ELSEVIER

Journal of Statistical Planning and Inference 112 (2003) 63–76

journal of statistical planning and inference

www.elsevier.com/locate/jspi

On measures of uncertainty of empirical Bayes small-area estimators

Ferry Butar Butar[a,*], P. Lahiri[b]

[a] *Department of Mathematics, Computer Science and Statistics, Sam Houston State University, P.O. Box 2206, Huntsville, TX 77341-2206, USA*

[b] *Division of Statistics, Department of Mathematics and Statistics, 922 Oldfather Hall, University of Nebraska-Lincoln, Lincoln, NE 68588-0323, USA*

Abstract

Small area typically refers to a small geographic area or a demographic group for which very little information is obtained from the sample surveys. An empirical Bayes (EB) method uses sample survey data in conjunction with relevant supplementary data which are obtained from various administrative sources. The method has been found to be very useful in many applications of small-area estimation and related problems.

In this paper, a method based on bootstrap samples is proposed to measure the accuracy of the proposed EB estimator of a small-area characteristic. A simple approximation of the method which does not require any bootstrap simulation is also proposed. The model expectation of the proposed measure of uncertainty of the EB estimator is equal to the integrated Bayes risk of the EB estimator up to the order $O(m^{-1})$. Two well-known data sets are considered to compare the proposed method with some existing methods. A Monte Carlo simulation is conducted to compare the performances of different measures of uncertainty.

MSC: 62C12; 62C25; 62D05; 62F15

Keywords: Best linear unbiased prediction; Bootstrap; Borrow strength; Integrated Bayes risk; Monte Carlo

1. Introduction

Model-based small-area estimation has received a considerable importance in the last two decades. Small area (domain) generally refers to a subgroup of a population from which samples are drawn. The subgroup may refer to a small geographical

* Corresponding author.

E-mail address: mth_fbb@shsu.edu (F.B. Butar).

0378-3758/03/$ - see front matter
PII: S0378-3758(02)00323-3

region (e.g., state, county, municipality, etc.) or a particular group obtained by a cross-classification of various demographic factors such as age, gender, race, etc. Small-area statistics are needed in regional planning and fund allocation in many government programs and thus the importance of producing reliable small-area statistics cannot be over-emphasized. See Ghosh and Rao (1994) and Rao (1999) for a thorough review of different small-area estimation techniques.

Parametric empirical Bayes (EB) point estimators have been found to be very useful in small-area estimation and related problems. See Carter and Rolph (1974) and Fay and Herriot (1979), among others. The EB data analysis considered in Efron and Morris (1975) can be related to a small-area estimation problem. Asymptotic optimality results of EB estimators have been given in Ghosh and Meeden (1986) and Ghosh and Lahiri (1987), among others.

Although parametric EB point estimators are attractive, very little research has been done regarding measuring their uncertainty, especially for complex small-area models. Morris (1983) and Laird and Louis (1987) are among handful of researchers who addressed this important issue. Their measures can be viewed as approximated posterior variances. The main focus of this paper is to develop a simple measure of uncertainty of a parametric EB estimator for a complex small-area model which does not require any specification of prior distributions for the hyper-parameters. Instead of approximating the posterior variance, we approximate the integrated Bayes risk (same as mean-squared error or MSE) of the proposed parametric EB estimator. Thus, our measure of uncertainty of EB can be compared to those of empirical best linear predictors given in Prasad and Rao (1990), Datta and Lahiri (2000) and Rao (2001), among others.

In Section 2, we propose a general longitudinal mixed linear normal model which covers many small-area models used in the literature and consider point estimation of small-area parameters of interest. In Section 3, we propose a method to measure the uncertainty of the EB estimator given in Section 2. The method is applicable for various methods of estimating the variance components (e.g., ANOVA, ML, REML, etc.). Our measure approximates the true integrated Bayes risk of the proposed EB estimator very accurately, the order of the neglected terms being $o(m^{-1})$, where m denotes the number of small areas. Thus, our measure enjoys a good frequentist property similar to that of Prasad and Rao (1990). It turns out that for a very special case of our model, our measure is the same as that of Morris (1983), up to the order $o(m^{-1})$. In Section 4, we consider two examples to illustrate our method. Results from a Monte Carlo simulation study are presented in Section 5.

2. Empirical Bayes point estimation

Let X_i and Z_i be $n_i \times p$ and $n_i \times k_i$ matrices of known constants. Let $n = \sum_{i=1}^m n_i$ and $k = \sum_{i=1}^m k_i$. Consider the following Bayesian model:

Model 1: (i) Conditional on a $k_i \times 1$ random vector U_i, Y_i's are independent with $Y_i | U_i \sim N_{n_i}(X_i\beta + Z_i U_i, R_i)$, $i = 1, \dots, m$ and

(ii) a priori, $U_i \overset{\text{ind}}{\sim} N_{k_i}(0, G_i)$, $i = 1, \dots, m$, where β is a $p \times 1$ column vector of unknown regression coefficients, $R_i = R_i(\psi)$ and $G_i = G_i(\psi)$ are, respectively, $n_i \times n_i$ and $k_i \times k_i$ matrices which possibly depend on ψ, a $s \times 1$ vector of unknown variance components.

Consider the estimation of $\theta_i = l_i'\beta + \lambda_i' U_i$, where l_i and λ_i are $p \times 1$ and $k_i \times 1$ vector of known constants respectively. Under the above model and squared error loss function, the Bayes estimator of θ_i is given by

$$\begin{aligned} \hat{\theta}_i^{\mathrm{B}} &= E[\theta_i | Y_i; \beta, \psi] = l_i'\beta + \lambda_i' G_i(\psi) Z_i' V_i^{-1}(\psi)(Y_i - X_i\beta) \\ &= \hat{\theta}_i(Y_i; \beta, \psi) \quad \text{say}, \end{aligned} \tag{1}$$

where $V_i(\psi) = R_i + Z_i G_i Z_i'$ $(i = 1, \ldots, m)$.

When ψ is known, but β is unknown, β is estimated by the maximum-likelihood estimator $\hat{\beta}(\psi)$, where $\hat{\beta}(\psi) = [\sum_{i=1}^m X_i' V_i^{-1}(\psi) X_i]^{-1} [\sum_{i=1}^m X_i' V_i^{-1}(\psi) Y_i]$. Plugging in $\hat{\beta}(\psi)$ for β in $\hat{\theta}_i^{\mathrm{B}}$, we get the following EB estimator of θ_i:

$$\tilde{\theta}_i^{\mathrm{EB}} = \hat{\theta}_i(Y_i; \hat{\beta}(\psi), \psi) = l_i'\hat{\beta}(\psi) + \lambda_i' G_i(\psi) Z_i' V_i^{-1}(\psi)[Y_i - X_i \hat{\beta}(\psi)]. \tag{2}$$

Note that $\hat{\theta}_i^{\mathrm{EB}}$ is a robust estimator in the sense that it can be viewed as the best linear unbiased predictor (see Prasad and Rao, 1990) when normality is relaxed in Model 1 or a linear EB estimator under the assumption of posterior linearity (see Ghosh and Lahiri, 1987).

In practice β and ψ are both unknown. In this case, an EB estimator of θ_i is obtained as $\hat{\theta}_i^{\mathrm{EB}} = \hat{\theta}_i(Y_i; \hat{\beta}(\hat{\psi}), \hat{\psi})$, where $\hat{\psi}$ is an estimator of ψ which satisfies the regularity conditions (RC) given in the Appendix. We shall assume that $E(\hat{\psi} - \psi) = -m^{-1}B(\psi) + o(m^{-1})$, where the functional form of $B(\psi)$ is known.

Example. Model 1 covers the following model (see Fay and Herriot, 1979):

(i) $Y_i | \theta_i \overset{\text{ind}}{\sim} N(\theta_i, D_i)$, $i = 1, \ldots, m$ and
(ii) a priori, $\theta_i \overset{\text{ind}}{\sim} N(x_i'\beta, A)$, $i = 1, \ldots, m$,

where D_i's are known and x_i's are $p \times 1$ vector of known constants. In the notations of Model 1, $n_i = k_i = 1, Z_i = 1, U_i = \theta_i - x_i'\beta, \psi = A, R_i(\psi) = D_i$ and $G_i(\psi) = A (i = 1, \ldots, m)$.

In this case, we may consider the following EB estimator of θ_i:

$$\hat{\theta}_i^{\mathrm{EB}} = x_i'\hat{\beta} + (1 - \hat{B}_i)(Y_i - x_i'\hat{\beta}),$$

where

$$\hat{\beta} = \left(\sum_{j=1}^m \frac{1}{\hat{A} + D_j} x_j x_j' \right)^{-1} \left(\sum_{j=1}^m \frac{1}{\hat{A} + D_j} x_j Y_j \right),$$

$$\hat{A} = \max\left(0, \tilde{A} = \frac{1}{m-p}\left[\sum_{j=1}^{m}(Y_j - x_j'\hat{\beta}_{\text{ols}})^2 - \sum_{j=1}^{m}(1-h_{jj})D_j\right]\right),$$

$$\hat{\beta}_{\text{ols}} = \left(\sum_{j=1}^{m} x_j x_j'\right)^{-1}\left(\sum_{j=1}^{m} x_j Y_j\right), \quad h_{ii} = x_i'\left(\sum_{j=1}^{m} x_j x_j'\right)^{-1} x_i, \quad \hat{B}_i = \frac{D_i}{\hat{A}+D_i}.$$

In this case, $B(\psi) = 0$.

For other particular cases of Model 1, see Carter and Rolph (1974) and Battese et al. (1988), among others.

3. A measure of uncertainty of $\hat{\theta}_i^{\text{EB}}$

Define the integrated Bayes risk of $\hat{\theta}_i^{\text{EB}}$ as $r(\hat{\theta}_i^{\text{EB}}) = E(\hat{\theta}_i^{\text{EB}} - \theta_i)^2$, where the expectation is taken with respect to Model 1. Note that $r(\hat{\theta}_i^{\text{EB}})$ is identical to the MSE of $\hat{\theta}_i^{\text{EB}}$ as defined in Prasad and Rao (1990). An appropriate estimator of $r(\hat{\theta}_i^{\text{EB}})$ can be taken as a measure of uncertainty of $\hat{\theta}_i^{\text{EB}}$. First note that

$$r(\hat{\theta}_i^{\text{EB}}) = g_{1i}(\psi) + g_{2i}(\psi) + E[\hat{\theta}_i^{\text{EB}} - \tilde{\theta}_i^{\text{EB}}]^2, \tag{3}$$

where $g_{1i}(\psi) = \lambda_i' G_i(\psi)[I_{k_i} - Z_i' V_i^{-1}(\psi) Z_i G_i(\psi)]\lambda_i$, ($I_{k_i}$ being the identity matrix of order $k_i \times k_i$), $g_{2i}(\psi) = (l_i - X_i' V_i^{-1}(\psi) Z_i G_i(\psi)\lambda_i)'(\sum_{i=1}^{m} X_i' V_i^{-1}(\psi) X_i)^{-1}(l_i - X_i' V_i^{-1}(\psi) Z_i G_i(\psi)\lambda_i)$ (see, Kackar and Harville, 1984; Prasad and Rao, 1990). Note that $g_{1i}(\psi)$ is the measure of uncertainty of the Bayes estimator $\hat{\theta}_i^{\text{B}}$, $g_{2i}(\psi)$ is the uncertainty due to the estimation of β and the third term is the uncertainty due to the estimation of ψ. One may naively approximate $r(\hat{\theta}_i^{\text{EB}})$ by $g_{1i}(\psi) + g_{2i}(\psi)$, which ignores the uncertainty due to the estimation of ψ. Datta and Lahiri (2000) showed that under certain mild RC

$$E(\hat{\theta}_i^{\text{EB}} - \tilde{\theta}_i^{\text{EB}})^2 = g_{3i}(\psi) + \text{o}(m^{-1}), \tag{4}$$

where $g_{3i}(\psi) = \text{trace}[L_i(\psi) V_i(\psi) L_i'(\psi) \sum(\psi)]$, $L_i(\psi) = \text{col}_{1 \leqslant j \leqslant s} L_{ij}'(\psi)$, $L_{ij}'(\psi) = \partial/\partial\psi_j(\lambda_i' G_i(\psi) Z_i' V_i^{-1}(\psi))$ and $\sum(\psi) = E(\hat{\psi} - \psi)(\hat{\psi} - \psi)'$. The expression for $\sum(\psi)$ for some standard methods of estimation of ψ (e.g., ANOVA, ML, REML) are given in Prasad and Rao (1990) and Datta and Lahiri (2000). Thus, the naive approximation, i.e., $g_{1i}(\psi) + g_{2i}(\psi)$, could lead to a serious underestimation since $g_{3i}(\psi)$ is of order $\text{O}(m^{-1})$, same as the order of $g_{2i}(\psi)$. A naive estimator of $r(\hat{\theta}_i^{\text{EB}})$ is obtained from the naive approximation and is given by $V_i^{\text{N}} = g_{1i}(\hat{\psi}) + g_{2i}(\hat{\psi})$. This introduces additional bias of the order $\text{O}(m^{-1})$ since

$$E\{g_{1i}(\hat{\psi}) + g_{2i}(\hat{\psi})\} = g_{1i}(\psi) + g_{2i}(\psi) - m^{-1}B'(\psi)\nabla g_{1i}(\psi) - g_{3i}(\psi) + \text{o}(m^{-1}), \tag{5}$$

where $\nabla g_{1i}(\psi) = \text{col}_{1 \leqslant j \leqslant s}(\partial/\partial\psi_j) g_{1i}(\psi)$. See Datta and Lahiri (2000).

To introduce the parametric bootstrap method, consider the following bootstrap model:

Model 2:

(i) $Y_i^\star | U_i^\star \overset{\text{ind}}{\sim} N_{n_i}(X_i\hat{\beta} + Z_iU_i^\star, \hat{R}_i)$, $i = 1, \ldots, m$ and

(ii) a priori, $U_i^\star \overset{\text{ind}}{\sim} N_{k_i}(0, \hat{G}_i)$, $i = 1, \ldots, m$,

where $\hat{R}_i = R_i(\hat{\psi})$ and $\hat{G}_i = G_i(\hat{\psi})$.

We use the proposed parametric bootstrap twice—once to estimate the first two terms of (3) by correcting the bias of $g_{1i}(\hat{\psi}) + g_{2i}(\hat{\psi})$ and again to estimate the third term of (3). We propose to estimate $g_{1i}(\psi) + g_{2i}(\psi)$ by

$$g_{1i}(\hat{\psi}) + g_{2i}(\hat{\psi}) - E_\star[g_{1i}(\hat{\psi}^\star) + g_{2i}(\hat{\psi}^\star) - g_{1i}(\hat{\psi}) - g_{2i}(\hat{\psi})]$$

and the third term of (3) by

$$E_\star[\hat{\theta}_i(Y_i; \hat{\beta}(\hat{\psi}^\star), \hat{\psi}^\star) - \hat{\theta}_i(Y_i; \hat{\beta}(\hat{\psi}), \hat{\psi})]^2,$$

where $E_\star$ is the expectation with respect to Model 2, and the calculation of $\hat{\psi}^\star$ is the same as that of $\hat{\psi}$ except that it is based on $Y_i^\star$'s instead of Y_i's. Our proposed estimator of $r(\hat{\theta}_i^{\text{EB}})$ is then given by

$$V_i^{\text{BOOT}} = 2[g_{1i}(\hat{\psi}) + g_{2i}(\hat{\psi})] - E_\star[g_{1i}(\hat{\psi}^\star) + g_{2i}(\hat{\psi}^\star)] + E_\star[\hat{\theta}_i(Y_i; \hat{\beta}((\hat{\psi}^\star), \hat{\psi}^\star) - \hat{\theta}_i(Y_i; \hat{\beta}(\hat{\psi}), \hat{\psi})]^2. \quad (6)$$

In Theorem A.2, we proved that

$$E[V_i^{\text{BOOT}}] = r(\hat{\theta}_i^{\text{EB}}) + \text{o}(m^{-1}).$$

Theorem A.1 can be used to justify the following approximation to V_i^{BOOT}:

$$V_i^{\text{BOOT}} \approx g_{1i}(\hat{\psi}) + m^{-1}B'(\hat{\psi})\nabla g_{1i}(\hat{\psi}) + g_{2i}(\hat{\psi}) + g_{3i}(\hat{\psi}) + g_{4i}(\hat{\psi}; Y_i), \quad (7)$$

where $g_{4i}(\hat{\psi}; Y_i) = \text{trace}[L_i(\hat{\psi})[Y_i - X_i\hat{\beta}(\hat{\psi})][Y_i - X_i\hat{\beta}(\hat{\psi})]'L_i'(\psi)\sum(\hat{\psi})]$.

The above measure was first given in the unpublished dissertation of Butar (1997). Note that the second-order term g_{4i} in (7) is area specific in Y_i and not the leading term g_{1i}. This is also the case for other area specific measures given in Remark 1 of the paper. However, for nonlinear models the leading term could be area specific (see Meza, 2003). For the Fay–Herriot model, Rao (2001) provided a simple way to produce two different area specific MSE estimators. One of them is identical to (7) and the other one is more area specific than (7). We note that the above measure can be also obtained by extending an approach of Rao (2001) to Model 1.

Laird and Louis (1987) proposed a measure of uncertainty of an EB estimator for a very special case of Model 1, specifically the Fay–Herriot model with $x_i'\beta = \mu$ and $D_i = D (i = 1, \ldots, m)$. For the general model, one may extend their measure as $V_i^{\text{LL}} = E_\star[g_{1i}(\hat{\psi}^\star)] + V_\star[\hat{\theta}_i(Y_i; \hat{\beta}^\star(\hat{\psi}^\star), \hat{\psi}^\star)]$, where $\hat{\beta}^\star(\hat{\psi}^\star) = [\sum_{i=1}^m X_i'V_i^{-1}(\hat{\psi}^\star)X_i]^{-1} [\sum_{i=1}^m X_i'V_i^{-1}(\hat{\psi}^\star)Y_i^\star]$, $E_\star$ and $V_\star$ being the expectation and the variance with respect to Model 2.

It is shown in Theorem A.1 that $V_i^{\text{LL}} = g_{1i}(\hat{\psi}) + g_{2i}(\hat{\psi}) + g_{4i}(\hat{\psi}; Y_i) - m^{-1}B'(\hat{\psi})\nabla g_{1i}(\hat{\psi}) - g_{3i}(\hat{\psi}) + \text{o}_\text{p}(m^{-1})$. Since $E\, g_{4i}(\hat{\psi}; Y_i) = g_{3i}(\psi) + \text{o}(m^{-1})$, it is quite possible that for some

i, V_i^{LL} could give us a measure which is less than the naive measure V_i^{N}, at least for large m.

In Theorem A.2, we show that under Model 1 and RC

$$E[V_i^{\mathrm{LL}}] = r(\hat{\theta}_i^{\mathrm{EB}}) - 2m^{-1}B'(\psi)\nabla g_{1i}(\psi) - g_{3i}(\psi) + \mathrm{o}(m^{-1}). \tag{8}$$

Thus, unlike V_i^{BOOT}, V_i^{LL} could lead to an underestimation of $r(\hat{\theta}_i^{\mathrm{EB}})$ since the order of bias for V_i^{LL} is $\mathrm{O}(m^{-1})$. This is because of the fact that Laird and Louis (1987) attempted to approximate the posterior variance and not the MSE.

Remark 1. Consider a special case of the Fay–Herriot model when $x_i'\beta = \mu$ and $D_i = D (i = 1,\dots,m)$. An EB estimator of θ_i is given by $\hat{\theta}_i^{\mathrm{EB}} = \bar{Y} + (1 - \hat{B}_1)(Y_i - \bar{Y})$, where $\bar{Y} = 1/m \sum Y_i$, $\hat{B}_1 = \min((m-3)/(m-1), \hat{B})$ and $\hat{B} = D(m-3)/\sum(Y_i - \bar{Y})^2$. See Morris (1983) and Laird and Louis (1987). Using a flat improper prior distributions on μ and $B = D/(A + D)$, Morris (1983) suggested an approximation to the posterior variance of θ_i as a measure of uncertainty of $\hat{\theta}_i^{\mathrm{EB}}$.

The measure is given by

$$V_i^{\mathrm{M}} = (1 - \hat{B}_1)D + \frac{D\hat{B}_1}{m} + \frac{2\hat{B}_1^2}{m-3}(Y_i - \bar{Y})^2. \tag{9}$$

Note that in this case $g_{1i}(B) = (1 - B)D$ can be unbiasedly estimated by $g_{1i}(\hat{B})$. When $\hat{B}_1$ is used to estimate B, the order of the bias is lower than $\mathrm{O}(m^{-1})$. Thus, this is a special situation where the Prasad–Rao (PR)-type bias correction of $g_{1i}(\hat{\psi})$ is not needed. Note that in this case the proposed and the P–R measures of uncertainty are given by

$$V_i^{\mathrm{BOOT}} = (1 - \hat{B}_1)D + \frac{D\hat{B}_1}{m} + \frac{2\hat{B}_1^2}{m}(Y_i - \bar{Y})^2 \tag{10}$$

and

$$V_i^{\mathrm{PR}} = (1 - \hat{B}_1)D + \frac{D\hat{B}_1}{m} + \frac{2D\hat{B}_1}{m}, \tag{11}$$

respectively.

Note that the measure given by (10) is identical to Butar (1997) and Rao (2001). Laird and Louis (1987) proposed the following measure of uncertainty of $\hat{\theta}_i^{\mathrm{EB}}$:

$$V_i^{\mathrm{LL}} = (1 - \hat{B}_0)D + \frac{m-1}{m-5}\frac{D\hat{B}_0}{m} + \frac{2\hat{B}_0^2}{m-5}(Y_i - \bar{Y})^2, \tag{12}$$

where $\hat{B}_0 = (m-1)/(m-3)\hat{B}_1$. Thus, V_i^{BOOT} and V_i^{M} are identical up to order $\mathrm{O_p}(m^{-1})$. However, the difference between V_i^{M} and V_i^{LL} is of order $\mathrm{O_p}(m^{-1})$. The PR measure V_i^{PR} cannot match a hierarchical Bayes solution since a hierarchical Bayes solution must be of the form $(1 - E(B|Y))D + DE(B|Y)/m + (Y_i - \bar{Y})^2 V(B|Y)$, where $E(B|Y)$ and $V(B|Y)$ are the posterior mean and variance of B, under a suitable prior on the hyper-parameters μ and B. We note that for this particular case the Prasad–Rao method gives exactly the same measure of uncertainty for all the small areas unlike the other methods since the PR method does not depend on the individual Y_i. However, if covariates or unequal variances are involved the PR method is area specific even if area specific response data are not involved.

4. Two examples

Efron and Morris (1975) successfully demonstrated the superiority of the EB estimator (see Section 3) over the classical estimator Y_i using the famous baseball data which contains the batting averages of 18 major league baseball players. It is instructive to compare various measures of uncertainty of their EB estimator. Table 1 presents various measures of uncertainty given in Remark 1. Amount of inflation of the measures which incorporate the uncertainty due to estimation of A is substantial when compared with the naive measure V_i^{N}. Note that V_i^{PR} and V_i^{N} are constant (0.353) and (0.153), respectively, for all the baseball players. All the other measures of uncertainty change from player to player since they depend on the individual Y_i.

For the above example, $m = 18$ may be considered to be small. We now consider another example where $m = 51$ is moderately large. The US Department of Health and Human Services uses estimates of median income of four-person families at the state level to formulate its energy assistance program to low-income families. Such data are provided by the US Census Bureau for all the states (including the District of Columbia) on an annual basis. The current estimates are produced by an empirical Bayes procedure (see Ghosh et al., 1996). The data we analyse provide the usual design-based estimates of four-person families (Y_i) and its sampling variances D_i for

Table 1
Comparison of different measures of uncertainty of empirical Bayes estimates for the baseball data (Efron and Morris, 1975)

Player name	Naive	Prasad–Rao	Laird–Louis	BOOT
Clemente (Pitts, NL)	0.153	0.353	0.646	0.725
F. Robinson (Balt, AL)	0.153	0.353	0.512	0.590
F. Howard (Wash, AL)	0.153	0.353	0.396	0.474
Johnstone (Cal, AL)	0.153	0.353	0.301	0.380
Berry (Chi, AL)	0.153	0.353	0.232	0.310
Spencer (Cal, AL)	0.153	0.353	0.232	0.310
Kessinger (Chi, NL)	0.153	0.353	0.188	0.266
L. Alvarado (Bos, AL)	0.153	0.353	0.169	0.248
Santo (Chi, NL)	0.153	0.353	0.179	0.258
Swoboda (NY, NL)	0.153	0.353	0.179	0.258
Unser (Wash, AL)	0.153	0.353	0.218	0.297
Williams (Chi, AL)	0.153	0.353	0.218	0.297
Scott (Bos, AL)	0.153	0.353	0.218	0.297
Petrocelli (Bos, AL)	0.153	0.353	0.218	0.297
E. Rodriguez (KC, AL)	0.153	0.353	0.218	0.297
Campaneries (Oak, AL)	0.153	0.353	0.293	0.372
Munson (NY, AL)	0.153	0.353	0.402	0.481
Alvis (Mill, AL)	0.153	0.353	0.558	0.636

Table 2
Comparison of different measures of uncertainty of empirical Bayes estimates of median incomes of four-person family for 50 States and the District of Columbia for the year 1988

State no.	$\sqrt{D_i}$	Naive	Prasad–Rao	Laird–Louis	BOOT	State no.	$\sqrt{D_i}$	Naive	Prasad–Rao	Laird–Louis	BOOT
1	2183	1265	1284	1266	1276	27	1765	1166	1186	1173	1183
2	3248	1426	1439	1426	1433	28	2632	1335	1352	1336	1344
3	2908	1371	1387	1374	1382	29	3790	1442	1454	1444	1450
4	1989	1246	1266	1248	1258	30	1577	1109	1128	1134	1143
5	3040	1393	1408	1446	1453	31	3094	1380	1395	1381	1388
6	3655	1470	1482	1474	1479	32	3089	1385	1400	1386	1393
7	1972	1231	1250	1232	1242	33	2766	1350	1366	1369	1377
8	1705	1173	1192	1179	1189	34	3006	1368	1384	1371	1378
9	1636	1129	1149	1133	1143	35	2031	1226	1245	1263	1272
10	1576	1108	1127	1108	1118	36	2723	1342	1358	1343	1352
11	3037	1385	1400	1390	1397	37	3122	1381	1395	1383	1390
12	1642	1135	1154	1138	1148	38	2492	1316	1334	1317	1326
13	1744	1166	1185	1167	1176	39	2586	1324	1342	1342	1350
14	1718	1154	1174	1178	1188	40	2867	1356	1372	1356	1364
15	2644	1348	1365	1349	1357	41	2685	1335	1352	1354	1363
16	2169	1260	1279	1264	1274	42	3116	1395	1410	1403	1410
17	3587	1427	1439	1432	1438	43	2733	1340	1357	1340	1348
18	2301	1278	1297	1314	1323	44	3951	1447	1458	1457	1463
19	2329	1281	1300	1317	1326	45	2521	1318	1335	1322	1331
20	2778	1351	1368	1352	1360	46	3019	1387	1402	1395	1403
21	2400	1301	1319	1341	1350	47	2827	1366	1382	1367	1375
22	3393	1425	1438	1435	1442	48	2780	1354	1370	1374	1382
23	3489	1452	1464	1457	1463	49	1731	1164	1183	1173	1182
24	8264	1533	1536	1533	1534	50	3885	1467	1478	1474	1480
25	3106	1403	1418	1407	1415	51	3856	1452	1464	1458	1463
26	1866	1187	1207	1206	1216						

all the 50 states and the District of Columbia for the year 1988. As a covariate for the Fay–Herriot model, we choose 1979 census median income of four-person families updated by the change in per-capita income obtainable from the Bureau of Economic Analysis. Our focus here is on the comparison of different measures of uncertainty of the EB estimator. Table 2 presents the standard error of Y_i (i.e., $\sqrt{D_i}$, $\sqrt{V_i^{\mathrm{N}}}$, $\sqrt{V_i^{\mathrm{PR}}}$, $\sqrt{V_i^{\mathrm{LL}}}$ and $\sqrt{V_i^{\mathrm{BOOT}}}$). All the different measures of uncertainty of the EB estimates are smaller than the measure of uncertainty of Y_i, the design-based estimates of the median income of the four-person families. In fact, there is a considerable gain in using EB estimator. Generally, both V_i^{PR} and V_i^{BOOT} are more conservative than V_i^{LL}. It appears that V_i^{PR} is generally slightly more conservative than V_i^{BOOT}.

5. A Monte Carlo simulation study

In this section, we compare the finite-sample accuracy of the proposed parametric bootstrap MSE estimator (BOOT) with the naive (N), Prasad–Rao (PR) and Laird–Louis (LL) measures by Monte Carlo simulation. We first generated 10,000 independent

θ_i from a normal distribution with mean $\mu=0$ and variance $A=1.5$, and then for each θ_i we generated Y_i from a normal distribution with mean θ_i and variance $D=1 (i=1,\ldots,m)$. We studied three different values of $m=20$, 30 and 50. This is the special case discussed in Remark 1.

We study percent average relative biases (ARB) and coverage probabilities for different measures uncertainty of EB when the average is taken over small areas for which $Y_i^2/(A+D) \geqslant \chi^2_{1;\alpha}$, where $\chi^2_{1;\alpha}$ denotes point on the χ^2_1 such that right hand tail area is α. For example, when $\alpha = 1$, the average is taken over all the small areas and when $\alpha = 0.05$ the average is taken over roughly 5% of all the small areas for which $Y_i^2/(A+D) \geqslant \chi^2_{1;0.05}$, i.e., these are areas which deviate from the prior mean considerably.

We summarize our simulation results using Figs. 1 and 2. For both the figures, the x-axis represents α. In Fig. 1, we plot the percent ARB of different MSE estimators which are calculated by 100× [average E(estimator of MSE)−average MSE]/(average of MSE), where the average is taken over all the small areas such that $Y_i^2/(A+D) \geqslant \chi^2_{1;\alpha}$ and E denotes the simulated average. Note that when α is small, the performance of our measure is the best among all the measures considered. The difference between our measure and the Prasad–Rao measure diminishes as α increases and their performances become almost identical when $\alpha = 1$, i.e., when we consider average over all areas. Also, the differences among various measures diminishes as m increases.

In Fig. 2, we plot the coverage probabilities for different α. For each replication, we found confidence intervals, i.e., $e_i^{\mathrm{EB}} \pm z_{0.025}\sqrt{V_i^j}$, where $z_{0.025}$ is the upper 2.5% point of the standard normal deviate and $j = \mathrm{N}, \mathrm{LL}, \mathrm{PR}, \mathrm{BOOT}$, and checked whether θ_i belonged to the confidence interval $(i = 1,\ldots,m)$. As in Fig. 1, we plot the average coverage probabilities for each method when the average is taken over a group of small areas such that $Y_i^2/(A+D) \geqslant \chi^2_{1;\alpha}$ for various values of α.

Here the conclusion is similar to that of Fig. 1.

Acknowledgements

The authors thank a referee for helpful comments and constructive suggestions. The second author's research was supported in part by National Science Foundation Grants SES-9511202 and SES 9978145 and a grant from the Gallup Organization.

Appendix

We shall assume the following regularity conditions throughout the paper. The regularity conditions will be referred to as (RC).

Regularity conditions (*RC*): (a) The elements of X_i and Z_i are uniformly bounded such that $\sum_{i=1}^m X_i' V_i^{-1}(\psi) X_i = [\mathrm{O}(m)]_{p\times p}$; $i = 1,\ldots,m$.

(b) $\sup_{i\geqslant 1} n_i < \infty$ and $\sup_{i\geqslant 1} k_i < \infty$.

(c) $l_i - X_i' V_i^{-1}(\psi) Z_i G_i(\psi) \lambda_i = [\mathrm{O}(1)]_{p\times 1}$.

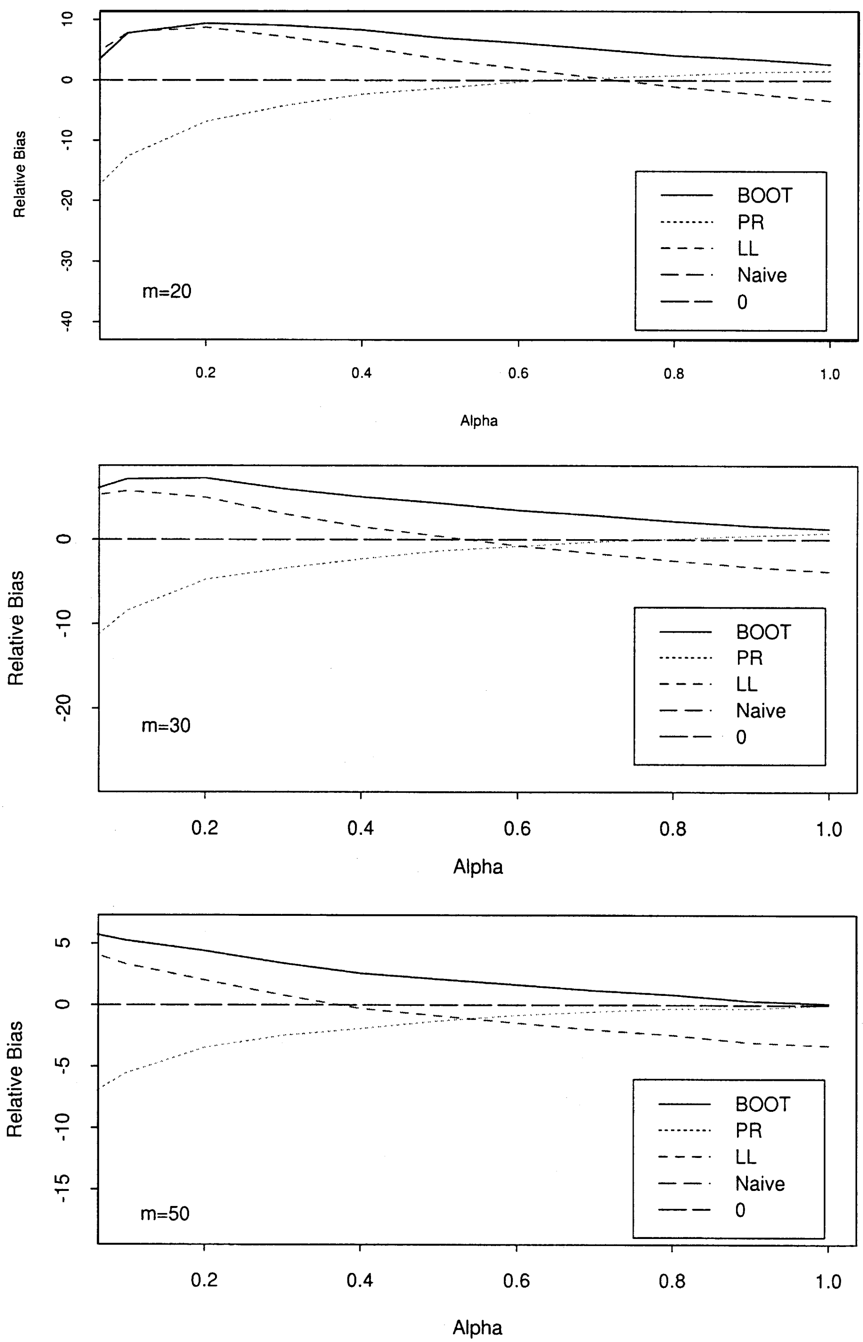

Fig. 1. Percent average relative biases of different measures of uncertainty of EB estimator.

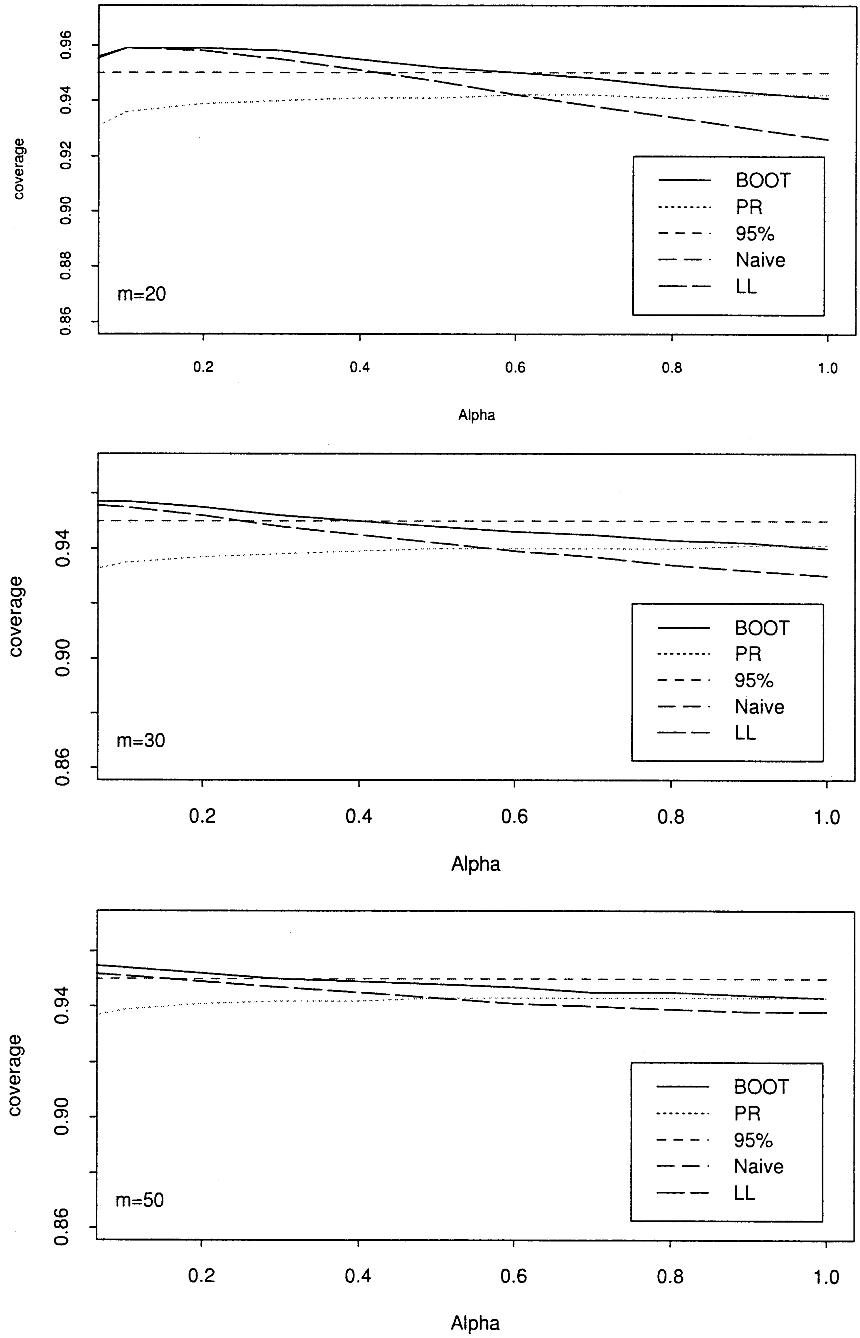

Fig. 2. Coverage probabilities of different confidence intervals with nominal 95%.

(d) $\partial/\partial\psi_j[X_i' V_i^{-1}(\psi)Z_i G_i(\psi)\lambda_i] = [\mathrm{O}(1)]_{p\times 1}$ for $j = 1,\dots,s$.

(e) $R_i(\psi) = \sum_{j=0}^{s} \psi_j D_{ij} D_{ij}'$ and $G_i(\psi) = \sum_{j=0}^{s} \psi_j F_{ij} F_{ij}'$, where $\psi_0 = 1$, D_{ij} and F_{ij} ($i=1,\dots,m$, $j=0,\dots,s$) are known matrices of order $n_i \times k_i$ and $k_i \times k_i$, respectively, and the elements are uniformly bounded known constants such that $R_i(\psi)$ and $G_i(\psi)$ ($i = 1,\dots,m$) are all positive definite matrices. In special cases, some of D_{ij} and F_{ij} may be null matrices.

(f) $\hat\psi$ is an estimator of ψ which satisfies (i) $\hat\psi - \psi = \mathrm{O}_\mathrm{p}(m^{-1/2})$, (ii) $\hat\psi = \hat\psi_{\mathrm{ML}} = \mathrm{O}_\mathrm{p}(m^{-1})$ (iii) $\hat\psi(-Y) = \hat\psi(Y)$ and (iv) $\hat\psi(Y + Xb) = \hat\psi(Y)$, for any $b \in R^p$ and for all Y, where $Y = \mathrm{col}_{1\leqslant i\leqslant m} Y_i$, $X = \mathrm{col}_{1\leqslant i\leqslant m} X_i$ and $\hat\psi_{\mathrm{ML}}$ is the maximum-likelihood estimator of ψ. Assume that $E(\hat\psi - \psi) = -m^{-1}B(\psi) + \mathrm{o}(m^{-1})$.

(g) $E(\hat\beta(\psi) - \beta)(\hat\psi - \psi)' = \mathrm{o}(m^{-1})$.

Note that conditions (a)–(f) were also needed in Datta and Lahiri (2000) (also see Prasad and Rao, 1990). Condition (g) is reasonable following an argument of Cox and Reid (1987).

Let $R_m = \mathrm{O}_\mathrm{p}(m^{-1})$ and $R_m^\star = \mathrm{O}_{\mathrm{p}^\star}(m^{-1})$ denote sequences of random variables such that mR_m and $mR_m^\star$ are bounded in probability under Models 1 and 2, respectively.

Lemma A.1. *Under Model* 1 *and the RC*:

(i) $E_\star[g_{1i}(\hat\psi^\star)] = g_{1i}(\hat\psi) - m^{-1}B'(\hat\psi)\nabla g_{1i}(\hat\psi) - g_{3i}(\hat\psi) + \mathrm{o}_\mathrm{p}(m^{-1})$,
(ii) $E_\star[g_{2i}(\hat\psi^\star)] = g_{2i}(\hat\psi) + \mathrm{o}_\mathrm{p}(m^{-1})$,
(iii) $E_\star[\hat\theta_i(Y_i;\hat\beta(\hat\psi^\star),\hat\psi^\star) - \hat\theta_i(Y_i;\hat\beta(\hat\psi),\hat\psi)]^2 = g_{4i}(\hat\psi;Y_i) + \mathrm{o}_\mathrm{p}(m^{-1})$ *and*
(iv) $V_\star[\hat\theta_i(Y_i;\hat\beta^\star(\hat\psi^\star),\hat\psi^\star)] = g_{4i}(\hat\psi;Y_i) + \mathrm{o}_\mathrm{p}(m^{-1})$.

Proof of Lemma A.1. Parts (i) and (ii) follow immediately after an application of parts (a) and (b) of Theorem A.2 of Datta and Lahiri (2000) on Model 2. To prove part (iii), first note that (use Eq. (A.2) of Datta and Lahiri, 2000 on Model 2)

$$\hat\theta_i(Y_i;\hat\beta(\hat\psi^\star),\hat\psi^\star) - \hat\theta_i(Y_i;\hat\beta(\hat\psi),\hat\psi) = (\hat\psi^\star - \hat\psi)' L_i(\hat\psi)(Y_i - X_i\hat\beta(\hat\psi)) + \mathrm{O}_{\mathrm{p}^\star}(m^{-1}), \tag{A.1}$$

and then use algebra.

Let us now turn our attention to the proof of part (iv). Using (A.1), $E_\star\hat\beta^\star(\hat\psi) = \hat\beta(\hat\psi)$ and $E_\star(\hat\psi^\star - \hat\psi) = \mathrm{O}_\mathrm{p}(m^{-1})$, we get

$$E_\star[\hat\theta_i(Y_i;\hat\beta^\star(\hat\psi^\star),\hat\psi^\star)] = \hat\theta_i(Y_i;\hat\beta(\hat\psi),\hat\psi) + \mathrm{O}_\mathrm{p}(m^{-1}). \tag{A.2}$$

Now using (A.2), we get

$$\begin{aligned}
&V_\star[\hat\theta_i(Y_i;\hat\beta^\star(\hat\psi^\star),\hat\psi^\star)] \\
&\quad = E_\star[\hat\theta_i(Y_i;\hat\beta^\star(\hat\psi^\star),\hat\psi^\star) - E_\star\{\hat\theta_i(Y_i;\hat\beta^\star(\hat\psi^\star),\hat\psi^\star)\}]^2 \\
&\quad = E_\star[\hat\theta_i(Y_i;\hat\beta^\star(\hat\psi^\star),\hat\psi^\star) - \hat\theta_i(Y_i;\hat\beta(\hat\psi),\hat\psi) + \mathrm{O}_\mathrm{p}(m^{-1})]^2 \\
&\quad = E_\star[\hat\theta_i(Y_i;\hat\beta^\star(\hat\psi^\star),\hat\psi^\star) - \hat\theta_i(Y_i;\hat\beta(\hat\psi),\hat\psi)]^2 + \mathrm{O}_\mathrm{p}(m^{-2})
\end{aligned} \tag{A.3}$$

since the cross-product term in (A.3) is also order $\mathrm{O}_\mathrm{p}(m^{-2})$.

Define $Q_1 = \hat{\theta}_i(Y_i; \hat{\beta}^\star(\hat{\psi}), \hat{\psi}), -\hat{\theta}_i(Y_i; \hat{\beta}(\hat{\psi}), \hat{\psi})$ and $Q_2 = \hat{\theta}_i(Y_i; \hat{\beta}^\star(\hat{\psi}^\star), \hat{\psi}^\star) - \hat{\theta}_i(Y_i; \hat{\beta}^\star(\hat{\psi}), \hat{\psi})$. Then the first term in the r.h.s of (A.3) is equal to

$$E_\star Q_1^2 + E_\star Q_2^2 + E_\star Q_1 Q_2. \tag{A.4}$$

Using (1) and (2) and, some algebra, we get

$$Q_1 = \{l_i - X_i' V_i^{-1}(\hat{\psi}) Z_i G_i(\hat{\psi}) \lambda_i\}' [\hat{\beta}^\star(\hat{\psi}) - \hat{\beta}(\hat{\psi})], \tag{A.5}$$

which gives $E_\star Q_1^2 = g_{2i}(\hat{\psi})$.

Using argument similar to the one given in the proof of Theorem A.1 of Datta and Lahiri (2000), we get $Q_2 = (\hat{\psi}^\star - \hat{\psi})' L_i(\hat{\psi})(Y_i - X_i \hat{\beta}(\hat{\psi})) + O_{p^\star}(m^{-1})$. Using this and some algebra, we get $E_\star Q_2^2 = g_{4i}(\hat{\psi}; Y_i) + o_p(m^{-1})$. Using (A.1) and (A.5), we get

$$\begin{aligned} E_\star Q_1 Q_2 &= [l_i - X_i' V_i^{-1}(\hat{\psi}) Z_i G_i(\hat{\psi}) \lambda_i]' E_\star \{(\hat{\beta}^\star(\hat{\psi}) - \hat{\beta}(\hat{\psi})) \\ &\quad \times (\hat{\psi}^\star - \hat{\psi})'\} L_i(\hat{\psi})(Y_i - X_i \hat{\beta}(\hat{\psi})) + o_p(m^{-1}) \\ &= o_p(m^{-1}), \end{aligned} \tag{A.6}$$

since $E_\star[(\hat{\beta}^\star(\hat{\psi}) - \hat{\beta}(\hat{\psi}))(\hat{\psi}^\star - \hat{\psi})'] = o_p(m^{-1})$ by (g) of RC.

Now using (A.3)–(A.6), we get part (iv). □

Theorem A.1. *Under Model* 1 *and the RC, we have*

(i) $V_i^{\text{BOOT}} = g_{1i}(\hat{\psi}) + m^{-1} B'(\hat{\psi}) \nabla g_{1i}(\hat{\psi}) + g_{2i}(\hat{\psi}) + g_{3i}(\hat{\psi}) + g_{4i}(\hat{\psi}; Y_i) + o_p(m^{-1})$ *and*

(ii) $V_i^{\text{LL}} = g_{1i}(\hat{\psi}) + g_{2i}(\hat{\psi}) + g_{4i}(\hat{\psi}; Y_i) - g_{3i}(\hat{\psi}) - m^{-1} B'(\hat{\psi}) \nabla g_{1i}(\hat{\psi}) + o_p(m^{-1})$.

Proof. Part (i) follows from parts (i)–(iii) of Lemma A.1. Part (ii) follows from parts (i) and (iv) of Lemma A.1. □

Theorem A.2. *Under Model* 1 *and the RC, we have*

(i) $E[V_i^{\text{BOOT}}] = r(\hat{\theta}_i^{\text{EB}}) + o(m^{-1})$ *and*

(ii) $E[V_i^{\text{LL}}] = r(\hat{\theta}_i^{\text{EB}}) - 2m^{-1} B'(\psi) \nabla g_{1i}(\psi) - g_{3i}(\psi) + o(m^{-1})$.

Proof. Since $\hat{\psi} - \psi = o_p(1)$ and $\Sigma(\psi) = O(m^{-1})$, we have $L_i(\hat{\psi}) = L_i(\psi) + o_p(1)$, $\hat{\beta}(\hat{\psi}) = \beta + o_p(1)$ and $\Sigma(\hat{\psi}) = \Sigma(\psi) + o_p(m^{-1})$. Thus,

$$\begin{aligned} &L_i(\hat{\psi})[(Y_i - X_i \hat{\beta}(\hat{\psi}))(Y_i - X_i \hat{\beta}(\hat{\psi}))'] L_i'(\hat{\psi}) \Sigma(\hat{\psi}) \\ &\quad = L_i(\psi)[(Y_i - X_i \beta)(Y_i - X_i \beta)'] L_i'(\psi) \Sigma(\psi) + o_p(m^{-1}). \end{aligned} \tag{A.7}$$

Using (A.7) and the expressions for $g_{3i}(\psi)$ and $g_{4i}(\hat{\psi}; Y_i)$, we get

$$E[g_{4i}(\hat{\psi}; Y_i)] = g_{3i}(\psi) + o(m^{-1}). \tag{A.8}$$

Using (5), (A.8), $m^{-1} B'(\hat{\psi}) \nabla g_{1i}(\hat{\psi}) = m^{-1} B'(\psi) \nabla g_{1i}(\psi) + o_p(m^{-1})$ (which follows from $\hat{\psi} - \psi = o_p(1)$), $g_{3i}(\hat{\psi}) = g_{3i}(\psi) + o_p(m^{-1})$ and the expression for V_i^{BOOT}, part (i) of the theorem follows. Now, using part (ii) of Theorem A.1, (5), (A.8) and part (i) of this theorem, part (ii) of the theorem follows. □

References

Battese, G.E., Harter, R.M., Fuller, W.A., 1988. An error-components Model for prediction of county crop areas using survey and satellite data. J. Amer. Statist. Assoc. 80, 28–36.

Butar, B.F., 1997. Empirical Bayes methods in survey sampling. Ph.D. Thesis, Department of Mathematics and Statistics, University of Nebraska-Lincoln, unpublished.

Carter, G.M., Rolph, J.E., 1974. Empirical Bayes methods applied to estimating fire alarm probabilities. J. Amer. Statist. Assoc. 69, 269–277.

Cox, D.R., Reid, N., 1987. Parameter orthogonality and approximate conditional inference. J. Roy. Statist. Soc. B 49, 1–39 (with discussion).

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

Efron, B., Morris, C., 1975. Data analysis using stein's estimator and its generalizations. J. Amer. Statist. Assoc. 70, 311–319.

Fay, R.E., Herriot, R.A., 1979. Estimates of income for small places: an application of James–Stein procedure to census data. J. Amer. Statist. Assoc. 74, 269–277.

Ghosh, M., Lahiri, P., 1987. Robust empirical Bayes estimation of means from stratified samples. J. Amer. Statist. Assoc. 82, 1153–1162.

Ghosh, M., Meeden, G., 1986. Empirical Bayes estimation in finite population sampling. J. Amer. Statist. Assoc. 81, 1058–1062.

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

Ghosh, M., Rao, J.N.K., 1994. Small area estimation: an appraisal. Statist. Sci. 9 (1), 55–93 (with discussion).

Kackar, R.N., Harville, D.A., 1984. Approximations for standard errors of estimators of fixed and random effects in mixed linear models. J. Amer. Statist. Assoc. 79, 853–862.

Laird, N.M., Louis, T.A., 1987. Empirical Bayes confidence intervals based on bootstrap samples. J. Amer. Statist. Assoc. 82, 739–750.

Meza, J.L., 2003. Empirical Bayes estimation smoothing of relative risks in disease mapping. J. Statist. Plann. Inference 112, this issue.

Morris, C.N., 1983. Parametric empirical Bayes inference: theory and applications. J. Amer. Statist. Assoc. 78, 47–59.

Prasad, N.G.N., Rao, J.N.K., 1990. The estimation of mean squared errors of small area estimators. J. Amer. Statist. Assoc. 85, 163–171.

Rao, J.N.K., 1999. Some recent advances in model-based small area estimation. Surv. Methodol. 25, 175–186.

Rao, J.N.K., 2001. EB and EBLUP in small area estimation. In: Ahmed, S.E., Reid, N. (Eds.), Empirical Bayes and Likelihood Inference. Springer, Berlin, pp. 33–44.