4-1-1999

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

Balgobin Nandram
Worcester Polytechnic Institute, balnan@wpi.edu

Statistica Sinica **9**(1999), 325-343

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 μ_i and variance σ_i^2. Then, given σ_i^2, the μ_i are independently distributed with each μ_i having a normal distribution with mean θ and variance $\sigma_i^2\tau$, and the σ_i^2 are a random sample from an inverse gamma distribution with index η and scale $(\eta-1)\delta$. First, assuming θ, τ, δ and η are fixed and known, we obtain the highest posterior density (HPD) interval for the finite population mean of the ℓth area. Second, we obtain the EB interval by "substituting" point estimators for the fixed and unknown parameters θ, τ, δ and η 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 ℓ.

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 $\underset{\sim}{Y}_\ell$ denote the vector of all values from the ℓth area, $\gamma(\underset{\sim}{Y}_\ell)$ denote the finite population mean of the ℓth area, $\underset{\sim}{Y}_s$ denote the vector of sample values from the ℓth area, and let $I(\underset{\sim}{Y}_s) = [P_L(\underset{\sim}{Y}_s), P_U(\underset{\sim}{Y}_s)]$ represent an interval for $\gamma(\underset{\sim}{Y}_\ell)$ which depends on the sample data $\underset{\sim}{Y}_s$. If $P\{\gamma(\underset{\sim}{Y}_\ell) \in I(\underset{\sim}{Y}_s)\} = 1-\alpha$, then $I(\underset{\sim}{Y}_s)$ is an exact $100(1-\alpha)\%$ prediction interval for $\gamma(\underset{\sim}{Y}_\ell)$. If $P\{\gamma(\underset{\sim}{Y}_\ell) \in I(\underset{\sim}{Y}_s)\}$ is approximately $1-\alpha$, then $I(\underset{\sim}{Y}_s)$ is an approximate $100(1-\alpha)\%$ prediction interval. A $100(1-\alpha)\%$ credible set S of $\gamma(\underset{\sim}{Y}_\ell)$ values is a set such that $P\{\gamma(\underset{\sim}{Y}_\ell) \ \in S|\underset{\sim}{Y}_s\} = 1-\alpha$. The smallest $100(1-\alpha)\%$ set is the $100(1-\alpha)\%$ highest posterior density (HPD) prediction credible set. Moreover, if the posterior density of $\gamma(\underset{\sim}{Y}_\ell)$ 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 $\gamma(\underset{\sim}{Y}_\ell)$ given $\underset{\sim}{Y}_s$. (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 ith area follow the super population model

$$Y_{i1}, \ldots, Y_{iN_i} | \mu_i, \sigma_i^2 \overset{i.i.d.}{\sim} N(\mu_i, \sigma_i^2) \tag{1.1}$$

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

$$\mu_i | \sigma_i^2 \sim N(\theta, \tau\sigma_i^2) \tag{1.2}$$

$i = 1, \ldots, \ell$ 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 σ_i^2, and without loss of generality, for convenience, we reparameterize the inverted gamma distribution. Thus, finally we specify

$$\sigma_1^2, \ldots, \sigma_\ell^2 \overset{i.i.d.}{\sim} IG\{\eta, (\eta - 1)\delta\}, \tag{1.3}$$

where the inverse gamma density in (1.3) is given by $f(\sigma_i^2) = \{(\eta - 1)\delta\}^\eta (1/\sigma_i^2)^{\eta+1} e^{-(\eta-1)\delta/\sigma_i^2}/\Gamma(\eta), \sigma_i^2 > 0$ and $f(\sigma_i^2) = 0$ otherwise. We assume θ, τ, δ and η are fixed but unknown parameters. Nandram and Sedransk (1993) considered a similar model in Section 2 of their paper in which they assumed that η is *fixed* and *known*. Since $E(\sigma_i^2) = \delta$ and $\text{var}(\sigma_i^2) = \delta^2/(\eta - 2), \eta > 2$, for fixed δ small values of η express a belief that the variances are very different whereas large values express a belief that they are very similar.

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

$$(\underset{\sim}{Y}_i - \tilde{\underset{\sim}{\theta}})\{(\tau + 1)(\eta - 1)\delta/\eta\}^{-1/2} \sim t_{N_i}(\underset{\sim}{0},\ R,\ 2\eta), \tag{1.4}$$

where $t_{N_i}(\underset{\sim}{0},\ R,\ 2\eta)$ 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 2η degrees of freedom; see, for example, Box and Tiao (1992), pg. 117. Note that the model specifies that the $\underset{\sim}{Y}_i$ are independent, and that the components of the $\underset{\sim}{Y}_i$ are exchangeable. In fact, if the area sizes N_i are equal, the $\underset{\sim}{Y}_i$ have the same distribution with parameters θ, η, δ and τ. 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 (σ_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 μ_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 μ_i. Note that in our model the μ_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 $\gamma(\underset{\sim}{Y}_\ell) = \sum_{j=1}^{N_\ell} Y_{\ell j}/N_\ell$. Let s_i denote the set of n_i individuals sampled from the N_i individuals of the ith area. Then, $f_i = n_i/N_i$ is the sampling fraction, $\overline{Y}_i = \sum_{j \epsilon s_i} Y_{ij}/n_i$ is the sample mean and $S_i^2 = \sum_{j \epsilon s_i}(Y_{ij} - \overline{Y}_i)^2/(n_i - 1)$ is the sample variance.

We denote the shrinkage factor by $\omega_i = (1 + n_i\tau)^{-1}$ for the ith area, $i = 1, \ldots, \ell$. Note that larger shrinkage factor for the ith area corresponds to larger degree of pooling for the ith area (i.e., the EB estimator of the finite population mean of the ith area uses more information from the other areas). Also observe that, as expected, the shrinkage factor ω_i decreases as either n_i or τ 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 $\underset{\sim}{\phi} = (\theta, \tau, \delta, \eta)'$ be fixed, $\tilde{\sigma}_\ell^2 = \{(n_\ell - 1)S_\ell^2 + n_\ell \omega_\ell (\overline{Y}_\ell - \theta)^2 + 2(\eta - 1)\delta\}\kappa^{-1}$ where $\kappa = n_\ell + 2\eta$. Then, as in Nandram and Sedransk (1993), given the sampled data from the ℓth area, a $100(1-\alpha)\%$ highest posterior density (HPD) interval for $\gamma(\underset{\sim}{Y}_\ell)$ is

$$e_B \pm \nu_B t_{\kappa,\alpha/2}, \tag{1.5}$$

where, for convenience, we write

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

and $t_{\kappa,\alpha/2}$ is the $100(1-\alpha/2)$th percentile point of the Student t distribution with κ degrees of freedom. Under squared error loss, e_B is the Bayes estimator of $\gamma(\underset{\sim}{Y}_\ell)$. 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 $\underset{\sim}{\phi}$ is unknown, our objective is to construct a $100(1-\alpha)\%$ EB prediction interval for $\gamma(\underset{\sim}{Y}_\ell)$ 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 varability 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 $\underset{\sim}{\phi}$ based on $\overline{Y}_i$ and S_i^2 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 μ_i and $\sigma_i^2, i = 1, \ldots, \ell$ are eliminated (by integration) in the model given by (1.1), (1.2) and (1.3); all the components of $\underset{\sim}{\phi}$ are fixed but *unknown.* In Section 2, after obtaining the point estimators for $\underset{\sim}{\phi} = (\theta, \tau, \eta, \delta)'$, we develop a two-stage empirical Bayes prediction interval of $\gamma(\underset{\sim}{Y}_\ell)$ 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 ℓth 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 $\underset{\sim}{\phi}$ and investigate their asymptotic properties. We assume throughout that $2 \le \inf_{i\ge 1} n_i \le \sup_{i \ge 1} n_i \le k < \infty$. (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 μ_i and the σ_i^2. 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 $\underset{\sim}{\phi}$ and provide their relevant asymptotic properties.

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

$$\{(1-\omega_i)/(1-\eta^{-1})\delta\tau\}^{1/2}(\overline{Y}_i - \theta) \sim t_{2\eta} \text{ and } \eta S_i^2/(\eta-1)\delta \sim F(n_i - 1, 2\eta), \quad (2.1)$$

where $t_{2\eta}$ is a student t distribution on 2η 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 $\underset{\sim}{\phi}$.

First, assuming that τ, δ and η are known, we construct an estimator of θ. Using (2.1) it is easy to show that the best linear unbiased estimator of θ is $\hat{\theta}_*$ where

$$\hat{\theta}_* = \sum_{i=1}^{\ell}(1-\omega_i)\overline{Y}_i / \sum_{i=1}^{\ell}(1-\omega_i). \quad (2.2)$$

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

$$\hat{\delta} = WMS = (n_T - \ell)^{-1}\sum_{i=1}^{\ell}(n_i - 1)S_i^2 \quad (2.3)$$

which is an unbiased estimator of δ. Now, letting

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

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

$$\hat{\tau} = \max(0, \hat{\tau}_*), \tag{2.4}$$

where

$$\hat{\tau}_* = (\ell-1)\Big[n_T - n_T^{-1} \sum_{i=1}^{\ell} n_i^2\Big]^{-1} \{BMS/WMS - 1\}, \ell > 1. \tag{2.5}$$

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 ℓ.) Thus we use the estimator

$$\hat{\theta} = \begin{cases} \sum_{i=1}^{\ell} (1-\hat{\omega}_i)\overline{Y}_i / \sum_{i=1}^{\ell} (1-\hat{\omega}_i), & \hat{\tau} > 0 \\ \overline{\overline{Y}}, & \hat{\tau} = 0, \end{cases}$$

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 $\sum_{i=1}^{\ell}(1-\hat{\omega}_i)\overline{Y}_i / \sum_{i=1}^{\ell}(1-\hat{\omega}_i)$ is indeterminate. The second estimator is sensible because $\lim_{\hat{\tau}\to 0} \sum_{i=1}^{\ell}(1-\hat{\omega}_i)\overline{Y}_i / \sum_{i=1}^{\ell}(1-\hat{\omega}_i) = \overline{\overline{Y}}$.

Next, we construct an estimator for η. 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 δ 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 η we consider

$$\hat{\eta} = 2 + \{\max(\ell^{-1}, \hat{\eta}_*^{-1})\}^{-1}, \tag{2.6}$$

where

$$\hat{\eta}_*^{-1} = \Big\{\sum_{i=1}^{\ell} h_i(1-h_i) + 2(\ell-1)(n_T-\ell)^{-1}\Big\}^{-1} \Big\{\Big[\sum_{i=1}^{\ell} h_i (S_i^2 \hat{\delta}^{-1} - 1)^2\Big] - 2(\ell-1)(n_T-\ell)^{-1}\Big\}.$$

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 $\underset{\sim}{\phi} = (\theta, \delta, \tau, \eta)'$.

Lemma 1. *Assume* $2 \le \inf_{i\ge 1} n_i \le \sup_{i\ge 1} n_i \le k < \infty$. *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 $\max_{i=1,2,\ldots,\ell} |\hat{\omega}_i - \omega_i| \overset{a.s.}{\to} 0$, (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 $\underset{\sim}{\phi}$ is known. Then under the marginal distribution of the Y_{ij}

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

where e_B and ν_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 τ, δ, η are known, and consider the pivotal quantity

$$(\gamma(\underset{\sim}{Y}_\ell) - \hat{e}_B^*)/\nu_{Ba}, \tag{2.8}$$

where

$$\begin{aligned}
\hat{e}_B^* &= \overline{Y}_\ell - (1 - f_\ell)\omega_\ell(\overline{Y}_\ell - \hat{\theta}_*),\\
\nu_{Ba}^2 &= (1 - 2\kappa^{-1})\text{var}\{\gamma(\underset{\sim}{Y}_\ell) - \hat{e}_B^*\} = \nu_B^2 + \nu_{\hat{\theta}_*}^2,\\
\nu_{\hat{\theta}_*}^2 &= (1 - 2\kappa^{-1})\delta(1 - f_\ell)^2\omega_\ell^2 / \sum_{i=1}^{\ell} n_i\omega_i,
\end{aligned}$$

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

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

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, ν_{Ba}^2 in (2.8) contains an additional variability $\nu_{\hat{\theta}_*}^2$ beyond ν_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 $\gamma(\underset{\sim}{Y}_\ell)$ as

$$\hat{e}_B \pm \hat{\nu}_B t_{\hat{\kappa},\alpha/2}. \tag{2.10}$$

In (2.10)

$$\hat{e}_B = \overline{Y}_\ell - (1-f_\ell)\hat{\omega}_\ell(\overline{Y}_\ell - \hat{\theta}), \hat{\nu}_B^2 = \tilde{\nu}_B^2 + \hat{\nu}_{\hat{\theta}_*}^2,$$
$$\tilde{\nu}_B^2 = (1-2\hat{\kappa}^{-1})(1-f_\ell)\{f_\ell + (1-f_\ell)(1-\hat{\omega}_\ell)\}\hat{\sigma}_\ell^2 n_\ell^{-1},$$
$$\hat{\nu}_{\hat{\theta}_*}^2 = (1-2\hat{\kappa}^{-1})\hat{\delta}(1-f_\ell)^2\hat{\omega}_\ell^2 / \sum_{i=1}^{\ell} n_i\hat{\omega}_i,$$

where

$$\hat{\sigma}^2 = \{(n_\ell - 1)S_\ell^2 + n_\ell\hat{\omega}_\ell(\bar{Y} - \hat{\theta})^2 + 2(\hat{\eta}-1)\hat{\delta}\}\hat{\kappa}^{-1} \text{ and } \hat{\kappa} = n_\ell + 2\hat{\eta}.$$

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 σ_i^2 are fixed but unknown, and that $2(\eta-1)\delta\sigma_\ell^{-2}$ has a chi-square distribution on 2η degrees of freedom independently of S_ℓ^2, $\bar{Y}_\ell$ and $\gamma(\underset{\sim}{Y}_\ell) - e_B$. Then, under the marginal distributions of the y_{ij}, independently $\{\gamma(\underset{\sim}{Y}_\ell) - e_B\}\{(1-f_\ell)\{f_\ell + (1-f_\ell)(1-\omega_\ell)\}\sigma_\ell^2 n_\ell^{-1}\}^{-1/2} \sim N(0,1)$ and $\kappa\tilde{\sigma}_\ell^2/\sigma_\ell^2 \sim \chi_\kappa^2$. Consequently, $(\gamma(\underset{\sim}{Y}_\ell) - e_B)/\nu_B \sim t_\kappa$, and the $100(1-\alpha)\%$ shortest prediction interval (non-Bayesian) for $\gamma(\underset{\sim}{Y}_\ell)$ 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 σ_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 $Pr(e_B - \nu_B t_{\kappa,\alpha/2} \le \gamma(\underset{\sim}{Y}_\ell) \le e_B + \nu_B t_{\kappa,\alpha/2}) = E_{\underset{\sim}{Y}_\ell}(Pr(e_B - \nu_B t_{\kappa,\alpha/2} \le \gamma(\underset{\sim}{Y}_\ell) \le e_B + \nu_B t_{\kappa,\alpha/2} | \underset{\sim}{Y}_\ell)) = E_{\underset{\sim}{Y}_\ell}(1-\alpha) = 1-\alpha$. 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 $\underset{\sim}{Y}_\ell$. 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 $\hat{W}_B = 2t_{\hat{\kappa},\alpha/2}\hat{\nu}_B$. Also letting

$$\hat{P}_B = \mathcal{T}\{(\hat{e}_B - e_B + t_{\hat{\kappa},\alpha/2}\hat{\nu}_B)\nu_B^{-1}\} - \mathcal{T}\{(\hat{e}_B - e_B - t_{\hat{\kappa},\alpha/2}\hat{\nu}_B)\nu_B^{-1}\},$$

the coverage probability of the EB interval is $P_B = E_{\underset{\sim}{Y}}(\hat{P}_B)$ where expectation is taken over the joint marginal distribution of Y_{ij} obtained from (1.1), (1.2) and (1.3) and $\mathcal{T}(\cdot)$ is the cumulative distribution function of a Student t on 2η 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 $\gamma(\underset{\sim}{Y}_\ell)$ respectively.

First, we present Lemma 2.

Lemma 2. *Assume* $2 \le \inf_{i\ge 1} n_i \le \sup_{i\ge 1} n_i \le k < \infty$. *Then as* $\ell \to \infty$
(a) $\hat{\nu}^2_{\hat{\theta}_*} \overset{a.s.}{\to} 0$ and $E(\hat{\nu}^2_{\hat{\theta}_*}) \to 0$, (b) $\hat{\nu}_B - \nu_B \overset{a.s.}{\to} 0$, (c) $E \mid \hat{\nu}_B - \nu_B \mid \to 0$.

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

(b) It is easy to show

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

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

(c) By Lemma 1(c), $\hat{P}_B \overset{a.s.}{\to} \mathcal{T}(t_{\kappa,\alpha/2}) - \mathcal{T}(-t_{\kappa,\alpha/2}) = 1 - \alpha$ 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 $\gamma(\underset{\sim}{Y}_\ell)$ under squared error loss, $r(e)$, is $r(e) = E_{\underset{\sim}{Y}}\{e - \gamma(\underset{\sim}{Y}_\ell)\}^2$, where expectation is taken over the marginal distribution of $\underset{\sim}{Y}$ 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

$$Y_{ij} = \theta + \mu_i + \epsilon_{ij} \tag{3.1}$$

$i = 1, \ldots, \ell$ and $j = 1, \ldots, n_i$, where the μ_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 RNNOA- 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 σ_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 θ, apart from the estimated center, the sampling distributions of all quantities we study are invariant to choices of θ. 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_ℓ). To ensure a large range of degrees of borrowing (i.e., values of ω_ℓ), we take $\tau = 0.05, 0.25, 1.25$; see Table 1. Thus there are 81 combinations of ℓ, n_ℓ, τ and η.

Table 1. Values of ω_ℓ (i.e., degree of borrowing) by τ and n_ℓ

	τ		
n_ℓ	0.05	0.25	1.25
5	.80	.44	.14
10	.67	.29	.07
15	.57	.21	.05

NOTE: The degree of borrowing ($\omega_\ell = (1+n_\ell\tau)^{-1}$) depends on ℓ only through n_ℓ, and smaller τ 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

$$\bar{W}_{EB} = G^{-1} \sum_{i=1}^{G} \hat{W}_B^{(i)},$$

and its standard error

$$SE_{EB} = \{\sum_{i=1}^{G}(\hat{W}_B^{(i)} - \bar{W}_{EB})^2/G(G-1)\}^{1/2}, \quad (3.2)$$

where $\hat{W}_B^{(i)}$ is the width of the EB interval for the ith 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 $\bar{W}_{EB}$ 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 $R(EB,B) = r(\hat{e}_B)/r(e_B) - 1$. 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 ℓ, 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 ℓ, n_ℓ, τ or η 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 η, but these decrease as ℓ, n_ℓ or τ 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 ℓth area. This prediction interval, denoted by I_o, for the finite population mean of the ℓth area is

$$\bar{Y}_\ell \pm \frac{S_\ell}{\sqrt{n_\ell}}(1-f_\ell)^{1/2} t_{n_\ell - 1, \alpha/2}, \quad (3.3)$$

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 ℓ, n_ℓ, τ and η. 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 ℓ, n_ℓ and τ

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

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 τ, for the EB interval we also present the median (in parentheses) in column 7. However, the distributions become more symmetric as τ increases. That is, for small values of τ (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 η increases from 5 to 15.

For all values of ℓ, n_ℓ and τ, 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 ℓ, n_ℓ and τ. 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 ℓ, n_ℓ or τ increases. While there are small changes as τ increases, it seems that the EB interval performs better in terms of coverage probabilities for larger values of τ. 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 ℓ, n_ℓ and τ, the EB interval is remarkably close to the HPD interval. Further simulations for larger values of ℓ 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 ℓ and $n_i, i = 1, \ldots, \ell$ 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 $\ell > 20$, $n_l > 5$ 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. Fortunately, 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 ℓ small areas, all $\ell+1$ areas follow (1.1), (1.2) and (1.3), and interest is in $\gamma(\underset{\sim}{Y}_{\ell+1}) = \sum_{j=1}^{N_{\ell+1}} Y_{\ell+1,j}/N_{\ell+1}$ where $N_{\ell+1} \geq 1$ is the size of the small area. Then the 100(1-α)% HPD prediction interval for $\gamma(\underset{\sim}{Y}_{\ell+1})$ 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 Polytechnic 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* $var(S_i^2) = [(2+(n_i+1)/(\eta-2)]\delta^2/(n_i-1) \leq [(k+1)/(\eta-2)+2]\delta^2 = A$, *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* $BMS = (\ell-1)^{-1}\{\sum_{i=1}^{\ell} n_i\overline{Y}_i^2 - n_T\overline{\overline{Y}}^2\}$, *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* $\mid \hat{\omega}_i - \omega_i \mid \leq \mid 1 - \hat{\tau}\tau^{-1} \mid$, $i = 1, 2, \ldots, \ell, \max_{i=1,2,\ldots,\ell} \mid \hat{\omega}_i - \omega_i \mid \overset{a.s.}{\to} 0$ *as* $\ell \to \infty$.

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

$$\hat{\eta}_*^{-1} = \delta^2 \hat{\delta}^{-2} \{H_1(\delta, \hat{\delta}) - H_2(\delta, \hat{\delta}) + (\eta - 2)^{-1}\}, \tag{A.1}$$

where

$$H_1(\delta, \hat{\delta}) = \delta^{-2} \{\sum_{i=1}^{\ell} h_i (S_i^2 - \hat{\delta})^2 - E\{\sum_{i=1}^{\ell} h_i (S_i^2 - \hat{\delta})^2\}\}(n_T - \ell)(\ell - 1)^{-1} A_\ell,$$
$$H_2(\delta, \hat{\delta}) = 2(\hat{\delta}^2 \delta^{-2} - 1) A_\ell,$$

and $A_\ell = (\ell - 1) / \sum_{i=1}^{\ell} (1 - h_i)(n_i + 1)$. *Now both* A_ℓ *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 $Q_1 = \sum_{i=1}^{\ell} h_i S_i^4 - E(\sum_{i=1}^{\ell} h_i S_i^4)$ and $Q_2 = \hat{\delta}^2 - \delta^2 - \text{var}(\hat{\delta})$. 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 $\hat{\eta}_*^{-1} \overset{a.s}{\to} (\eta - 2)^{-1}$ as $\ell \to \infty$.

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

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

If $\hat{\tau} > 0, |\hat{\theta} - \theta| = |\sum_{i=1}^{\ell} (1 - \hat{\omega}_i)(\overline{Y}_i - \theta)| / \sum_{i=1}^{\ell} (1 - \hat{\omega}_i) \leq (k + \hat{\tau}^{-1})\{|\ell^{-1} \sum_{i=1}^{\ell} (1 - \omega_i)(\overline{Y}_i - \theta)| + \{\max_{i=1,2,\ldots,\ell} |\hat{\omega}_i - \omega_i|\} \ell^{-1} \sum_{i=1}^{\ell} |\overline{Y}_i - \theta|\}$. *By SLLN* $\ell^{-1} \sum_{i=1}^{\ell} (1 - \omega_i)(\overline{Y}_i - \theta) \overset{a.s.}{\to} 0$ *as* $\ell \to \infty$. *Since* $E(\ell^{-1} \sum_{i=1}^{\ell} |\overline{Y}_i - \theta|) \leq \delta\tau(1 + k\tau)^{1/2} < \infty$, *it follows that* $\ell^{-1} \sum_{i=1}^{\ell} |\overline{Y}_i - \theta|$ *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

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

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

First, by (B.1),

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

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

Second,

$$(\hat\theta - \theta)^2 \le \{n_T^{-1}\sum_{i=1}^{\ell} n_i(\overline{Y}_i - \theta)\}^2 + \{\sum_{i=1}^{\ell}(1-\hat\omega_i)(\overline{Y}_i - \theta)/\sum_{i=1}^{\ell}(1-\hat\omega_i)\}^2. \quad \text{(B.3)}$$

Using (B.3) it is easy to show that

$$(\hat\theta - \theta)^2 \le k^2\ell^{-1}\sum_{i=1}^{\ell}(\overline{Y}_i - \theta)^2. \quad \text{(B.4)}$$

Then because $\ell^{-1}\sum_{i=1}^{\ell}(\overline{Y}_i - \theta)^2$ 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 $\hat\nu^2_{\hat\theta_*} \le \hat\delta(1+k\hat\tau)/2\ell, \hat\nu^2_{\hat\theta_*} \overset{a.s.}{\to} 0$ as $\ell \to \infty$. Now $E(\hat\nu^2_{\hat\theta_*}) \le \{E\hat\delta(1+k\hat\tau)\}/2\ell$. Thus by Lemma 1 again, there exists $A < \infty$ s.t. $\hat\delta(1+k\hat\tau) < A$ a.e. Thus $E(\hat\nu^2_{\hat\theta_*}) \to 0$ as $\ell \to \infty$.*

Lemma 2(b). *Using the triangle inequality $\mid \hat\nu^2_B - \nu^2_B \mid \le \hat\nu^2_{\hat\theta_*} + \mid \tilde\nu^2_B - \nu^2_B \mid$. Thus by Lemma 2(a) it is only required to show that $\tilde\nu^2_B - \nu^2_B \overset{a.s.}{\to} 0$ as $\ell \to \infty$; see Appendix D.*

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

Appendix D: Completion of the Proof of Lemma 2

We show that $\tilde\nu^2_B - \nu^2_B \overset{a.s.}{\to} 0$ as $\ell \to \infty$ and $E(\tilde\nu^2_B - \nu^2_B)^2 \to 0$ as $\ell \to \infty$.

Letting $\Delta_1 = \tilde\sigma^2_\ell \max_{i=1,2,\ldots,\ell} \mid \hat\omega_i - \omega_i \mid$ and $\Delta_2 = \mid (1 - 2\hat\kappa^{-1})\hat\sigma^2_\ell - \tilde\sigma^2_\ell \mid$, we have by the triangle inequality, $\mid \tilde\nu^2_B - \nu^2_B \mid \le \Delta_1 + \Delta_2$, and by Minkowski's inequality $E(\tilde\nu^2_B - \nu^2_B)^2 \le [\{E(\Delta_1^2)\}^{1/2} + E(\Delta_2^2)^{1/2}]^2$.

First we show that $\Delta_1 \overset{a.s.}{\to} 0$ as $\ell \to \infty$ and $E(\Delta_1^2) \to 0$ as $\ell \to \infty$. As $\tilde\sigma^2_\ell$ has finite expectation, it is bounded a.e.. Thus, by Lemma 1(b), $\Delta_1 \overset{a.s.}{\to} 0$ as $\ell \to \infty$. Also since $E(\Delta_1^2) \le AE\{\max_{i=1,2,\ldots,\ell} \mid \hat\omega_i - \omega_i \mid\}^2$ where $A < \infty$ and $\max_{i=1,2,\ldots,\ell} \mid \hat\omega_i - \omega_i \mid\} \le 1$ (i.e., uniformly bounded), by Lemma 1(b) again, $E(\Delta_1^2) \to 0$ as $\ell \to \infty$.

Second, we note that

$$\Delta_2 \le \sum_{i=1}^{4} A_i \quad \text{(D.1)}$$

where

$$A_1 = \frac{1}{2}(n_\ell - 1)S_\ell^2 \mid \hat{\eta}^{-1} - \eta^{-1} \mid,$$
$$A_2 = 2n_\ell[\{(\bar{Y}_\ell - \theta)^2 + (\hat{\theta} - \theta)^2\} \max_{i=1,2,\ldots,\ell} \mid \hat{\omega}_i - \omega_i \mid],$$
$$A_3 = 2n_\ell\omega_\ell \mid \hat{\theta} - \theta \mid\mid \bar{Y}_\ell - \theta \mid + \frac{1}{2}n_\ell\omega_\ell(\bar{Y} - \theta)^2 \mid \hat{\eta}^{-1} - \eta^{-1} \mid,$$
$$A_4 = \frac{1}{2}\{(\hat{\eta} - 1)\hat{\delta} + \hat{\eta}\} \mid \hat{\eta}^{-1} - \eta^{-1} \mid + \frac{1}{2} \mid \hat{\delta} - \delta \mid .$$

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

$$E(\Delta_2^2) \leq 4 \sum_{i=1}^{4} E(A_i^2). \tag{D.2}$$

By using Minkowkski'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

Arora, V., Lahiri, P. and Mukherjee, K. (1997). Empirical Bayes estimation of finite population means from complex surveys. *J. Amer. Statist. Assoc.* **92**, 1555-1562.

Box, G. E. P. and Tiao, G. C. (1992). *Bayesian Inference in Statistical Analysis.* John Wiley, New York.

Carlin, B. P. and Gelfand, A. E. (1990). Approaches for empirical Bayes confidence intervals. *J. Amer. Statist. Assoc.* **85**, 105-114.

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

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

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

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

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

Hulting, F. L. and Harville, D. A. (1991). Some Bayesian and non-Bayesian procedures for the analysis of comparative experiments and for small-area estimation: computational aspects, frequentist properties and relationships. *J. Amer. Statist. Assoc.* **86**, 557- 568.

Kass, R. E. and Steffey, D. (1989). Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *J. Amer. Statist. Assoc.* **84**, 717-726.

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

Laird, N. M. and Louis, T. A. (1987). Empirical Bayes confidence intervals based on bootstrap samples (with comments). *J. Amer. Statist. Assoc.* **82**, 739-757.

Laird, N. M. and Louis, T. A. (1989). Empirical Bayes confidence intervals for a series of related experiments. *Biometrics.* **45**, 481-495.

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

Morris, C. (1983b). Parametric empirical Bayes confidence intervals. In *Scientific Inference Data Analysis, and Robustness* (G. E. P. Box, T. Leonard and C. F. J. Wu, eds.), 25-49. Academic Press, New York.

Nandram, B. and Sedransk, J. (1993). Empirical Bayes estimation for the finite population mean on the current occasion. *J. Amer. Statist. Assoc.* **88**, 994-1000.

Prasad, N. G. N. and Rao J. N. K. (1990). The estimation of the mean squared error of small area estimators. *J. Amer. Statist. Assoc.* **85**, 163-171.

Raghunathan, T. E. (1993). A quasi-empirical Bayes method for small area estimation. *J. Amer. Statist. Assoc.* **88**, 1444-1448.

Robbins, H. (1955). An empirical Bayes approach to statistics. In *Proc. 3rd Berkeley Symp. Math. Statist. Prob.* (Vol. 1), 159-163. University of California Press, Berkeley, California.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley, New York.

Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609-2280, U.S.A.

E-mail: balnan@wpi.edu

(Received September 1996; accepted March 1998)