An Empirical Bayes Prediction Interval for the Finite Population Mean of a Small Area

Balgobin Nandram

Worcester Polytechnic Institute, balnan@wpi.edu

Follow this and additional works at: https://digitalcommons.wpi.edu/mathematicalsciences-pubs

Part of the Mathematics Commons

Suggested Citation


This Article is brought to you for free and open access by the Department of Mathematical Sciences at DigitalCommons@WPI. It has been accepted for inclusion in Mathematical Sciences Faculty Publications by an authorized administrator of DigitalCommons@WPI. For more information, please contact akgold@wpi.edu.
AN EMPIRICAL BAYES PREDICTION INTERVAL FOR THE
FINITE POPULATION MEAN OF A SMALL AREA

Balgobin Nandram

Worcester Polytechnic Institute

Abstract: We construct an empirical Bayes (EB) prediction interval for the finite population mean of a small area when data are available from many similar small areas. We assume that the individuals of the population of the $i^{th}$ area are a random sample from a normal distribution with mean $\mu_i$ and variance $\sigma_i^2$. Then, given $\sigma_i^2$, the $\mu_i$ are independently distributed with each $\mu_i$ having a normal distribution with mean $\theta$ and variance $\sigma_i^2\tau$, and the $\sigma_i^2$ are a random sample from an inverse gamma distribution with index $\eta$ and scale $(\eta - 1)\delta$. First, assuming $\theta, \tau, \delta$ and $\eta$ are fixed and known, we obtain the highest posterior density (HPD) interval for the finite population mean of the $\ell^{th}$ area. Second, we obtain the EB interval by “substituting” point estimators for the fixed and unknown parameters $\theta, \tau, \delta$ and $\eta$ into the HPD interval, and a two-stage procedure is used to partially account for underestimation of variability. Asymptotic properties (as $\ell \to \infty$) of the EB interval are obtained by comparing its center, width and coverage probability with those of HPD interval. Finally, by using a small-scale numerical study, we assess the asymptotic properties of the proposed EB interval, and we show that the EB interval is a good approximation to the HPD interval for moderate values of $\ell$.

Key words and phrases: Asymptotic, Bayes risk, Monte Carlo, HPD interval, simulation, uniform integrability.

1. Introduction

Many federal government agencies are required to obtain estimates of population counts, unemployment rates, per capita income, health needs, crop yields, and livestock numbers for state and local government areas. In the US the National Health Planning and Resources Development Act of 1974 created a need for accurate small area estimates. The Health Systems Agencies, mandated by the Planning Act, are required to collect and analyze data related to the health status of the residents and to the health delivery systems in their health service areas. Consequently, there is a growing demand for reliable statistics for small areas. Our objective is to construct an interval estimator of the finite population mean of a small area using data from many other similar areas as well, and then to assess its properties.

Ghosh and Rao (1994) gave a comprehensive review of the growing small area literature. They described three recent approaches which have made significant
impact on small area estimation during the past decade: empirical Bayes (EB), hierarchical Bayes (HB) and empirical best linear unbiased predictor (EBLUP). They also described several small area models and many different point estimators, along with several applications. Unfortunately, even in their extensive review there is virtually no discussion on interval estimation, of critical importance to practitioners in small area estimation. In this paper we consider mainly the naive EB approach with some adjustments for underestimation of variability; see Morris (1983a) for an excellent account of the EB approach with several important applications.

Hulting and Harville (1991) described and compared frequentist and Bayesian methods for constructing approximate prediction intervals for a small area population mean when a mixed linear model might hold. However, their models do not include the type of models we wish to consider in this paper. In addition, there is the current concern that interval estimation is receiving relatively little attention in the small area literature.

Typically there is great variability among the sample sizes of the small areas with small samples dominating. Thus, if a prediction interval for the finite population mean of a small area is based only on its own data, it is likely to be too wide. Narrower intervals can be obtained by “borrowing strength” from similar areas. The primary objective is to provide the best possible estimates for areas that contain few, if any, sampling units (the “small” areas). For example, the allocation of federal funds to local governments is based in part on per capita income (PCI) (Fay and Herriot (1979)). In practice the distribution of these monies is based on estimates of PCI determined from a national survey sample. Thus, good estimates of PCI are required, even when the sample information for a local government is sparse. In particular, empirical Bayes methods have been proposed for use in such situations.

Let $Y_{\ell}$ denote the vector of all values from the $\ell$th area, $\gamma(Y_{\ell})$ denote the finite population mean of the $\ell$th area, $Y_s$ denote the vector of sample values from the $\ell$th area, and let $I(Y_s) = [P_L(Y_s), P_U(Y_s)]$ represent an interval for $\gamma(Y_{\ell})$ which depends on the sample data $Y_s$. If $P\{\gamma(Y_{\ell}) \in I(Y_s)\} = 1 - \alpha$, then $I(Y_s)$ is an exact $100(1 - \alpha)$% prediction interval for $\gamma(Y_{\ell})$. If $P\{\gamma(Y_{\ell}) \in I(Y_s)\}$ is approximately $1 - \alpha$, then $I(Y_s)$ is an approximate $100(1 - \alpha)$% prediction interval. A $100(1 - \alpha)$% credible set $S$ of $\gamma(Y_{\ell})$ values is a set such that $P\{\gamma(Y_{\ell}) \in S|Y_s\} = 1 - \alpha$. The smallest $100(1 - \alpha)$% set is the $100(1 - \alpha)$% highest posterior density (HPD) prediction credible set. Moreover, if the posterior density of $\gamma(Y_{\ell})$ is unimodal, the HPD prediction set is a $100(1 - \alpha)$% HPD prediction interval. In both cases the probability measure is taken over the marginal distribution of $\gamma(Y_{\ell})$ given $Y_s$. (In the Bayesian context this marginal distribution is the posterior predictive distribution.)
As a basis for inference, we assume that the \( N_i \) individuals in the population in the \( i \)th area follow the super population model

\[
Y_{i1}, \ldots, Y_{iN_i} | \mu_i, \sigma_i^2 \overset{i.i.d.}{\sim} N(\mu_i, \sigma_i^2)
\]  

(1.1)

with independence over \( i = 1, \ldots, \ell \). Next, we specify

\[
\mu_i | \sigma_i^2 \sim N(\theta, \tau \sigma_i^2)
\]

(1.2)

\( i = 1, \ldots, \ell \) with independence across areas. For small area estimation, it is not entirely unreasonable to assume that the variances share an effect (i.e., they have a common distribution). Quite naturally we specify an inverted gamma distribution for the \( \sigma_i^2 \), and without loss of generality, for convenience, we reparameterize the inverted gamma distribution. Thus, finally we specify

\[
\sigma_1^2, \ldots, \sigma_{\ell}^2 \overset{i.i.d.}{\sim} IG\{\eta, (\eta - 1)\delta\},
\]

(1.3)

where the inverse gamma density in (1.3) is given by

\[
f(\sigma_i^2) = \frac{(\eta - 1)\delta}{\eta (\eta - 1)\delta/\Gamma(\eta)}, \quad \sigma_i^2 > 0 \quad \text{and} \quad f(\sigma_i^2) = 0 \quad \text{otherwise.}
\]

We assume \( \theta, \tau, \delta \) and \( \eta \) are fixed but unknown parameters. Nandram and Sedransk (1993) considered a similar model in Section 2 of their paper in which they assumed that \( \eta \) is fixed and known. Since \( E(\sigma_i^2) = \delta \) and \( \text{var}(\sigma_i^2) = \delta^2/(\eta - 2), \eta > 2 \), for fixed \( \delta \) small values of \( \eta \) express a belief that the variances are very different whereas large values express a belief that they are very similar.

Let \( \tilde{Y}_i = (Y_{i1}, \ldots, Y_{iN_i})' \) be the vector of all values from the \( i \)th area, \( i = 1, \ldots, \ell \). Also, let \( \tilde{\theta} = (\theta, \ldots, \theta)' \) be a \( N_i \times 1 \) vector with each component \( \theta \). Then, it is straightforward to show that

\[
(Y_i - \tilde{\theta})\{(\tau + 1)(\eta - 1)\delta/\eta\}^{-1/2} \overset{d}{\sim} t_{N_i}(0, R, 2\eta),
\]

(1.4)

where \( t_{N_i}(0, R, 2\eta) \) is a \( N_i \)-variate Student \( t \) distribution located at the origin with correlation matrix \( R = (r_{jj'}) \), \( r_{jj'} = 1 \), \( j = j' \) and \( r_{jj'} = \tau/(1 + \tau) \), \( j \neq j' \), \( j, j' = 1, \ldots, N_i \), on \( 2\eta \) degrees of freedom; see, for example, Box and Tiao (1992), pg. 117. Note that the model specifies that the \( Y_i \) are independent, and that the components of the \( Y_i \) are exchangeable. In fact, if the area sizes \( N_i \) are equal, the \( Y_i \) have the same distribution with parameters \( \theta \), \( \eta \), \( \delta \) and \( \tau \). The multivariate Student \( t \) distribution in (1.4) will be used as the basis for the asymptotics. This is the general set up for many small area models when there are no covariates.

While (1.1), (1.2) and (1.3) provide a simple specification, it is expected to hold within strata (or clusters) of the entire population of small areas. We note that in more complex surveys there will be covariates.
Details of more complicated small area models incorporating covariates are presented by Prasad and Rao (1990). They assume that the error variances ($\sigma_i^2$) are equal. Our approach is similar to theirs in that we use a two-stage procedure to help account for underestimation in variability, an issue well-known in empirical Bayes statistics. However, while their analysis is motivated by the EBLUP approach, our approach is motivated by the EB approach. Note that in the spirit of Prashad and Rao (1990), the issue of how to obtain the mean squared error for the EBLUP of the $\mu_i$ in a model with unequal error variances is discussed by Kleffe and Rao (1992). Neither of these papers address the construction of an empirical prediction interval for the finite population mean. Also note that the problem of constructing a good empirical Bayes prediction interval is much more difficult than the problem of obtaining an accurate mean squared error. The issue of constructing an empirical Bayes prediction interval for the finite population mean of a small area is of greatest concern in this paper.

More recently, Arora, Lahiri and Mukherjee (1997) described empirical Bayes estimation of finite population means from complex surveys. Their model assumes that the error variances are unequal but distinct from the prior variance of the $\mu_i$. Note that in our model the $\mu_i$ do not share an effect. The EB estimator obtained by Arora, Lahiri and Mukherjee (1997) does not exist in closed form; one needs to perform a one-dimensional numerical integration. While Arora, Lahiri and Mukherjee (1997) incorporated covariates into their model, we consider a model without covariates. However, because it is not possible to obtain a closed form for the EB estimator with their model, it is clearly impossible to obtain a closed form interval estimator for the finite population mean of a small area.

Finally, we note that our model can be motivated using the posterior linearity assumption made by Ghosh and Lahiri (1987), see the fifth paragraph of Section 2 of their paper.

It is required to construct a $100(1 - \alpha)\%$ prediction interval estimator for $\gamma(Y_{\ell}) = \sum_{j=1}^{N_{\ell}} Y_{ij}/N_{\ell}$. Let $s_i$ denote the set of $n_i$ individuals sampled from the $N_i$ individuals of the $i$th area. Then, $f_i = n_i/N_i$ is the sampling fraction, $\bar{Y}_i = \sum_{j \in s_i} Y_{ij}/n_i$ is the sample mean and $S_i^2 = \sum_{j \in s_i} (Y_{ij} - \bar{Y}_i)^2/(n_i - 1)$ is the sample variance.

We denote the shrinkage factor by $\omega_i = (1 + n_i \tau)^{-1}$ for the $i$th area, $i = 1, \ldots, \ell$. Note that larger shrinkage factor for the $i$th area corresponds to larger degree of pooling for the $i$th area (i.e., the EB estimator of the finite population mean of the $i$th area uses more information from the other areas). Also observe that, as expected, the shrinkage factor $\omega_i$ decreases as either $n_i$ or $\tau$ increases. It may be desirable to use a model that makes the shrinkage factor depend on the data. This is an advantage of the model proposed by Arora, Lahiri and Mukherjee (1997) but the cost is a one-dimensional numerical integration problem.
Also, let \( \phi = (\theta, \tau, \delta, \eta)' \) be fixed, \( \tilde{\sigma}^2 = \{(n_\ell - 1)S^2_\ell + n_\ell\omega_\ell(\bar{Y}_\ell - \theta)^2 + 2(\eta - 1)\delta\}k^{-1} \) where \( k = n_\ell + 2\eta \). Then, as in Nandram and Sedransk (1993), given the sampled data from the \( \ell \)th area, a $100(1 - \alpha)$% highest posterior density (HPD) interval for \( \gamma(Y_\ell) \) is

\[
e_B \pm \nu_B t_{\kappa,\alpha/2},
\]

where, for convenience, we write

\[
e_B = \bar{Y}_\ell - (1 - f_\ell)\omega_\ell(\bar{Y}_\ell - \theta), \quad \nu_B^2 = (1 - f_\ell)(f_\ell + (1 - f_\ell)(1 - \omega_\ell))\tilde{\sigma}^2_\ell n_\ell^{-1}
\]

and \( t_{\kappa,\alpha/2} \) is the 100(1 - \( \alpha/2 \))th percentile point of the Student \( t \) distribution with \( \kappa \) degrees of freedom. Under squared error loss, \( e_B \) is the Bayes estimator of \( \gamma(Y_\ell) \). We note that since the HPD interval in (1.5) is based on the Student \( t \) distribution, there is some degree of robustness against outliers.

However, since \( \phi \) is unknown, our objective is to construct a $100(1 - \alpha)$% EB prediction interval for \( \gamma(Y_\ell) \) by appropriately weighting the sampled data from all areas.

There are two approaches to EB interval estimation. The first is simple. It requires virtually no computation and simple estimators are used for model parameters; see, for example, Nandram and Sedransk (1993). The second approach is a set of more sophisticated methods to account for underestimation of variability. Morris (1983a,b) used flat priors on hyperparameters while Laird and Lewis (1987, 1989) used a bootstrap method. It should be noted that Arora, Lahiri and Mukherjee (1997) extended the Laird-Lewis bootstrap method to finite population sampling to provide a measure of variability for their EB estimator. A third approach uses an asymptotic approximation to the appropriate posterior variance (Kass and Steffey (1989)) with a positive correction term added to the estimated posterior variance. A fourth method due to Carlin and Gelfand (1990) uses a bias-corrected technique which requires at least moderate computational effort. There is also a fifth method (Raghunathan (1993)) which provides a quasi-empirical Bayes approach for small area estimation based on a specification of a set of conditionally independent hierarchical mean and variance functions describing the first two moments of the process generating the data. However, because we prefer simplicity and reasonable accuracy over sophistication, we use the first approach to “substitute” point estimators of the components of \( \phi \) based on \( \bar{Y}_\ell \) and \( S^2_\ell \) into (1.5). The underestimation in variability is partially reduced by using a two-stage approach as in Nandram and Sedransk (1993).

Henceforth, maintaining the EB spirit, all analyses are based on the joint marginal distribution of the \( Y_{ij} \). That is, the parameters \( \mu_i \) and \( \sigma^2_i, i = 1, \ldots, \ell \) are eliminated (by integration) in the model given by (1.1), (1.2) and (1.3); all the components of \( \phi \) are fixed but unknown. In Section 2, after obtaining the point estimators for \( \phi = (\theta, \tau, \eta, \delta)' \), we develop a two-stage empirical Bayes prediction
interval of $\gamma(Y_\ell)$ and, using asymptotic theory, we compare it with the HPD interval in (1.5). In Section 3 we describe a small-scale numerical study to assess the properties of the EB interval for moderate sample sizes, and compare the EB interval with the interval based only on the data from the $\ell$th area. Section 4 has concluding remarks, and sketches of proofs are given in the appendix.

2. Empirical Bayes Prediction Interval

First, we construct point estimators for $\phi$ and investigate their asymptotic properties. We assume throughout that $2 \leq \inf_{i \geq 1} n_i \leq \sup_{i \geq 1} n_i \leq k < \infty$. (This assumption is realistic in many applications including small area estimation, and it is used mainly for theoretical reasons.) Then, we obtain the EB prediction interval and study its asymptotic properties. All results are obtained under the joint marginal distributions of the $Y_{ij}$. In particular, Lemma 1, Lemma 2, Theorem 1 and Corollary 1 are proved after integrating out the $\mu_i$ and the $\sigma^2_i$.

Also, the abbreviations almost everywhere (a.e.) and almost surely (a.s.) refer to the probability measure associated with the joint marginal distributions of the $Y_{ij}$.

2.1. Parametric point estimators and asymptotic properties

We construct point estimators of $\phi$ and provide their relevant asymptotic properties.

We note that the bivariate statistics $(\bar{Y}_i, S^2_i)$ are independently distributed over $i$ but individually $\bar{Y}_i$ and $S^2_i$ are not independently distributed for each $i$. Moreover, for $i = 1, \ldots, \ell$, it is easy to show that

\[
\{(1 - \omega_i)/(1 - \eta^{-1})\delta \tau\}^{1/2}(\bar{Y}_i - \theta) \sim t_{2\eta} \quad \text{and} \quad \eta S^2_i/(\eta - 1)\delta \sim F(n_i - 1, 2\eta),
\]

where $t_{2\eta}$ is a student $t$ distribution on $2\eta$ degrees of freedom and $F(n_i - 1, 2\eta)$ is an $f$-distribution on $(n_i - 1, 2\eta)$ degrees of freedom. We use (2.1) to obtain and study point estimators of $\phi$.

First, assuming that $\tau, \delta$ and $\eta$ are known, we construct an estimator of $\theta$. Using (2.1) it is easy to show that the best linear unbiased estimator of $\theta$ is $\hat{\theta}_s$ where

\[
\hat{\theta}_s = \frac{\ell}{\sum_{i=1}^{\ell}(1 - \omega_i)\bar{Y}_i/\sum_{i=1}^{\ell}(1 - \omega_i)}.
\]

Note that $\hat{\theta}_s$ depends only on $\tau$. Letting $n_T = \sum_{i=1}^{\ell} n_i$, we have

\[
\hat{\delta} = WMS = (n_T - \ell)^{-1}\sum_{i=1}^{\ell}(n_i - 1)S^2_i
\]
which is an unbiased estimator of $\delta$. Now, letting

$$
\bar{Y} = n_T^{-1} \sum_{i=1}^\ell n_i \bar{Y}_i \quad \text{and} \quad BMS = (\ell - 1)^{-1} \sum_{i=1}^\ell n_i (\bar{Y}_i - \bar{Y})^2,
$$

and using arguments similar to ones in Ghosh and Meeden (1986), an estimator of $\tau$ is

$$
\hat{\tau} = \max(0, \hat{\tau}^*),
$$

(2.4)

where

$$
\hat{\tau}^* = (\ell - 1) \left[ n_T - n_T^{-1} \sum_{i=1}^\ell n_i^2 \right]^{-1} \{ BMS/WMS - 1 \}, \ell > 1.
$$

(2.5)

Formula (2.5) differs from that in (2.8) of Ghosh and Meeden (1986) because the adjustment they made seems unnecessary in our case. (Their estimator has substantial positive bias for small $\ell$.) Thus we use the estimator

$$
\hat{\tau} = \left\{ \begin{array}{ll}
\frac{\sum_{i=1}^\ell (1 - \hat{\omega}_i) \bar{Y}_i}{\sum_{i=1}^\ell (1 - \hat{\omega}_i)} & , \hat{\tau} > 0 \\
\bar{Y} & , \hat{\tau} = 0
\end{array} \right.
$$

where $\hat{\omega}_i = (1 + n_i \hat{\tau})^{-1}$. We need a separate estimator when $\hat{\tau} = 0$ because in this case $\hat{\omega}_i = 1, i = 1, \ldots, \ell$, and $\sum_{i=1}^\ell (1 - \hat{\omega}_i) \bar{Y}_i / \sum_{i=1}^\ell (1 - \hat{\omega}_i)$ is indeterminate. The second estimator is sensible because $\lim_{\hat{\tau} \to 0} \sum_{i=1}^\ell (1 - \hat{\omega}_i) \bar{Y}_i / \sum_{i=1}^\ell (1 - \hat{\omega}_i) = \bar{Y}$.

Next, we construct an estimator for $\eta$. Consider $\sum_{i=1}^\ell h_i (S_i^2 - \hat{\delta})^2$ where $h_i = (n_i - 1)(n_T - \ell)^{-1}, i = 1, \ldots, \ell$ and $\hat{\delta}$ is defined in (2.3). Then it is easy to show that

$$
E\{ \sum_{i=1}^\ell h_i (S_i^2 - \hat{\delta})^2 \} = \delta^2 (n_T - \ell)^{-1} \sum_{i=1}^\ell (1 - h_i) \{ 2 + (n_i + 1)/(\eta - 2) \}, \eta > 2.
$$

Thus, as an estimator of $\eta$ we consider

$$
\hat{\eta} = 2 + \{ \max(\ell^{-1}, \hat{\eta}^{-1}) \}^{-1},
$$

(2.6)

where

$$
\hat{\eta}^{-1} = \left\{ \sum_{i=1}^\ell h_i (1 - h_i) + 2(\ell - 1)(n_T - \ell)^{-1} \right\}^{-1}
$$

$$
\left\{ \left[ \sum_{i=1}^\ell h_i (S_i^2 \hat{\delta}^{-1} - 1) \right]^2 - 2(\ell - 1)(n_T - \ell)^{-1} \right\}.
$$

Again it is necessary to have a truncated estimator of the form $\hat{\eta}$ in (2.6) because $\hat{\eta}$ could be negative or indeterminate.
For convenience we present in Lemma 1 asymptotic properties of the point estimators of $\hat{\phi} = (\hat{\theta}, \hat{\delta}, \tau, \eta)^T$.

**Lemma 1.** Assume $2 \leq \inf_{i \geq 1} n_i \leq \sup_{i \geq 1} n_i \leq k < \infty$. Then as $\ell \to \infty$

(a) $\hat{\delta} \overset{a.s.}{\to} \delta$ and $E(\hat{\delta} - \delta)^2 \to 0$, (b) $\hat{\tau} \overset{a.s.}{\to} \tau$ and $\max_{i=1,2,...,\ell} |\hat{\omega}_i - \omega_i| \overset{a.s.}{\to} 0$, (c) $\hat{\eta} \overset{a.s.}{\to} \eta$, (d) $\hat{\theta} \overset{a.s.}{\to} \theta$.

**Proof.** A sketch of the proof is presented in Appendix A.

### 2.2. Empirical Bayes interval and asymptotic properties

Suppose $\hat{\phi}$ is known. Then under the marginal distribution of the $Y_{ij}$

$$ (\gamma(Y_\ell) - e_B)/\nu_B \sim t_\kappa, \tag{2.7} $$

where $e_B$ and $\nu_B$ are given in (1.5). The two-stage procedure we consider in this section will help in a simple way to reduce the underestimation well known to be associated with naive EB methods.

At the first stage we assume $\tau, \delta, \eta$ are known, and consider the pivotal quantity

$$ (\gamma(Y_\ell) - \hat{e}_B^*)/\nu_{\hat{\theta}_*}, \tag{2.8} $$

where

$$ \hat{e}_B^* = \overline{Y}_\ell - (1 - f_\ell)\omega_\ell(\overline{Y}_\ell - \hat{\theta}_*), $$

$$ \nu_{\hat{\theta}_*}^2 = (1 - 2\kappa^{-1})\text{var}\{\gamma(Y_\ell) - \hat{e}_B^*\} = \nu_B^2 + \nu_{\hat{\theta}_*}^2, $$

$$ \nu_{\hat{\theta}_*}^2 = (1 - 2\kappa^{-1})\delta(1 - f_\ell)^2\omega_\ell^2/\sum_{i=1}^{\ell} n_i \omega_i, $$

and $\hat{\theta}_*$ is given by (2.2), $\nu_{\hat{\theta}_*}^2$ by (1.5). Acting as if the pivotal quantity in (2.8) has a Student $t$ distribution with $\kappa$ degrees of freedom, an approximate $100(1 - \alpha)$% confidence interval for $\gamma(Y_\ell)$ is

$$ \hat{e}_B^* \pm \nu_{\hat{\theta}_*}t_{\kappa,\alpha/2}, \tag{2.9} $$

where $t_{\kappa,\alpha/2}$ is the $100(1 - \alpha/2)$th percentile point of the Student $t$ distribution.

Note that there are two adjustments being made here to account for underestimation of variability. First, $\nu_{\hat{\theta}_*}^2$ in (2.8) contains an additional variability $\nu_{\hat{\theta}_*}^2$ beyond $\nu_B^2$. Second, to be on a par with the HPD interval, the $100(1 - \alpha)$% in (2.9) is obtained from the student $t$ distribution and not from the normal distribution.

At the second stage we substitute estimators $\hat{\tau}, \hat{\delta}, \hat{\eta}$ from (2.3), (2.4), (2.6) into (2.9) to obtain the proposed $100(1 - \alpha)$% EB prediction interval for $\gamma(Y_\ell)$ as

$$ \hat{e}_B \pm \hat{\nu}_B t_{\kappa,\alpha/2}. \tag{2.10} $$
In (2.10)
\[
\hat{e}_B = \overline{Y}_\ell - (1 - f_\ell)\hat{\omega}_\ell(\overline{Y}_\ell - \hat{\theta}), \quad \hat{\nu}_B^2 = \hat{\nu}_B^2 + \hat{\nu}_{\hat{\theta}_*}^2,
\]
\[
\hat{\nu}_B^2 = (1 - 2\hat{\kappa}^{-1})(1 - f_\ell)\{f_\ell + (1 - f_\ell)(1 - \hat{\omega}_\ell)}\hat{\sigma}_\ell^2 n_\ell^{-1},
\]
\[
\hat{\nu}_{\hat{\theta}_*}^2 = (1 - 2\hat{\kappa}^{-1})\hat{\delta}(1 - f_\ell)^2 \hat{\omega}_\ell^2 / \sum_{i=1}^{\ell} n_i \hat{\omega}_i,
\]
where
\[
\hat{\sigma}^2 = \{(n_\ell - 1)S_\ell^2 + n_\ell \hat{\omega}_\ell(\overline{Y} - \hat{\theta})^2 + 2(\hat{\eta} - 1)\hat{\kappa}^{-1} \text{ and } \hat{\kappa} = n_\ell + 2\hat{\eta}.
\]

Next, we consider how well the EB interval (2.10) approximates the HPD interval (1.5) as $\ell \to \infty$. As a basis of the asymptotics, using the joint marginal distribution of the $Y_{ij}$ obtained from (1.1), (1.2) and (1.3), we compare the centers, widths, and coverage probabilities of the two intervals.

Our approach can be motivated in the following way: Suppose instead of the gamma distribution in (1.3) we pretend that the $\sigma_i^2$ are fixed but unknown, and that $2(\hat{\eta} - 1)\hat{\delta}\sigma_\ell^{-2}$ has a chi-square distribution on $2\hat{\eta}$ degrees of freedom independently of $S_\ell^2$, $\overline{Y}_\ell$ and $\gamma(Y_\ell) - e_B$. Then, under the marginal distributions of the $y_{ij}$, independently $\{\gamma(Y_\ell) - e_B\}((1 - f_\ell)\{f_\ell + (1 - f_\ell)(1 - \omega_\ell)}\sigma_\ell^2 n_\ell^{-1})^{-1/2} \sim N(0, 1)$ and $\kappa\hat{\sigma}_\ell^2/\sigma_\ell^2 \sim \chi_\kappa^2$. Consequently, $(\gamma(Y_\ell) - e_B)/\nu_B \sim t_\kappa$, and the $100(1 - \alpha)$% shortest prediction interval (non-Bayesian) for $\gamma(Y_\ell)$ is $e_B \pm \nu_B t_{\kappa, \alpha/2}$, the same as the HPD interval in (1.5). In fact, if (1.3) is removed from the model and the $\sigma_i^2$ are kept fixed but unknown, the frequentist prediction interval is exactly the HPD prediction interval. Results in which the HPD intervals are exactly the same as the frequentist intervals are familiar; see, for example, Box and Tiao (1992), Sec. 2.2.

Finally, note that $Pr(e_B - \nu_B t_{\kappa, \alpha/2} \leq \gamma(Y_\ell) \leq e_B + \nu_B t_{\kappa, \alpha/2}) = EY_{\ell_0}(Pr(e_B - \nu_B t_{\kappa, \alpha/2} \leq \gamma(Y_\ell) \leq e_B + \nu_B t_{\kappa, \alpha/2}|Y_\ell) = EY_{\ell}(1 - \alpha) = 1 - \alpha$. That is, rather interestingly, the probability content of the $100(1 - \alpha)$% HPD interval is $(1 - \alpha)$ when the probability measure is based on the joint marginal distribution of the $Y_\ell$. Thus, since the HPD interval is optimal, it is sensible to demonstrate that the EB interval is a good approximation to the HPD interval.

Let $W$ denote the width and $P$ the coverage probability of an interval. Then $W_B = 2t_{\kappa, \alpha/2}\nu_B$ and $\hat{W}_B = 2\hat{t}_{\hat{\kappa}, \alpha/2}\hat{\nu}_B$. Also letting
\[
\hat{P}_B = T\{(\hat{e}_B - e_B + t_{\kappa, \alpha/2}\hat{\nu}_B)\nu_B^{-1}\} - T\{(\hat{e}_B - e_B - t_{\kappa, \alpha/2}\hat{\nu}_B)\nu_B^{-1}\},
\]
the coverage probability of the EB interval is $P_B = E_Y(\hat{P}_B)$ where expectation is taken over the joint marginal distribution of $Y_{ij}$ obtained from (1.1), (1.2) and (1.3) and $T(\cdot)$ is the cumulative distribution function of a Student $t$ on $2\hat{\eta}$ degrees
of freedom. The centers of the two intervals are $e_B$ and $\hat{e}_B$, which are the Bayes and the empirical Bayes estimators of $\gamma(Y_\ell)$ respectively.

First, we present Lemma 2.

**Lemma 2.** Assume $2 \leq \inf_{i \geq 1} n_i \leq \sup_{i \geq 1} n_i \leq k < \infty$. Then as $\ell \to \infty$

(a) $\hat{\nu}^2_{\hat{\theta}_i} \xrightarrow{a.s.} 0$ and $E(\hat{\nu}^2_{\hat{\theta}_i}) \to 0$, (b) $\hat{\nu}_B - \nu_B \xrightarrow{a.s.}$, (c) $E | \hat{\nu}_B - \nu_B | \to 0$.

**Proof.** A sketch of the proof is presented in Appendix C and Appendix D.

Theorem 1 gives a neat summary of our main results and it establishes that for a large number of small areas the EB interval is expected to be approximately the same as the HPD interval for the finite population mean. A key idea in the proof of Theorem 1 is the use of uniform integrability (see Serfling (1980), Sec. 1.4.)

**Theorem 1.** Assume $2 \leq \inf_{i \geq 1} n_i \leq \sup_{i \geq 1} n_i \leq k < \infty$. Then as $\ell \to \infty$

(a) $E(\hat{e}_B - e_B) \to 0$, (b) $E(\hat{W}_B - W_B) \to 0$, (c) $E(\hat{P}_B) \to 1 - \alpha$.

**Proof.**

(a) Since $E(\hat{e}_B - e_B) \leq \{E((\hat{e}_B - e_B)^2)^{1/2}$, we show that $\hat{e}_B - e_B \xrightarrow{a.s.} 0$ as $\ell \to \infty$ and $(\hat{e}_B - e_B)^2$ is uniformly integrable; see Serfling (1980), Sec. 1.4. Because $|\hat{e}_B - e_B| \leq |\hat{\theta} - \theta| + |\hat{Y}_\ell - \theta| \max_{i=1,2,...,\ell} |\hat{\omega}_i - \omega_i|$ and $|\hat{Y}_\ell - \theta|$ is finite a.e., by Lemma 1 (b) and (d), $\hat{e}_B - e_B \xrightarrow{a.s.} 0$ as $\ell \to \infty$. Appendix B shows that $(\hat{e}_B - e_B)^2$ is uniformly integrable.

(b) It is easy to show

$$E|\hat{W}_B - W_B| \leq 2E[t_{2k,\alpha/2}||\hat{\nu}_B - \nu_B|| + 2\nu_B E |t_{k,\alpha/2} - t_{k,\alpha/2}|].$$

By using Lemma 1 and since $t_{a,\alpha/2}$ is continuous in $a$ for any positive real number $a$, $t_{k,\alpha/2} \xrightarrow{a.s.} t_{k,\alpha/2}$ as $\ell \to \infty$. But since $t_{k,\alpha/2} \leq t_{4,\alpha/2} = A < \infty, t_{k,\alpha/2} - t_{k,\alpha/2}$ is uniformly bounded and $E(t_{k,\alpha/2} - t_{k,\alpha/2}) \to 0$. Thus, by Lemma 2(c), $E|\hat{W}_B - W_B| \to 0$ as $\ell \to \infty$.

(c) By Lemma 1(c), $\hat{P}_B \xrightarrow{a.s.} T(t_{k,\alpha/2}) - T(-t_{k,\alpha/2}) = 1 - \alpha$ and, since $\hat{P}_B$ is uniformly bounded, $E(\hat{P}_B) \to 1 - \alpha$ as $\ell \to \infty$.

Finally, we present Corollary 1. The Bayes risk of any estimator $e$ of $\gamma(Y_\ell)$ under squared error loss, $r(e)$, is $r(e) = E_Y \{e - \gamma(Y_\ell)\}^2$, where expectation is taken over the marginal distribution of $Y$ obtained from (1.1), (1.2) and (1.3). As in Lemma 3 of Ghosh and Meeden (1986), we have $r(e) = r(e_B) = E(e - e_B)^2$.

**Corollary 1.** Under the conditions of Theorem 1, $r(\hat{e}_B) - r(e_B) \to 0$ as $\ell \to \infty$.

The proof follows immediately from Theorem 1(a).

Corollary 1 shows that $\hat{e}_B$ is asymptotically optimal in the sense of Robbins (1955). This adds credence to the center of the EB interval as an approximation to the center of the HPD interval.
3. A Small-Scale Numerical Study

We investigate the properties of the EB interval in (2.10) relative to the HPD interval in (1.5) by performing a small-scale Monte Carlo study when the model given by (1.1), (1.2) and (1.3) holds. We select repeated samples from

\[ Y_{ij} = \theta + \mu_i + \epsilon_{ij} \quad (3.1) \]

\( i = 1, \ldots, \ell \) and \( j = 1, \ldots, n_i \), where the \( \mu_i \) are drawn independently from \( N(0, \tau \sigma_i^2) \), independently for each \( i, \epsilon_{i1}, \epsilon_{i2}, \ldots, \epsilon_{in_i} \) is a random sample from \( N(0, \sigma_i^2) \), while \( \sigma_1^2, \ldots, \sigma_\ell^2 \) is a random sample from \( IG\{\eta, (\eta - 1)\delta \} \). Random deviates are generated using RNOOA- and RNGAM-generating functions of the IMSL library.

Throughout we take \( f = .05, \ell = 20, 30, 40 \) and \( \eta = 5, 10, 15 \). We restrict the simulations to the case in which the \( \sigma_i^2 \) have unit variance (i.e., \( \delta = (\eta - 2)^{1/2} \)). Now observing that the distributions of \( \bar{Y}_i - \theta \) are invariant to choices of \( \theta \), apart from the estimated center, the sampling distributions of all quantities we study are invariant to choices of \( \theta \). Thus, for convenience, we take \( \theta = 10 \). Also we take \( n_\ell = 5, 10, 15 \) while, using RNGDA of the IMSL library, the \( n_i, i = 1, 2, \ldots, \ell - 1 \), are drawn at random from a discrete uniform on \( (2, n_\ell) \). To ensure a large range of degrees of borrowing (i.e., values of \( \omega_\ell \)), we take \( \tau = 0.05, 0.25, 1.25 \); see Table 1. Thus there are 81 combinations of \( \ell, n_\ell, \tau \) and \( \eta \).

Table 1. Values of \( \omega_\ell \) (i.e., degree of borrowing) by \( \tau \) and \( n_\ell \)

<table>
<thead>
<tr>
<th>( n_\ell )</th>
<th>( \tau )</th>
<th>0.05</th>
<th>0.25</th>
<th>1.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.80</td>
<td>0.44</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.67</td>
<td>0.29</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.57</td>
<td>0.21</td>
<td>0.05</td>
<td></td>
</tr>
</tbody>
</table>

NOTE: The degree of borrowing \( (\omega_\ell = (1 + n_\ell \tau)^{-1}) \) depends on \( \ell \) only through \( n_\ell \), and smaller \( \tau \) means larger degree of borrowing.

We study the width, center and coverage probability of the EB interval relative to the Bayes interval with nominal coverage probability of .95. We compute the expected width, center, and coverage probability of the EB interval over repeated samples at each combination of \( (\ell, n_\ell, \tau, \eta) \). For \( G \) replications of the Monte Carlo experiment, we estimate the expected value of the width of an EB interval by computing the sample mean of the \( G \) replications

\[ \bar{W}_{EB} = G^{-1} \sum_{i=1}^G \bar{W}_{EB}^{(i)}, \]
and its standard error

\[ SE_{EB} = \left\{ \sum_{i=1}^{G} (\hat{W}_{EB}^{(i)} - W_{EB})^2 / G(G-1) \right\}^{1/2}, \]  

(3.2)

where \( \hat{W}_{EB}^{(i)} \) is the width of the EB interval for the \( i \)th replication of the Monte Carlo experiment. As in (3.2), for the coverage probability of the EB interval we compute \( \bar{P}_{EB} \) and for the center we compute \( \bar{e}_{EB} \). (We take \( G = 10,000 \).)

Over all choices of \((\ell, n_{\ell}, \tau, \eta)\), the standard error of \( \bar{P}_{EB} \) is at most 0.001, the standard error of \( \bar{e}_{EB} \) is at most 0.005, and the standard error of \( \bar{e}_{EB} \) ranges from 0.003 to 0.022 with first, second and third quartiles 0.004, 0.008 and 0.019. With this precision we study all quantities of interest.

Next, we compare the risk of \( \hat{e}_B \) with \( e_B \) by using \( R(EB, B) = \tau(\hat{e}_B)/\tau(e_B) - 1 \). Among the 81 values the first, second and third quartiles of \( R(EB, B) \) are 0.003, 0.013 and 0.046. For all values of \( \ell \), when \( \tau = 1.25, R(EB, B) \) ranges from 0.000 to 0.005; when \( \tau = .25, R(EB, B) \) has quantiles 0.010, 0.013 and 0.038; when \( \tau = .05, R(EB, B) \) has quantiles 0.044, 0.060 and 0.084. Of course, as \( \ell, n_{\ell}, \tau \) or \( \eta \) increases, \( R(EB, B) \) decreases. Further, we compute estimates of \( |E(\hat{e}_B - e_B)/SE(\hat{e}_B - e_B)| \). The first, second and third quartiles of these estimates are 0.005, 0.007 and 0.013, and the smallest and largest values are 0.000 and 0.025.

We also compute estimates of \( |E(\hat{W}_B - W_B)/SE(\hat{W}_B - W_B)| \). The first, second and third quartiles are 0.143, 0.262 and 0.351 with the smallest and largest values 0.000 and 0.495, larger values occurring at larger values of \( \eta \), but these decrease as \( \ell, n_{\ell} \) or \( \tau \) increases. Thus, in terms of these measures the EB interval performs well when compared with the HPD interval for almost all combinations.

For comparison we consider a third interval based on the data from only the \( \ell \)th area. This prediction interval, denoted by \( I_o \), for the finite population mean of the \( \ell \)th area is

\[ \bar{Y}_\ell \pm \frac{S_{\ell}}{\sqrt{n_{\ell}}} (1 - f_{\ell})^{1/2} t_{n_{\ell} - 1, \alpha/2}, \]  

(3.3)

where \( t_{n_{\ell} - 1, \alpha/2} \) is the 100(1 - \( \alpha \)/2)th percentile point of the Student t distribution on \( n_{\ell} - 1 \) degrees of freedom. By considering the estimated width, center and coverage probabilities we compare the EB, HPD and \( I_o \) intervals. (For the interval \( I_o \) we compute the estimated width, center, and coverage probability under our general model.)

Let \( R_w(I_o, B) = W_{I_o}/W_B \) denote the ratio of the widths of the \( I_o \) interval to the Bayes interval with similar meanings for \( R_w(EB, B) \), \( R_c(I_o, B) \) and \( R_c(EB, B) \) where \( EB, B \) and \( I_o \) denote the EB interval in (2.10), the HPD interval in (1.5) and the interval in (3.3) respectively. We found that \( R_c(I_o, B) \approx 1.000 \) and \( R_c(EB, B) \approx 1.000 \) for all combinations of \( \ell, n_{\ell}, \tau \) and \( \eta \). In Table 2 we
present the values for \( R_w(I_o, B) \) and \( R_w(EB, B) \) in columns 4 and 5. We also present the estimated coverage probabilities in columns 6 for the \( I_o \) interval, and in column 7 for the \( EB \) interval. Recall that coverage is computed using the average of the coverage probabilities over the replications.

Table 2. Ratios of estimated widths of \( I_o \) and \( EB \) intervals to the HPD interval and coverage probabilities of nominal 95% \( I_o \) and \( EB \) intervals by \( \ell \), \( n_{\ell} \) and \( \tau \)

<table>
<thead>
<tr>
<th>( \ell )</th>
<th>( n_{\ell} )</th>
<th>( \tau )</th>
<th>( R_w(I_o, B) )</th>
<th>( R_w(EB, B) )</th>
<th>( I_o ) Coverage Probabilities</th>
<th>( EB ) Coverage Probabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>5</td>
<td>0.05</td>
<td>2.841</td>
<td>1.056</td>
<td>0.952</td>
<td>0.888 (0.923)</td>
</tr>
<tr>
<td>20</td>
<td>0.25</td>
<td>1.831</td>
<td>0.951</td>
<td>0.955</td>
<td>0.895 (0.938)</td>
<td></td>
</tr>
<tr>
<td>1.25</td>
<td>1.493</td>
<td>0.997</td>
<td>0.958</td>
<td>0.943 (0.946)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>0.05</td>
<td>1.917</td>
<td>0.965</td>
<td>0.953</td>
<td>0.866 (0.925)</td>
</tr>
<tr>
<td>0.25</td>
<td>1.360</td>
<td>0.968</td>
<td>0.957</td>
<td>0.918 (0.942)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.25</td>
<td>1.204</td>
<td>0.998</td>
<td>0.959</td>
<td>0.946 (0.947)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>15</td>
<td>0.05</td>
<td>1.631</td>
<td>0.940</td>
<td>0.953</td>
<td>0.861 (0.931)</td>
</tr>
<tr>
<td>0.25</td>
<td>1.233</td>
<td>0.976</td>
<td>0.956</td>
<td>0.928 (0.944)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.25</td>
<td>1.130</td>
<td>0.998</td>
<td>0.958</td>
<td>0.947 (0.948)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>5</td>
<td>0.05</td>
<td>2.846</td>
<td>1.032</td>
<td>0.953</td>
<td>0.883 (0.923)</td>
</tr>
<tr>
<td>0.25</td>
<td>1.834</td>
<td>0.959</td>
<td>0.956</td>
<td>0.902 (0.941)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.25</td>
<td>1.496</td>
<td>1.005</td>
<td>0.959</td>
<td>0.946 (0.948)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>10</td>
<td>0.05</td>
<td>1.917</td>
<td>0.953</td>
<td>0.953</td>
<td>0.867 (0.935)</td>
</tr>
<tr>
<td>0.25</td>
<td>1.360</td>
<td>0.985</td>
<td>0.957</td>
<td>0.933 (0.946)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.25</td>
<td>1.204</td>
<td>1.004</td>
<td>0.959</td>
<td>0.949 (0.950)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>15</td>
<td>0.05</td>
<td>1.631</td>
<td>0.933</td>
<td>0.954</td>
<td>0.871 (0.938)</td>
</tr>
<tr>
<td>0.25</td>
<td>1.233</td>
<td>0.990</td>
<td>0.957</td>
<td>0.940 (0.947)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.25</td>
<td>1.130</td>
<td>1.002</td>
<td>0.958</td>
<td>0.949 (0.950)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>5</td>
<td>0.05</td>
<td>2.848</td>
<td>1.016</td>
<td>0.952</td>
<td>0.880 (0.932)</td>
</tr>
<tr>
<td>0.25</td>
<td>1.836</td>
<td>0.973</td>
<td>0.956</td>
<td>0.914 (0.944)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.25</td>
<td>1.497</td>
<td>1.009</td>
<td>0.959</td>
<td>0.948 (0.949)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>10</td>
<td>0.05</td>
<td>1.920</td>
<td>0.938</td>
<td>0.952</td>
<td>0.863 (0.936)</td>
</tr>
<tr>
<td>0.25</td>
<td>1.363</td>
<td>0.986</td>
<td>0.957</td>
<td>0.935 (0.946)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.25</td>
<td>1.207</td>
<td>1.004</td>
<td>0.960</td>
<td>0.949 (0.950)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>15</td>
<td>0.05</td>
<td>1.629</td>
<td>0.931</td>
<td>0.953</td>
<td>0.871 (0.939)</td>
</tr>
<tr>
<td>0.25</td>
<td>1.231</td>
<td>0.993</td>
<td>0.956</td>
<td>0.943 (0.947)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.25</td>
<td>1.129</td>
<td>1.004</td>
<td>0.958</td>
<td>0.950 (0.950)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: Here \( R_w(I, B) = W_I/W_B \) denotes the ratio of the widths for interval \( I \) versus the HPD interval, and \( I \) is either \( I_o \) or \( EB \). (Median coverage probabilities for the EB interval are in parentheses.)
Because the distributions of these estimated probabilities are skewed to the left for smaller values of $\tau$, for the $EB$ interval we also present the median (in parentheses) in column 7. However, the distributions become more symmetric as $\tau$ increases. That is, for small values of $\tau$ (heavy pooling) the mean is an underestimate and the median is preferred. Finally, note that we present results for only $\eta = 5$ because all quantities either decrease very slowly or remain steady as $\eta$ increases from 5 to 15.

For all values of $\ell, n_\ell$ and $\tau$, the coverage probabilities of the $I_o$ interval are larger than the nominal 95% value, and they are always wider than the HPD interval. The shortest $I_o$ interval has $R_w(I_o, B) = 1.114$, at $\ell = 40, n_\ell = 15, \tau = 1.25$ and $\eta = 15$. Thus, on the average, the centers and the coverage probabilities of the $I_o$ interval are closer to the centers of the HPD interval than the widths are. In terms of coverage the $I_o$ interval is conservative.

The widths of the $EB$ intervals are very close to the HPD interval for all values $\ell, n_\ell$ and $\tau$. The coverage probabilities of the EB interval, as measured by the sample median, are very close to the HPD interval. As expected, the EB interval gets closer to the HPD interval as $\ell, n_\ell$ or $\tau$ increases. While there are small changes as $\tau$ increases, it seems that the EB interval performs better in terms of coverage probabilities for larger values of $\tau$. That is, when there is a small to moderate amount of borrowing among the small areas, the EB interval performs just fine. This is precisely the situation in which the EB interval is required to work well as our experience suggests too much borrowing is not a good practice. Even for small values of $\ell, n_\ell$ and $\tau$, the EB interval is remarkably close to the HPD interval. Further simulations for larger values of $\ell$ showed that the EB interval is again a good approximation to the HPD interval.

4. Concluding Remarks

We provide a simple prediction interval for the finite population mean of a small area when data are available from a large number of areas, say at least about 20. It will be useful for practitioners who prefer a less sophisticated method or who need quick answers. Because the properties of the interval are based on the multivariate Student distribution, it should be more robust to outliers than the interval provided by Nandram and Sedransk (1993) in Sections 3 and 4 of their paper.

We have shown that the EB interval works fairly well for reasonable values of $\ell$ and $n_\ell, i = 1, \ldots, \ell$ when pooling data from these areas is reasonable. Our numerical study shows that if the model holds, the EB interval is a reasonable approximation to the HPD interval, and much better than the interval based on only the data of a particular small area. In particular, the width and center of
the EB interval are comparable to the HPD interval, and its coverage probability
is near the nominal 95% value for \( \ell > 20, n_i > 5 \) and \( \tau > .25 \).

We used a simple method to partially take care of the underestimation in
variability while Prasad and Rao (1990) used a more sophisticated method to
provide asymptotically valid confidence interval for a small-area mean. Fortu-
nately, extensive theoretical and empirical investigations demonstrate that our
prediction intervals have good coverage properties. We believe that there is a
need for more exploration in small area interval estimation when empirical Bayes
methods are used, but see Carlin and Gelfand (1990). If more sophistication is
brought in, one might prefer to use a full Bayesian approach that would require
use of a method such as the Gibbs sampler (Gelfand and Smith 1990).

One can also construct a prediction interval for the \((\ell+1)\)th area that has not
been sampled. Thus, assume observations are obtained from \( \ell \) small areas, all \( \ell+1 \) areas follow (1.1), (1.2) and (1.3), and interest is in \( \gamma(Y_{\ell+1}) = \sum_{j=1}^{N_{\ell+1}} Y_{\ell+1,j}/N_{\ell+1} \) where \( N_{\ell+1} \geq 1 \) is the size of the small area. Then the 100(1-\( \alpha \))% HPD prediction
interval for \( \gamma(Y_{\ell+1}) \) is

\[
\theta \pm \{\delta(1 - \eta^{-1})(N_{\ell+1}^{-1} + \tau)\}^{1/2} t_{2\eta, \alpha/2}.
\]

As an approximation to the HPD interval, the EB interval is

\[
\hat{\theta} \pm \{\hat{\delta}(1 - \hat{\eta}^{-1})(N_{\ell+1}^{-1} + \hat{\tau} + 1/\sum_{i=1}^{\ell} n_i \hat{\omega}_i)\}^{1/2} t_{2\hat{\eta}, \alpha/2}
\]

and Theorem 1 still holds.

Acknowledgement

This research was supported in part by a grant from the Worcester Polytech-
nic Research Development Council. The author is grateful to the referee and the
Associate Editor for their comments.

Appendix: Completion of Proofs

Appendix A: Proof of Lemma 1

**Lemma 1 (a):** Since \( 2 \leq \inf_{i \geq 1} n_i \leq \sup_{i \geq 1} n_i \leq k < \infty \) and \( \text{var}(S_i^2) = [(2 + (n_i + 1)/(\eta - 2))]^2/(n_i - 1) \leq [(k + 1)/(\eta - 2) + 2]\delta^2 = A \), by the Kolmogorov
strong law of large numbers (SLLN), \( \hat{\delta} \overset{a.s.}{\to} \delta \); see Serfling (1980), pg. 27. Also, since \( E(\hat{\delta} - \delta)^2 \leq A\ell^{-1}, E(\hat{\delta} - \delta)^2 \to 0 \) as \( \ell \to \infty \).

**Lemma 1 (b):** By using (a) and applying the SLLN to each term in \( BMS = (\ell - 1)^{-1}\{\sum_{i=1}^\ell n_i Y_i^2 - n\overline{Y}^2\} \), we have \( \hat{\tau} \overset{a.s.}{\to} \tau \). Thus, by continuity, \( \max(0, \hat{\tau}) \overset{a.s.}{\to} \tau \).
as \( \ell \to \infty \). Also since \( |\hat{\omega}_i - \omega_i| \leq 1 - \hat{\tau}^{-1} \), \( i = 1, 2, \ldots, \ell, \max_{i=1,2,\ldots,\ell} |\hat{\omega}_i - \omega_i| \to 0 \) as \( \ell \to \infty \).

**Lemma 1 (c):** We show \( \hat{\eta}_n^{-1} \overset{a.s.}{\to} (\eta - 2)^{-1} \) as \( \ell \to \infty \). First note that

\[
\hat{\eta}_n^{-1} = \delta^2 \hat{\delta}^{-2} \{ H_1(\delta, \hat{\delta}) - H_2(\delta, \hat{\delta}) + (\eta - 2)^{-1} \}, \tag{A.1}
\]

where

\[
H_1(\delta, \hat{\delta}) = \delta^{-2} \{ \sum_{i=1}^{\ell} h_i(S_i^2 - \hat{\delta})^2 - E \{ \sum_{i=1}^{\ell} h_i(S_i^2 - \hat{\delta})^2 \} \} (n_T - \ell)(\ell - 1)^{-1} A_\ell, \]

\[
H_2(\delta, \hat{\delta}) = 2(\delta^2 \hat{\delta}^{-2} - 1) A_\ell,
\]

and \( A_\ell = (\ell - 1)/\sum_{i=1}^{\ell} (1 - h_i)(n_i + 1) \). Now both \( A_\ell \) and \( (n_T - \ell)(\ell - 1)^{-1} \) are bounded. It follows by Lemma 1(a) that \( H_2(\delta, \hat{\delta}) \overset{a.s.}{\to} 0 \) and \( \delta^2 \hat{\delta}^{-2} \overset{a.s.}{\to} 1 \) as \( \ell \to \infty \). Thus in (A.1) we only need to show \( \sum_{i=1}^{\ell} h_i(S_i^2 - \hat{\delta})^2 - E \{ \sum_{i=1}^{\ell} h_i(S_i^2 - \hat{\delta})^2 \} \overset{a.s.}{\to} 0 \) as \( \ell \to \infty \).

Now

\[
\sum_{i=1}^{\ell} h_i(S_i^2 - \hat{\delta})^2 - E \{ \sum_{i=1}^{\ell} h_i(S_i^2 - \hat{\delta})^2 \} = Q_1 - Q_2,
\]

where \( Q_1 = \sum_{i=1}^{\ell} h_i(S_i^2 - \hat{\delta})^2 - E(\sum_{i=1}^{\ell} h_iS_i^2) \) and \( Q_2 = \delta^2 - \delta^2 - \text{var}(\hat{\delta}) \). By SLLN, provided that \( \eta > 4, Q_1 \overset{a.s.}{\to} 0 \) as \( \ell \to \infty \). Also by Lemma 1(a), \( Q_2 \to 0 \) as \( \ell \to \infty \). Thus \( \hat{\eta}_n^{-1} \overset{a.s.}{\to} (\eta - 2)^{-1} \) as \( \ell \to \infty \).

Since \( \hat{\eta} = 2 + \{ \max(\ell^{-1}, \hat{\eta}_n^{-1}) \}^{-1} \), by continuity \( \hat{\eta} \overset{a.s.}{\to} \eta \) as \( \ell \to \infty \).

**Lemma 1 (d):** If \( \hat{\tau} < 0, \hat{\theta} - \theta = n_T^{-1} \sum_{i=1}^{\ell} n_i(\hat{Y}_i - \theta) \), and by SLLN, \( \hat{\theta} - \theta \overset{a.s.}{\to} 0 \) as \( \ell \to \infty \).

If \( \hat{\tau} > 0, |\hat{\theta} - \theta| = \{ \sum_{i=1}^{\ell} (1 - \hat{\omega}_i)(\hat{Y}_i - \theta) \}/\sum_{i=1}^{\ell} (1 - \omega_i)(\bar{Y}_i - \theta) \} \leq (k + \hat{\tau}^{-1}) \{ \ell^{-1} \sum_{i=1}^{\ell} (1 - \omega_i)(\bar{Y}_i - \theta) \} + \{ \max_{i=1,2,\ldots,\ell} |\hat{\omega}_i - \omega_i| \} \ell^{-1} \sum_{i=1}^{\ell} |\bar{Y}_i - \theta| \}. By SLLN \( \ell^{-1} \sum_{i=1}^{\ell} (1 - \omega_i)(\bar{Y}_i - \theta) \overset{a.s.}{\to} 0 \) as \( \ell \to \infty \). Since \( E(\ell^{-1} \sum_{i=1}^{\ell} |\bar{Y}_i - \theta|) \leq \delta \tau (1 + k \tau) \) as \( \ell \to \infty \), it follows that \( \ell^{-1} \sum_{i=1}^{\ell} |\bar{Y}_i - \theta| \) is finite, a.e.. Using Lemma 1(b) and assuming \( \hat{\tau} > 0, \hat{\theta} - \theta \overset{a.s.}{\to} 0 \) as \( \ell \to \infty \).

**Appendix B: Proof of the Uniform Integrability of \( (\hat{e}_B - e_B)^2 \)**

Since

\[
(\hat{e}_B - e_B)^2 < 2\{ (\bar{Y}_\ell - \theta)^2 + (\hat{\theta} - \theta)^2 \}, \tag{B.1}
\]

we show that \( (\bar{Y}_\ell - \theta)^2 \) and \( (\hat{\theta} - \theta)^2 \) are both uniformly integrable (u.i.).

First, by (B.1),

\[
(\bar{Y}_\ell - \theta)^2 \overset{d}{=} (1 - \eta^{-1})\delta \tau (1 - \omega_\ell)^{-1} F(1, 2\eta) \tag{B.2}
\]

where \( F(1, 2\eta) \) has an \( f \) distribution. Then by (B.2), recalling \( \sup_{i \geq 1} n_i \leq k < \infty \), \( (\bar{Y}_\ell - \theta)^2 \overset{u.i.}{\to} \delta (k \tau + 1) F(1, 2\eta)/2 \) and, since \( \eta > 1 \), \( (\bar{Y}_\ell - \theta)^2 \) is bounded by a random
variable with finite expectation. Thus \((Y_i - \theta)^2\) is u.i., see Serfling (1980), Sec. 1.4. It follows that \(\ell^{-1} \sum_{i=1}^{\ell} (Y_i - \theta)^2\) is also u.i.

Second,\[
(\hat{\theta} - \theta)^2 \leq \{n_T^{-1} \sum_{i=1}^{\ell} n_i (Y_i - \theta)\}^2 + \{(\sum_{i=1}^{\ell} (1 - \hat{\omega}_i) (Y_i - \theta)/\sum_{i=1}^{\ell} (1 - \hat{\omega}_i)\}^2. \tag{B.3}
\]
Using (B.3) it is easy to show that \((\hat{\theta} - \theta)^2 \leq k^2 \ell^{-1} \sum_{i=1}^{\ell} (Y_i - \theta)^2. \tag{B.4}
\]
Then because \(\ell^{-1} \sum_{i=1}^{\ell} (Y_i - \theta)^2\) is u.i., by (B.4), \((\hat{\theta} - \theta)^2\) is u.i.

**Appendix C: Proof of Lemma 2**

**Lemma 2(a).** Using Lemma 1 and the inequality \(\hat{\nu}_B^2 \leq \hat{\nu}_B^2' \leq 2 \ell, \hat{\nu}_B^2 \stackrel{a.s.}{\rightarrow} 0\) as \(\ell \rightarrow \infty\). Now \(E(\hat{\nu}_B^2) \leq \{E\hat{\nu}_B^2 + \theta \hat{\nu}_B^2\}/2\ell\). Thus by Lemma 1 again, there exists \(A < \infty\) s.t. \(\theta(1 + k\hat{\tau}) < A\) a.e. Thus \(E(\hat{\nu}_B^2) \rightarrow 0\) as \(\ell \rightarrow \infty\).

**Lemma 2(b).** Using the triangle inequality \(|\hat{\nu}_B^2 - \nu_B^2| \leq \hat{\nu}_B^2 + |\hat{\nu}_B^2 - \nu_B^2|\). Thus by Lemma 2(a) it is only required to show that \(\hat{\nu}_B^2 - \nu_B^2 \stackrel{a.s.}{\rightarrow} 0\) as \(\ell \rightarrow \infty\); see Appendix D.

**Lemma 2 (c):** It is easy to show that \(E |\hat{\nu}_B^2 - \nu_B^2| \leq \sqrt{3} \{E(\hat{\nu}_B^2 - \nu_B^2)^2\}^{1/2}\). Since \(E |\hat{\nu}_B^2 - \nu_B^2| \leq \{E(\hat{\nu}_B^2 - \nu_B^2)^2\}^{1/2} + E(\hat{\nu}_B^2),\) by Lemma 2(a) it is only required to show that \(E(\hat{\nu}_B^2 - \nu_B^2)^2 \rightarrow 0\) as \(\ell \rightarrow \infty\); see Appendix D.

**Appendix D: Completion of the Proof of Lemma 2**

We show that \(\hat{\nu}_B^2 - \nu_B^2 \stackrel{a.s.}{\rightarrow} 0\) as \(\ell \rightarrow \infty\) and \(E(\hat{\nu}_B^2 - \nu_B^2)^2 \rightarrow 0\) as \(\ell \rightarrow \infty\).

Letting \(\Delta_1 = \sigma_\ell^2 \max_{i=1,2,...,\ell} |\hat{\omega}_i - \omega_i|\) and \(\Delta_2 = |(1 - 2\hat{\kappa}^{-1})\sigma_\ell^2 - \sigma_\ell^2|\), we have by the triangle inequality, \(|\hat{\nu}_B^2 - \nu_B^2| \leq \Delta_1 + \Delta_2\), and by Minkowski’s inequality \(E(\hat{\nu}_B^2 - \nu_B^2)^2 \leq \{(E(\Delta_1^2))^{1/2} + E(\Delta_2^2)^{1/2}\}^2\).

First we show that \(\Delta_1 \stackrel{a.s.}{\rightarrow} 0\) as \(\ell \rightarrow \infty\) and \(E(\Delta_1^2) \rightarrow 0\) as \(\ell \rightarrow \infty\). As \(\sigma_\ell^2\) has finite expectation, it is bounded a.e.. Thus, by Lemma 1(b), \(\Delta_1 \stackrel{a.s.}{\rightarrow} 0\) as \(\ell \rightarrow \infty\). Also since \(E(\Delta_1^2) \leq AE\{\max_{i=1,2,...,\ell} |\hat{\omega}_i - \omega_i|\}^2\) where \(A < \infty\) and \(\max_{i=1,2,...,\ell} |\hat{\omega}_i - \omega_i|\) \(\leq 1\) (i.e., uniformly bounded), by Lemma 1(b) again, \(E(\Delta_1^2) \rightarrow 0\) as \(\ell \rightarrow \infty\).

Second, we note that

\[
\Delta_2 \leq \sum_{i=1}^{4} A_i \tag{D.1}
\]
where

\[ A_1 = \frac{1}{2} (n\ell - 1) S_\ell^2 \mid \hat{\eta}^{-1} - \eta^{-1} \mid, \]

\[ A_2 = 2n\ell \{(\bar{Y}_\ell - \theta)^2 + (\hat{\theta} - \theta)^2 \} \max_{i=1,2,\ldots,\ell} \mid \hat{\omega}_i - \omega_i \mid, \]

\[ A_3 = 2n\ell \omega \ell \mid \hat{\theta} - \theta \mid \bar{Y}_\ell - \theta + \frac{1}{2} n\ell \omega (\bar{Y} - \theta)^2 \mid \hat{\eta}^{-1} - \eta^{-1} \mid, \]

\[ A_4 = \frac{1}{2} \{(\hat{\eta} - 1)\hat{\delta} + \hat{\eta} \} \mid \hat{\eta}^{-1} - \eta^{-1} \mid + \frac{1}{2} \mid \hat{\delta} - \delta \mid. \]

As \( \bar{Y}_\ell \) and \( S^2_\ell \) are bounded a.e., by Lemma 1 \( A_i \xrightarrow{a.s.} 0 \) as \( \ell \to \infty \), \( i = 1, 2, 3, 4 \), and by (D.1), \( \Delta_2 \xrightarrow{a.s.} 0 \) as \( \ell \to \infty \). Now,

\[ E(\Delta_2^2) \leq 4 \sum_{i=1}^{4} E(A_i^2) \].

(D.2)

By using Minkowski’s inequality, Lemma 1 and boundedness repeatedly, it follows that \( E(A_i^2) \to 0 \) as \( \ell \to \infty \), \( i=1, 2, 3, 4 \), and by (D.2), \( E(\Delta_2^2) \to 0 \) as \( \ell \to \infty \).

References


Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609-2280, U.S.A.
E-mail: balnan@wpi.edu

(Received September 1996; accepted March 1998)