

ESTIMATING VARIANCE OF RANDOM EFFECTS TO SOLVE MULTIPLE PROBLEMS SIMULTANEOUSLY

BY MASAYO YOSHIMORI HIROSE^{*,†}, PARTHA LAHIRI^{†,§}

Institute of Statistical Mathematics[‡] and University of Maryland, College Park[§]

The two-level normal hierarchical model (NHM) has played a critical role in statistical theory for the last several decades. In this paper, we propose random effects variance estimator that simultaneously (i) improves on the estimation of the related shrinkage factors, (ii) protects empirical best linear unbiased predictors (EBLUP) (same as empirical Bayes (EB)) of the random effects from the common overshrinkage problem, (iii) avoids complex bias correction in generating strictly positive second-order unbiased mean square error (MSE) (same as integrated Bayes risk) estimator either by the Taylor series or single parametric bootstrap method. The idea of achieving multiple desirable properties in an EBLUP or EB method through a suitably devised random effects variance estimator is the first of its kind and holds promise in providing good inferences for random effects under the EBLUP or EB framework. The proposed methodology is also evaluated using a Monte Carlo simulation study and real data analysis.

1. Introduction. We advance EBLUP/EB theory for the following widely applied two-level Normal hierarchical model:

A Two-Level Normal Hierarchical Model (NHM)

For $i = 1, \dots, m$,

Level 1 (sampling model): $y_i | \theta_i \stackrel{\text{ind}}{\sim} N(\theta_i, D_i)$;

Level 2 (linking model): $\theta_i \stackrel{\text{ind}}{\sim} N(x_i' \beta, A)$.

In the above model, level 1 is used to account for the sampling distribution of unbiased estimates y_i . For example (see Morris 1983; Morris and Tang 2011), y_i could be a sample mean based on n_i observations taken from the i th population (e.g., a small geographic area, a hospital or a school) and

*Supported by Grant-in-Aid for Research Activity start-up, JSPS Grant Number 26880011.

†Supported in part by the National Science Foundation Grant Number SES-1534413.
MSC 2010 subject classifications: Primary 62C12; secondary 62F40

Keywords and phrases: Adjusted maximum likelihood method, Empirical Bayes, Empirical best linear unbiased prediction, Linear mixed model, Second-order unbiasedness

$D_i = \sigma^2/n_i$, where the common σ^2 is either known or accurately estimated using data from all populations. In some NHM applications when n_i 's are moderately large, y_i 's represent variance stabilizing transformed direct estimates so that D_i 's are known; see Efron and Morris (1975), Carter and Rolph (1974), Casas-Cordero (2015), and others. When n_i 's are small as in many small area estimation problems, smoothed estimates of D_i 's are obtained using empirical variance modeling (Fay and Herriot 1979, Bell and Otto 1995), but assumed known for the subsequent application of NHM. Bell (2008) examined how error in D_i estimates can affect inference of the random effects θ_i . There are some attempts to incorporate additional variability in the estimates of D_i through variance modeling; see Otto and Bell (1995), Arora and Lahiri (1997), Arora et al. (1997), You and Chapman (2006), Liu et al. (2015), Ha et al. (2015), and others. In order to focus on the central theme of this paper, that is, estimation of random effects variance of NHM satisfying multiple properties, we do not entertain the complexity involving the estimation of D_i and assume known D_i throughout the paper like many other papers written on NHM.

Level 2 links the random effects θ_i to a vector of p known auxiliary variables $x_i = (x_{i1}, \dots, x_{ip})'$, often obtained from various alternative data sources (e.g., administrative records, severity index for a hospital, school register, etc.). The parameters β and A of the linking model, commonly referred to as hyperparameters, are generally unknown and are estimated from the available data. We assume that $\beta \in R^p$, the p -dimensional Euclidian space, and $A > 0$.

The NHM model can be viewed as the following simple linear mixed model:

$$(1.1) \quad y_i = \theta_i + e_i = x_i' \beta + v_i + e_i, \quad i = 1, \dots, m,$$

where the v_i 's and e_i 's are independent with $v_i \stackrel{\text{iid}}{\sim} N(0, A)$ and $e_i \stackrel{\text{ind}}{\sim} N(0, D_i)$; see Prasad and Rao (1990). NHM can be also called a Bayesian model where level 1 and level 2 correspond to the sampling and prior distributions, respectively. In small area estimation literature (see Rao and Molina 2015), NHM is commonly referred to as the Fay-Herriot model.

NHM is particularly effective in combining different sources of information and explaining different sources of errors. Some earlier applications of NHM include the estimation of: (i) false alarm probabilities in New York city (Carter and Rolph 1974), (ii) the batting averages of major league baseball players (Efron and Morris 1975), and (iii) prevalence of toxoplasmosis in El Salvador (Efron and Morris 1975). More recently, NHM was used: to estimate poverty rates for the US states, counties, and school districts (Citro and

Kalton 2000) and Chilean municipalities (Casas-Codero et al. 2015), and to estimate proportions at the lowest level of literacy for states and counties (Mohadjer et al. 2007).

The MSE of a given predictor $\hat{\theta}_i$ of θ_i is defined as $M_i(\hat{\theta}_i) = E(\hat{\theta}_i - \theta_i)^2$, where the expectation is with respect to the joint distribution of $y = (y_1, \dots, y_m)'$ and $\theta = (\theta_1, \dots, \theta_m)'$ under the NHM. In the Bayesian terminology, $M_i(\hat{\theta}_i)$ is called the integrated Bayes risk under the squared error loss function. The best linear unbiased predictor (BLUP) $\hat{\theta}_i^{BLUP}$ of θ_i , which minimizes $M_i(\hat{\theta}_i)$ among all linear unbiased predictors $\hat{\theta}_i$, is given by:

$$\hat{\theta}_i^{BLUP}(A) = (1 - B_i)y_i + B_i x_i' \hat{\beta}(A),$$

where $B_i \equiv B_i(A) = D_i/(A + D_i)$ is the shrinkage factor and $\hat{\beta}(A) = (X'V^{-1}X)^{-1}X'V^{-1}y$ is the weighted least square estimator of β when A is known. Here we use the following notation: $X' = (x_1, \dots, x_m)$, a $p \times m$ matrix of known auxiliary variables; $V = \text{diag}(A + D_1, \dots, A + D_m)$, a $m \times m$ diagonal matrix. By plugging in an estimator \hat{A} for A (e.g., ML, REML, ANOVA) in the BLUP, one gets an empirical BLUP: $\hat{\theta}_i^{EB} \equiv \hat{\theta}_i^{BLUP}(\hat{A})$. Note that $\hat{\theta}_i^{EB}$ can be also interpreted as an empirical Bayes estimator of θ_i . In this paper, we use EB to refer to either empirical BLUP or empirical Bayes. We find it convenient to use the superscript *EB* in $\hat{\theta}_i^{EB}$ to denote either empirical best linear unbiased predictor or empirical Bayes estimator of θ_i .

In the context of an empirical Bayesian approach, Morris (1983) noted that for making inferences about the random effects θ_i , estimation of B_i is more important than that of A because the posterior means and variances of θ_i are linear in B_i , not in A , when the hyperparameters β and A are known. He also noted that, even if an exact unbiased estimator of A is plugged in $B_i \equiv B_i(A)$, one may estimate B_i with large bias. For that reason, to motivate the James-Stein estimator of θ_i , Efron and Morris (1973) used an exact unbiased estimator of B and *not* maximum likelihood estimator of A . For small m , maximum likelihood estimator of A (even with the REML correction) frequently produces estimate of A at the boundary (that is, 0) resulting in $B_i = 1$ for all i , even when some of the true B_i are not close to 1. This causes an overshrinkage problem in EB. That is, for each i , EB of θ_i reduces to the regression estimator. To overcome the overshrinkage problem, Morris (1983) suggested the fraction $(m - p - 2)/(m - 1)$ when estimator of B_i is 1. Li and Lahiri (2010) and Yoshimori and Lahiri (2014) avoided the overshrinkage problem by considering strictly positive consistent estimators of A , but did not devise their estimators of A to obtain nearly unbiased estimator of B_i ; that is, biases of their estimators of B_i , like all

other existing estimators (e.g., ML or REML), are of the order $O(m^{-1})$ and not $o(m^{-1})$. This is an important research gap, which we will fill in this paper.

An estimator $\hat{M}_i(\hat{\theta}_i^{EB})$ of $M_i(\hat{\theta}_i^{EB})$ is called second-order unbiased if $E[\hat{M}_i(\hat{\theta}_i^{EB})] = M_i(\hat{\theta}_i^{EB}) + o(m^{-1})$, for large m , under suitable regularity conditions. Let $M_{i;approx}(A)$ be a second-order approximation to $M_i(\hat{\theta}_i^{EB})$. That is, $M_i(\hat{\theta}_i^{EB}) = M_{i;approx}(A) + o(m^{-1})$, for large m , under regularity conditions. Prasad and Rao (1990) proposed a second-order unbiased estimator of $M_i(\hat{\theta}_{i;MOM}^{EB})$, where $\hat{\theta}_{i;MOM}^{EB}$ is EB of θ_i when method-of-moments (MOM) estimator \hat{A}_{MOM} of A is used and obtained $M_{i;approx}(A) = g_{1i}(A) + g_{2i}(A) + g_{3i;PR}(A)$, where $g_{1i}(A) = AD_i/(A + D_i)$, $g_{2i}(A) = D_i^2 x_i'(X'V^{-1}X)^{-1}x_i/(A + D_i)^2$, $g_{3i;PR}(A) = 2D_i^2 tr[V^2]/[m^2(A + D_i)^3]$. They noticed that the simple plug-in estimator $M_{i;approx}(\hat{A}_{MOM})$ is not second-order unbiased estimator of $M_i(\hat{\theta}_{i;MOM}^{EB})$. They showed that

$$E[M_{i;approx}(\hat{A}_{MOM})] = M_i[\hat{\theta}_i^{EB}(\hat{A}_{MOM})] + O(m^{-1}),$$

for large m , under regularity conditions. Note that the expression for $M_{i;approx}(A)$ could be different for different consistent estimators of A . For example, if REML estimator of A is used, $M_{i;approx}(A) = g_{1i}(A) + g_{2i}(A) + g_{3i}(A)$, where $g_{3i}(A) = 2D_i^2/[(A + D_i)^3 tr\{V^{-2}\}]$. We stress that $M_{i;approx}(\hat{A})$ is not second-order unbiased estimator of $M_i(\hat{\theta}_i^{EB})$ for any variance component estimators proposed in the literature. Bias correction is usually applied to achieve second-order unbiasedness. However, some bias-correction can even yield negative estimates of MSE. See Jiang (2007) and Molina and Rao (2015) for further discussions.

Mimicking a Bayesian hyperprior calculation, Laird and Louis (1987) introduced a parametric bootstrap method for measuring uncertainty of an empirical Bayes estimator. While their point estimator is identical to EBLUP, their measure of uncertainty has more of a Bayesian flavor rather than MSE. Butar (1997) [see also Butar and Lahiri 2003] was the first to introduce parametric bootstrap method to produce a second-order unbiased MSE estimator in the small area estimation context. Since Butar's work, a number of papers on parametric bootstrap MSE estimation methods appeared in the SAE literature; see Pfeiffermann and Glickman (2004), Chatterjee and Lahiri (2007); Hall and Maiti (2006); Pfeiffermann and Correra (2012). Some of them are the second-order unbiased but not strictly positive. Some adjustments were proposed to make the second-order unbiased double parametric bootstrap MSE estimators strictly positive, but adjusted MSE estimators were not claimed to have the dual property of second-order

unbiasedness and strict positivity. As pointed out in Jiang et al. (2016), a proof is not at all trivial and it is not even clear if the adjustments for positivity retain the second-order unbiasedness of the MSE estimators.

In this paper, we focus on the estimation of two important area-specific functions of A — the shrinkage factor B_i and the MSE of the EB $M_i(\hat{\theta}_i^{EB})$. We propose a area specific estimator of A , say \hat{A}_i , that simultaneously satisfies the following multiple desirable properties under certain mild regularity conditions:

- Property 1: Obtain a second-order unbiased estimator of B_i , that is, $E(\hat{B}_i) = B_i + o(m^{-1})$, among the class of estimators of B_i with identical variance, up to the order $O(m^{-1})$, where $\hat{B}_i = D_i/(\hat{A}_i + D_i)$.
- Property 2: $0 < \inf_{m \geq 1} \hat{B}_i \leq \sup_{m \geq 1} \hat{B}_i < 1$. That is, it protects EB from overshrinking to the regression estimator, a common problem encountered in the EB method;
- Property 3: Obtain second-order unbiased Taylor series MSE estimator of EB without any bias correction; that is, $E[M_{i;approx}(\hat{A}_i)] = M_i(\hat{\theta}_i^{EB}) + o(m^{-1})$.
- Property 4: Produce a strictly positive second-order unbiased single parametric bootstrap MSE estimator without any bias-correction.

Note that the variance component A in the NHM is not area specific, but to satisfy the above properties simultaneously for a given area, we propose an area specific estimator of A . This introduces an area specific bias, but interestingly the order of bias is $O(m^{-1})$, same as the bias of the ML estimator of A but higher than that of REML in the higher-order asymptotic sense. This seems to be a reasonable approach as our main targets are area specific parameters and *not* the global parameter A . Obviously, if A is the main target, we would recommend a standard variance component method. We stress that in general *none* of the existing methods for estimating A satisfies any of all the four properties simultaneously.

In Section 2, we propose a new adjusted maximum likelihood estimator of A that satisfies all the four desirable properties listed above. The balanced case has been heavily studied in the literature. We consider the balanced case in Section 3 and show how our results are related to the ones in the literature. [In Section 4, we illustrate our methodology by analyzing a real life data from the U.S. Census Bureau. Results from a Monte Carlo simulation study are described in Section 5.](#) All the technical proofs are deferred to the Appendix.

2. A New Adjusted Maximum Likelihood Estimator of A . The residual maximum likelihood estimator of A is defined as:

$$\hat{A}_{RE} = \arg \max_{A \in [0, \infty)} L_{RE}(A),$$

where $L_{RE}(A)$ is the residual likelihood of A given by

$$L_{RE}(A) = |X'V^{-1}X|^{-\frac{1}{2}} |V|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}y'Py\right)$$

with $P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}$. Note that \hat{A}_{RE} does not satisfy any of the four desirable properties listed in the introduction.

In an effort to find a likelihood-based estimator of A that satisfies all the four desirable properties, we start by deriving an adjusted maximum likelihood estimator of A defined as:

$$(2.1) \quad \hat{A}_i = \arg \max_{A \in [0, \infty)} h_i(A)L_{RE}(A),$$

where $h_i(A)$ is a factor to be suitably chosen so that all the four desirable properties are satisfied.

We first find $h_i(A)$ so that the resulting estimator of A results in a nearly unbiased estimator of B_i that also protects EB from overshrinking. In other words, we first find the adjustment factor $h_i(A)$ that simultaneously satisfies Properties 1 and 2. Interestingly, it turns out that such an adjusted maximum likelihood estimator also satisfies Properties 3 and 4.

Using Lemma 1 in Appendix A and Taylor series expansion, we have

$$(2.2) \quad \text{Var}(\hat{B}_i) = \frac{2D_i^2}{(A + D_i)^4 \text{tr}[V^{-2}]} + o(m^{-1}),$$

for large m . We restrict ourselves to the class of estimators of A that satisfies (2.2).

Using Lemma 1 and Taylor series expansion, we have

$$(2.3) \quad E(\hat{B}_i) = B_i + \left[\frac{\partial B_i}{\partial A} \frac{\partial \log h_i(A)}{\partial A} + \frac{1}{2} \frac{\partial^2 B_i}{\partial A^2} \right] \frac{2}{\text{tr}[V^{-2}]} + o(m^{-1}).$$

Thus, Property 1 is satisfied if we have

$$(2.4) \quad \frac{\partial B_i}{\partial A} \frac{\partial \log h_i(A)}{\partial A} + \frac{1}{2} \frac{\partial^2 B_i}{\partial A^2} = 0.$$

Now the differential equation (2.4) simplifies to:

$$(2.5) \quad \frac{\partial \log h_i(A)}{\partial A} = \frac{1}{A + D_i}.$$

Thus, an adjustment factor that satisfies (2.5) is given by

$$h_{i0}(A) = (A + D_i).$$

This adjustment factor is indeed the unique solution to (2.4) up to the order $O(m^{-1})$. If we substitute $h_{i0}(A)$ for $h_i(A)$ in (2.1), the resulting estimator \hat{A}_i will satisfy only two of the four desirable properties (properties 3 and 4). To rectify the problem, we propose our final estimator of A as:

$$\hat{A}_{i;MG} = \arg \max_{A \in [0, \infty)} \tilde{h}_i(A) L_{RE}(A),$$

where $\tilde{h}_i(A) = h_+(A)h_{i0}(A)$ with the additional adjustment $h_+(A)$ satisfying regularity conditions R6 and R7. The choice of $h_+(A)$ is not unique in general. One can use the choice given in Yoshimori and Lahiri (2014): $h_+(A) = \{\tan^{-1}[\text{tr}(I_m - B)]\}^{1/m}$, where $B = \text{diag}\{B_1, \dots, B_m\}$.

Our proposed estimator of B_i and EB are given by

$$\hat{B}_{i;MG} = B_i(\hat{A}_{i;MG}), \quad \hat{\theta}_{i;MG}^{EB} = \hat{\theta}_i^{BLUP}(\hat{A}_{i;MG}),$$

respectively.

Unlike the common practice, we avoid bias correction in obtaining both Taylor series and parametric bootstrap MSE estimators of our proposed EB. Interestingly, our approach ensures the important dual property of MSE estimator — second-order unbiasedness and strict positivity. This kind of MSE estimators is the first of its kind in the small area estimation literature.

We obtain our Taylor series estimator of MSE of EB by simply plugging in the proposed estimator $\hat{A}_{i;MG}$ for A in the second-order MSE approximation $M_{i;approx}(A)$ and is given by:

$$(2.6) \quad \hat{M}_{i;MG} \equiv M_{i;approx}(\hat{A}_{i;MG}) = g_{1i}(\hat{A}_{i;MG}) + g_{2i}(\hat{A}_{i;MG}) + g_{3i}(\hat{A}_{i;MG}),$$

Our proposed parametric bootstrap MSE estimator retains the simplicity of bootstrap originally intended in Efron (1979). It is given by

$$(2.7) \quad \hat{M}_{i;MG}^{boot} \equiv E_*[\hat{\theta}_i(\hat{A}_{i;MG}^*, y^*) - \theta_i^*]^2,$$

where $\theta_i^* = x_i' \hat{\beta}(\hat{A}_{1;MG}, \dots, \hat{A}_{m;MG}) + v_i^*$ with $v_i^* \stackrel{ind}{\sim} N(0, \hat{A}_{i;MG})$. Note that the new bootstrap MSE estimator does not require any bias correction.

The following theorem states that our proposed adjusted maximum likelihood estimator of A satisfies all the four desirable properties.

THEOREM 2.1. *Under the regularity conditions R1 – R7, we have, for large m ,*

- (i) $Bias(\hat{B}_{i;MG}) = o(1)$; $Var(\hat{B}_{i;MG}) = \frac{2D_i^2}{(A+D_i)^4 tr[V^{-2}]} + o(m^{-1})$;
- (ii) $0 < \inf_{m \geq 1} \hat{B}_{i;MG} \leq \sup_{m \geq 1} \hat{B}_{i;MG} < 1$, for $m > p + 2$;
- (iii) $E(\hat{M}_{i;MG}) - M_i(\hat{\theta}_{i;MG}^{EB}) = o(m^{-1})$;
- (iv) $E(\hat{M}_{i;MG}^{boot}) - M_i(\hat{\theta}_{i;MG}^{EB}) = o(m^{-1})$.

For proof of Theorem 2.1, see Appendix B.

3. The balanced case: $D_i = D$, $i = 1, \dots, m$. In this section, we show how the proposed adjusted maximum likelihood estimator of A is related to the problem of simultaneous estimation of several independent normal means, a topic for intense research activities, especially in the 60's, 70's and 80's, since the introduction of the celebrated James-Stein estimator (James and Stein 1961).

Let $y_i | \theta_i \stackrel{iid}{\sim} N(\theta_i, 1)$, $i = 1, \dots, m$. James and Stein (1961) showed that for $m \geq 3$ the maximum likelihood (also unbiased) estimator of θ_i is inadmissible under the sum of squared error loss function $L(\hat{\theta}, \theta) = \sum_{j=1}^m (\hat{\theta}_j - \theta_j)^2$ and is dominated by the James-Stein estimator: $\hat{\theta}_i^{JS} = (1 - \hat{B}_{JS})y_i$, where $\hat{B}_{JS} = (m - 2) / \sum_{j=1}^m y_j^2$. That is,

$$(3.1) \quad E \left[\sum_{j=1}^m (\hat{\theta}_j^{JS} - \theta_j)^2 | \theta \right] \leq E \left[\sum_{j=1}^m (y_j - \theta_j)^2 | \theta \right], \quad \forall \theta \in R^m,$$

where R^m is the m -dimensional Euclidean space, with strict inequality holding for at least one point θ . The dominance result, however, does not hold for individual components.

Efron and Morris (1973) offered an empirical Bayesian justification of the James-Stein estimator under the prior $\theta_i \stackrel{iid}{\sim} N(0, A)$, $i = 1, \dots, m$. Their model is indeed a special case of NHM with $D_i = 1$, $x'_i \beta = 0$, $i = 1, \dots, m$, and thus the James-Stein estimator of θ_i can be also viewed as an EB.

Morris (1983) discussed an empirical Bayesian estimation of θ_i for a Bayesian model that is equivalent to the balanced case of NHM, that is, when $D_i = D$ implying $B_i = B$, $i = 1, \dots, m$. In this case, he noted that $\hat{B}_U = (m - p - 2)D/S$ is an exact unbiased estimator of B , using the fact that, under NHM, $S = \sum_{j=1}^m (y_j - x'_j \hat{\beta}_{ols})^2 \sim (D + A)\chi_{m-p}^2$, where $\hat{\beta}_{ols}$ is the ordinary least square estimator of β . We can write $\hat{B}_U \equiv B(\hat{A}_{Morris}) = D/(D + \hat{A}_{Morris})$, where $\hat{A}_{Morris} = S/(m - p - 2) - D$. One can alternatively estimate B by a simple plug-in estimator: $\hat{B}_{plug} \equiv B(\hat{A}_U) = D/(D + \hat{A}_U)$,

where $\hat{A}_U = S/(m-p) - D$ is an unbiased estimator of A . Note that for $m > p + 4$

$$E(\hat{B}_U - B) = 0, \quad E(\hat{B}_{plug} - B) = \frac{2}{m-p-2}B = O(m^{-1}),$$

$$V(\hat{B}_U) = \left(\frac{m-p-2}{m-p}\right)^2 V(\hat{B}_{plug}) \leq V(\hat{B}_{plug}).$$

Thus, \hat{B}_U is better than \hat{B}_{plug} both in terms of bias and variance properties. We can write $\hat{B}_U = \hat{B}_{plug}(m-p-2)/(m-p)$. As pointed out by Morris (1983), the factor $(m-p-2)/(m-p)$ helps correct for the curvature dependence of B on A .

Consider the following empirical Bayes estimator of θ_i :

$$(3.2) \quad \hat{\theta}_i^{EB}(\hat{A}_{Morris}) = (1 - \hat{B}_U)y_i + \hat{B}_U x_i' \hat{\beta}_{ols}.$$

In this case, exact MSE and exact unbiased estimator of MSE can be obtained. Componentwise, for $m \geq p + 3$, we have

$$E[(\hat{\theta}_i^{EB}(\hat{A}_{Morris}) - \theta_i)^2] \leq D.$$

Thus, $\hat{\theta}_i^{EB}(\hat{A}_{Morris})$ dominates y_i in terms of unconditional MSE for $m \geq p + 3$. Such a componentwise dominance property, however, does not hold for conditional MSE (conditional on θ); see Morris (1983) for details.

Since $B < 1$, using Stein's argument, Morris (1983) suggested the following estimator of B : $\hat{B}_{Morris} = D/(D + \hat{A}_{Morris}^+)$, where $\hat{A}_{Morris}^+ = S/(m-p-2) - D$ if $S > (m-p-2)D$ and $\hat{A}_{Morris}^+ = 2D/(m-p-2)$ otherwise. This improves on the estimation of both B and θ_i . It is straightforward to show that in this special case \hat{A}_{Morris}^+ satisfies all the four properties. Moreover, under the regularity conditions R6-R8 and $m > p + 2$, \hat{A}_{MG} , our proposed estimator of A , is unique (see Appendix C for a proof) and is equivalent to \hat{A}_{Morris}^+ in the higher-order asymptotic sense, that is, $E(\hat{A}_{MG} - \hat{A}_{Morris}^+) = o(m^{-1})$.

Let $\hat{\theta}_i^{EB} = \hat{\theta}_i^{EB}(\hat{A})$ denote an EB of θ_i , where \hat{A} could be \hat{A}_{MG} , \hat{A}_{Morris}^+ or the REML $\hat{A}_{RE} = \max(0, \hat{A}_U)$. We can write $M_{i,approx}(A) = g_1(A) + g_2(A) + g_3(A)$ as the second-order approximation to $M_i(\hat{\theta}_i^{EB}) = MSE(\hat{\theta}_i^{EB})$ for any of the three choices of the estimator of A . The traditional second-order unbiased MSE estimator is obtained by correcting bias of $M_{i,approx}(\hat{A}_{RE})$, up to the order $O(m^{-1})$. It is given by $\hat{M}_{i,RE} = g_1(\hat{A}_{RE}) + g_2(\hat{A}_{RE}) + 2g_3(\hat{A}_{RE})$; see Prasad and Rao (1990), Datta and Lahiri (2000), Das et al. (2004). In this paper, we suggest an alternative second-order unbiased MSE estimator without bias-correction, that is, $\hat{M}_{i,MG} = g_1(\hat{A}_{MG}) + g_2(\hat{A}_{MG}) + g_3(\hat{A}_{MG})$.

We can show that

$$\begin{aligned} V(\hat{M}_{i,RE}) &= a_m + o(m^{-1}), \\ V[\hat{M}_{i,MG}] &= b_m + o(m^{-1}), \end{aligned}$$

where

$$\begin{aligned} a_m &= \left[\frac{(m-4-mq_i)(m-p)}{m(m-p-2)} \right]^2 \frac{2D^2B^2}{m-p-4}, \\ b_m &= \left(\frac{m-2-mq_i}{m} \right)^2 \frac{2D^2B^2}{m-p-4}, \quad q_i = x_i'(X'X)^{-1}x_i. \end{aligned}$$

It is straightforward to check that for $m > p + 4$ and $p \geq 3$, $b_m \leq a_m$. Thus, in the higher-order asymptotic sense, $\hat{M}_{i,MG}$ is a better second-order unbiased estimator of $M_i(\hat{\theta}_i^{EB})$ than $\hat{M}_{i,RE}$.

4. SAIPE data analysis. For purposes of evaluation, we consider the problem of estimating the percentages of school-age (aged 5-17) children in poverty for the fifty states and the District of Columbia using the same data set considered by Bell (1999). We choose two years (1992 and 1993) of state level data from the U.S. Census Bureau's Small Area Income and Poverty Estimates (SAIPE) program. In 1992, the REML estimate of A is zero while in year 1993 it is positive. Thus, these years would provide two different scenarios for evaluating estimation methods.

We assume the standard SAIPE state level model in which survey-weighted estimates (y_i) of the percentages of 5-17-year-old (related) children in poverty follow the Fay-Herriot model (1.1). The survey-weighted percentages are obtained using the Current Population Survey (CPS) data with their sampling variances D_i estimated by a Generalized Variance Function (GVF) method, following Otto and Bell (1995). However, as in any data analysis that use the Fay-Herriot model, we assume the sampling variances to be known throughout the estimation procedure. For this application, the small areas are 50 states and the District of Columbia of the United States and so $m = 51$. We consider the following $p = 5$ auxiliary variables:

- x_1 : A dummy variable for the intercept.
- x_2 : Ratio of the number of child tax exemptions for poor households and the total number of child tax exemptions.
- x_3 : The tax non-filer rates tabulated from IRS tax data, defined as the difference between the estimated population and number of tax exemptions under age 65, divided by the estimated population under age 65.

x_4 : The average monthly number of individuals receiving food stamps over a 12-month period, as a percentage of the population.

x_5 : The residuals obtained by fitting a Fay-Herriot model to the estimates of children in poverty from the 1990 census, with analogous covariates to those used here but for the year 1989.

See Bell et al. (2015) for more details.

Table 1 displays REML and our proposed estimates (HL) of the shrinkage parameters B_i and A for Washington DC (DC), Hawaii (HI) and California (CA) for the year 1992 and DC, Oregon (OR) and CA for the year 1993. They have the largest, median and smallest sampling variances D_i among all the states and DC, respectively. For 1992, REML estimate of A is zero yielding a B_i estimate of 1 for all the states and DC. This overshrinkage problem reduces EBs for all the states to regression synthetic estimates. Thus, even for states with reliable direct estimates (e.g., CA), there is no contribution of direct estimates in the EB formula. Our proposed estimates of shrinkage parameters offer a sensible solution. For DC, our shrinkage estimate is very close to 1 (giving nearly zero weight to the survey-weighted direct estimate in the EB formula), but for California survey estimate gets considerable weight (about 28%). In 1993, we do not have overshrinkage problem for REML estimates of the shrinkage factors, but our proposed estimates of B_i always gives more weights to the survey-weighted direct estimates than the corresponding REML estimates. Both REML and proposed estimates of B_i for all the states and DC are displayed in the left panel of Figure 1. Overall, our proposed estimates of B_i are smaller than REML.

Table 2 displays different MSE estimates of EBs for the selected three states for both years. The right panel of Figure 1 displays different MSE estimates for all the states in both years, where states are arranged in decreasing order of the sampling variances. For this study, we included the following MSE estimators of EB:

- (a) Naive MSE estimator (naive.RE) given by $g_{1i}(\hat{A}_{RE}) + g_{2i}(\hat{A}_{RE})$, where \hat{A}_{RE} denotes the REML estimator of A . This MSE estimator neither incorporates the extra uncertainty due to the estimation of A nor adjusts bias of the estimator $g_{1i}(\hat{A}_{RE})$ and is not second-order unbiased;
- (b) Single parametric bootstrap MSE estimator (PB.RE) that is obtained from (2.7) when REML estimator of A is used in the EB formula and is not a second-order unbiased.
- (c) Two second-order unbiased MSE estimators based on Taylor-series:
 - (i) DL.RE: $g_{1i}(\hat{A}_{RE}) + g_{2i}(\hat{A}_{RE}) + 2g_{3i}(\hat{A}_{RE})$; see Datta and Lahiri (2000).

- (ii) Taylor.HL: the proposed Taylor series MSE estimator given by (2.6).
- (d) Two second-order unbiased single parametric bootstrap MSE estimators:
- (i) **PB.BL**: $2\{g_{1i}(\hat{A}_{RE}) + g_{2i}(\hat{A}_{RE})\} - E_*[g_{1i}(\hat{A}_{RE}^*) + g_{2i}(\hat{A}_{RE}^*)] + E_*[\{\tilde{\theta}_i^*(y_i, \hat{A}_{RE}^*, \hat{\beta}(\hat{A}_{RE}^*, y_i)) - \tilde{\theta}_i^*(y_i, \hat{A}_{RE}, \hat{\beta}(\hat{A}_{RE}, y_i))\}^2]$; see Butar and Lahiri (2003).
- (ii) **PB.HL**: our proposed single parametric bootstrap MSE estimator given by (2.7).

For this application, **there is difference between the naive MSE estimates and some of the MSE estimates** that attempt to capture additional variability due to the estimation of A . In most of the cases, naive MSE estimates are slightly lower than both the first-order and second-order MSE estimates. The first-order unbiased MSE estimates (**PB.RE**) are generally slightly smaller than the second-order unbiased MSE estimates. The **PB.BL** MSE estimates can take negative values because of the adjustment needed to make it second-order unbiased. Except for large states (e.g., CA), MSE estimates for **EBs** are considerably lower than the corresponding sampling variances D_i indicating possible improvements by **EBs** over the direct estimates.

For the year 1992, REML estimate of A is zero. This is probably causing unusual behavior for DL.RE or **PB.BL** MSE estimates. **For the same application, Bell (1999) was the first to point out this problem.** For example, DL.RE MSE estimate for a large state like CA is more than that for a small state DC (similar behavior can be observed for **PB.BL**). For CA, DL.RE MSE estimate is even higher than the corresponding sampling variance of the direct estimate while all the other MSE estimates are showing opposite results. Overall, our proposed MSE estimates appear reasonable for both years.

TABLE 1

Estimates of shrinkage factors B_i for 3 areas (corresponding to minimum, median and max B_i values) using SAIPE data for 1992 and 1993; REML estimates of A for 1992 and 1993 are 0 and 1.7, respectively; HL area specific estimates of A are given within parenthesis.

1992 year				1993 year			
States	D_i	RE	HL	States	D_i	RE	HL
DC	31.694	1	0.997 (0.102)	DC	38.226	0.957	0.955 (1.819)
HI	11.347	1	0.989 (0.130)	OR	12.188	0.878	0.856 (2.045)
CA	1.883	1	0.723 (0.723)	CA	2.156	0.559	0.428 (2.876)

TABLE 2
*Estimates of MSEs for 3 areas (corresponding to minimum, median and max B_i values)
 using 1992 and 1993 SAIPE data*

1992 data							
States	D_i	naive.RE	PB.RE	DL.RE	Taylor.HL	PB.BL	PB.HL
DC	31.69	1.81	1.80	1.91	2.08	1.19	2.07
HI	11.35	1.19	1.30	1.45	1.48	0.88	1.57
CA	1.88	1.26	1.34	2.82	1.72	1.20	1.37
1993 data							
States	D_i	naive.RE	PB.RE	DL.RE	Taylor.HL	PB.BL	PB.HL
DC	38.23	4.07	4.14	4.23	4.41	4.97	4.33
OR	12.19	3.02	2.91	3.39	3.52	3.13	3.21
CA	2.16	1.64	1.74	2.19	1.87	1.72	1.60

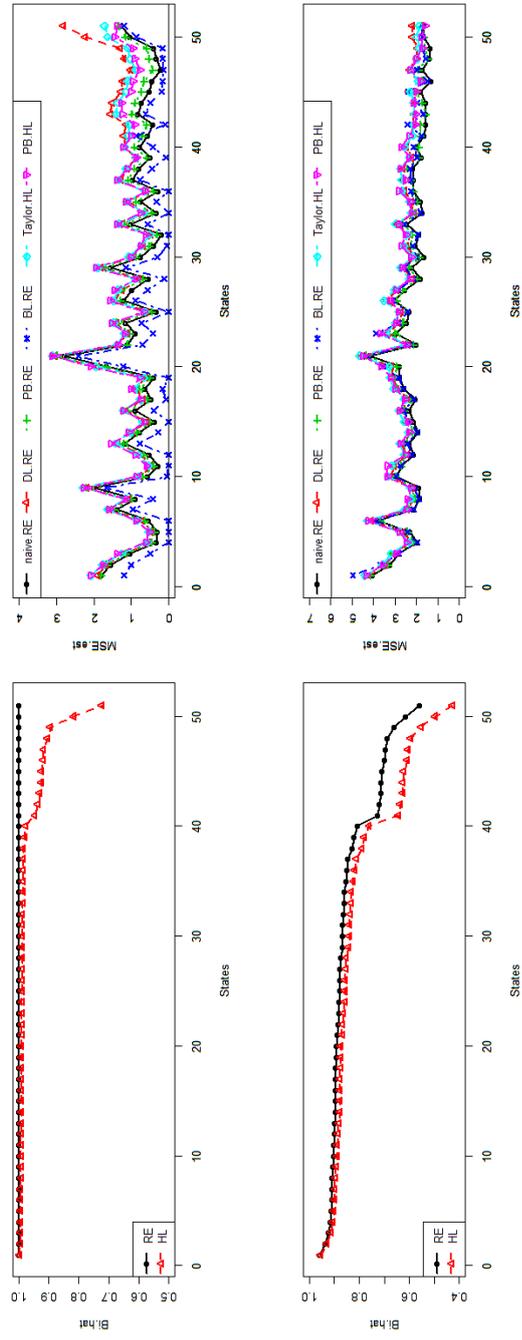


FIG 1. Estimates of B_i and MSE using all SAIPE data for 1992 (above) and 1993(bottom) year; states are arranged in decreasing order of the sampling variances

5. Monte Carlo simulation. In this section, we report results from a Monte Carlo simulation study. In particular, we evaluate finite sample performances of two different estimators of A — the commonly used REML \hat{A}_{RE} and the proposed estimator $\hat{A}_{i,MG}$ — in estimating the shrinkage parameters B_i , small area means θ_i and MSE of EBs of θ_i . To understand the effect of small m on different estimation problems, we set $m = 15$ and generate $\{(y_i, \theta_i), i = 1, \dots, m\}$ using the Fay-Herriot model (1.1).

We use the 1992 SAIPE data described in the previous section to design our simulation study. The 15 areas correspond to states with largest sampling variances D_i . In the simulation, we use x_i and D_i for these states from the 1992 SAIPE data and use $A = 15.94$, which is the median of D_i for the 15 states. The weighted least squared estimates of β from the real data including all 50 states and DC are treated as true β for the simulation.

We define the relative bias (RB) and relative root mean squared error (RRMSE) of an estimator \hat{B}_i of B_i as:

$$\begin{aligned} \text{RB of } \hat{B}_i &: \frac{E(\hat{B}_i - B_i)}{B_i} \times 100; \\ \text{RRMSE of } \hat{B}_i &: \frac{\sqrt{\text{MSE}(\hat{B}_i)}}{B_i} \times 100, \end{aligned}$$

where $\text{MSE}(\hat{B}_i) = E(\hat{B}_i - B_i)^2$. The expectations in the definitions of RB and RRMSE are approximated by Monte Carlo 1,000 independent samples from the Fay-Herriot model. The RB and RRMSE of an estimator \hat{M}_i of $M_i = \text{MSE}(\hat{\theta}_i) = E(\hat{\theta}_i - \theta_i)^2$, where $\hat{\theta}_i$ is an estimator of θ_i , are defined similarly. For the parametric bootstrap method, we use 1,000 bootstrap samples.

Table 3 displays simulated RBs and RRMSEs of two estimators of B_i for three selected states: DC, North Dakota (ND), Oklahoma (OK) corresponding to maximum, median and minimum values of D_i . These three states correspond to the maximum (0.67), median (0.50) and minimum values (0.46) of B_i 's among the 15 states. The two estimators of B_i are simple plug-in estimators – one obtained from REML \hat{A}_{RE} (denoted by RE) and the other from the proposed estimator $\hat{A}_{i,MG}$ (denoted by HL). For these states, RE consistently overestimates B_i while HL underestimates. The absolute values of the RB for HL are always smaller than those of RE. Moreover, variation of RBs for different B_i is much lower than that of RE. In terms of RRMSE, HL outperforms RE, especially for small values of B_i . Figure 2 displays the RB and RRMSE behavior for RE and HL for all the 15 selected states demonstrating superiority of HL over RE.

Figure 3 displays the simulated MSEs of two EBs of θ_i for each of the 15 states, where two EBs are obtained using the REML \hat{A}_{RE} (RE in the figure) and estimator $\hat{A}_{i,MG}$ (HL in the figure). There is hardly any difference between the simulated MSEs of the two EBs supporting the theory that these two MSEs are identical up to the order $O(m^{-1})$.

Table 4 reports simulated RBs and RRMSEs of different MSE estimators of EB that uses REML estimator of A . As mentioned earlier, all MSE estimators except naive.RE and PB.RE are second-order unbiased. The naive estimator naive.RE consistently underestimates. All the other MSE estimators improve on naive.RE. The parametric bootstrap estimator PB.RE that uses REML and does not use bias correction continues to underestimate. The second-order unbiased parametric bootstrap MSE estimator PB.BL that uses bias correction also underestimates although the amount of underestimation is generally smaller than that of PB.RE. The proposed second-order unbiased MSE estimators — Taylor.HL and PB.HL — are quite competitive to the second-order unbiased Taylor series MSE estimator, DL.RE, which overestimates for the state with smallest D_i . Our single parametric bootstrap second-order unbiased MSE estimator (PB.HL) that does not involve any bias correction is remarkably better than single parametric bootstrap MSE PB.RE (without bias correction) and even second-order unbiased parametric bootstrap MSE estimator PB.BL (with bias correction). All MSE estimators except PB.BL have lower RRMSE than naive.RE. It is interesting to note that the second-order unbiased PB.BL has more RRMSE than naive.RE for all the three states. This is probably due to the poor performance of REML of A that PB.BL uses. The REML of A produces zero estimates 12.4% of the times although true A is 15.94. The performances of DL.RE, Taylor.HL and PB.HL are similar and all are better than PB.RE. The performances of the MSE estimators of EB using the proposed estimator of A is similar to the results of Table 4; see Table 5. The RB and RRMSE behavior of all the MSE estimators for all the 15 states are given in Figure 5.

TABLE 3
RB and RRMSE of \hat{B}_i for 3 areas (corresponding to minimum, median and maximum values of B_i)

States	B_i	RB		RRMSE	
		RE	HL	RE	HL
DC	0.67	6.64	-2.86	28.70	28.49
ND	0.50	16.95	-5.28	50.29	41.96
OK	0.46	20.31	-6.09	56.90	44.79

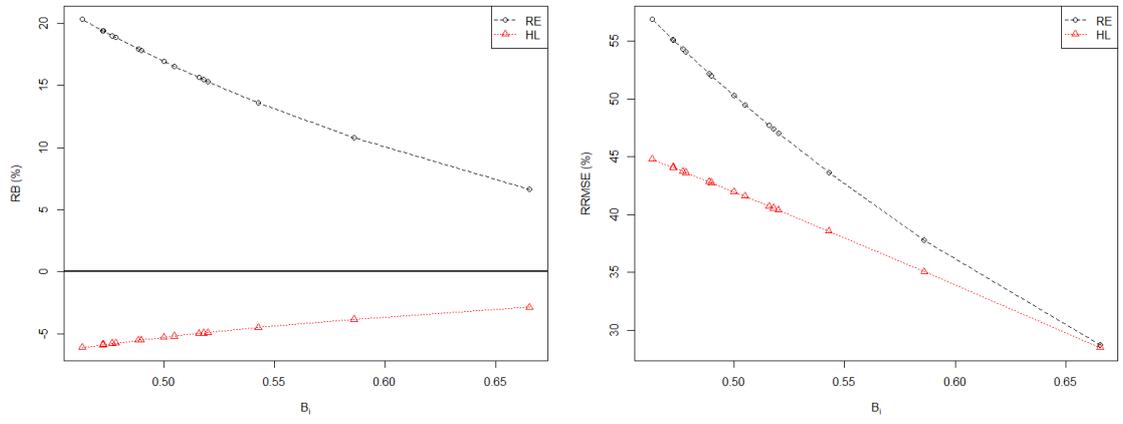


FIG 2. RB and $RRMSE$ of \hat{B}_i

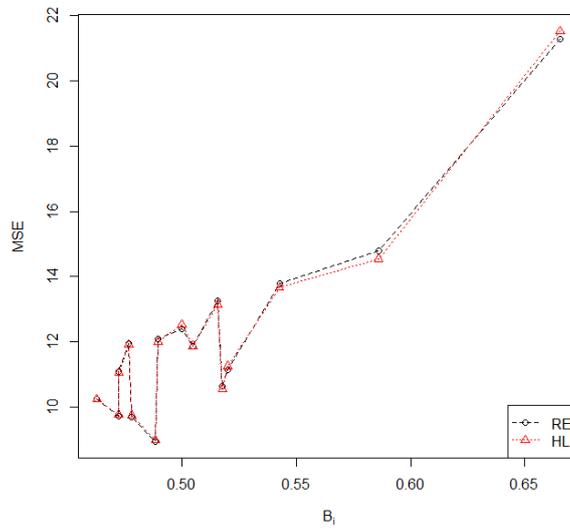


FIG 3. MSE of EB with RE and HL

TABLE 4
*RB and RRMSE of \hat{M}_i for MSE of EB with REML;
 results for 3 areas (corresponding to minimum, median and maximum values of B_i)*

RB							
States	B_i	naive.RE	PB.RE	DL.RE	Taylor.HL	PB.BL	PB.HL
DC	0.67	-10.10	-4.90	1.52	4.31	-2.01	3.83
ND	0.50	-17.50	-11.81	3.39	-0.35	-6.57	-2.63
OK	0.46	-14.94	-8.41	10.48	4.43	-2.51	1.96
RRMSE							
States	B_i	naive.RE	PB.RE	DL.RE	Taylor.HL	PB.BL	PB.HL
DC	0.67	21.33	20.60	19.07	18.33	26.88	18.30
ND	0.50	25.51	22.54	10.64	12.57	29.28	15.48
OK	0.46	25.68	22.91	13.07	13.52	31.91	16.47

TABLE 5
*RB and RRMSE of \hat{M}_i for MSE of EB with HL;
 results for 3 areas (corresponding to minimum, median and maximum value B_i)*

RB							
States	B_i	naive.RE	PB.RE	DL.RE	Taylor.HL	PB.BL	PB.HL
DC	0.67	-11.09	-5.95	0.40	3.16	-3.09	2.68
ND	0.50	-18.39	-12.76	2.27	-1.43	-7.57	-3.68
OK	0.46	-14.91	-8.38	10.51	4.46	-2.48	1.99
RRMSE							
States	B_i	naive.RE	PB.RE	DL.RE	Taylor.HL	PB.BL	PB.HL
DC	0.67	21.64	20.66	18.81	17.90	26.68	17.90
ND	0.50	25.98	22.88	10.23	12.51	29.22	15.53
OK	0.46	25.67	22.91	13.10	13.54	31.92	16.48

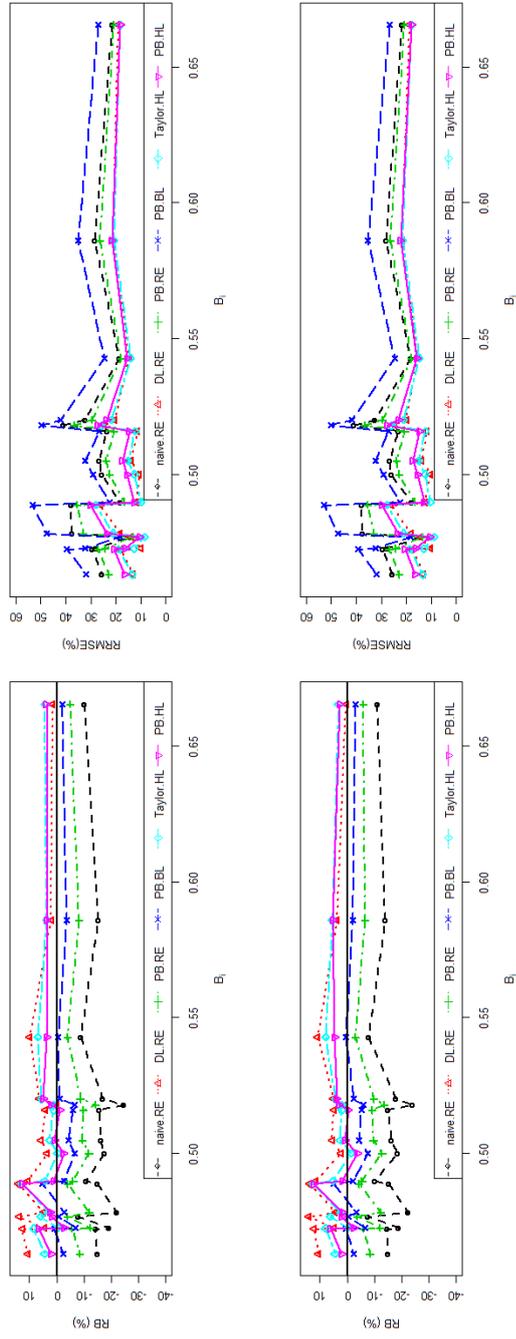


FIG 4. RB and RRMSE of MSE estimators for MSE of EB using REML (above) and HL (bottom); states are arranged in decreasing order of the sampling variances

6. Concluding Remarks. In this paper, we have solved a set of important problems for NHM through a suitably devised adjusted maximum likelihood estimator of the model variance parameter. Overall, we demonstrated that our proposed method offers reasonable results and it outperforms the existing methods in estimating shrinkage factors. In the future, we will explore the proposed methodology for other NHM with multiple variance components.

APPENDIX A: REGULARITY CONDITIONS AND LEMMA 1

R1: $\text{rank}(X) = p$ is bounded for large m ;

R2: The elements of X are uniformly bounded, implying $\sup_{j \geq 1} x_j (X'X)^{-1} x_j = O(m^{-1})$;

R3: $0 < \inf_{j \geq 1} D_j \leq \sup_{j \geq 1} D_j < \infty$, $A \in (0, \infty)$;

R4: $\log h_i(A)$ is free of y and four times continuously differentiable with respect to A . Moreover, $\frac{\partial^k \log h_i(A)}{\partial A^k}$ is of order $O(1)$, respectively, for large m with $k = 0, 1, 2, 3, 4$;

R5: $|\hat{A}_i| < C_{ad} m^\lambda$, where C_{ad} a generic positive constant and λ is small positive constant.

In addition to *R4*, the adjustment factor $h_+(A)$ satisfy the following regularity conditions:

R6: $\log h_+(A)$ is free of y and four times continuously differentiable with respect to A . Moreover, $\frac{\partial^k \log h_+(A)}{\partial A^k}$ is of order $o(1)$, for large m with $k = 0, 1, 2, 3, 4$;

R7: $h_+(A)$ is a strictly positive on $A > 0$ satisfying that $h_+(A)|_{A=0} = 0$ and $h_+(A) < C$ on $A > 0$ with a generic positive constant C ;

R8: In balanced case, that is, $D_i = D$ for all i , $(A + D)^2 \frac{\partial \log h_+(A)}{\partial A}$ is a monotonically decreasing function of $A > 0$ with $\lim_{A \rightarrow +0} (A + D)^2 \frac{\partial \log h_+(A)}{\partial A} = \infty$. When we assume that $\frac{\partial \log h_+(A)}{\partial A} > 0$, then, $\lim_{A \rightarrow \infty} (A + D)^2 \frac{\partial \log h_+(A)}{\partial A} = C$ for fixed m , where C is a generic positive constant.

We first present the following Lemma that provides properties of \hat{A}_i of A . The proof of the theorem is immediate from Theorem 1 of Yoshimori and Lahiri (2014) and Das et al. (2004).

LEMMA 1. Under the regularity conditions *R1* – *R5*, we have, for large m ,

$$(i) E(\hat{A}_i - A) = \frac{\partial \log h_i(A)}{\partial A} \frac{2}{\text{tr}[V^{-2}]} + o(m^{-1});$$

$$(ii) E(\hat{A}_i - A)^2 = \frac{2}{\text{tr}[V^{-2}]} + o(m^{-1});$$

(iii) $E[\hat{\theta}_i^{EB}(\hat{A}_i) - \theta_i]^2 \equiv M_i[\hat{\theta}_i^{EB}(\hat{A}_i)] = M_{i,approx}(A) + o(m^{-1})$, where *the logarithm of adjusted maximum likelihood* $l_{ad}(A) = \log L_{RE}(A|y) + \log h_i(A)$ and $M_{i,approx}(A) = g_{1i}(A) + g_{2i}(A) + g_{3i}(A)$ with $g_{1i}(A) = AD_i/(A + D_i)$, $g_{2i}(A) = D_i^2 x_i'(X'V^{-1}X)^{-1}x_i/(A + D_i)^2$, $g_{3i}(A) = 2D_i^2/[(A + D_i)^3 tr\{V^{-2}\}]$.

APPENDIX B: PROOFS OF THEOREM 2.1

B.1. Proof of part (i). First note that the adjustment factor $h_i(A)$ satisfies regularity condition R4. Then part (i) follows from the construction and (2.2).

B.2. Proof of part (ii). It suffices to show the strictly positivity for $\hat{A}_{i,MG}$. Note that $h_+(A)h_{i0}(A)L_{RE}(A)|_{A=0} = 0$ and $h_+(A)h_{i0}(A)L_{RE}(A) \geq 0$ for $A \geq 0$ using R6-R7. Thus, we are left to show that

$$\lim_{A \rightarrow \infty} h_+(A)h_{i0}(A)L_{RE}(A) = 0.$$

Let C be a generic constant. Using regularity conditions and $m \geq 1$, we have

$$\begin{aligned} h_+(A)h_{i0}(A) &< C(A + \sup_{i \geq 1} D_i), \\ L_{RE}(A) &< C(A + \sup_{i \geq 1} D_i)^{\frac{p}{2}} |X'X|^{-\frac{1}{2}} (A + \inf_{i \geq 1} D_i)^{-\frac{m}{2}}, \end{aligned}$$

which imply

$$\begin{aligned} 0 &\leq h_+(A)h_{i0}(A)L_{RE}(A) \\ &< C(A + \sup_{i \geq 1} D_i)^{1+p/2} (A + \inf_{i \geq 1} D_i)^{-m/2} |X'X|^{-1/2} \approx A^{-\frac{1}{2}(m-p-2)}, \end{aligned}$$

for large A . Thus, $\hat{A}_{i,MG}$ is strictly positive if $m > p + 2$.

B.3. Proof of part (iii). Using part (iii) of Lemma 1, we get

$$M_i(\hat{\theta}_{i,MG}^{EB}) = M_{i,approx}(A) + o(m^{-1}).$$

Note that using part (i) of Lemma 1 we have: $E[g_{2i}(\hat{A}_{i,MG})] = g_{2i}(A) + o(m^{-1})$, $E[g_{3i}(\hat{A}_{i,MG})] = g_{3i}(A) + o(m^{-1})$. Since $g_{1i}(A) = (1 - B_i)D_i$, we have $E[g_{1i}(\hat{A}_{i,MG})] = g_{1i}(A) + o(m^{-1})$, using part (i). This proves part (iii).

B.4. Proof of part (iv). Using part (iii), we have

$$\begin{aligned}\hat{M}_{i;MG}^{boot} &= g_1(\hat{A}_{i;MG}) + g_2(\hat{A}_{i;MG}) + g_3(\hat{A}_{i;MG}) + R, \\ &= M_i(\hat{A}_{i;MG}) + R,\end{aligned}$$

where $E[|R|] = o(m^{-1})$. The result now follows from part (iii).

APPENDIX C: PROOF OF THE UNIQUENESS OF \hat{A}_{MG} IN BALANCED CASE

In the balanced case, we have

$$\frac{\partial \log L_{RE}(A)}{\partial A} = \frac{1}{2(A+D)^2} \left[y' \{ I_m - X(X'X)^{-1}X' \} y - (m-p)(A+D) \right].$$

Thus, $(A+D)^2 \frac{\partial \log L_{RE}(A)}{\partial A}$ is a linear function of A . Therefore, our estimate of A is obtained as a solution of:

$$(C.1) \quad -(m-p-2)(A+D) + 2(A+D)^2 \frac{\partial \log h_+(A)}{\partial A} + y' \{ I_m - X(X'X)^{-1}X' \} y = 0.$$

Define $K(A)$ as the left hand of (C.1). For $A > 0$, using the regularity condition R6-R8 and $m > p + 2$, we show that $\lim_{A \rightarrow +0} K(A) = \infty$, $\lim_{A \rightarrow \infty} K(A) = -\infty$ and $K(A)$ is a strictly monotonically decreasing function of A on $A > 0$. Hence, there exist A_+ and A_- such that $K(A_+) = -\varepsilon$ and $K(A_-) = \varepsilon$ with small $\varepsilon > 0$ and $0 < A_- < A_+ < \infty$. Thus, using the intermediate value theorem, we conclude that the adjustment term $h_+(A)$ leads to a unique estimate of A on $A > 0$.

ACKNOWLEDGMENTS

We are grateful to two referees, an Associate Editor and Editor, for making a number of constructive suggestions, which led to a significant improvement of our paper.

REFERENCES

- [1] Arora, V. and Lahiri, P. (1997), On the Superiority of the Bayesian Method Over the BLUP in Small Area Estimation Problems, *Statistica Sinica*, 7, 1053-1063.
- [2] Arora, V., Lahiri, P. and Mukherjee, K. (1997), Empirical Bayes estimation of finite population means from complex surveys, *Journal of the American Statistical Association*, 92, 1555-1562.
- [3] Bell, W. R. (1999) Accounting for Uncertainty About Variances in Small Area Estimation, *Bulletin of the International Statistical Institute*, 52nd Session, Helsinki.

- [4] Bell, William R. (2008), Examining Sensitivity of Small Area Inferences to Uncertainty About Sampling Error Variances, *Proceedings of the American Statistical Association, Survey Research Methods Section*, [CD-ROM], Alexandria, VA: American Statistical Association, 327–334.
- [5] Bell, W. R., Basel, W., Cruse, C., Dalzell, L., Maples, J., OHara, B., and Powers, D. (2007), “Use of ACS Data to Produce SAIPE Model-Based Estimates of Poverty for Counties,” *Census Report*.
- [6] Bell, William R., Basel, Wesley W., and Maples, Jerry J., (2015) “An Overview of the U.S. Census Bureau’s Small Area Income and Poverty Estimates (SAIPE) Program,” in *Analysis of Poverty Data by Small Area Methods*, ed. Monica Pratesi, Wiley.
- [7] Butar, B.F., (1997). “Empirical Bayes methods in survey sampling,” *Ph.D. Thesis, Department of Mathematics and Statistics, University of Nebraska-Lincoln, unpublished*.
- [8] Butar, B.F., and Lahiri, P. (2003), “On measures of uncertainty of empirical Bayes small area estimators,” Special issue II: Model Selection, Model Diagnostics, Empirical Bayes and Hierarchical Bayes, *Journal of Statistical Planning and Inference*, 112, 63-76.
- [9] Casas-Cordero, C., Encina, J. and Lahiri, P. (2015), “Poverty Mapping for the Chilean Comunas,” In *Analysis of Poverty Data by Small Area Estimation*, ed. Monica Pratesi, Wiley Series in Survey Methodology.
- [10] Carter, G. M., and Rolph, J. F. (1974), “Empirical Bayes methods applied to estimating fire alarm probabilities,” *Journal of the American Statistical Association* 69 880-885.
- [11] Chatterjee, S., and Lahiri, P. (2007), “A simple computational method for estimating mean squared prediction error in general small-area model,” in *Proceedings of the section on survey research methods, American Statistical Association Alexandria VA*, pp. 3486-3493.
- [12] Citro, C. and Kalton, G. (Eds.). (2000), “Small-Area Income and Poverty Estimates: Priorities for 2000 and Beyond,” Panel on Estimates of Poverty for Small Geographic Area, Committee on National Statistics. Washington, DC: National Academy Press.
- [13] Datta, G. S., and Lahiri, P. (2000), “A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems,” *Statistica Sinica*, 10, 613-627.
- [14] Datta, G. S., Rao, J. N. K., and Smith, D. D. (2005), “On measuring the variability of small area estimators under a basic area level model,” *Biometrika*, 92, 183-196.
- [15] Das, K., Jiang, J., and Rao, J.N.K. (2004), “Mean squared error of empirical predictor,” *The Annals of Statistics*, 32, 818-840.
- [16] Efron, B. (1979), “Bootstrap methods: Another look at the jackknife,” *The Annals of Statistics*, 7, 1-26.
- [17] Efron, B., and Morris, C. (1973), “Stein’s Estimation Rule and Its Competitors—An Empirical Bayes Approach,” *Journal of the American Statistical Association*, 68, 117-130.
- [18] Efron, B., and Morris, C. (1975), “Data Analysis Using Stein’s Estimator and Its Generalizations,” *Journal of the American Statistical Association*, 70, 311-319.
- [19] Fay, R. E., and Herriot, R. A. (1979), “Estimates of income for small places: an application of James-Stein procedures to census data,” *Journal of the American Statistical Association*, 74, 269-277.

- [20] Ha, N. S., Lahiri, P. and Parsons, V. (2014). Methods and results for small area estimation using smoking data from the 2008 National Health Interview Survey, *Statistics in Medicine*. 33. 22.
- [21] Hall, P. and Maiti, T. (2006), “On parametric bootstrap methods for small area prediction,” *Journal of the Royal Statistical Society: Series B*, 68, 221-238.
- [22] James, W., and Stein, C. (1961), “Estimation with Quadratic Loss,” in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability (Vol. I)*, University of California Press, Berkeley, 361-379.
- [23] Jiang, J., Lahiri, P., and Nguyen, T. (2016), “A Unified Monte-Carlo Jackknife for Small Area Estimation after Model Selection,” *arXiv preprint*.
- [24] Laird, N. M., and Louis, T. A. (1987), “Empirical Bayes confidence intervals based on bootstrap samples,” *Journal of the American Statistical Association*, 82, 739-757.
- [25] Li, H., and Lahiri, P. (2010), “An adjusted maximum likelihood method for solving small area estimation problems,” *Journal of Multivariate Analysis*, 101, 882-892.
- [26] Liu, B., Lahiri, P. and Kalton, G. (2014). Hierarchical Bayes Modeling of Survey-Weighted Small Area Proportions. *Survey Methodology*. 40. 1-13.
- [27] Mohadjer, L., Rao, J.N.K., Liu, B., Krenzke, T., and Van De Kerckhove, W. (2012), “Hierarchical Bayes small area estimates of adult literacy using unmatched sampling and linking models,” *Journal of the Indian Society of Agricultural Statistics*, 66 (1), 55-63.
- [28] Morris, C. N. (1983), “Parametric empirical Bayes inference: Theory and applications,” *Journal of the American Statistical Association*, 78, 47-65.
- [29] Morris, C. and Tang, R. (2011). Estimating random effects via adjustment for density maximization. *Statistical Science*, 271-287.
- [30] Otto, M. C., and Bell, W. R. (1995), “Sampling Error Modeling of Poverty and Income Statistics for States,” in *Proceedings of the Section on Government Statistics*, American Statistical Association, Alexandria, VA. pp.160-165.
- [31] Pfeffermann, D. (2013), “New important developments in small area estimation,” *Statistical Science*, 28, 40-68.
- [32] Pfeffermann, D., and Glickman, H. (2004), “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, pp.4167-4178.
- [33] Pfeffermann, D., and Correa, S. (2012), “Empirical bootstrap bias correction and estimation of prediction mean square error in small area estimation,” *Biometrika*, 99, 457-472.
- [34] Prasad, N. G. N., and Rao, J. N. K. (1990), “The estimation of the mean squared error of small area estimators,” *Journal of the American Statistical Association*, 85, 163-171.
- [35] Rao, J. N. K., and Molina, I. (2015), *Small Area Estimation*, 2nd ed., Wiley, New York.
- [36] Yoshimori, M., and Lahiri, P. (2014), “A new adjusted maximum likelihood method for the Fay–Herriot small area model,” *Journal of Multivariate Analysis*, 124, 281-294.
- [37] You, Yong and Chapman, Beatrice (2006), Small Area Estimation Using Area Level Models and Estimated Sampling Variances, *Survey Methodology*, 32, 97-103.

M. Y. HIROSE
INSTITUTE OF STATISTICAL MATHEMATICS
10-3 MIDORI-CHO, TACHIKAWA, TOKYO, 190-8562
JAPAN

masayo@ism.ac.jp

P. LAHIRI
THE JOINT PROGRAM IN SURVEY METHODOLOGY
UNIVERSITY OF MARYLAND
1218 LEFRAK HALL, COLLEGE PARK, MARYLAND 20742
USA

plahiri@survey.umd.edu