Journal of Multivariate Analysis 124 (2014) 281–294

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

A new adjusted maximum likelihood method for the Fay–Herriot small area model

Masayo Yoshimori *, Partha Lahiri

The Joint Program in Survey Methodology, 1218 Lefrak Hall, University of Maryland, College Park, MD 20742, USA

ARTICLE INFO

Article history:
Received 25 October 2012
Available online 20 November 2013

AMS subject classifications:
62F12
62J99

Keywords:
Empirical Bayes
Linear mixed model
Profile likelihood
Residual likelihood
Shrinkage

ABSTRACT

In the context of the Fay–Herriot model, a mixed regression model routinely used to combine information from various sources in small area estimation, certain adjustments to a standard likelihood (e.g., profile, residual, etc.) have been recently proposed in order to produce strictly positive and consistent model variance estimators. These adjustments protect the resulting empirical best linear unbiased prediction (EBLUP) estimator of a small area mean from the possible over-shrinking to the regression estimator. However, in certain cases, the existing adjusted likelihood methods can lead to high biases in the estimation of both model variance and the associated shrinkage factors and can even produce a negative second-order unbiased mean square error (MSE) estimate of an EBLUP. In this paper, we propose a new adjustment factor that rectifies the above-mentioned problems associated with the existing adjusted likelihood methods. In particular, we show that our proposed adjusted residual maximum likelihood and profile maximum likelihood estimators of the model variance and the shrinkage factors enjoy the same higher-order asymptotic bias properties of the corresponding residual maximum likelihood and profile maximum likelihood estimators, respectively. We compare performances of the proposed method with the existing methods using Monte Carlo simulations.

1. Introduction

For the last few years, there has been an increasing demand to produce reliable estimates for small geographic areas, commonly referred to as small areas, since such estimates are routinely used for fund allocation and regional planning. The primary data, usually a survey data, are usually too sparse to produce reliable direct small area estimates that use the data from the small area under consideration. To improve upon direct estimates, different small area estimation techniques that use multi-level models to combine information from relevant auxiliary data have been proposed in the literature. The readers are referred to [23] for a comprehensive review of small area estimation.

The following two-level model, commonly referred to as the Fay–Herriot model (see [11]), has been extensively used in different small area applications (see, e.g., [4,10,11], etc.).

The Fay–Herriot Model. For $i = 1, \dots, m$,

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

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

* Correspondence to: Division of Mathematical Science, Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama-cho, Toyonaka, Osaka, 560-8531, Japan.

E-mail address: masayo@sigmath.es.osaka-u.ac.jp (M. Yoshimori).

0047-259X/$ – see front matter
http://dx.doi.org/10.1016/j.jmva.2013.10.012

In the above model, level 1 is used to account for the sampling distribution of the direct estimates y_i, which are weighted averages of observations from small area i. Level 2 links the true small area means θ_i to a vector of p known auxiliary variables $x_i = (x_{i1}, \ldots, x_{ip})'$, often obtained from various administrative records. The parameters β and A of the linking model are generally unknown and are estimated from the available data. We assume that $\beta \in R^p$, the p-dimensional Euclidean space, and $A \geq 0$. As in other papers on the Fay–Herriot model (e.g., [9]), the sampling variances D_i are assumed to be known. The assumption of the known sampling variances D_i often follows from the asymptotic variances of transformed direct estimates [10,4] and/or from the empirical variance modeling [11].

The Fay–Herriot two-level model can be viewed as the following simple linear mixed model:

$$y_i = \theta_i + e_i = x_i'\beta + v_i + e_i, \quad i = 1, \ldots, m,$$

where the v_i's and e_i's are independent with $v_i \overset{\text{iid}}{\sim} N(0, A)$ and $e_i \overset{\text{ind}}{\sim} N(0, D_i)$; see [22]. Fay and Herriot [11] called the model a Bayesian model where level 1 and level 2 correspond to the sampling and prior distributions, respectively. We define the mean squared error (MSE) of an estimator $\hat{\theta}_i$ of θ_i as $E(\hat{\theta}_i - \theta_i)^2$, where the expectation is with respect to the joint distribution of y and θ under the Fay–Herriot model. The best prediction (BP) estimator of θ_i, which minimizes the MSE, is given by:

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

where $B_i = \frac{D_i}{D_i + A}$ $(i = 1, \ldots, m)$. The purpose of using the superscript 'B' in $\hat{\theta}_i^{\text{B}}$ is to indicate that $\hat{\theta}_i^{\text{B}}$ is also the Bayes estimator of θ_i under the squared error loss function.

Define $y = (y_1, \ldots, y_m)'$; $X' = (x_1, \ldots, x_m)$, $V = \text{diag}(A + D_1, \ldots, A + D_m)$. If A is known, β can be estimated by the standard weighted least squares estimator:

$$\hat{\beta}(A) = (X'V^{-1}X)^{-1}X'V^{-1}y.$$

Replacing β by $\hat{\beta}(A)$, one gets the following best linear unbiased prediction (BLUP) estimator of θ_i:

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

In the most realistic case when both β and A are unknown, an empirical best linear unbiased prediction (EBLUP) estimator of θ_i is given by

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

where $\hat{B}_i = D_i/(\hat{A} + D_i)$, $i = 1, \ldots, m$, $\hat{\beta} = \hat{\beta}(\hat{A})$, and $\hat{A}$ is a consistent estimator of A, for large m. The superscript 'EB' in $\hat{\theta}_i^{\text{EB}}$ is a natural notation to indicate that $\hat{\theta}_i^{\text{EB}}$ is also an empirical Bayes estimator of θ_i under the squared error loss function.

Rao [23] and Jiang and Lahiri [14] list several consistent estimators of A. They include different method-of-moments estimators (see [11,22]) and different maximum likelihood estimators such as residual maximum likelihood (REML) and (profile) maximum likelihood (ML) estimators (see, e.g., [8,7]). In terms of the MSE up to the order $O(m^{-1})$, the REML and ML estimators of A are equivalent and are better than those of the method-of-moments estimators [9]. It is also known that REML is superior to the ML method in terms of the higher-order asymptotic bias; for example, the bias of REML is $o(m^{-1})$ while that of ML is $O(m^{-1})$, under certain regularity conditions, for large m.

The literature on higher-order approximation to the MSE of EBLUP is quite rich; see [23,14] for a review. Under regularity conditions of Theorem 1 given in Section 2,

$$\text{MSE}[\hat{\theta}_i^{\text{EB}}] = g_{1i}(A) + g_{2i}(A) + g_{3i}(A) + o(m^{-1}),$$

where $g_{1i}(A) = \frac{AD_i}{A+D_i}$, $g_{2i}(A) = \frac{D_i^2}{(A+D_i)^2}\text{Var}(x_i'\hat{\beta})$, $g_{3i}(A) = \frac{D_i^2}{(A+D_i)^3}\text{Var}_{\text{approx}}(\hat{A})$, $\text{Var}(\hat{\beta}) = x_i'\left(\sum_{j=1}^m \frac{1}{A+D_j}x_jx_j'\right)^{-1}x_i$, and $\text{Var}_{\text{approx}}(\hat{A})$ is the variance of $\hat{A}$ correct up to the order $O(m^{-1})$. The term $g_{1i}(A)$ is the dominating term [of order $O(1)$], capturing the uncertainty of the BP. The additional terms $g_{2i}(A)$ and $g_{3i}(A)$, which are of order $O(m^{-1})$, capture the uncertainty due to estimation of β and A, respectively. It is interesting to note that the method of estimation of A does not affect the terms $g_{1i}(A)$ and $g_{2i}(A)$, but it affects the term $g_{3i}(A)$ term through $\text{Var}_{\text{approx}}(\hat{A})$ — the more the variability in the estimator $\hat{A}$ the more the $g_{3i}(A)$ and thus more the second-order approximation to MSE of EBLUP. For example, for both REML and ML, $\text{Var}_{\text{approx}}(\hat{A}) = 2\left\{\sum_{j=1}^m \frac{1}{(A+D_j)^2}\right\}^{-1}$, smaller than the asymptotic variance of the Prasad–Rao (PR) and Fay–Herriot (FH) method-of-moments estimators of A; see [9].

Note that, the second-order approximation involves unknown A and thus cannot be used to assess the uncertainty of EBLUP for a given data set. A MSE estimator, denoted as $\widehat{\text{MSE}}(\hat{\theta}_i^{\text{EB}})$, is second-order unbiased (or nearly unbiased) if $\text{E}[\widehat{\text{MSE}}(\hat{\theta}_i^{\text{EB}})] = \text{MSE}(\hat{\theta}_i^{\text{EB}}) + o(m^{-1})$. The second-order approximation given above is useful for obtaining the following second-order unbiased MSE estimator of EBLUP:

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

where $\widehat{\text{Bias}}(\hat{A})$ is a second-order unbiased estimator of $\text{Bias}(\hat{A})$. Note that $\widehat{\text{MSE}}(\hat{\theta}_i^{\text{EB}})$ is non-negative when REML or ML estimator of A is used in EBLUP since $\widehat{\text{Bias}}(\hat{A})$ is zero for REML and negative for ML.

It is well-known that the standard method-of-moments and maximum likelihood methods can yield a zero estimate of A, especially when A is small relative to the sampling variances and m is small. As a consequence, the corresponding EBLUPs of θ_i over-shrink, that is, they reduce to regression estimates $x_i'\hat{\beta}$ for all areas, and thus EBLUPs may differ from the direct estimates substantially, even for areas with moderately large samples.

In order to avoid the zero estimate of A, Bell [1] proposed a mean likelihood estimator, which is the mean of a probability distribution proportional to the REML likelihood. Lahiri [17] commented that other descriptive statistics such as the median of the probability distribution proportional to the REML likelihood can be explored. Note that the mean likelihood method or Lahiri's suggestion is essentially different descriptive statistics of the posterior distribution of A under a flat prior on A. Wang and Fuller [25] suggested a strictly positive method-of-moments estimator for the model variance.

In the context of a general linear mixed model, [18] introduced a generalized maximum likelihood method in order to produce strictly positive and consistent estimators of variance components. Their generalized maximum likelihood estimators of variance components are obtained by maximizing the standard residual likelihood after multiplying by an adjustment term intended to keep the variance component estimators away from zero. Li and Lahiri [20] proposed a specific adjustment term for the Fay–Herriot model and studied asymptotic and small sample properties of adjusted REML (AR.LL) and adjusted ML (AM.LL) estimators of A under the assumed Fay–Herriot model. The order of bias for AR.LL and AM.LL is $O(m^{-1})$, same as the order of bias for the ML but higher than the REML. In a Monte Carlo simulation study, [19] observed that AR.LL and the mean likelihood method can be subjected to high bias, especially when m is small and B_i is close to 1, under the assumed Fay–Herriot model. However, in terms of higher-order MSE, the Li–Lahiri adjusted REML and ML methods are equivalent to the standard likelihood methods for estimating A, shrinkage factors B_i or BLUPs of θ_i.

Li and Lahiri [20] proposed second-order unbiased MSE estimators of EB estimators, which use their adjusted maximum likelihood estimators of A. However, the Li–Lahiri MSE estimators are not guaranteed to be positive. It is not straightforward at all to handle the negative value of a variance or MSE estimator. For example, the simple approach of truncating the negative value to zero would alternate the asymptotic behavior of the MSE estimator such as the order of the bias. On the other hand, although some procedures such as the parametric bootstrap are known to produce strictly positive MSE estimators, the latter are not second-order unbiased; and a bias correction procedure such as the one based on a single parametric bootstrap due to [3] or a more computationally intensive double-bootstrap due to [13] or [5], may again run into the possibility of negative MSE estimates (e.g., [15, p. 745]). It is, however, possible to ensure a non-negative second-order unbiased MSE estimate using an additional step proposed by Hall and Maiti [13] or Chatterjee and Lahiri [5]. In any case, the methods based on the single or double parametric bootstrap are computer intensive and increase the variability of the MSE estimators due to the Monte Carlo step.

In Section 2, we propose a new adjustment factor for the adjusted residual (AR.YL) and (profile) maximum likelihood (AM.YL) methods. The purpose of introducing this new adjustment factor is to reduce the bias of AR.LL and AM.LL in estimating A and the corresponding shrinkage factors B_i, under the assumed Fay–Herriot model. In this section, we prove that AR.YL and AM.YL are strictly positive while maintaining higher-order asymptotic properties of the REML and ML estimators, respectively, in estimating A, B_i and θ_i. The results stated in this section also suggest that the second-order unbiased MSE estimators of the EBLUPs, which use AR.YL or AM.YL, are always non-negative—this is not necessarily true when AR.LL or AM.LL is used in the EBLUPs. In Section 3, we examine small sample performances of different estimators of A, B_i, θ_i using a Monte Carlo simulation experiment.

2. A new adjustment for the adjusted maximum likelihood method

Following [18], we define an adjusted likelihood as

$$L_{ad}(A) \propto h(A) \times L(A),$$

where $h(A)$ is an adjustment factor and $L(A)$ is a standard likelihood function (e.g., profile likelihood, residual likelihood, etc.). Adjusted maximum likelihood estimator of A is obtained by maximizing $L_{ad}(A)$ with respect to A over $[0, \infty)$. The adjustment factor $h(A)$ was introduced so that the resulting adjusted maximum likelihood estimate of A is strictly positive for any $y,\ X$ and $m > p$.

As in [20], we choose $L(A)$ to be the profile likelihood or the residual likelihood of A, which are given by

$$L_P(A) \propto |V|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}y'Py\right\},$$

and

$$L_R(A) = |X'V^{-1}X|^{-\frac{1}{2}} L_P(A),$$

respectively, where $P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}$.

Let $\hat{A}_{AM.g}$ and $\hat{A}_{AR.g}$ denote the adjusted profile maximum likelihood and adjusted residual maximum likelihood estimator of A for a general adjustment factor $h(A)$, respectively. The following theorem provides the higher-order bias and MSE properties of $\hat{A}_{AM.g}$ and $\hat{A}_{AR.g}$. The proof of the theorem is straightforward and uses results given in [18] and algebra.

Theorem 1. *Assume the following regularity conditions:*

1. *the adjustment factor $h(A)$ is free of y and four times continuously differentiable with respect to A over $[0, \infty)$;*
2. *the kth derivative of $\log h(A)$ with respect to A is bounded for $k > 0$;*
3. *$rank(X) = p$ is fixed;*
4. *$\sup_{i\geq 1} h_{ii} = O(m^{-1})$, where $h_{ii} = x_i'(X'X)^{-1}x_i$;*
5. *$0 < \inf_{i\geq 1} D_i \leq \sup_{i\geq 1} D_i < \infty$ and $A > 0$.*

Then, we have

(i) $E[\hat{A}_{AR.g} - A] = \frac{2\tilde{l}_{ad}^{(1)}}{tr[V^{-2}]} + o(m^{-1})$,
(ii) $E[\hat{A}_{AM.g} - A] = \frac{tr[P-V^{-1}]+2\tilde{l}_{ad}^{(1)}}{tr[V^{-2}]} + o(m^{-1})$,
(iii) $E[(\hat{A}_{Ad.g} - A)^2] = \frac{2}{tr[V^{-2}]} + o(m^{-1})$,

for large m, where $\tilde{l}_{ad}^{(1)}$ denotes the derivative of $\log h(A)$ with respect to A and $\hat{A}_{Ad.g}$ is either $\hat{A}_{AM.g}$ or $\hat{A}_{AR.g}$.

Note that Theorem 1 implies the consistency of $\hat{A}_{AM.g}$ and $\hat{A}_{AR.g}$ for any adjustment factor $h(A)$ satisfying the regularity conditions of Theorem 1. There are many choices for $h(A)$ that satisfy the regularity conditions of Theorem 1 and yield strictly positive estimator of A. One such choice is $h(A) = A$, which yields the Li–Lahiri adjusted residual maximum likelihood ($\hat{A}_{AR.LL}$) and adjusted profile maximum likelihood ($\hat{A}_{AM.LL}$) estimators of A. Since the BP of θ_i, the main parameter of interest, is a linear function of B_i, it is important to estimate B_i accurately (similar comments can be found in [21], in the Bayesian context). Lahiri and Pramanik [19] observed that, for small m, the choice $h(A) = A$ could result in an adjusted REML estimator of B with substantially higher asymptotic bias than the corresponding REML when B is close to 1, under the assumed Fay–Herriot model for the balanced case $D_i = D$ $(i = 1, \ldots, m)$.

Note that $\tan^{-1}(A) \leq A$ for $A \geq 0$. Thus, the adjusted residual (or profile) likelihood based on the adjustment $\tan^{-1}(A)$ is closer to the residual (or profile) likelihood than the one proposed by Li and Lahiri [20]. Moreover, like the Li–Lahiri adjustment factor, $\tan^{-1}(0) = 0$ so that adjusted residual (or profile) likelihood function cannot have a maximum at 0. One problem with the adjustment factor $\tan^{-1}(A)$ is that the adjusted residual (or profile) maximum likelihood estimator of A based on this adjustment factor does not satisfy the following desirable property:

Property A. *Suppose that we multiply the original data $\{y_i, i = 1, \ldots, m\}$ by a large constant, say c, to yield new data, say $\{y_i^* = cy_i, i = 1, \ldots, m\}$. Then the new data will follow the Fay–Herriot model with θ_i, D_i, x_i, and A replaced by $\theta_i^* = c\theta_i, D_i^* = c^2D_i, x_i^* = cx_i$, and $A^* = c^2A$. We say that an estimator, say $\hat{A}$, of A satisfies Property A, if it satisfies the following natural relationship: $\hat{A}^* = c^2\hat{A}$, where $\hat{A}^*$ is the same estimator based on the new data.*

Note that commonly used estimators of A such as the residual maximum likelihood, profile maximum likelihood, and adjusted maximum likelihood estimators due to [20] satisfy Property A. It is easy to check that the adjusted maximum likelihood estimators based on the adjustment $\tan^{-1}(A)$ do not satisfy Property A because of the fact that $\tan^{-1}(x)$ is almost flat for large x. In order to satisfy Property A, we need to devise an appropriate scaling for A in the adjustment factor $\tan^{-1}(A)$. Replacement of A by a function of $A/D_i,\ i = 1, \ldots, m$, in $\tan^{-1}(A)$ does the job since the factor c^2 cancels from the numerator and denominator. The choice of the function needs careful attention.

We would like our adjustment factor to have more influence on the adjusted likelihood when the probability of residual (or profile) maximum likelihood estimator of A being zero is high. Note that, using the central limit theorem on the REML, the probability of the REML estimator of A being zero can be approximated by $\Phi\left(-\sqrt{\frac{1}{2}tr[I-B]^2}\right)$, where Φ is the cumulative distribution function of the standard normal deviate and $B \equiv B(A) = \mathrm{diag}(B_1, \ldots, B_m)$. Since the probability of REML of A being zero depends on $tr[I - B]$, we modify the adjustment factor $\tan^{-1}(A)$ as: $\tan^{-1}\{tr[I - B(A)]\}$.

Finally, we propose the adjustment factor $h_{YL}(A) = \left(\tan^{-1}\{tr[I - B(A)]\}\right)^{1/m}$ in order to make the adjustment negligible for large m when standard maximum likelihood methods yield zero estimates with low probability due to their consistency property. Needless to say, $h_{YL}(A)$ is a function of $A/D_i,\ i = 1, \ldots, m$ ensuring Property A of the resulting adjusted residual or profile maximum likelihood estimator of A.

One can raise the following question: does the adjustment factor $h_{YL}(A)$ yield an adjusted likelihood that is closer to the standard likelihood than the one using the adjustment factor A? The answer is yes since it can be shown that $h_{YL}(A) < A \times K$, where K is a generic constant free of A. The proof is as follows:

$$\begin{aligned}
\log[h_{YL}(A)] &= \frac{1}{m}\log\left[\tan^{-1}\{tr(I-B)\}\right] \\
&< \frac{1}{m}\log\left[tr(I-B)\right] \\
&= \frac{1}{m}\log\left[A\sum_{j=1}^{m}\frac{D_j}{A+D_j}\frac{1}{D_j}\right]
\end{aligned}$$

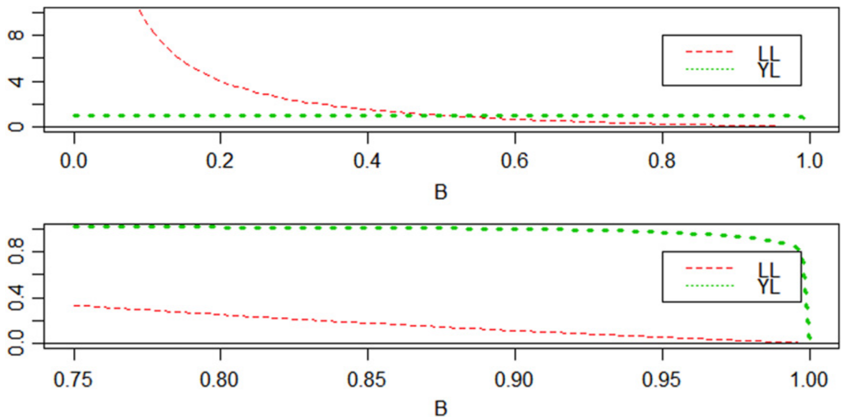

Fig. 1. Graph of $h(B)$ against B; $B \in [0, 1]$ in the upper graph and $B \in [0.75, 1]$ in the lower graph.

$$\leq \frac{1}{m} \log \left[Am \frac{1}{\inf_{i \geq 1} D_i} \right]$$
$$\leq \frac{1}{m} \log(A) + K.$$

Thus, $h_{YL}(A) < A \times K$. Note that the generic constant does not play a role in the adjusted maximum likelihood estimator of A.

Note that for the balanced case $D_i = 1$ $(i = 1, \ldots, m)$, we can express the Li–Lahiri and the proposed adjustment factors as $h_{LL}(B) = \frac{1-B}{B}$ and $h_{YL}(B) = \left(\tan^{-1}\{m(1-B)\}\right)^{1/m}$, respectively. We plot $h_{LL}(B)$ and $h_{YL}(B)$ against $B \in [0, 1]$ in Fig. 1.

Note that both adjustment factors are identical to 0 at $B = 1$, which prevents EBLUPs from over-shrinking. However, $h_{YL}(B)$ is much flatter than $h_{LL}(B)$ over most of the range except near 1 when $h_{YL}(B)$ bends, initially slowly and then rapidly, to zero, making the proposed adjusted maximum likelihood estimators closer to the standard maximum likelihood estimators than the corresponding Li–Lahiri maximum likelihood estimators while maintaining strict positivity of the corresponding A estimates.

With the above intuitive justification for our choice of $h(A)$ in the background, we have verified that $h_{YL}(A)$ satisfies the regularity conditions of Theorem 1 so that the resulting adjusted profile maximum likelihood estimator ($\hat{A}_{AM.YL}$) and the adjusted residual maximum likelihood estimator ($\hat{A}_{AR.YL}$) are consistent.

Next, we show that $\hat{A}_{AM.YL}$ and $\hat{A}_{AR.YL}$ are both strictly positive. To this end, first note that

$$h_{YL}(A)L(A)|_{A=0} = 0, \tag{1}$$

because $h_{YL}(0) = 0$ and $L(A) = O(1)$, for fixed m and $A \in [0, \infty)$.

Since $0 < L_{ad}(A) = h_{YL}(A)L(A) < \frac{\pi}{2}L(A)$, for $A > 0$, we have, under the regularity conditions of Theorem 1,

$$\lim_{A\to\infty} h_{YL}(A)L_R(A) = 0 \quad \textit{for } m > p, \tag{2}$$

$$\lim_{A\to\infty} h_{YL}(A)L_P(A) = 0. \tag{3}$$

Using (1)–(3), and the continuity and non-negativity of $h_{YL}(A)$ for $A \geq 0$, we can claim that $\hat{A}_{AM.YL}$ and $\hat{A}_{AR.YL}$ are both strictly positive.

The following result, which readily follows from Theorem 1, compares the higher-order bias and MSE properties of $\hat{A}_{AM.YL}$ and $\hat{A}_{AR.YL}$ with those of $\hat{A}_{AM.LL}$ and $\hat{A}_{AR.LL}$, respectively.

Result 1. Under regularity conditions of Theorem 1,

(i) $E[\hat{A}_{AR.YL}] - A = o(m^{-1})$; $E[\hat{A}_{AR.LL}] - A = \frac{2/A}{tr[V^{-2}]} + o(m^{-1})$;

(ii) $E[\hat{A}_{AM.YL}] - A = \frac{tr[P-V^{-1}]}{tr[V^{-2}]} + o(m^{-1})$; $E[\hat{A}_{AM.LL}] - A = \frac{tr[P-V^{-1}]+2/A}{tr[V^{-2}]} + o(m^{-1})$;

(iii) $E[(\hat{A}_{Ad.g} - A)^2] = \frac{2}{tr[V^{-2}]} + o(m^{-1})$,

where $\hat{A}_{Ad.g} \in \{\hat{A}_{AM.YL}, \hat{A}_{AR.YL}, \hat{A}_{AM.LL}, \hat{A}_{AR.LL}\}$.

Thus, overall $\hat{A}_{AR.YL}$ emerges as the best among all the estimators of A considered in Result 1—this has the best higher order asymptotic bias property, which matches the higher order asymptotic bias property of REML. Note that we are primarily interested in estimating the small area means θ_i. Part (iii) of Result 1 implies that the EBLUPs of θ_i based on all the estimators considered in Result 1 are equivalent in terms of the MSE criterion, up to order $O(m^{-1})$. Applying results in [16], we can also say that these EBLUPs are all equivalent in terms of the bias property as well. Since the biases of AR.YL (AM.YL) and REML (ML) estimators of A are of the same order, the formulae for the second-order unbiased MSE estimators of the corresponding EBLUPs of θ_i, given in the introduction, are identical and hence always non-negative. These desirable properties are not shared by AR.LL and AM.LL.

Let us now turn our attention to the estimation of the shrinkage factors B_i, which provide some idea about the effectiveness of Level 2 of the Fay–Herriot model. Note that the variances of $B_i(\hat{A})$ for all likelihood methods considered are equivalent, up to order $O(m^{-1})$. Thus we can compare different estimators of B_i using the bias to the variance (BV) ratio. Using [18], we have

$$\frac{Bias[B(\hat{A})]}{Var[B(\hat{A})]} = \frac{A+D_i}{D_i}\left[1-\tilde{l}_*^{(1)}(A+D_i)\right]+o(1), \tag{4}$$

where $\tilde{l}_*^{(1)} = \tilde{l}_{ad}^{(1)} + \frac{\partial \log |X'V^{-1}X|^{\frac{1}{2}}}{\partial A}$, for $L(A) = L_P(A)$, and $\tilde{l}_*^{(1)} = \tilde{l}_{ad}^{(1)}$, for $L(A) = L_R(A)$.

The following result is obtained directly from (4), after some algebra.

Result 2. Under the regularity conditions of Theorem 1, we have, for large m,

(i) $BV_{RE} = \frac{1}{B_i} + o(1),\ BV_{AR.YL} = \frac{1}{B_i} + o(1),\ BV_{AR.LL} = -\frac{1}{1-B_i} + o(1)$,

(ii) $BV_{ML} = \frac{1}{B_i}[1+(A+D_i)\frac{H}{2}] + o(1),\ BV_{AM.YL} = \frac{1}{B_i}[1+(A+D_i)\frac{H}{2}] + o(1), BV_{AM.LL} = -\frac{1}{1-B_i}\left[1-\frac{A}{D_i}(A+D_i)\frac{H}{2}\right] + o(1)$,

where $H = tr[V^{-1}X(X'V^{-1}X)^{-}X'V^{-1}] > 0$.

Result 2 demonstrates that our new adjusted REML of B_i generally overestimates while the Li–Lahiri adjusted REML underestimates in the higher-order asymptotic sense. Also, for the estimation of B_i, the leading term of the bias of our proposed adjusted REML is smaller than the Li–Lahiri adjusted REML when B_i is greater than 1/2, but greater than the Li–Lahiri adjusted REML when B_i is less than 1/2. When $B_i = 1/2$, the absolute values of the leading term of the bias of the two adjusted REML methods are exactly the same. Since $H > 0$, AR.YL has less asymptotic bias than AM.YL, for estimating the shrinkage factors.

3. A Monte Carlo simulation study

In this section, we design a Monte Carlo simulation study to compare finite sample performances of different estimators of the model variance A, the shrinkage parameters B_i, different EBLUP estimators corresponding to different estimators of A for the Fay–Herriot model. In Section 3.1, we consider a common mean Fay–Herriot model extensively studied in the literature (e.g., [9,5]). In Section 3.2, we consider a general Fay–Herriot model where we use the sampling variances D_i and the auxiliary variables x_i from a real life data.

3.1. The Fay–Herriot model with a common mean

Throughout this subsection, we consider a common mean model, that is, $x_i'\beta = \mu$. Since the mean squared error is translation invariant (that, it remains the same if y_i is changed to $y_i - \mu$), we set $\mu = 0$ without loss of generality. However, to account for the unknown mean parameter that arises in practice, we still estimate the zero mean. We generate $R = 10^4$ independent replicates $\{Y_i,\ v_i,\ i = 1, \ldots, m\}$ using the following Fay–Herriot model:

$$Y_i = v_i + e_i,$$

where v_i and e_i are mutually independent with $v_i \overset{iid}{\sim} N(0, A)$, $e_i \overset{ind}{\sim} N(0, D_i),\ i = 1, \ldots, m$. We set $A = 1$.

We consider both the balanced and unbalanced cases for the sampling variances D_i in this subsection. For the balanced case, in order to examine effects of both m and A/D on the performances of different estimators, we consider $m = 15, 45$, and $D \in \{0.05, 0.1, 1, 10, 20\}$. We have also investigated performances of different estimators when we vary $A \in \{0.05, 0.1, 1, 10, 20\}$, keeping D fixed. But our findings about the relative performances of different estimators are similar and so to save space we do not report the results for this case.

In the unbalanced case, for $m = 15$, we consider five groups, say $G \equiv (G_1, G_2, G_3, G_4, G_5)$, of small areas, each with three small areas, such that the sampling variances D_i are the same within a given area. We consider the following three patterns of the sampling variances: pattern (a) (4.0, 0.6, 0.5, 0.4, 0.1), pattern (b) (3.5, 3, 2.5, 2, 1.5) and pattern (c) (20, 6, 5, 4, 2). Note that in pattern (a) all but one small area have sampling variances less than A. In contrast, in patterns (b) and (c), sampling variances of all the areas are more than A with pattern (c) representing a case for extremely small areas. The patterns (a)

Table 1
Percentage of zero estimates of A.

m	A/D	RE	ML	AR.LL	AM.LL	AR.YL	AM.YL	WF
15	0.05	49.65	56.36	0	0	0	0	0
	0.1	45.13	52.42	0	0	0	0	0
	1	6.48	8.52	0	0	0	0	0
	10	0	0	0	0	0	0	0
	20	0	0	0	0	0	0	0
45	0.05	44.22	48.44	0	0	0	0	0
	0.1	35.79	39.61	0	0	0	0	0
	1	0.22	0.38	0	0	0	0	0
	10	0	0	0	0	0	0	0
	20	0	0	0	0	0	0	0

Table 2
Percentage of zero estimates of A (unbalanced case).

m	Pattern	RE	ML	AR.LL	AM.LL	AR.YL	AM.YL	PR	WF
15	a	0.91	1.56	0	0	0	0	11.6	0
	b	21.67	27.36	0	0	0	0	23.7	0
	c	34.09	42.62	0	0	0	0	47.27	0
45	a	0	0	0	0	0	0	0.94	0
	b	6.32	7.7	0	0	0	0	8.33	0
	c	17.12	20.33	0	0	0	0	35.49	0

and (c) correspond to the sampling variance pattern (c) of [9] and (b) of [6]. For $m = 45$, we simply add six more small areas in each group.

We first compare following estimators of A: residual maximum likelihood (RE), maximum likelihood (ML), Li–Lahiri adjusted residual maximum likelihood (AR.LL), Li–Lahiri adjusted maximum likelihood (AM.LL), Wang–Fuller method-of-moments (WF), proposed adjusted residual maximum likelihood (AR.YL), and proposed adjusted maximum likelihood (AM.YL). Note that in the balanced case, the residual maximum likelihood and the Prasad–Rao [22] estimators of A are identical. We then compare the corresponding estimators of $B = D/(A + D)$ and EBLUP of θ_i.

Tables 1 and 2 display simulated probabilities of obtaining zero estimates of A by different methods for the balanced and unbalanced cases, respectively. Note that only the PR, RE and ML could yield zero estimates and the probability of getting zero estimates is high when A/D is small for both methods, although RE is relatively less prone to the zero estimate than the PR and ML. The performances of PR, RE and ML improve as m increases. In the unbalanced case, patterns (b) and (c) yield high probabilities of the zero estimate. The situation for pattern (b) improves much faster than pattern (c) when m is increased from 15 to 45.

We define the percent relative bias (RB) of a given estimator, say $\hat{A}$, of A by $\frac{1}{R}\sum_{r=1}^{R}(\hat{A}^{(r)} - A)/A \times 100$, where $\hat{A}^{(r)}$ denotes an estimate of A for the rth replication, $r = 1, \ldots, R$. The relative bias of the corresponding estimator of B is defined in a similar way. Tables 3 and 4 display the percent relative bias (RB) of different estimators of A for the balanced and unbalanced cases. For the balanced case, when $A/D \geq 1$, ML and AM.YL have some tendencies for underestimation. In this case, performances of RE, WF and AR.YL are almost identical and are generally better ML, AM.YL, AR.LL and AM.LL. For small values of A/D, all the methods overestimate A. It is interesting to note that ML performs the best, followed by AM.YL and they are both better than RE. In this case, both AM.YL and AR.YL perform better than WF and substantially better than AM.LL and AR.LL. Overall, in terms of relative bias, AR.YL (AM.YL) is tracking RE (ML) well supporting the theoretical result given in Result 1. As m increases, the performances of all the estimators improve. In the unbalanced case, AR.YL (AM.YL) tracks REML (ML) well for patterns (a) and (b), but not (c) for small m. In general, AR.LL, AM.LL, PR and WF overestimate and substantially so for pattern (c).

We define the simulated percent MSE of an estimator $\hat{B}_i$ of B_i as:

$$MSE(\hat{B}_i) = 100 \times \frac{1}{R}\sum_{r=1}^{R}(\hat{B}_i^{(r)} - B_i)^2,$$

where $\hat{B}_i^{(r)}$ denotes an estimate of B_i for the rth replication, $r = 1, \ldots, R$. The simulated MSE of an EBLUP is defined in a similar way. We report the relative biases and MSEs of different estimators of B_i in Tables 5–7. For the balanced case and small A/D, AR.LL and AM.LL are subject to severe underestimation and large MSE, even when $m = 45$. The proposed adjustment factor cuts down this underestimation and MSE substantially, but suffers from an overestimation problem for large A/D, which diminishes when $m = 45$. It appears that in terms of bias and MSE, the new adjustment factor works better than the one proposed by Li and Lahiri [20] when A/D is small. However, the opposite is true when A/D is large. In the unbalanced case, AR.LL and AM.LL underestimate while AM.YL and AR.YL overestimate. It appears that AR.YL performs better than PR and RE for all patterns, and WF for patterns (a) and (c). Overall, for both the balanced and unbalanced cases, AR.YL (AM.YL) is

Table 3
The percent relative bias of different estimators of A.

m	A/D	RE	ML	AR.LL	AM.LL	AR.YL	AM.YL	WF
15	0.05	269.29	181.93	1290.13	1110.01	322.91	235.88	416.71
	0.1	112.07	63.71	632.42	537.77	138.14	90.37	182.55
	1	1.62	−11.32	63.88	45.33	2.37	−10.47	3.02
	10	−0.48	−7.78	19.61	9.65	−0.47	−7.77	−0.48
	20	−0.22	−7.2	18.12	8.65	−0.22	−7.2	−0.22
45	0.05	126.17	97.27	553.31	514.16	137.58	108.99	223.78
	0.1	48.03	30.88	262.46	241.06	53.02	36.11	88.85
	1	0.03	−4.41	18.98	14.1	0.04	−4.39	0.08
	10	−0.09	−2.53	5.67	2.98	−0.09	−2.53	−0.09
	20	0.09	−2.24	5.35	2.78	0.09	−2.24	0.09

Table 4
The percent relative bias of different estimators of A (unbalanced case).

m	Pattern	RE	ML	AR.LL	AM.LL	AR.YL	AM.YL	PR	WF
15	a	1.2	−9.8	41.6	25.1	1.3	−9.8	3.1	10.4
	b	11.4	−8.3	145.3	114.0	15.3	−4.0	13.6	41.2
	c	32.3	1.5	293.3	233.3	41.7	11.3	98.7	233.1
45	a	0.2	−3.4	11.2	7.1	0.2	−3.4	0.4	1.0
	b	1.2	−6.0	47.8	40.0	1.5	−5.6	1.7	9.0
	c	8.7	−2.5	100.2	87.0	9.9	−1.2	41.6	109.5

Table 5
The percent relative bias and MSE of different estimators of B.

m	A/D	RE	ML	AR.LL	AM.LL	AR.YL	AM.YL	WF
				Relative Bias × 100 of $B_i(\hat{A})$				
15	0.05	−7.5	−4.9	−35.8	−32.6	−10.1	−7.5	−14.2
	0.1	−4.8	−1.8	−34.1	−30.6	−7.4	−4.5	−11.5
	1	12.7	19.7	−17.3	−11.4	11.6	18.3	10.3
	10	17.2	25.5	−1.6	6.5	17.2	25.5	17.2
	20	16.9	25.2	−0.8	7.4	16.9	25.2	16.9
45	0.05	−4.1	−3	−19.8	−18.7	−4.7	−3.6	−8.7
	0.1	−2.2	−0.9	−18.1	−16.8	−2.7	−1.4	−6.2
	1	4.9	7.2	−5.1	−2.9	4.8	7.2	4.8
	10	4.8	7.2	−0.5	1.9	4.8	7.2	4.8
	20	4.8	7.1	−0.2	2.1	4.8	7.1	4.8
				MSE × 100 of $B_i(\hat{A})$				
15	0.05	2.97	2.2	12.79	10.7	2.96	2.17	3.21
	0.1	2.9	2.27	10.78	8.87	2.7	2.04	2.63
	1	4.5	5.21	2.2	1.88	4.04	4.66	3.52
	10	0.25	0.31	0.15	0.17	0.25	0.31	0.25
	20	0.07	0.09	0.04	0.05	0.07	0.09	0.07
45	0.05	1.32	1.14	4.24	3.82	1.29	1.1	1.31
	0.1	1.45	1.3	3.45	3.06	1.38	1.23	1.13
	1	1.45	1.58	0.96	0.94	1.44	1.57	1.41
	10	0.05	0.05	0.04	0.04	0.05	0.05	0.05
	20	0.01	0.02	0.01	0.01	0.01	0.02	0.01

tracking RE (ML) well supporting our theory. Tables 8 and 9 display the simulated MSEs of EBLUPs using different estimators of A. In the balanced case, for large A/D, all the methods provide similar results. The new adjustment factor performs better than the Li–Lahiri adjustment factor for small A/D, even when $m = 45$. In the unbalanced case, AR.YL and AM.YL usually perform better than AR.LL, AM.LL, PR and WF, especially so for $m = 15$ and pattern (c). The new adjustment factor provides results similar to the standard likelihood method.

3.2. The Fay–Herriot model with D_i and x_i from real life data

In this subsection, we compare AM.LL and AM.YL with AR.YL using a real life data on the sampling variances D_i and auxiliary variables x_i from the Small Area Income and Poverty Estimates (SAIPE) program of the U.S. Census Bureau. We actually included AR.LL in this study, but it did not perform well, especially for small A, and so in order to produce comparable

Table 6
The percent relative bias of different estimators of B (unbalanced case).

m	Pattern	G	RE	ML	AR.LL	AM.LL	AR.YL	AM.YL	PR	WF
15	a	a	1.1	3.2	−6.2	−3.5	1.0	3.2	2.4	0.3
		b	13.4	21.4	−10.7	−3.0	13.2	21.1	32.1	13.5
		c	15.8	24.7	−10.6	−2.4	15.6	24.4	39.0	15.8
		d	19.1	29.2	−10.4	−1.5	18.8	28.9	48.9	18.8
		e	46.7	68.3	−7.7	4.7	44.6	65.1	171.6	35.9
	b	a	2.6	6.4	−21.3	−17.3	1.3	5.0	2.6	−5.5
		b	3.6	7.9	−23.1	−18.8	2.1	6.2	3.6	−5.9
		c	5.1	10.0	−25.2	−20.6	3.2	7.9	5.1	−6.3
		d	7.5	13.3	−27.7	−22.7	5.0	10.6	7.6	−6.7
		e	11.8	19.0	−30.8	−25.3	8.3	15.1	12.1	−7.1
	c	a	−1.0	0.4	−11.6	−9.5	−1.4	−0.1	−2.9	−9.2
		b	−0.4	3.1	−26.8	−22.5	−2.0	1.4	−3.3	−21.8
		c	0.2	4.2	−29.6	−25.0	−1.8	2.1	−2.7	−24.3
		d	1.2	5.9	−33.2	−28.2	−1.3	3.2	−1.5	−27.5
		e	8.3	16.0	−43.9	−38.1	2.8	10.1	7.3	−37.6
45	a	a	0.4	1.1	−1.8	−1.0	0.4	1.1	1.0	0.9
		b	4.0	6.4	−2.8	−0.5	4.0	6.4	12.2	10.5
		c	4.6	7.2	−2.7	−0.2	4.6	7.2	14.4	12.2
		d	5.4	8.1	−2.6	0.1	5.4	8.1	17.4	14.5
		e	9.4	13.3	−1.4	2.2	9.4	13.3	43.0	28.6
	b	a	2.0	3.5	−8.1	−6.7	1.8	3.4	2.1	−0.2
		b	2.5	4.3	−8.9	−7.3	2.4	4.1	2.7	0.0
		c	3.4	5.4	−9.8	−8.0	3.2	5.2	3.6	0.3
		d	4.7	7.1	−10.9	−8.9	4.5	6.9	5.1	0.8
		e	7.0	10.0	−12.2	−9.9	6.7	9.7	7.7	1.8
	c	a	−0.2	0.3	−4.4	−3.8	−0.3	0.2	−1.3	−4.6
		b	0.5	2.0	−11.4	−10.0	0.3	1.7	−1.2	−11.7
		c	0.9	2.6	−12.8	−11.3	0.6	2.3	−0.7	−13.2
		d	1.6	3.7	−14.8	−12.9	1.2	3.3	0.2	−15.1
		e	6.2	9.7	−21.1	−18.6	5.4	8.8	7.0	−21.6

nice scaled graphs, we have not displayed the results for AR.LL. The sampling variances D_i were obtained from a sampling error model of Otto and Bell [2] that involved fitting a generalized variance function (GVF) to five years of direct variance and covariance estimates for each state produced by [12]. The state level auxiliary variables were derived from the U.S. Internal Revenue Service income tax return files, state participation rates in the food stamp poverty assistance program of the U.S. Department of Agriculture and the residual from regressing 5–17 state poverty rates from the previous 1990 decennial census on the other regression variables for 1989 (the census income reference year). For further details on SAIPE, the readers are referred to [1] and the website: http://www.census.gov/hhes/www/saipe.html.

We generate the simulated data $\{(y_i, \theta_i),\ i = 1, \dots, 51\}$ using the general Fay–Herriot model with x_i and D_i taken from the above-mentioned SAIPE data and using extremely small $A = D_{\min}/10$, moderate $A = D_{\text{med}}$ and large $A = 10 \times D_{\max}$, where $D_{\min}$, D_{med} and $D_{\max}$ denote the minimum, median and maximum of the sampling variances D_i for the 50 states and the District of Columbia. For each of these three simulation conditions, we compute simulated MSE of different EBLUPs corresponding to AM.LL, AM.YL, and AR.YL and then compute the ratio $\frac{\text{MSE}[\hat{\theta}_i^{\text{EB}}(\hat{A})]}{\text{MSE}[\hat{\theta}_i^{\text{EB}}(\hat{A}_{AR.YL})]}$. The ratios are plotted in Fig. 2(a)–(c) against states arranged in order of D_i. If the ratio is more than 1, EBLUP with A estimated by $\hat{A}$ is worse than the EBLUP with A estimated by $\hat{A}_{AR.YL}$. On the other hand, if the ratio is less than 1, EBLUP with A estimated by $\hat{A}$ is better than the EBLUP with A estimated by $\hat{A}_{AR.YL}$. It is clear from the Fig. 2(a) that, for small A, AR.YL performs better than AM.LL but worse than AM.YL. For large A, there is virtually no difference among the three methods. We also observe that AM.LL performs better than the other two for moderate A, although AR.YL and AM.YL are quite competitive. Overall, for this simulation, AM.YL performed the best.

4. Concluding remarks

The generalized (same as adjusted) maximum likelihood method, originally proposed by Lahiri and Li [18], is a frequentist alternative to the Bayesian method in producing strictly positive consistent estimators of variance components in linear mixed model. However, the choice of the adjustment factor needed for this method is not trivial. In this paper, we propose a new adjustment factor for the adjusted maximum likelihood method for the Fay–Herriot model and contrast it with the one proposed by Li and Lahiri [20]. Our theoretical and simulation results show that the use of the Li–Lahiri adjusted maximum likelihood estimator of the model variance could result in some efficiency loss in different estimation problems when

Table 7
Percent MSE of different estimators of B (unbalanced case).

m	Pattern	G	RE	ML	AR.LL	AM.LL	AR.YL	AM.YL	PR	WF
15	a	a	0.8	0.8	1.1	0.9	0.8	0.8	1.8	1.3
		b	2.9	3.6	1.6	1.6	2.8	3.5	8.5	2.8
		c	2.9	3.7	1.4	1.5	2.8	3.6	9.1	2.6
		d	2.9	3.7	1.2	1.3	2.8	3.6	9.9	2.4
		e	1.8	2.7	0.2	0.3	1.6	2.3	12.0	0.6
	b	a	2.8	2.8	4.1	3.1	2.5	2.4	3.0	1.5
		b	3.3	3.3	4.4	3.4	2.9	2.8	3.5	1.6
		c	4.0	4.1	4.7	3.6	3.4	3.4	4.3	1.7
		d	5.0	5.2	4.9	3.8	4.1	4.3	5.3	1.9
		e	6.4	7.0	4.8	3.8	5.1	5.5	6.8	1.9
	c	a	0.5	0.4	1.7	1.2	0.4	0.3	1.2	1.3
		b	2.4	2.1	6.6	5.0	2.1	1.8	4.5	4.7
		c	2.9	2.6	7.5	5.7	2.6	2.2	5.3	5.3
		d	3.7	3.4	8.5	6.5	3.2	2.8	6.3	6.0
		e	7.1	7.3	9.8	7.8	5.3	5.3	10.2	7.2
45	a	a	0.3	0.3	0.3	0.3	0.3	0.3	0.7	0.7
		b	0.7	0.7	0.5	0.6	0.7	0.7	2.6	2.0
		c	0.6	0.7	0.5	0.5	0.6	0.7	2.6	2.0
		d	0.5	0.6	0.4	0.4	0.5	0.6	2.6	1.8
		e	0.1	0.1	0.1	0.1	0.1	0.1	1.8	0.5
	b	a	1.4	1.4	1.2	1.1	1.3	1.4	1.5	1.0
		b	1.6	1.7	1.3	1.2	1.6	1.6	1.8	1.1
		c	2.0	2.1	1.5	1.3	1.9	2.0	2.2	1.3
		d	2.5	2.6	1.6	1.5	2.4	2.5	2.8	1.5
		e	3.2	3.4	1.7	1.6	3.0	3.3	3.6	1.7
	c	a	0.2	0.2	0.3	0.3	0.2	0.2	0.5	0.4
		b	1.2	1.1	1.7	1.4	1.1	1.1	2.6	1.9
		c	1.5	1.5	2.0	1.7	1.4	1.4	3.2	2.2
		d	2.0	2.0	2.4	2.0	1.9	1.9	4.0	2.6
		e	4.2	4.4	3.3	2.9	3.9	4.0	7.5	3.3

Table 8
Percent MSE of different EBLUPs.

m	A/D	RE	ML	AR.LL	AM.LL	AR.YL	AM.YL	WF
15	0.05	3.17	2.95	5.34	4.87	3.18	2.95	3.22
	0.1	1.97	1.87	2.9	2.67	1.96	1.85	1.95
	1	0.59	0.6	0.59	0.58	0.59	0.59	0.59
	10	0.09	0.09	0.09	0.09	0.09	0.09	0.09
	20	0.05	0.05	0.05	0.05	0.05	0.05	0.05
45	0.05	1.74	1.69	2.41	2.31	1.73	1.68	1.74
	0.1	1.3	1.28	1.55	1.5	1.29	1.27	1.27
	1	0.53	0.54	0.53	0.53	0.53	0.54	0.53
	10	0.09	0.09	0.09	0.09	0.09	0.09	0.09
	20	0.05	0.05	0.05	0.05	0.05	0.05	0.05

there is strong faith in the assumed Fay–Herriot model. In this paper, we rectify this problem associated with the adjusted maximum likelihood method by proposing a new adjustment factor. Using theory and simulations, we demonstrate that our new adjustment factor leads to estimation methodologies that are similar to the ones when we use the corresponding standard maximum likelihood methodologies. Thus our new method rectifies the over-shrinking problem associated with the standard likelihood-based methods and at the same time enhances the previously adjusted likelihood-based method, especially when one has strong a faith in the Fay–Herriot model. In the higher-order asymptotic sense, our proposed adjusted REML has less bias than the adjusted ML in estimating both the model variance and the shrinkage factors. Although, in the higher-order asymptotic sense, all the likelihood methods are equivalent in estimating the small area means, our simulation based on the real life data suggests superiority of our proposed adjusted ML over the other methods. In the future, we plan to investigate what causes our proposed ML to outperform our adjusted REML in Section 3.2. We are currently investigating extensions of the proposed method to more complex small area models.

Acknowledgments

The first author's research was supported by JSPS KAKENHI Grant Number 242742. She conducted this research while visiting the University of Maryland, College Park, USA as a research scholar under the supervision of the second author. The

Table 9
Percent MSE of different EBLUPs (unbalanced case).

m	Pattern	G	RE	ML	AR.LL	AM.LL	AR.YL	AM.YL	PR	WF
15	a	a	0.92	0.92	0.94	0.92	0.92	0.92	1.00	0.98
		b	0.42	0.43	0.42	0.41	0.42	0.43	0.49	0.43
		c	0.38	0.39	0.38	0.38	0.38	0.39	0.46	0.39
		d	0.31	0.32	0.31	0.31	0.31	0.32	0.39	0.32
		e	0.10	0.10	0.09	0.09	0.10	0.10	0.16	0.10
	b	a	1.07	1.06	1.17	1.12	1.06	1.05	1.09	1.05
		b	1.00	0.99	1.09	1.05	0.99	0.98	1.01	0.96
		c	0.96	0.95	1.03	0.99	0.94	0.94	0.96	0.91
		d	0.89	0.89	0.95	0.92	0.88	0.87	0.90	0.84
		e	0.80	0.80	0.83	0.80	0.78	0.78	0.82	0.74
	c	a	1.44	1.41	1.80	1.66	1.44	1.40	1.94	1.96
		b	1.35	1.31	1.74	1.60	1.34	1.29	1.52	1.53
		c	1.32	1.28	1.68	1.55	1.30	1.26	1.45	1.46
		d	1.26	1.23	1.57	1.46	1.24	1.21	1.37	1.37
		e	1.00	0.99	1.14	1.09	0.98	0.96	1.08	1.02
45	a	a	0.83	0.83	0.83	0.83	0.83	0.83	0.85	0.85
		b	0.39	0.39	0.39	0.39	0.39	0.39	0.41	0.41
		c	0.35	0.35	0.35	0.35	0.35	0.35	0.37	0.36
		d	0.29	0.29	0.29	0.29	0.29	0.29	0.32	0.31
		e	0.09	0.09	0.09	0.09	0.09	0.09	0.10	0.10
	b	a	0.88	0.88	0.88	0.88	0.88	0.88	0.89	0.87
		b	0.85	0.85	0.86	0.85	0.85	0.85	0.86	0.84
		c	0.82	0.82	0.81	0.81	0.81	0.82	0.82	0.80
		d	0.77	0.77	0.76	0.75	0.76	0.77	0.77	0.75
		e	0.69	0.69	0.68	0.67	0.69	0.69	0.71	0.67
	c	a	1.11	1.11	1.15	1.13	1.11	1.10	1.25	1.24
		b	1.04	1.04	1.09	1.07	1.04	1.03	1.15	1.10
		c	1.01	1.01	1.06	1.04	1.01	1.00	1.12	1.07
		d	0.98	0.98	1.03	1.01	0.98	0.97	1.09	1.02
		e	0.82	0.82	0.84	0.83	0.82	0.82	0.93	0.82

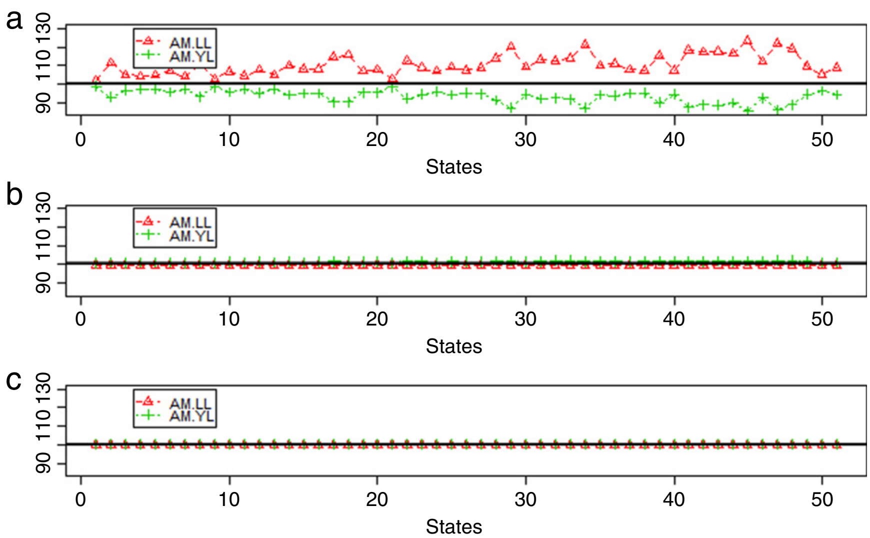

Fig. 2. Plots of $\frac{\mathrm{MSE}[\hat{\theta}_i^{\mathrm{EB}}(\hat{A})]}{\mathrm{MSE}[\hat{\theta}_i^{\mathrm{EB}}(\hat{A}_{AR.YL})]}$ for different estimators of A; in (a), $A = D_{\min}/10$; in (b), $A = D_{\mathrm{med}}$; in (c), $A = 10 \times D_{\max}$.

second author's research was supported in part by the National Science Foundation Grant Number SES-085100 and National Institute of Health Grant Number R01 CA 129101. The authors thank Dr. William R. Bell of the U.S. Census Bureau for reading an earlier draft of the paper and making constructive suggestions. We are also grateful to a referee for making a number of constructive suggestions, which led to a significant improvement of our paper.

Appendix. Proofs

We use the following notations in the proof:

l_{ad}: adjusted log-likelihood with general adjustment term $h(A)$; additional suffixes such as RE and PML are used when the log-likelihood functions refer to residual and profile log-likelihood, respectively.
$l_{ad}^{(j)}$: jth derivative of l_{ad} with respect to A $(j = 1, 2, 3)$;
$\tilde{l}_{ad}^{(j)}$: jth derivative of $\log h(A)$ with respect to A $(j = 1, 2, 3)$.

Lemma 1. *Let $y \sim N(0, \Sigma)$. Then for symmetric matrices F and G,*

$$Cov(y'Fy, y'Gy) = 2tr[F\Sigma G\Sigma];$$
$$Cov[(y'Fy)^2, y'Gy] = 8tr[F\Sigma F\Sigma G\Sigma] + 2tr[F\Sigma G\Sigma]tr[F\Sigma].$$

See [24] for a proof of the above lemma.

Lemma 2 stated below provides expressions for $l_{ad}^{(1)}, l_{ad}^{(2)}$ and $l_{ad}^{(3)}$.

Lemma 2. *Under regularity conditions of Theorem 1, we have*

$$\text{(i) } l_{ad.RE}^{(1)} = \frac{1}{2}[y'P^2y - tr(P)] + \tilde{l}_{ad}^{(1)},$$
$$l_{ad.PML}^{(1)} = \frac{1}{2}[y'P^2y - tr(V^{-1})] + \tilde{l}_{ad}^{(1)},$$
$$\text{(ii) } l_{ad}^{(2)} = -y'P^3y + \frac{1}{2}tr(V^{-2}) + \tilde{l}_{ad}^{(2)} + O(1),$$
$$\text{(iii) } l_{ad}^{(3)} = 3y'P^4y + tr(V^{-3}) + \tilde{l}_{ad}^{(3)} + O(1),$$

where $\tilde{l}_{ad}^{(1)} = O(1)$ and $l_{ad} \in \{l_{ad.RE}, l_{ad.PML}\}$.

Proof of Lemma 2. First note that $l_{ad.RE}$ and $l_{ad.PML}$ are given by

$$l_{ad.RE} = \log L_R(A) + \log h(A)$$
$$l_{ad.PML} = \log L_P(A) + \log h(A).$$

Part (i) is obtained by taking the derivatives of $l_{ad.RE}$ and $l_{ad.PML}$ with respect to A. Using the fact that $|tr(P^k) - tr(V^{-k})| = O(1)$ for $k \geq 1$, we get parts (ii) and (iii).

Remark. For the choice $h(A) \equiv h_{YL}(A)$, we have $\tilde{l}_{ad}^{(1)} = \frac{1}{m}\frac{\sum_{i=1}^{m}(D_i/(A+D_i)^2)}{\tan^{-1}(tr(I-B))(1+tr(I-B)^2)} = O(m^{-1})$ because $\frac{1}{\tan^{-1}(tr(I-B))(1+tr(I-B)^2)} \leq \frac{1}{tr(I-B)}$ for $A \geq 0$. Using the property of $h_{YL}(A)$, we can similarly show that $\tilde{l}_{ad}^{(2)}$ and $\tilde{l}_{ad}^{(3)}$ are of the order $O(m^{-1})$.

A straightforward application of Lemmas 1 and 2 yields the following lemma.

Lemma 3. *Under regularity conditions of Theorem 1, we have*

$$E[l_{ad.RE}^{(1)}] = \tilde{l}_{ad}^{(1)}, \quad (5)$$
$$E[l_{ad.PML}^{(1)}] = \frac{1}{2}[tr(P) - tr(V^{-1})] + \tilde{l}_{ad}^{(1)}, \quad (6)$$
$$E[(l_{ad}^{(1)})^2] = \frac{1}{2}tr(V^{-1}) + O(1), \quad (7)$$
$$E[l_{ad}^{(2)}] = -\frac{1}{2}tr(V^{-1}) + O(1), \quad (8)$$
$$E[l_{ad}^{(3)}] = 2tr(V^{-3}) + O(1), \quad (9)$$
$$Cov(l_{ad}^{(1)}, l_{ad}^{(2)}) = -tr[V^{-3}] + O(1), \quad (10)$$
$$Cov((l_{ad}^{(1)})^2, l_{ad}^{(3)}) = O(m). \quad (11)$$

An outline of proof of Theorem 1 and Result 1: proofs of Theorem 1 and Result 1 follow along the lines of the proof given in [20] and repeated applications of Lemmas 1–3.

Using the fact that $h(A)$ and its derivatives are both of order $O(1)$, we verify regularity conditions of Theorem 2.1 given in [7]. Now an application of Theorem 2.1 of [7] yields

$$E(\hat{A}_{Ad.g} - A)^2 = E\left(-\frac{l_{ad}^{(1)}}{E[l_{ad}^{(2)}]} + r\right)^2 + o(m^{-1})$$
$$= \frac{E[(l_{ad}^{(1)})^2]}{E[l_{ad}^{(2)}]^2} + R_1 + R_2 + o(m^{-1}), \quad (12)$$

where $R_1 = -2\frac{E[l_{ad}^{(1)} r]}{E[l_{ad}^{(2)}]}$, $R_2 = E[r^2]$ and $|r| \leq m^{-\rho}U$ with $E[U^g]$ bounded for any fixed $0 < \rho < 1$ and $g > 0$. Hence $R_2 = o(m^{-1})$ under regularity conditions. Using the Cauchy–Schwarz inequality, (7), (8) and the fact that $|tr(P^k) - tr(V^{-k})| = O(1)$ for $k \geq 1$, it is straightforward to prove that $R_1 = o(m^{-1})$. Thus, we get part (iii) of Theorem 1 (also that of Result 1) using

$$(12) = \frac{2}{tr(V^{-2})} + o(m^{-1}).$$

Finally, we show parts (i) and (ii) of Theorem 1 (and those in Result 1), which are related to the bias of $\hat{A}$. Using the fact that $h(A)$ is four times continuously differentiable with respect to A ($A > 0$), we verify regularity conditions of Theorem 4.1 of [7] for proving the bias of $\hat{A}$. We, therefore, obtain

$$E[\hat{A} - A] = I + II + III + o(m^{-1}),$$

where

$$I = -\frac{E[l_{ad}^{(1)}]}{E[l_{ad}^{(2)}]},$$
$$II = \frac{Cov(l_{ad}^{(1)},\ l_{ad}^{(2)})}{E[l_{ad}^{(2)}]^2},$$
$$III = -\frac{1}{2}\frac{Cov[(l_{ad}^{(1)})^2, l_{ad}^{(3)}] + E[(l_{ad}^{(1)})^2]E[l_{ad}^{(3)}]}{E[l_{ad}^{(2)}]^3}.$$

Using from (5) to (11) and following the proof given in [20], we obtain

$$I = \frac{2E[l_{ad}^{(1)}]}{tr(V^{-2})},$$
$$II = -\frac{4tr(V^{-3})}{[V^{-2}]^2} + o(m^{-1}),$$
$$III = \frac{4tr(V^{-3})}{[V^{-2}]^2} + o(m^{-1}).$$

This completes the proof of parts (i) and (ii) of Theorem 1. We can also get parts (i) and (ii) of Result 1 using $\tilde{l}_{ad}^{(1)} = O(m^{-1})$.

References

[1] W.R. Bell, Accounting for uncertainty about variances in small area estimation, in: Bulletin of the International Statistical Institute, 52nd Session, Helsinki, 1999. Available at www.census.gov/hhes/www/saipe under "Publications".

[2] W.R. Bell, M.C. Otto, Sampling error modelling of poverty and income statistics for states, in: Proceedings of the Section on Government Statistics, American Statistical Association, Alexandria, VA, 1995, pp. 160–165.

[3] F. Butar, P. Lahiri, Empirical Bayes estimation of several population means and variances under random sampling variances model, J. Statist. Plann. Inference 102 (2003) 59–69.

[4] G.M. Carter, J.F. Rolph, Empirical Bayes methods applied to estimating fire alarm probabilities, J. Amer. Statist. Assoc. 69 (1974) 880–885.

[5] S. Chatterjee, P. Lahiri, 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, pp. 3486–3493.

[6] S. Chen, P. Lahiri, On mean squared prediction error estimation in small area estimation problems, Commun. Stat.—Theory Methods 37 (2008) 1792–1798.

[7] K. Das, J. Jiang, J.N.K. Rao, Mean squared error of empirical predictor, Ann. Statist. 32 (2004) 818–840.

[8] G.S. Datta, P. Lahiri, A unified measure of uncertainty of estimated best linear unbiased predictions in small area estimation problems, Statist. Sinica 10 (2000) 613–627.

[9] 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.

[10] B. Efron, C. Morris, Data analysis using Stein's estimator and its generalizations, J. Amer. Statist. Assoc. 70 (1975) 311–319.

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

[12] R.E. Fay, G.F. Train, Small domain methodology for estimating income and poverty characteristics for states in 1993, in: Proceedings of the Social Statistics Section, American Statistical Association, 1997, pp. 183–188.

[13] P. Hall, T. Maiti, Nonparametric estimation of mean-squared prediction error in nested-error regression models, Ann. Statist. 34 (2006) 1733–1750.
[14] J. Jiang, P. Lahiri, Mixed model prediction and small area estimation, Test 15 (2006) 1–96. Editor's invited discussion paper.
[15] J. Jiang, T. Nguyen, J.S. Rao, Best predictive small area estimation, J. Amer. Statist. Assoc. 106 (2011) 732–745.
[16] R.N. Kackar, D.A. Harville, Approximations for standard errors of estimators of fixed and random effects in mixed linear models, J. Amer. Statist. Assoc. 79 (1984) 853–862.
[17] P. Lahiri, A review of empirical best linear unbiased prediction for the Fay–Herriot small area model, Philippine Statist 52 (1–4) (2003) 1–15.
[18] P. Lahiri, H. Li, Generalized maximum likelihood method in linear mixed models with an application in small area estimation, in: Proceedings of the Federal Committee on Statistical Methodology Research Conference, 2009. Available at http//www.fcsm.gov/events/papers2009.html.
[19] P. Lahiri, S. Pramanik, in: C. Morris, R. Tang (Eds.), Discussion of "Estimating Random Effects via Adjustment for Density Maximization", in: Statistical Science, vol. 26, 2011, pp. 291–295.
[20] H. Li, P. Lahiri, An adjusted maximum likelihood method for solving small area estimation problems, J. Multivariate Anal. 101 (2010) 882–892.
[21] C.N. Morris, R. Tang, Estimating random effects via adjustment for density maximization, Statist. Sci. 26 (2011) 271–287.
[22] N.G.N. Prasad, J.N.K. Rao, The estimation of the mean squared error of small area estimators, J. Amer. Statist. Assoc. 85 (1990) 163–171.
[23] J.N.K. Rao, Small Area Estimation, Wiley, New York, 2003.
[24] V.K. Srivastava, R. Tiwari, Evaluation of expectation of products of stochastic matrics, Scand. J. Stat. 3 (1976) 135–138.
[25] J. Wang, W.A. Fuller, The mean squared error of small area predictors constructed with estimated area variances, J. Amer. Statist. Assoc. 98 (2003) 716–723.