Journal of Multivariate Analysis 101 (2010) 882–892

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

An adjusted maximum likelihood method for solving small area estimation problems

Huilin Li [a,*], P. Lahiri [b]

[a] Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD 20892, United States

[b] Joint Program of Survey Methodology, University of Maryland, College Park, MD 20742, United States

ARTICLE INFO

Article history:
Received 25 September 2008
Available online 31 October 2009

AMS subject classifications:
62F12
62F40
62J99

Keywords:
Adjusted density maximization estimator
The Fay–Herriot model
Parametric bootstrap
Prediction intervals

ABSTRACT

For the well-known Fay–Herriot small area model, standard variance component estimation methods frequently produce zero estimates of the strictly positive model variance. As a consequence, an empirical best linear unbiased predictor of a small area mean, commonly used in small area estimation, could reduce to a simple regression estimator, which typically has an overshrinking problem. We propose an adjusted maximum likelihood estimator of the model variance that maximizes an adjusted likelihood defined as a product of the model variance and a standard likelihood (e.g., a profile or residual likelihood) function. The adjustment factor was suggested earlier by Carl Morris in the context of approximating a hierarchical Bayes solution where the hyperparameters, including the model variance, are assumed to follow a prior distribution. Interestingly, the proposed adjustment does not affect the mean squared error property of the model variance estimator or the corresponding empirical best linear unbiased predictors of the small area means in a higher order asymptotic sense. However, as demonstrated in our simulation study, the proposed adjustment has a considerable advantage in small sample inference, especially in estimating the shrinkage parameters and in constructing the parametric bootstrap prediction intervals of the small area means, which require the use of a strictly positive consistent model variance estimate.

Published by Elsevier Inc.

1. Introduction

The Fay–Herriot model [1], widely used in small area estimation, consists of two levels. In Level 1, we have the sampling model,

$$Y_i|\theta_i \sim N(\theta_i, D_i), \quad i = 1, \ldots, m,$$

independently for each i. In Level 2, we have the linking model,

$$\theta_i \sim N(x_i'\beta, A), \quad i = 1, \ldots, m,$$

also independently for each i.

Level 1 accounts for the sampling variability of the regular survey estimates Y_i of true small area means θ_i. Level 2 links θ_i to a vector of p known auxiliary variables $x_i = (x_{i1}, \ldots, x_{ip})'$, often obtained from administrative and census records. The sampling variances D_i are assumed to be known.

The Fay–Herriot model has been widely used in small area estimation and related problems for a variety of reasons, including its simplicity, its ability to protect confidentiality of microdata and its ability to produce design-consistent

* Corresponding author.
E-mail addresses: lih5@mail.nih.gov (H. Li), plahiri@survey.umd.edu (P. Lahiri).

0047-259X/$ – see front matter. Published by Elsevier Inc.
doi:10.1016/j.jmva.2009.10.009

estimators. Some earlier applications of the Fay–Herriot model include the estimation of: (i) false alarm probabilities in New York city [2]; (ii) the batting averages of major league baseball players [3]; and (iii) prevalence of toxoplasmosis in El Salvador [3]. More recently, the Fay–Herriot model was used: to estimate poverty rates for the US states, counties, and school districts [4] and to estimate proportions at the lowest level of literacy for states and counties [5]. For a comprehensive review of the theory and applications of the above model, see [6, Chapter 7].

The best predictor (BP) of θ_i and the associated mean squared prediction error (MSPE) are given by

$$\hat{\theta}_i^{\mathrm{BP}} = Y_i - B_i(Y_i - x_i'\beta),$$

and

$$\mathrm{MSPE}[\hat{\theta}_i^{\mathrm{BP}}] = \mathrm{E}[\hat{\theta}_i^{\mathrm{BP}} - \theta_i]^2 = g_{1i}(B_i),$$

where $0 < B_i = D_i/(A+D_i) < 1$, $i = 1, \ldots, m$; E is the expectation with respect to the joint distribution of Y and θ induced by the Fay–Herriot model and $g_{1i}(B_i) = D_i(1 - B_i)$.

The best predictor shrinks the direct estimator Y_i towards the regression surface $x_i'\beta$, the amount of shrinkage being determined by the shrinkage factor B_i. The closer the value of B_i to 1, the greater the strength of the Level 2 model and hence the greater the efficiency of the best predictor, as reflected by a smaller value of the mean squared prediction error of the best predictor. When $A = 0$, that is when the Level 2 model is perfect, $B_i = 1$ for all $i = 1, \ldots, m$. In this case, the best predictor is identical to the regression estimator. This situation, however, is unrealistic, since Level 2 modeling, just like any modeling, cannot be perfect, that is A should be always greater than 0. Thus, throughout the paper we assume $A > 0$.

In practice, both β and the B_i's are unknown and need to be estimated from the data. The regression parameter β is estimated by the weighted least square estimator $\tilde{\beta}^{\mathrm{w}} = (\sum_{j=1}^m x_j x_j' B_j/D_j)^{-1} \sum_{j=1}^m x_j Y_j B_j/D_j$. When this estimator of β is plugged into the best predictor, the best linear unbiased predictor (BLUP) of θ_i is obtained and is denoted by $\hat{\theta}_i^{\mathrm{BLUP}}$. It is now clear that the shrinkage factors B_i are important parameters to estimate. They are needed for a good evaluation of the Level 2 model and to carry out the necessary prediction analyses. When estimates of B_i are plugged into the best linear unbiased predictor formula, one obtains an empirical best linear unbiased predictor (EBLUP) of θ_i, denoted by $\hat{\theta}_i^{\mathrm{EBLUP}}$.

From Jensen's inequality and the convexity of B_i as a function of A, it follows that $\hat{B}_i$ overestimates B_i even when an exactly unbiased estimator of A is used, and the extent of the overestimation may be severe for small m. In addition, standard methods of estimation of A considered in the literature, including using the Prasad–Rao simple method-of-moments estimator, $\hat{A}^{\mathrm{PR}}$ [7], the Fay–Herriot method-of-moments estimator, $\hat{A}^{\mathrm{FH}}$ [1,8], the maximum likelihood estimator, $\hat{A}^{\mathrm{ML}}$, and the residual maximum likelihood estimator, $\hat{A}^{\mathrm{RE}}$, are all subject to zero estimates resulting in undesirable estimates $\hat{B}_i = 1$ for all $i = 1, \ldots, m$. In real life data analyses, this problem is quite frequent (see, e.g., [9] and [10]).

In Section 2, we introduce an adjustment to the maximum (profile or residual) likelihood estimator of A in order to produce a strictly positive estimate of A, for small m. The proposed adjustment increases the order of bias of the residual maximum likelihood estimator, but not the mean squared error, up to the order $O(m^{-1})$. However, the mean squared error or the bias property of the maximum profile likelihood estimator of A remains unaffected, up to order $O(m^{-1})$. In terms of the estimation of the shrinkage factors B_i, the adjustment does not increase the order of the bias or the mean squared error, irrespective of whether a profile or residual likelihood function is used for the adjustment. While there is no clear advantage of using the proposed adjusted maximum likelihood methods for large m, they have a clear edge over the standard methods for small m in terms of preventing the full shrinkage.

Morris [11] proposed a method, known as the adjusted density maximization (ADM) method, as an intermediary step in approximating a hierarchical Bayes solution. Recently, Morris and Tang [12] (also see [13]) pursued the ADM method for the Fay–Herriot model. The ADM approximations to the posterior means of A and B_i, under an (improper) uniform prior on β and superharmonic prior [14] on A, are identical to the corresponding adjusted maximum residual likelihood estimators given in this paper. However, unlike Morris and Tang [12], we consider a classical prediction approach, which does not assume a prior distribution for β and A, in measuring the uncertainty of the proposed EBLUP and the associated prediction interval. Moreover, for the Fay–Herriot model, Morris and Tang [12] did not suggest the adjusted maximum profile likelihood estimator, which appears to perform better than the adjusted maximum residual likelihood estimator in our simulation study.

The mean squared prediction errors of empirical best linear unbiased predictors of θ_i that use the proposed adjusted maximum likelihood estimators are presented in Section 3. In this section, we also provide the second-order (or nearly) unbiased estimators of the mean squared prediction errors of empirical best linear unbiased predictors when the proposed adjusted maximum likelihood estimators are used. The use of the proposed adjusted maximum likelihood estimators of A does not affect the mean squared prediction errors of empirical best linear unbiased predictors, up to the order $O(m^{-1})$. However, the expressions for the proposed nearly unbiased mean squared prediction error estimators are different for different methods of estimating A.

Cox [15] and Morris [16] proposed normality based empirical Bayes confidence intervals of θ_i. The coverage errors of such intervals are typically of order $O(m^{-1})$. Chatterjee, Lahiri and Li [17] proposed an improved interval estimation using the parametric bootstrap method. The method requires repeated generation of a pivotal quantity from several bootstrap samples. A strictly positive estimate of A is absolutely needed for this method since the pivotal quantity is undefined when the A estimate is zero. A crude fix is to take a small positive number whenever the A estimate turns out to be zero. But, in

our simulation study, the results are found to be sensitive to the choice of this ad hoc positive truncation point. The adjusted maximum likelihood method offers a sensible solution to this important problem in the parametric bootstrap method. In Section 4, we obtain the coverage accuracy of the parametric bootstrap prediction intervals proposed by Chatterjee, Lahiri and Li [17] when the regression coefficients are estimated by the weighted least square method with A estimated by the adjusted maximum likelihood method.

In Section 5, we present results from a Monte Carlo simulation. All the technical proofs are presented in the Appendix.

2. Adjusted maximum likelihood methods

Define $X = (x_1, \ldots, x_m)'$, $v = (v_1, \ldots, v_m)'$, $e = (e_1, \ldots, e_m)'$, and $Y = (Y_1, \ldots, Y_m)'$. The Fay–Herriot model can be rewritten as

$$Y = X\beta + v + e, \tag{1}$$

a special case of the general linear mixed model with block diagonal variance–covariance structure: $\Sigma = D + A$, where $D = \mathrm{diag}(D_1, \ldots, D_m)$.

Let $y = (y_1, \ldots, y_m)'$ denote the observed value of Y. We propose an adjusted likelihood of A as

$$L_{\mathrm{adj}}(A) = A \times L(A),$$

where $L(A)$ is a given likelihood function. In this paper, we consider two choices for $L(A)$: the profile likelihood function given by

$$L_{\mathrm{P}}(A) = c|\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}y'Py\right\},$$

and the residual likelihood given by

$$L_{\mathrm{RE}}(A) = c|X'\Sigma^{-1}X|^{-1/2}|\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}y'Py\right\},$$

where c is a generic constant free from A and $P = \Sigma^{-1} - \Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}$.

The adjusted maximum likelihood estimator of A is obtained by maximizing the proposed adjusted likelihood $L_{\mathrm{adj}}(A)$ or, equivalently, the corresponding log-likelihood function of A, denoted by $l_{\mathrm{adj}}(A)$. The adjusted maximum profile likelihood and adjusted maximum residual likelihood estimators of A are obtained by choosing $L_{\mathrm{P}}(A)$ and $L_{\mathrm{RE}}(A)$ for $L(A)$, respectively.

The adjusted maximum likelihood estimators of A are strictly positive (see the Appendix), even for small m. Under the Fay–Herriot model (1) and regularity conditions (r.1)–(r.3) given in the Appendix, we have

$$\mathrm{E}(\hat{A} - A)^2 = \frac{2}{\mathrm{tr}(\Sigma^{-2})} + o(m^{-1}), \tag{2}$$

for both adjusted maximum likelihood estimators of A, and

$$\begin{aligned}\mathrm{Bias}(\hat{A}) &= \mathrm{E}[\hat{A}] - A \\ &= \begin{cases} \dfrac{\mathrm{tr}(P - \Sigma^{-1}) + 2/A}{\mathrm{tr}(\Sigma^{-2})} + o(m^{-1}) & \text{if } \hat{A} = \hat{A}^{\mathrm{AM}} \\ \dfrac{2/A}{\mathrm{tr}(\Sigma^{-2})} + o(m^{-1}) & \text{if } \hat{A} = \hat{A}^{\mathrm{AR}}. \end{cases}\end{aligned} \tag{3}$$

Thus, under the regularity conditions, the adjusted maximum likelihood estimators of A are consistent, for large m. In addition, under the same asymptotic setting, the mean squared errors of the adjusted and unadjusted maximum (profile or residual) likelihood estimators of A are all equivalent, up to order $O(m^{-1})$. The biases of the adjusted maximum (profile or residual) likelihood estimators of A are of order $O(m^{-1})$, the same as the order of the corresponding profile maximum likelihood estimator, but higher than that of the residual maximum likelihood estimator.

The plugged-in estimators of the shrinkage factors B_i, i.e. $\hat{B}_i^{\mathrm{AM}} = D_i/(\hat{A}^{\mathrm{AM}} + D_i)$ and $\hat{B}_i^{\mathrm{AR}} = D_i/(\hat{A}^{\mathrm{AR}} + D_i)$, are all strictly less than 1, even for small m, and are also consistent, for large m and under the same regularity conditions. In our simulation study, $\hat{B}_i^{\mathrm{AM}}$ performed the best, in terms of both bias and mean squared error criteria, among all the well-known estimators of the shrinkage factors B_i. Moreover, in our simulation, adjusted maximum likelihood estimators usually produced negative biases while unadjusted maximum likelihood estimators produced positive biases. Thus, direct estimators tend to get more weights in the EBLUP method if adjusted maximum likelihood estimators of B_i are used.

3. Estimation of the mean squared prediction error

An empirical best linear unbiased predictor of θ_i is given by

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

where $\hat{B}_i = D_i/(\hat{A} + D_i)$, $i = 1, \ldots, m$, and $\hat{\beta}^{\mathrm{w}} = (\sum_{j=1}^m x_j x_j'\hat{B}_j/D_j)^{-1} \sum_{j=1}^m x_j Y_j \hat{B}_j/D_j$, the weighted least square estimator of β with B_j replaced by the adjusted maximum likelihood estimators $\hat{B}_j$.

We can measure the uncertainty of the proposed empirical best linear unbiased predictor of θ_i by its mean squared prediction error, defined as

$$\mathrm{MSPE}(\hat{\theta}_i^{\mathrm{EBLUP}}) = \mathrm{E}(\hat{\theta}_i^{\mathrm{EBLUP}} - \theta_i)^2,$$

where E is the expectation with respect to the joint distribution of Y and θ induced by the Fay–Herriot model (1). In small area estimation, there is continuing interest in obtaining second-order (or nearly) unbiased estimators of $\mathrm{MSPE}(\hat{\theta}_i^{\mathrm{EBLUP}})$ following the pioneering work of Prasad and Rao [7]. An estimator, denoted as $\mathrm{mspe}(\hat{\theta}_i^{\mathrm{EBLUP}})$, is called a second-order unbiased (or nearly unbiased) estimator of $\mathrm{MSPE}(\hat{\theta}_i^{\mathrm{EBLUP}})$ if

$$\mathrm{E}[\mathrm{mspe}(\hat{\theta}_i^{\mathrm{EBLUP}})] = \mathrm{MSPE}(\hat{\theta}_i^{\mathrm{EBLUP}}) + o(m^{-1}).$$

Some important papers that use the Taylor series linearization method in obtaining second-order unbiased MSPE estimators include [18–20]. See [6,21] for a detailed account of the Taylor linearization method in small area estimation. Since the Taylor linearization technique is fairly well-known in small area estimation, we omit the technical details in obtaining second-order unbiased mean squared prediction error estimators of our proposed empirical best linear unbiased predictors and refer the interested readers to Li [22] for details.

Using [23], we have

$$\begin{aligned}\mathrm{MSPE}(\hat{\theta}_i^{\mathrm{EBLUP}}) &= \mathrm{E}(\hat{\theta}_i^{\mathrm{BLUP}} - \theta_i)^2 + \mathrm{E}(\hat{\theta}_i^{\mathrm{EBLUP}} - \hat{\theta}_i^{\mathrm{BLUP}})^2 \\ &= g_{1i}(B_i) + g_{2i}(B) + \mathrm{E}(\hat{\theta}_i^{\mathrm{EBLUP}} - \hat{\theta}_i^{\mathrm{BLUP}})^2, \end{aligned} \tag{4}$$

where

$$g_{1i}(B_i) = D_i(1 - B_i),$$

and

$$g_{2i}(B) = B_i^2 x_i' \left(\sum_{j=1}^{m} \frac{B_j}{D_j} x_j x_j' \right)^{-1} x_i,$$

with $B = (B_1, \ldots, B_m)'$. Note that $g_{1i}(B_i)$ and $g_{2i}(B)$ do not depend on the method of estimating the shrinkage factors B_i.

The third term in (4) captures the additional uncertainty due to the estimation of A or, equivalently, the shrinkage factors B_i. It has no closed-form expression, but one can approximate it up to order $O(m^{-1})$, the same as the order of $g_{2i}(B)$, by the standard Taylor linearization method. For both adjusted methods, we obtain

$$\mathrm{E}(\hat{\theta}_i^{\mathrm{EBLUP}} - \hat{\theta}_i^{\mathrm{BLUP}})^2 = g_{3i}(B) + o(m^{-1}),$$

where

$$g_{3i}(B) = 2\frac{B_i^3}{D_i} \left\{ \sum_{j=1}^{m} \left(\frac{B_j}{D_j} \right)^2 \right\}^{-1},$$

under regularity conditions (r.1)–(r.3). Thus,

$$\mathrm{MSPE}[\hat{\theta}_i^{\mathrm{EBLUP}}] = \widetilde{\mathrm{MSPE}}[\hat{\theta}_i^{\mathrm{EBLUP}}] + o(m^{-1}),$$

where

$$\widetilde{\mathrm{MSPE}}[\hat{\theta}_i^{\mathrm{EBLUP}}] = g_{1i}(B_i) + g_{2i}(B) + g_{3i}(B), \tag{5}$$

is the second-order approximation to $\mathrm{MSPE}[\hat{\theta}_i^{\mathrm{EBLUP}}]$. Datta et al. [20] provided the corresponding second-order approximations to the MSPE of EBLUP for four different methods of estimating A: using the maximum likelihood and residual maximum likelihood, Fay–Herriot and Prasad–Rao methods. They noted that in terms of the MSPE of EBLUP, the maximum profile likelihood and residual maximum likelihood methods are equivalent and superior to both the Fay–Herriot and the Prasad–Rao methods, up to the order $O(m^{-1})$.

It is interesting to note that the proposed adjusted maximum likelihood methods retain the same level of efficiency as those of the maximum likelihood and residual maximum likelihood methods in terms of the mean squared prediction error, up to the order $O(m^{-1})$. Moreover, the adjusted maximum likelihood methods protect the empirical best linear unbiased predictors from the overshrinkage problem since they both ensure strictly positive estimators of A even for small m—this nice property is not shared by the usual unadjusted maximum (profile or residual) likelihood or other methods considered in the literature.

Note that the second-order approximation $\widetilde{\mathrm{MSPE}}[\hat{\theta}_i^{\mathrm{EBLUP}}]$ involves the unknown shrinkage factors B_i and thus cannot be used to assess the uncertainty of empirical best linear unbiased predictors for a given data set. However, this second-order approximation is useful in obtaining a second-order unbiased estimator of the mean squared prediction error of our empirical best linear unbiased predictor.

Using algebra similar to that of Datta et al. [20], we obtain the following second-order (or nearly) unbiased estimator of the mean squared prediction error of $\hat{\theta}_i^{\text{EBLUP}}$:

$$\text{mspe}(\hat{\theta}_i^{\text{EBLUP}}) = g_{1i}(\hat{B}_i) + g_{2i}(\hat{B}) + 2g_{3i}(\hat{B}) - (\hat{B}_i)^2\widehat{\text{Bias}}(\hat{A}),$$

where $\widehat{\text{Bias}}(\hat{A})$ is obtained from (3).

4. The parametric bootstrap prediction interval

Following [15], one can propose the following empirical best prediction interval for θ_i : $\hat{\theta}_i^{\text{EBLUP}} \pm z_{\alpha/2}\hat{\sigma}_i$, where $\hat{\sigma}_i^2 = g_{1i}(\hat{B}_i)$ and $z_{\alpha/2}$ is the upper $100(1-\alpha/2)\%$ point of the standard normal distribution. This prediction interval is asymptotically correct in the sense that the coverage probability converges to $1-\alpha$, for large m and regularity conditions (r.1)–(r.3). However, for small m, this is not efficient since the coverage error of this interval is of order $O(m^{-1})$, which is not accurate enough for most small area applications. See [17] for a review of different attempts to improve on Cox's prediction interval.

For a general linear mixed model, Chatterjee et al. [17] proposed a parametric bootstrap method for obtaining a prediction interval directly from the bootstrap histogram of the pivot $\hat{\sigma}_i^{-1}(\theta_i - \hat{\theta}_i^{\text{EBLUP}})$. However, they developed their theory for the ordinary least square estimator of β. In Theorem 4.1, presented below, we extend the results of Chatterjee, Lahiri and Li [17] to the case where β is estimated by the weighted least squares method and the model variance A by the adjusted maximum (profile or residual) likelihood method. Hall and Maiti [24] proposed an alternative parametric bootstrap approach, but, as Rao [25] pointed out, their method is more like a synthetic method.

A prediction interval of θ_i can be constructed based on $\mathcal{L}_i$, the distribution of $\hat{\sigma}_i^{-1}(\theta_i - \hat{\theta}_i^{\text{EBLUP}})$. Cox's interval is based on the assumption that we can adequately approximate $\mathcal{L}_i$ by a standard normal distribution. In this paper, we improve on this approximation to $\mathcal{L}_i$ using a parametric bootstrap method. Let

$$Y_i^* = x_i\hat{\beta}^{\text{w}} + v_i^* + e_i^*,$$

where $v_i^* \overset{iid}{\sim} N(0, \hat{A})$ and $e_i^* \overset{ind}{\sim} N(0, D_i)$, for $i = 1, \ldots, m$, are independent of one another. Let $\hat{\beta}^{\text{w}*}$, $\hat{A}^*$, $\hat{\theta}_i^{\text{EBLUP}*}$, and $\hat{\sigma}_i^*$ be based on Y^*, and $\theta_i^* = x_i\hat{\beta}^{\text{w}} + v_i^*$. The distribution of $\hat{\sigma}_i^{*-1}(\theta_i^* - \hat{\theta}_i^{\text{EBLUP}*})$, conditional on the data Y, is the parametric bootstrap approximation $\mathcal{L}_i^*$ of $\mathcal{L}_i$.

Our parametric bootstrap prediction interval is given in the following theorem.

Theorem 4.1. *With $\hat{\beta}^{\text{w}}$ and $\hat{A}$, for a preassigned $\alpha \in (0, 1)$ and arbitrary $i = 1, \ldots, m$, let q_1 and q_2 be real numbers such that*

$$\mathcal{L}_i^*(q_2) - \mathcal{L}_i^*(q_1) = 1 - \alpha.$$

Then, under the regularity conditions (r.1)–(r.3) *given in the Appendix, we have*

$$\text{pr}[\hat{\theta}_i^{\text{EBLUP}} - q_1\hat{\sigma}_i \le \theta_i \le \hat{\theta}_i^{\text{EBLUP}} + q_2\hat{\sigma}_i] = 1 - \alpha + O(m^{-3/2}).$$

The details of the proof are lengthy and quite technical in nature. We refer the readers to [22] for details.

5. Monte Carlo simulation

The finite-sample accuracy of different estimators of the shrinkage factors B_i and different prediction intervals of θ_i are investigated in this section through a Monte Carlo simulation study. Such a simulation exercise is common in small area estimation literature; see, for example, [20,26–28], and others.

In order to study the effect of m, we consider $m = 15$ and 45. With the increase of m, all methods considered in this simulation study get better and come closer to one another, supporting our asymptotic theory. However, it is interesting to note that even for $m = 45$, the adjusted likelihood estimators, especially the adjusted profile maximum likelihood estimator, perform considerably better than the standard methods for most of the groups. In the following, we provide detailed discussions only for $m = 15$, but the results for $m = 45$ are provided in Tables 1–5.

For $m = 15$, we consider an unbalanced case that corresponds to the Type III sampling variance pattern of Datta et al. [20]. Thus, there are five groups G_t ($t = 1, \ldots, 5$) of small areas such that within each group the sampling variances are the same. Specifically, $D_j = 4.0j \in G_1$; $D_j = 0.6j \in G_2$; $D_j = 0.5j \in G_3$; $D_j = 0.4j \in G_4$; $D_j = 0.2j \in G_5$. For $m = 45$, we simply increase the number of small areas from 3 to 9 in the above setting. Simulation results for other sampling variance patterns for $m = 15$ are reported in [22].

We generate $N = 10{,}000$ independent data sets $\{Y_i, i = 1, \ldots, 15\}$ using a simplified Fay–Herriot model $Y_i = v_i + e_i$, where v_i and e_i ($i = 1, \ldots, m$) are all independent with $v_i \sim \text{N}(0, A = 1)$, and $e_i \sim \text{N}(0, D_i)$. Since the estimators of A considered here are translation invariant, we take $x_i'\beta = 0$, without any loss of generality. However, to account for the uncertainty in the estimation of the common mean that arises in practice, we still estimate the zero mean. Since all the small areas are exchangeable in each group, we only report the group means for all the criteria.

Table 1
Percentages of zero estimates of A for different estimation methods.

m	PR	FH	RE	ML	AR	AM
15	12.15	4.11	0.99	3.96	0	0
45	1.28	0.09	0	0.01	0	0

Table 2
Comparison of different estimators of B_i for $\mathbf{m} = \mathbf{15}$ and $m = 45$.

Percentage relative biases						
Group	PR	FH	RE	ML	AR	AM
G1	**2.5**	**1.1**	**1.4**	**3.5**	**−5.9**	**−3.2**
	1.0	0.2	0.4	1.1	−1.7	−0.9
G2	**32.5**	**15.2**	**14.1**	**22.2**	**−10.2**	**−2.5**
	12.3	4.0	4.2	6.5	−2.7	−0.3
G3	**39.4**	**18.2**	**16.6**	**25.6**	**−10.1**	**−1.8**
	14.5	4.7	4.8	7.3	−2.6	0.0
G4	**49.3**	**22.4**	**19.9**	**30.2**	**−9.9**	**−1.0**
	17.6	5.5	5.6	8.3	−2.4	0.3
G5	**171.7**	**67.6**	**47.9**	**69.4**	**−7.2**	**5.3**
	45.1	10.5	9.8	13.7	−1.1	2.5
Mean squared error						
Group	PR	FH	RE	ML	AR	AM
G1	**0.018**	**0.009**	**0.008**	**0.008**	**0.010**	**0.008**
	0.007	0.003	0.003	0.003	0.003	0.003
G2	**0.085**	**0.036**	**0.029**	**0.035**	**0.015**	**0.015**
	0.027	0.008	0.007	0.007	0.006	0.006
G3	**0.091**	**0.038**	**0.029**	**0.037**	**0.013**	**0.014**
	0.028	0.007	0.006	0.007	0.005	0.005
G4	**0.099**	**0.039**	**0.029**	**0.037**	**0.012**	**0.013**
	0.028	0.006	0.006	0.006	0.004	0.004
G5	**0.119**	**0.041**	**0.019**	**0.027**	**0.002**	**0.003**
	0.020	0.002	0.001	0.001	0.001	0.001

Table 3
Simulated mean squared prediction error of $\hat{\theta}_i^{\mathrm{EBP}}(Y; \hat{A})$.

	PR	FH	RE	ML	AR	AM
G1	**0.99**	**0.93**	**0.91**	**0.91**	**0.93**	**0.92**
	0.87	0.84	0.84	0.84	0.84	0.84
G2	**0.48**	**0.42**	**0.42**	**0.43**	**0.41**	**0.41**
	0.42	0.39	0.39	0.39	0.39	0.39
G3	**0.44**	**0.38**	**0.37**	**0.38**	**0.37**	**0.37**
	0.37	0.35	0.35	0.35	0.35	0.35
G4	**0.39**	**0.32**	**0.32**	**0.33**	**0.31**	**0.31**
	0.32	0.30	0.30	0.30	0.29	0.29
G5	**0.16**	**0.1**	**0.1**	**0.1**	**0.09**	**0.09**
	0.11	0.09	0.09	0.09	0.09	0.09

5.1. Comparison of different estimators of B_i

The percentages of zero estimates for A are reported in Table 1. All methods except the adjusted maximum likelihood methods could produce a zero estimate of A, or equivalently, $\hat{B}_i = 1$ for all $i = 1, \ldots, m$. For $m = 15$, the Prasad–Rao method-of-moments approach yields as high as about 12% zero estimates of A. The residual maximum likelihood method is the best among all non-adjusted methods, producing only about 1% zero estimates of A.

We consider six different estimators of B_i derived from the following estimators of A: the Prasad–Rao simple method-of-moments estimator, the Fay–Herriot method-of-moments estimator, the profile and residual maximum likelihood estimators, the adjusted maximum profile and residual likelihood estimators.

We use the following criteria to compare different estimators of B_i. Let $\hat{B}_i^{(j)}$ be an estimator of B_i for the jth simulation run. Define:

- Relative bias: $\mathrm{RB}(\hat{B}_i) = \mathrm{Bias}(\hat{B}_i)/B_i$, where $\mathrm{Bias}(\hat{B}_i) = \sum_{j=1}^{N}(\hat{B}_i^{(j)} - B_i)/N$.
- Mean squared error: $\mathrm{MSE}(\hat{B}_i) = \sum_{j=1}^{N}(\hat{B}_i^{(j)} - B_i)^2/N$.

Table 4
Comparison of different mean square prediction error estimators.

Percentage relative biases							
Group	PR	FH	RE	ML	AR	AM	RE-*N*
G1	**2.0**	**−3.7**	**−1.0**	**−3.4**	**4.0**	**2.7**	**−9.1**
	−0.2	−0.8	−0.8	−1.0	−0.2	−0.4	−3.8
G2	**51.5**	**−3.4**	**−1.8**	**−4.6**	**1.8**	**1.5**	**−14**
	4.0	−0.4	−0.5	−0.6	−0.2	−0.2	−5.0
G3	**64.0**	**−4.4**	**−2.6**	**−5.2**	**0.3**	**0.2**	**−14.6**
	4.6	−0.8	−0.8	−0.9	−0.5	−0.6	−5.2
G4	**86.5**	**−4.0**	**−2.0**	**−4.7**	**0.7**	**0.7**	**−13.7**
	6.2	−0.2	−0.1	−0.2	0.0	0.0	−4.3
G5	**713.4**	**−0.1**	**1.8**	**−3.5**	**0.5**	**0.8**	**−7.6**
	43.4	−0.1	0.2	0.2	0.3	0.3	−1.7
Relative root mean squared error							
Group	PR	FH	RE	ML	AR	AM	RE-*N*
G1	**55.1**	**47.8**	**44.1**	**45.1**	**43.1**	**43.3**	**41.1**
	39.7	27.1	25.2	25.3	25.2	25.2	24.7
G2	**84.8**	**25.7**	**21.4**	**24.4**	**18.1**	**19.0**	**26.1**
	8.7	12.5	11.7	11.8	11.6	11.6	13.1
G3	**120.1**	**23.6**	**19.1**	**22.3**	**15.6**	**16.4**	**25.0**
	10.4	11.2	10.4	10.5	10.3	10.3	12.1
G4	**181.3**	**21.1**	**16.4**	**19.8**	**13.2**	**13.9**	**23.2**
	19.5	9.6	8.9	8.9	8.8	8.8	10.6
G5	**1946.2**	**24.7**	**5.4**	**12.3**	**3.3**	**3.1**	**13.2**
	331.5	2.9	3.0	2.9	3.1	3.1	4.2

Table 5
Average coverage and average length of different intervals (nominal coverage = 0.95).

Group	PR	FH	RE	Cox-RE	PB-RE	PB-AR	PB-AM
G1	**90.7** (**3.76**)	**89.6** (**3.55**)	**90.8** (**3.59**)	**88.1** (**3.31**)	**98.3** (**4.72**)	**94.4** (**4.01**)	**94.2** (**4.00**)
	92.3 (3.56)	93.9 (3.55)	94 (3.55)	93.4 (3.45)	95 (3.7)	94.6 (3.66)	94.7 (3.66)
G2	**98** (**3.30**)	**91.6** (**2.46**)	**93.3** (**2.49**)	**90** (**2.26**)	**97.9** (**3.34**)	**94.3** (**2.53**)	**94.5** (**2.53**)
	95.4 (2.58)	94.7 (2.44)	94.7 (2.44)	93.9 (2.37)	94.9 (2.47)	94.7 (2.44)	94.7 (2.44)
G3	**98.1** (**3.23**)	**92.1** (**2.32**)	**93.6** (**2.35**)	**90.5** (**2.14**)	**97.1** (**3.11**)	**94.7** (**2.36**)	**94.5** (**2.37**)
	95.5 (2.44)	94.8 (2.3)	94.7 (2.3)	94.2 (2.23)	94.9 (2.32)	94.6 (2.29)	94.7 (2.3)
G4	**98.1** (**3.15**)	**92.5** (**2.15**)	**93.7** (**2.17**)	**90.7** (**1.99**)	**97.4** (**2.82**)	**94.5** (**2.17**)	**94.4** (**2.18**)
	95.7 (2.28)	94.8 (2.13)	94.8 (2.12)	94.1 (2.07)	94.9 (2.14)	94.7 (2.12)	94.7 (2.12)
G5	**97.6** (**2.89**)	**95.3** (**1.25**)	**95.3** (**1.22**)	**93** (**1.15**)	**96.6** (**2.02**)	**94.6** (**1.19**)	**94.8** (**1.19**)
	95.8 (1.35)	95 (1.19)	95.1 (1.19)	94.7 (1.18)	94.8 (1.18)	94.6 (1.18)	94.7 (1.18)

The simulation results for the estimators of the shrinkage parameters are reported in Table 2. In terms of both bias and the mean squared error, the adjusted maximum profile likelihood method turns out to be the best among all the methods considered in the simulation experiment. The Prasad–Rao and Fay–Herriot method-of-moments approaches and the two unadjusted likelihood methods have the problem of overestimation and the extent of the overestimation depends on the method and the group. Among the unadjusted methods, the Prasad–Rao and the residual maximum likelihood methods exhibit the largest and smallest positive biases respectively. For the Prasad–Rao method, the relative bias could be as large as about 172%. For the residual maximum likelihood method, the maximum relative bias reported is about 48%. Overestimation of the shrinkage parameters implies that the methods give more weight to the synthetic part of the empirical best linear unbiased predictor formula. In contrast, the adjusted maximum likelihood methods are subject to underestimation of the shrinkage parameters and so they can be viewed as relatively conservative estimators of B_i, giving more weight to the direct estimator. The maximum negative bias reported by the adjusted residual likelihood method is about −10%. The adjusted maximum profile likelihood method corrects for the negative bias of the adjusted density maximization method, considerably.

5.2. Mean squared prediction error estimators

In this subsection, we examine first the efficiency of empirical best linear unbiased predictors when different estimators of A are used and then the performances of second-order mean squared prediction error estimators of different empirical best linear unbiased predictors corresponding to different estimators of A. We simulate the true mean squared prediction error of empirical best linear unbiased predictor estimators using different estimators of A. Let $\theta_i^{(s)}$ and $\hat{\theta}_i^{\mathrm{EBLUP}(s)}$ be the simulated true mean and the empirical best linear unbiased predictor of area i for the sth simulation respectively, $i = 1, \ldots, m$; $s = 1, \ldots, N = 10{,}000$. Then the simulated true mean squared prediction error of $\hat{\theta}_i^{\mathrm{EBLUP}}$ is given by $N^{-1}\sum_{s=1}^{N}[\hat{\theta}_i^{\mathrm{EBLUP}(s)} - \theta_i^{(s)}]^2$.

In Table 3, we report the group averages of the simulated true mean squared prediction errors of $\hat{\theta}_i^{\text{EBLUP}}$. The mean squared prediction errors of $\hat{\theta}_i^{\text{EBLUP}}$ corresponding to different adjusted and unadjusted maximum likelihood estimators of A are virtually the same, demonstrating the second-order equivalence of these empirical best linear unbiased predictors. Table 3 clearly demonstrates the inferiority of the Prasad–Rao method, especially for small $m = 15$, but the results for the empirical best linear unbiased predictor with the Fay–Herriot estimator of A are nearly the same as those for the likelihood based methods.

Table 4 displays the group average percentage relative biases and the average percentage relative root mean squared errors for each mean squared prediction error estimator. The relative bias of a mean squared prediction error estimator for a small area is calculated as RB = [E (mean squared prediction error estimator) − simulated true mean squared prediction error]/(simulated true mean squared prediction error), where E (the mean squared prediction error estimator) is the Monte Carlo expectation obtained by taking the average of mean squared prediction error estimates over $N = 10{,}000$ simulations. The relative root mean squared error of a mean squared prediction error estimator for a small area is calculated as RRMSE = [square root of MSE (mean squared prediction error estimator)]/(simulated true mean squared prediction error). In this table, we report seven simulated mean squared prediction error estimators: six of them are associated with six methods of estimating A, and one naive mean squared prediction error estimator is obtained using the restricted maximum likelihood method, and is denoted by "RE-N".

The naive estimator of the mean squared prediction error has a tendency toward underestimation. The mean squared prediction error estimator that uses the Prasad–Rao estimator of A tends to overestimate the true mean squared prediction error, especially for group G_5. For example, the overestimation could be as high as about 713% for $m = 15$. This is consistent with the findings of Datta et al. [20] and Pfeffermann and Glickman [26]. The mean squared prediction error estimators based on the Fay–Herriot, profile and residual maximum likelihood estimators are all subject to slight underestimation. In contrast, the mean squared prediction error estimators that use the adjusted maximum likelihood methods appear to have a slight overestimation problem.

5.3. Prediction interval

In Table 5, we compare coverage probabilities (with a nominal coverage of 0.95) and average lengths of the following seven prediction intervals of θ_i: Cox's empirical Bayes prediction interval with $\hat{A}^{\text{RE}}$ (Cox-RE), three traditional prediction intervals of the form $\hat{\theta}_i^{\text{EBLUP}} \pm 1.96\sqrt{\text{mspe}(\hat{\theta}_i^{\text{EBLUP}})}$ based on $\hat{A}^{\text{PR}}$, $\hat{A}^{\text{FH}}$ and $\hat{A}^{\text{RE}}$, and three prediction intervals based on the proposed parametric bootstrap method using $\hat{A}^{\text{AR}}$ (PB-AR), $\hat{A}^{\text{AM}}$ (PB-AM) and $\hat{A}^{\text{RE}}$ (PB-RE). For each parametric bootstrap method, we considered 1000 bootstrap samples and the shortest length prediction intervals.

Cox's prediction interval method consistently has an undercoverage problem. While the traditional prediction intervals that use second-order unbiased mean squared error prediction error estimates increase the coverage, undercoverage for the group G_1 with the largest shrinkage factor B_j is still noticeable. For the rest of the small area groups, these traditional prediction interval methods are subject to either undercoverage or overcoverage problems. The extent of the coverage errors appears to depend on the group and the method employed. The traditional method with the Prasad–Rao estimator of A is generally overly conservative at the expense of larger average length compared to the other rival methods. Thus, although the second-order mean squared prediction error estimators enjoy good theoretical properties, the resulting traditional prediction interval estimation methods have the coverage problem, possibly due to the enforced symmetry and the normality assumption.

The performance of the parametric bootstrap method depends on the estimator of A used. Among all the unadjusted standard methods, the residual maximum likelihood methods is the best and so we only report the parametric bootstrap prediction interval for the residual maximum likelihood estimator of A. The parametric bootstrap method with the residual maximum likelihood estimator of A is overly conservative at the expense of much larger average length compared to those using the adjusted maximum likelihood methods. This is perhaps due to the fact that the residual maximum likelihood method produces zero estimates. Since the estimator of A appears in the denominator of the pivot for our parametric bootstrap method, the pivot is undefined whenever $\hat{A}^{\text{RE}} = 0$. To get around the problem, we replaced those zero estimates by 0.01. In such situations, the pivotal values tend to be extraordinarily large, resulting in the large prediction intervals. In contrast, the performances of our parametric bootstrap methods based on the adjusted maximum likelihood estimators remain stable across all groups and always close to the target nominal level.

Acknowledgments

The second author's research was supported in part by National Science Foundation grant SES-0851001. The authors thank two anonymous referees for their constructive comments.

Appendix

We assume the following regularity conditions throughout the paper:

(r.1) $\text{rank}(X) = p$ is fixed;
(r.2) $\sup_{i\geq 1} h_{ii} = O(m^{-1})$, where $h_{ii} = x_i'(X'X)^{-1}x_i$;
(r.3) $0 < \inf_{i\geq 1} D_i \leq \sup_{i\geq 1} D_i < \infty$.

Positiveness of the adjusted maximum likelihood estimators of A. First note that $L_{\text{adj}}(0) = 0$ for $L(A) = L_P(A)$ or $L_{\text{RE}}(A)$. Moreover, it is easy to show that $\lim_{A\to\infty} AL_P(A) = 0$, for $m > 2$ and $\lim_{A\to\infty} AL_{\text{RE}}(A) = 0$, for $m > p + 2$. The strict positiveness of the adjusted profile maximum likelihood estimator (for $m \geq 2$) and adjusted residual maximum likelihood estimator (for $m \geq p + 2$) now follow from the continuity and non-negativeness of the adjusted likelihood functions.

To save space, we shall provide an outline of proof for Eqs. (2) and (3) only for the adjusted maximum profile likelihood estimator. The proof for the adjusted residual maximum estimator is similar and is given in [22].

The following lemma is used throughout the following proof. See [29] for the proof.

Lemma A.1. *Let $y \sim N(X\beta, \Sigma)$. Then for symmetric matrices F and G,*

$$\text{E}[(y'Fy)] = \text{tr}(F\Sigma) \tag{6}$$

$$\text{Var}(y'Fy) = 2\text{tr}(F\Sigma)^2 \tag{7}$$

$$\text{Cov}(y'Fy, y'Gy) = 2\text{tr}(F\Sigma G\Sigma) \tag{8}$$

$$\text{Cov}[(y'Fy)^2, y'Gy] = 8\text{tr}(F\Sigma F\Sigma G\Sigma) + 2\text{tr}(F\Sigma G\Sigma)\text{tr}(F\Sigma). \tag{9}$$

An outline of the proof for Eq. (2). First note that the adjusted profile log-likelihood is given by

$$l_{\text{adj,P}}(A) = c - (1/2)\left(\log|\Sigma| + y'Py\right) + \log(A).$$

Let $l^{(j)}$ denote the jth derivative of $l_{\text{adj,P}}(A)$ with respect to A ($j \geq 1$). For proving (2), we need expressions for $l^{(1)}$ and $l^{(2)}$, which are given by

$$l^{(1)} = \frac{1}{2}[y'P^2y - \text{tr}(\Sigma^{-1})] + \frac{1}{A}, \tag{10}$$

$$l^{(2)} = -y'P^3y + \frac{1}{2}\text{tr}(\Sigma^{-2}) - \frac{1}{A^2}. \tag{11}$$

For our case, we checked the regularity conditions of Theorem 2.1 of [19] (see [22] for details), which leads to

$$\text{E}[\hat{A}^{\text{AM}} - A]^2 = \text{E}\left\{-\frac{l^{(1)}}{\text{E}[l^{(2)}]} + r\right\}^2 + o(m^{-1}) = \frac{\text{E}[l^{(1)}]^2}{\{\text{E}[l^{(2)}]\}^2} + R_1 + R_2 + o(m^{-1}), \tag{12}$$

where

$$R_1 = -2\frac{\text{E}[l^{(1)}r]}{\text{E}[l^{(2)}]}$$

$$R_2 = \text{E}(r^2)$$

and

$$|r| \leq m^{-\rho}U \quad \text{with } \text{E}(|U|^g) \text{ bounded}, \tag{13}$$

for any fixed $0 < \rho < 1$ and $g > 0$. Evidently, $R_2 = o(m^{-1})$.

Using (6) and (7) in the lemma, we get

$$\text{E}[l^{(1)}] = \frac{1}{2}[\text{tr}(P) - \text{tr}(\Sigma^{-1})] + \frac{1}{A}, \tag{14}$$

$$\text{E}[l^{(1)}]^2 = \frac{1}{2}\text{tr}(\Sigma^{-2}) + O(1), \tag{15}$$

$$\text{E}[l^{(2)}] = -\frac{1}{2}\text{tr}(\Sigma^{-2}) + O(1). \tag{16}$$

Using the Cauchy–Schwarz inequality, (13), (15) and the fact that $|\text{tr}(P^k) - \text{tr}(\Sigma^{-k})| = O(1)$ for $k \geq 1$, we can claim $R_1 = o(m^{-1})$. Eq. (2) now follows from Eqs. (12), (15) and (16).

An outline of the proof for Eq. (3). We need the expression for $l^{(3)}$, which is given by

$$l^{(3)} = 3y'P^4y - \text{tr}(\Sigma^{-3}) + \frac{2}{A^3}. \tag{17}$$

Using (6), we have

$$\mathrm{E}[l^{(3)}] = 2\mathrm{tr}(\Sigma^{-3}) + O(1). \tag{18}$$

Noting that the adjustment factor $\log(A)$ and its derivatives are of order $O(1)$, it is straightforward to verify the regularity conditions of Theorem 4.1 of [19]. Thus, we get

$$\mathrm{E}[\hat{A}^{\mathrm{AM}} - A] = \mathrm{I} + \mathrm{II} + \mathrm{III}, \tag{19}$$

where

$$\mathrm{I} = -2\frac{\mathrm{E}[l^{(1)}]}{\mathrm{E}[l^{(2)}]},$$

$$\mathrm{II} = \frac{\mathrm{E}[l^{(1)}l^{(2)}]}{\{\mathrm{E}[l^{(2)}]\}^2} = \frac{\mathrm{Cov}(l^{(1)}, l^{(2)}) + \mathrm{E}[l^{(1)}]\mathrm{E}[l^{(2)}]}{\{\mathrm{E}[l^{(2)}]\}^2},$$

$$\mathrm{III} = -\frac{1}{2}\frac{\mathrm{E}[\{l^{(1)}\}^2 l^{(3)}]}{\{\mathrm{E}[l^{(2)}]\}^3} = -\frac{1}{2}\frac{\mathrm{Cov}(\{l^{(1)}\}^2, l^{(3)}) + \mathrm{E}[l^{(3)}]\mathrm{E}[l^{(2)}]^2}{\{\mathrm{E}[l^{(2)}]\}^3}.$$

First using (8) of Lemma A.1, we have

$$\mathrm{Cov}[l^{(1)}, l^{(2)}] = -\mathrm{tr}(\Sigma^{-3}) + O(1),$$

which when combined with (14) and (16) leads to

$$\mathrm{II} = \frac{-4\mathrm{tr}(\Sigma^{-3})}{[\mathrm{tr}(\Sigma^{-2})]^2} + o(m^{-1}). \tag{20}$$

Turning to term III, using (8) and (9) of Lemma A.1, we have

$$\mathrm{Cov}(\{l^{(1)}\}^2, l^{(3)}) = 6\mathrm{tr}(P^5) + 6\mathrm{tr}(P^4)/A + 3\mathrm{tr}(P^4)[\mathrm{tr}(P) - \mathrm{tr}(\Sigma^{-1})] = O(m),$$

which leads to

$$\frac{\mathrm{Cov}(\{l^{(1)}\}^2, l^{(3)})}{\{\mathrm{E}[l^{(2)}]\}^3} = o(m^{-1}). \tag{21}$$

Using (14), (16) and (18), we have

$$\mathrm{III} = \frac{4\mathrm{tr}(\Sigma^{-3})}{[\mathrm{tr}(\Sigma^{-2})]^2} + o(m^{-1}). \tag{22}$$

Combining terms I, II and III, we get (3).

References

[1] R.E. Fay, R.A. Herriot, Estimates of income for small places: An application of James–Stein procedure to census data, Journal of the American Statistical Association 74 (1979) 269–277.
[2] G. Carter, J. Rolph, Empirical Bayes methods applied to estimating fire alarm probabilities, Journal of the American Statistical Association 69 (1974) 880–885.
[3] B. Efron, C. Morris, Data analysis using Stein's estimator and its generalizations, Journal of the American Statistical Association 70 (1975) 311–319.
[4] C. Citro, G. Kalton (Eds.), Small-area estimates of school-age children in poverty, in: Evaluation of Current Methodology (National Research Council), National Academy Press, Washington DC, 2000.
[5] L. Mohadjer, J.N.K. Rao, B. Liu, T. Krenzke, W. VanDekerckhove, Hierarchical Bayes small area estimates of adult literacy using unmatched sampling and linking models, in: Proceedings of the Survey Research Methods Section, American Statistical Association [CD-ROM], American Statistical Association, Alexandria, VA, 2007, pp. 3203-3210.
[6] J.N.K. Rao, Small Area Estimation, Wiley, New York, 2003.
[7] N.G.N. Prasad, J.N.K. Rao, The estimation of mean squared errors of small area estimators, Journal of the American Statistical Association 85 (1990) 163–171.
[8] D. Pfeffermann, G. Nathan, Regression analysis of data from a cluster sample, Journal of the American Statistical Association 76 (1981) 681–689.
[9] W. Bell, (1999). Accounting for uncertainty about variances in small area estimation. Bulletin of the International Statistical Institute, 52nd Session, Helsinki, 1999, available at www.census.gov/hhes/www/saipe under "Publications.".
[10] B. Buelens, Comparing EBLUP and Bayesian small area estimates of Dutch unemployment rates, paper presented at the Conference on Sample Surveys and Bayesian Statistics, Southampton, August 27–29, 2008.
[11] C.N. Morris, Approximating posterior distributioins and posterior moments, in: J.M. Bernardo, M.H. DeGroot, A.F.M. Lindley, D.V. Smith (Eds.), in: Bayesian Statistics, Vol. 3, Oxford University Press, Oxford, 1988, pp. 327–344.
[12] C.N. Morris, R. Tang, Estimating random effects via adjustment for density maximization, Unpublished report, 2009.
[13] C.N. Morris, Mixed model prediction and small area estimation (with discussions), Test 15 (1) (2006) 72–76.
[14] C.M. Stein, Estimation of the mean of a multivariate normal distribution, The Annals of Statistics 9 (1981) 1135–1151.
[15] D.R. Cox, Prediction intervals and empirical Bayes confidence intervals, in: J. Gani (Ed.), Perspectives in Probability and Statistics, Papers in Honor of M.S. Bartlett, Academic Press, 1975, pp. 47–55.
[16] C.N. Morris, Parametric empirical Bayes confidence intervals, in: Scientific Inference, Data Analysis and Robustness, Academic Press, New York, 1983, pp. 25–50.

[17] A. Chatterjee, P. Lahiri, H. Li, Parametric bootstrap approximation to the distribution of EBLUP, and related prediction intervals in linear mixed models, The Annals of Statistics 36 (2008) 1221–1245.
[18] G.S. Datta, P. Lahiri, A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems, Statistica Sinica 10 (2000) 613–627.
[19] K. Das, J. Jiang, J.N.K. Rao, Mean squared error of empirical predictor, The Annals of Statistics 32 (2004) 818–840.
[20] G.S. Datta, J.N.K. Rao, D.D. Smith, On measuring the variability of small area estimators under a basic area level model, Biometrika 92 (2005) 183–196.
[21] J. Jiang, Linear and Generalized Linear Mixed Models and Their Applications, Springer, 2007.
[22] H. Li, Small area estimation: An empirical best linear unbiased prediction approach, Ph.D. Thesis, Dept. of Mathematics, University of Maryland, College Park, 2007.
[23] R.N. Kackar, D.A. Harville, Approximations for standard errors of estimators of fixed and random effects in mixed linear models, Journal of the American Statistical Association 79 (1984) 853–862.
[24] P. Hall, T. Maiti, On parametric bootstrap methods for small-area prediction, Journal of the Royal Statistical Society. Series B 68 (2006) 221–238.
[25] J.N.K. Rao, Inferential issues in small area estimation: Some new developments, Statistics in Transition 7 (2003) 513–526.
[26] D. Pfeffermann, H. Glickman, Mean square error approximation in small area estimation by use of parametric and nonparametric bootstrap, in: Proceedings of the Section on Survey Research Methods, American Statistical Association, 2004.
[27] E. Fabrizi, M.R. Ferrante, S. Pacei, Small area estimation of average household income based on unit level models for panel data, Survey Methodology 33 (2007) 187–198.
[28] L. Pereira, C. Pedro, Assessing uncertainty of the temporal EBLUP: A resampling-based approach, Journal of Statistical Computation and Simulation, (2009) (in press).
[29] V.K. Srivastava, R. Tiwari, Evaluation of expectation of products of stochastic matrics, Scandinavian Journal of Statistics 3 (1976) 135–138.