknown

Assuming a noninformative prior distribution for β, the posterior mean and variance of θ_i again coincide with the corresponding BLUP and MSE obtained under frequentist theory as defined by (2.4) and (2.5).

Case 3: λ and β Both Unknown

As discussed in the introduction, a common approach to dealing with this case is to substitute a suitable estimate of λ in the expression for the posterior mean of θ_i for the case where λ is known, thus yielding a PEB estimator. If $\tilde{\lambda}$ defines the "restricted maximum likelihood" (REML) estimator of λ, then it can be shown that for large k

$$E(\theta_i \mid y) = \hat{\theta}_i(y, \tilde{\lambda}) + O(k^{-1}) \tag{3.1}$$

regardless of the prior distribution on λ. Thus, the frequentist and Bayesian approaches give rise asymptotically to the same predictors under a quadratic loss function. Note that for the case of β known, the same results apply if $\tilde{\lambda}$ is taken to be the usual "maximum likelihood" estimator of λ, rather than the REML estimator. Thus it is of interest to consider approximations to the corresponding posterior variance, $V(\theta_i \mid y)$.

From here on in, as in section 2, we consider the case where only β and λ_1 are unknown; the case where λ_2 is also unknown is considered in the empirical study. Throughout the following discussion, we will assume a noninformative prior on β.

We began by noting that the posterior variance can be decomposed as follows:

$$\begin{aligned} V(\theta_i \mid y, \lambda_2) &= E_{\lambda_1 \mid y}[V(\theta_i \mid y, \lambda)] + V_{\lambda_1 \mid y} E[\theta_i \mid y, \lambda] \\ &= E_{\lambda_1 \mid y}[g_{1i}(\lambda) + g_{2i}(\lambda)] + V_{\lambda_1 \mid y}[\hat{\theta}_i(y, \lambda)] \end{aligned} \tag{3.2}$$

where $g_{1i}(\lambda)$ and $g_{2i}(\lambda)$ are defined by (2.2) and (2.5) and $\hat{\theta}_i(y, \lambda)$ is defined by (2.4).

3.1 Existing Methods

Use of the δ-method

First and second order approximations to the posterior variance (denoted KS-I and KS-II, respectively) in the context of hierarchical Bayes (HB) models have been developed by K-S using the δ-method. We briefly describe KS-I:

K-S show that

$$\begin{aligned} &E_{\lambda_1 \mid y}[g_{1i}(\lambda) + g_{2i}(\lambda)] \\ &= g_{1i}(\tilde{\lambda}_1, \lambda_2) + g_{2i}(\tilde{\lambda}_1, \lambda_2) + O(k^{-1}) \end{aligned} \tag{3.3}$$

and

$$\begin{aligned} V_{\lambda_1 \mid y}[\hat{\theta}_i(y, \lambda)] &= [d^*(\tilde{\lambda}_1)]^2 \tilde{Var}(\tilde{\lambda}_1) + o(k^{-1}) \\ &\equiv g_{3i}^*(\tilde{\lambda}_1, \lambda_2) + o(k^{-1}) \end{aligned} \tag{3.4}$$

where $d^*(\tilde{\lambda}_1) = (\partial \hat{\theta}_i(y, \lambda)/\partial \lambda_1)$ evaluated at $\lambda_1 = \tilde{\lambda}_1$ and $\tilde{Var}(\tilde{\lambda}_1)$ is minus the inverse of the second derivative of the log likelihood evaluated at $\lambda_1 = \tilde{\lambda}_1$. Substitution of (3.3) and (3.4) into (3.2) yields the K-S first order approximation to the posterior variance of θ_i:

$$\begin{aligned} &\hat{V}_{KS\text{-}I}(\theta_i \mid y, \lambda_2) \\ &= g_{1i}(\tilde{\lambda}_1, \lambda_2) + g_{2i}(\tilde{\lambda}_1, \lambda_2) + g_{3i}^*(\tilde{\lambda}_1, \lambda_2) + O(k^{-1}). \end{aligned} \tag{3.5}$$

In section 3.2, we modify the first order aproximation given by (3.5) to make it second order by adding an extra term of order $O(k^{-1})$ to (3.3). The modification (denoted KS-II*) is simpler than KS-II, and allows for a term by term comparison with the P-R approximation, unlike KS-II.

Use of MCI

Hamilton (1986) proposed an MCI approximation to the posterior variance of unobservable components of state-space time series models. We borrow his idea and apply it to the present context.

Now a general MCI procedure consists of approximating the two terms on the right hand side of (3.2) by drawing realizations λ_{1l}, $l = 1, \ldots, L$ from the posterior distribution $f(\lambda_1 \mid y)$. If we let $g_i(l) = [g_{1i}(\lambda_{1l}, \lambda_2) + g_{2i}(\lambda_{1l}, \lambda_2)]$, $\hat{\theta}_{i,l} = \hat{\theta}_i(y, \lambda_{1l}, \lambda_2)$ and $\bar{\theta}_i = \sum_{l=1}^{L} \hat{\theta}_{i,l}/L$, then,

$$E_{\lambda_1 \mid y}[g_{1i}(\lambda) + g_{2i}(\lambda)] \cong \textstyle\sum_{l=1}^{L} g_i(l)/L \tag{3.6}$$

and

$$V_{\lambda_1 \mid y}[\hat{\theta}_i(y, \lambda)] \cong \textstyle\sum_{l=1}^{L} (\hat{\theta}_{i,l} - \bar{\theta}_i)^2/L. \tag{3.7}$$

In approximating the posterior variance of unobservable components of state-space time series models, Hamilton (1986) approximated the posterior distribution $f(\lambda_1 \mid y)$ as multivariate normal with mean and variance given respectively by the REML estimator and its corresponding inverse information matrix. If we use the same aproximate posterior distribution in calculating (3.6) and (3.7), and then substitute these equations into (3.2), we obtain a Hamilton approximation to the posterior variance in terms of the present context. In section 3.2 ahead, we improve this approximation by reducing the magnitude of the bias in approximating the posterior mean of λ_1.

An attractive feature of the Hamilton procedure is its simplicity; without much difficulty, it can be applied to other models and/or to different predictors.

3.2 Proposed Modifications

The δ Method

As mentioned earlier (see equation (3.3)), the bias in approximating $E_{\lambda_1 \mid y}[g_{1i}(\lambda) + g_{2i}(\lambda)]$ by $g_{1i}(\tilde{\lambda}_1, \lambda_2) + g_{2i}(\tilde{\lambda}_1, \lambda_2)$ has order $O(k^{-1})$. In order to reduce the order of the bias, we propose the following procedure:

Consider the Taylor expansion of $g_{1i}(\lambda)$ about $\tilde{\lambda}_1$. This is identical to (2.9) except that the roles of $\tilde{\lambda}_1$ and λ_1 have been reversed. That is,

$$\begin{aligned} g_{1i}(\lambda) = {} & g_{1i}(\tilde{\lambda}_1, \lambda_2) + \lambda_2^2 m_i^{-2}(\tilde{\lambda}_1 + \lambda_2 m_i^{-1})^{-2}(\lambda_1 - \tilde{\lambda}_1) \\ & - \lambda_2^2 m_i^{-2}(\tilde{\lambda}_1 + \lambda_2 m_i^{-1})^{-3}(\lambda_1 - \tilde{\lambda}_1)^2 + o_p(k^{-1}) \end{aligned} \tag{3.8}$$

where the term $o_p(k^{-1})$ is with respect to the posterior distribution $f(\lambda_1 \mid y)$. It is known that,

$$E(\lambda_1 \mid y) = \tilde{\lambda}_1 + O(k^{-1}) \tag{3.9}$$

and

$$V(\lambda_1 \mid y) = \tilde{Var}(\tilde{\lambda}_1) + O(k^{-2}) \tag{3.10}$$

where $\tilde{Var}(\tilde{\lambda}_1)$ is defined as before. (c.f. equations (3.1) and (3.2) and Remark 1 of K-S.) Therefore, when taking expectations on both sides of (3.8) with respect to $f(\lambda_1 \mid y)$, the mean $E[(\lambda_1 - \tilde{\lambda}_1)^2 \mid y]$ can be approximated by $\tilde{Var}(\tilde{\lambda}_1)$ with neglected terms of order $o(k^{-1})$ (using (3.9) and (3.10)), but approximating the mean $E[(\lambda_1 - \tilde{\lambda}_1) \mid y]$ by $\tilde{\lambda}_1$ gives neglected terms of order $O(k^{-1})$ (using (3.9)). Thus, to correct the error in estimating $E_{\lambda_1 \mid y}[g_{1i}(\lambda) + g_{2i}(\lambda)]$ by the approximation given in (3.3), one needs to approximate the bias $E[(\lambda_1 - \tilde{\lambda}_1) \mid y]$, so that the neglected terms have order $o(k^{-1})$.

An appropriate approximation is attainable using the approach of Tierney and Kadane (1986) and Tierney, Kass and Kadane (1989). Using this approach, the log likelihood is modified by adding the term $\log(\lambda_1 + C)$ where C is a large positive constant. The modified likelihood is then maximized, resulting in a modified REML estimator λ_1^* of λ_1. An approximation to the posterior mean $E(\lambda_1 \mid y)$, with the correct order is calculated as

$$E(\lambda_1 \mid y) \simeq (\sigma^* / \sigma) \exp\{L^*(\lambda_1^*) - L(\tilde{\lambda}_1)\} - C \equiv \tilde{\tilde{\lambda}}_1, \quad (3.11)$$

say, where $L(\tilde{\lambda}_1)$ and $L^*(\lambda_1^*)$ are the log likelihood evaluated at $\tilde{\lambda}_1$, and modified log likelihood evaluated at λ_1^* respectively, and σ and σ^* are minus the inverse of the second derivatives of L and L^* with respect to λ_1 evaluated at $\tilde{\lambda}_1$ and λ_1^* respectively. By (3.8), (3.10) and (3.11),

$$E_{\lambda_1 \mid y}[g_{1i}(\lambda)] = g_{1i}(\tilde{\lambda}_1, \lambda_2) + g_{4i}(\tilde{\lambda}_1, \lambda_2) - g_{3i}(\tilde{\lambda}_1, \lambda_2) + o(k^{-1}) \quad (3.12)$$

where $g_{4i}(\tilde{\lambda}_1, \lambda_2) = \lambda_2^2 m_i^{-2} (\tilde{\lambda}_1 + \lambda_2 m_i^{-1})^{-2} (\tilde{\tilde{\lambda}}_1 - \tilde{\lambda}_1)$.
By (3.2), (3.3), (3.4) and (3.12), the proposed second order approximation to the posterior variance of θ_i is therefore

$$\hat{V}_{KS\text{-}II^*}(\theta_i \mid y, \lambda_2) = g_{1i}(\tilde{\lambda}_1, \lambda_2) + g_{2i}(\tilde{\lambda}_1, \lambda_2) + g_{3i}^*(\tilde{\lambda}_1, \lambda_2) + g_{4i}(\tilde{\lambda}_1, \lambda_2) - g_{3i}(\tilde{\lambda}_1, \lambda_2). \quad (3.13)$$

The MCI Method

The Hamilton procedure, defined in section 3.1 is correct to first order. The key to the effective use of this procedure is the correct specification of the posterior distribution $f(\lambda_1 \mid y)$. It is well known that under mild regularity conditions this posterior distribution is asymptotically normal, so that the specification of the posterior distribution reduces to the specification of the posterior mean and variance. If we use $V(\lambda_1 \mid y) \simeq \hat{Var}(\tilde{\lambda}_1)$ as an approximation to the posterior variance (see equation (3.10)), then the neglected terms have order $o(k^{-1})$, as desired. However, if we use $E(\lambda_1 \mid y) \simeq \tilde{\lambda}_1$ as an approximation to the posterior mean (see equation (3.9)), then the neglected terms are not of the corrrect order; if instead, we use $E(\lambda_1 \mid y) \simeq \tilde{\tilde{\lambda}}_1$ (see equation (3.11)), then the neglected terms are of order $o(k^{-1})$, as desired. Thus, a modified (second order) Hamilton procedure consists of generating observations λ_{1l} from $N[\tilde{\tilde{\lambda}}_1, \hat{Var}(\tilde{\lambda}_1)]$ and then computing (3.6) and (3.7)

4. MONTE CARLO STUDY

4.1 Design of the Monte Carlo Study

A Monte Carlo study was conducted to enable the examination of frequentist properties of the approximations to the MSE of the EBLUP and posterior variance, from the standpoints discussed in i) and ii) of D in the introduction to this paper; thus the percent relative errors with respect to the true MSE and the root mean squared errors of these approximations were examined. The model used for simulation purposes was the one-fold nested error regression model with one auxiliary variable:

$$y_{ij} = -16 + .494 x_{ij} + \nu_i + e_{ij} \quad (4.1)$$

where ν_i and e_{ij} were generated according to $e_{ij} \sim N(0, 150)$ and $\nu_i \sim N(0, \lambda_1)$, and where λ_1 was allowed to vary as $\lambda_1 = 30, 75, 150, 300$, giving rise to variance component ratios of $\lambda_1/\lambda_2 = .2, .5, 1, 2$. Note that $\beta' = (-16, .494)$ was assumed known; thus the formulae used in the simulation study differed from those presented in the main text of this paper. We took auxiliary data values (x_{ij}) from Battese, Harter and Fuller (1988), as well as the values for β', given above. Initially, there was enough data for $k = 12$ small areas, but three of these had $m_i = 1$, so they were pooled, resulting in $k = 10$ small areas. (m_i varied from 2 to 6 inclusively). We increased the number of small areas from $k = 10$ to $k = 20$, $k = 40$, $k = 100$ and $k = 200$ successively by duplicating (x_{ij}, m_i, X_i) as many times as was respectively needed. For each of the 16 combinations (k = 20, 40, 100, 200 × $\lambda_1/\lambda_2 = .2, .5, 1, 2$), we generated 4,000 independent sets of $\{e_{ij}; i = 1, \ldots, t; j = 1, \ldots, m_i\}$ and $\{\nu_i; i = 1, \ldots, t\}$ according to the distributions above, and thus generated 4,000 sets of $\{y_{ij}; i = 1, \ldots, t; j = 1, \ldots, m_i\}$ under the model (4.1), using the given x_{ij} values. For each set of $\{y_{ij}\}$, 1,000 realizations of $\lambda' = (\lambda_1, \lambda_2)$ were drawn for the purpose of Monte Carlo integration, for those methods that required them (MCI(I), MCI(II), Hamilton and Modified Hamilton). Note here that both λ_1 and λ_2 were assumed to be unknown, unlike the simpler case presented in the main text where only λ_1 was assumed to be unknown. For all nine approximations considered (Naive, K-H, P-R, MCI(I), MCI(II), KS-I, KS-II*, Hamilton and Modified Hamilton) the unknown variance components were estimated by MLE (rather than REML, since β was assumed to be known). Strictly speaking, for the frequentist-based approximations, λ should have been estimated by MFC, or alternatively, by any estimator $\hat{\lambda}$ for which $E(\hat{\lambda} - \lambda) = o(k^{-1})$; although we assumed this condition to hold for MLE, this assumption probably needs further investigation.

4.2 Results of the Monte Carlo Study

In the discussion that follows, we report only limited results from the Monte Carlo study since space is restricted; more detailed results can be obtained from the authors. The results for $k = 40$ small areas are given; at $k = 20$ small areas, the approximations behave somewhat erratically and at $k = 100$ and $k = 200$ small areas, all approximations behave similarly since all are asymptotically unbiased. In addition, we only report the results for the variance component ratios $\lambda_1/\lambda_2 = .2, .5$ and 2, and for those small areas having sample sizes $m_i = 2$ and 6, representing the extremes.

Table 1 ahead gives the percent relative error for the nine approximations considered. That is, it gives (as a percentage) the difference between the Monte Carlo expectation of the approximations and the true MSE, divided by the true MSE. As expected the Naive estimator was always negatively biased, the underestimation becoming less severe as λ_1/λ_2 increased. There was still underestimation in the case of K-H, but it was not as severe as in the case of the Naive estimator. A marked improvement could be noticed by using the P-R approximation. Not only did the sign of the bias change from negative to positive, making the P-R approximation conservative, but the magnitude of the bias decreased as well. The two MCI methods did a reasonable job of tracking the true MSE, notably, the second method when $\lambda_1/\lambda_2 = .2$ and the first method for the other values of λ_1/λ_2. KS-I behaved much like K-H, both underestimating the true MSE, which is not surprising since both are missing a term of order $O(k^{-1})$. For Hamilton's method, the underestimation was more severe than for KS-I, but still not as bad as for the Naive estimator; again, this is not surprising for the same reason as previously stated. The modified Hamilton method and KS-II* fared better than their unmodified counterparts in the sense that the sign of the bias generally turned from negative to positive, although the magnitude of the overestimation could be somewhat greater than hoped for in certain cases. For all approximations, the picture generally improved when m_i, the sample size within a small area, increased and/or when λ_1/λ_2 increased.

Table 2 ahead gives the Root Mean Squared Error (RMSE) for the nine approximations considered. Keeping in mind that MSE can be decomposed into two terms: a squared bias term and a variance term, it is of interest to consider the RMSE in the sense that any possible reduction in the magnitude of the bias from one approximation to another should not cause an increase in the corresponding RMSE. In perusing Table 2, first consider successively the Naive, K-H and P-R approximations, where each was an improvement over the last in the sense that they became successively more conservative (see Table 1). Fortunately, for the most part, there appeared to be successive drops in the RMSE as well. The two proposed MCI methods performed similarly, although they suffered from a slight increase in the RMSE over the P-R approximation. Finally, whereas the Modified Hamilton method and KS-II* turned the underestimation in their unmodified counterparts into overestimation (see Table 1), here they both suffered from a slight inflation in the RMSE over their unmodified counterparts. Finally, even though it is not surprising, it is worth noting that for all approximations, there was a rather dramatic drop in the RMSE as m_i, the sample size within a small area, increased from 2 to 6.

5. SUMMARY AND REMARKS

Under the model-based framework of small area estimation, both existing and proposed asymptotic Bayesian and frequentist methods for measuring uncertainty of estimators were considered. Frequentist properties of these methods were compared by means of a Monte Carlo study. It was found that the Prasad-Rao approximation performed best overall in terms of being able to reduce the bias without increasing the RMSE. The second MCI version of Prasad-Rao performed well for small values of the ratio of the variance components, whereas the first MCI version fared well for large such values. Thus, the proposed MCI versions can be useful in practice if the computation of the required derivatives is found to be too cumbersome for the application in question. The unmodified Bayesian methods of Kass-Steffey and Hamilton tended to be biased downward. However, the proposed modifications corrected them in the right direction and made then generally conservative; both, unfortunately, experienced an increase in RMSE over their unmodified counterparts. However, with the proposed modifications, one has the advantage of a dual interpretation of these approximations in both frequentist and Bayesian contexts.

REFERENCES

Battese, G.E., Harter, R.M. and Fuller, W.A. (1988), "An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data", *Journal of the American Statistical Association*, 83, 28-36.

Ghosh, M. and Rao, J.N.K. (1994), "Small Area Estimation: An Appraisal", to appear in *Statistical Science*.

Hamilton, J.D. (1986), "A Standard Error for the Estimated State Vector of a State-Space Model", *Journal of Econometrics*, 33, 387-397.

Kackar, R.N. and Harville, D.A. (1984), "Approximations for Standard Errors of Estimators of Fixed and Random Effects in Mixed Linear Models", *Journal of the American Statistical Association*, 79, 853-862.

Kass, R.E. and Steffey, D. (1989), "Approximate Bayesian Inference in Conditionally Independent Hierarchical Models (Parametric Empirical Bayes Models", *Journal of the American Statistical Association*, 84, 717-726.

Morris, C. (1983), "Parametric Empirical Bayes Inference: Theory and Applications (with discussion)", *Journal of the American Statistical Association*, 78, 47-65.

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.

Tierney, L. and Kadane, J.B. (1986), "Accurate Approximation for Posterior Moments and Marginal Densities", *Journal of the American Statistical Association*, 84, 82-86.

Tierney, L., Kass, R.E. and Kadane, J.B. (1989), "Fully Exponential Laplace Approximations to Expectations and Variances of Nonpositive Functions", *Journal of the American Statistical Association*, 84, 710-716.

Table 1: Percent Relative Error of the Approximations to the MSE of the EBLUP for k=40 Small Areas

Approximation	$\lambda_1/\lambda_2=.2$		$\lambda_1/\lambda_2=.5$		$\lambda_1/\lambda_2=2$	
	$m_i=2$	$m_i=6$	$m_i=2$	$m_i=6$	$m_i=2$	$m_i=6$
Naive	-11.63	-14.80	-7.46	-6.75	-2.48	-2.37
K-H	-4.50	-5.17	-3.44	-3.19	-1.01	-1.47
P-R	2.63	5.40	0.58	0.38	0.46	-0.78
MCI(I)	-5.50	-5.62	0.71	2.10	1.86	0.30
MCI(II)	-0.67	0.04	1.52	3.93	2.37	0.83
KS-I	-4.16	-5.86	-3.28	-3.34	-0.98	-1.49
KS-II*	3.61	-4.69	2.39	0.41	3.36	2.80
Hamilton	-6.19	-10.20	-6.54	-5.27	-1.97	-1.74
Modified Hamilton	7.15	-1.46	2.28	0.69	3.41	2.89

Table 2: Root Mean Squared Error of the Approximations to the MSE of the EBLUP for k=40 Small Areas

Approximation	$\lambda_1/\lambda_2=.2$		$\lambda_1/\lambda_2=.5$		$\lambda_1/\lambda_2=2$	
	$m_i=2$	$m_i=6$	$m_i=2$	$m_i=6$	$m_i=2$	$m_i=6$
Naive	8.58	4.51	7.59	2.79	7.18	2.95
K-H	8.16	3.19	7.13	2.51	7.25	3.02
P-R	8.11	2.67	6.98	2.41	7.35	3.04
MCI(I)	9.76	4.79	7.56	2.71	7.55	3.10
MCI(II)	8.67	3.93	7.45	2.77	7.64	3.14
KS-I	8.53	3.95	7.48	2.72	7.36	3.02
KS-II*	9.12	4.50	10.34	4.09	10.03	4.12
Hamilton	7.35	3.68	7.81	3.01	7.55	3.05
Modified Hamilton	7.75	3.21	9.81	3.84	9.99	4.11