The Annals of Statistics
2014, Vol. 42, No. 4, 1233–1261
DOI: 10.1214/14-AOS1219

A SECOND-ORDER EFFICIENT EMPIRICAL BAYES CONFIDENCE INTERVAL

BY MASAYO YOSHIMORI[1] AND PARTHA LAHIRI[2]

Osaka University and University of Maryland

We introduce a new adjusted residual maximum likelihood method (REML) in the context of producing an empirical Bayes (EB) confidence interval for a normal mean, a problem of great interest in different small area applications. Like other rival empirical Bayes confidence intervals such as the well-known parametric bootstrap empirical Bayes method, the proposed interval is second-order correct, that is, the proposed interval has a coverage error of order $O(m^{-3/2})$. Moreover, the proposed interval is carefully constructed so that it always produces an interval shorter than the corresponding direct confidence interval, a property not analytically proved for other competing methods that have the same coverage error of order $O(m^{-3/2})$. The proposed method is not simulation-based and requires only a fraction of computing time needed for the corresponding parametric bootstrap empirical Bayes confidence interval. A Monte Carlo simulation study demonstrates the superiority of the proposed method over other competing methods.

1. Introduction. Fay and Herriot (1979) considered empirical Bayes estimation of small area means θ_i using the following two-level Bayesian model and demonstrated, using real life data, that they outperform both the direct and synthetic (e.g., regression) estimators.

The Fay–Herriot model:

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

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

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

In the above model, level 1 is used to account for the sampling distribution of the direct survey estimates y_i, which are usually weighted averages of the sample observations in area i. Level 2 prior distribution links the true small area means θ_i to a vector of $p < m$ known area level auxiliary variables $x_i = (x_{i1}, \dots, x_{ip})'$, often obtained from various administrative records. The hyperparameters $\beta \in R^p$, the p-dimensional Euclidean space, and $A \in [0, \infty)$ of the linking model are generally unknown and are estimated from the available data.

Received March 2013; revised March 2014.

[1]Supported by the JSPS KAKENHI Grant Number 242742.

[2]Supported by the NSF SES-085100.

MSC2010 subject classifications. Primary 62C12; secondary 62F25.

Key words and phrases. Adjusted maximum likelihood, coverage error, empirical Bayes, linear mixed model.

It is often difficult or even impossible to retrieve all important sample data within small areas due to confidentiality or other reasons and the only data an analyst may have access to are aggregate data at the small area level. The Fay–Herriot model comes handy in such situations since only area level aggregate data are needed to implement the model. Even when unit level data are available within small areas, analysts may have some preference for the Fay–Herriot model over a more detailed (and perhaps more scientific) unit level model in order to simplify the modeling task. One good feature of the Fay–Herriot model is that the resulting empirical Bayes (EB) estimators of small area means are design-consistent. In the Fay–Herriot model, sampling variances D_i are assumed to be known, which often follows from the asymptotic variances of transformed direct estimates [Efron and Morris (1975), Carter and Rolph (1974)] and/or from empirical variance modeling [Fay and Herriot (1979)]. This known sampling variance assumption causes underestimation of the mean squared error (MSE) of the resulting empirical Bayes estimator of θ_i. Despite this limitation, the Fay–Herriot model has been widely used in different small area applications [see, e.g., Carter and Rolph (1974), Efron and Morris (1975), Bell et al. (2007), Fay and Herriot (1979), and others].

Note that the empirical Bayes estimator of θ_i obtained by Fay and Herriot (1979) can be motivated as an empirical best prediction (EBP) estimator [in this case same as the empirical best linear unbiased prediction (EBLUP) estimator] of the mixed effect $\theta_i = x_i'\beta + v_i$, under the following linear mixed model:

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

where the v_i's and e_i's are independent with $v_i \overset{\text{i.i.d.}}{\sim} N(0, A)$ and $e_i \overset{\text{ind}}{\sim} N(0, D_i)$; see Prasad and Rao (1990) and Rao (2003).

In this paper, we consider interval estimation of small area means θ_i. An interval, denoted by I_i, is called a $100(1-\alpha)\%$ interval for θ_i if $P(\theta_i \in I_i|\beta, A) = 1-\alpha$, for any fixed $\beta \in R^p$, $A \in (0, \infty)$, where the probability P is with respect to the Fay–Herriot model. Throughout the paper, $P(\theta_i \in I_i|\beta, A)$ is referred to as the coverage probability of the interval I_i; that is, coverage is defined in terms of the joint distribution of y and θ with fixed hyperparameters β and A. Most intervals proposed in the literature can be written as: $\hat{\theta}_i \pm s_\alpha \hat{\tau}_i(\hat{\theta}_i)$, where $\hat{\theta}_i$ is an estimator of θ_i, $\hat{\tau}_i(\hat{\theta}_i)$ is an estimate of the measure of uncertainty of $\hat{\theta}_i$ and s_α is suitably chosen in an effort to attain coverage probability close to the nominal level $1-\alpha$.

Researchers have considered different choices for $\hat{\theta}_i$. For example, the choice $\hat{\theta}_i = y_i$ leads to the direct confidence interval I_i^D, given by

$$I_i^D : y_i \pm z_{\alpha/2}\sqrt{D_i},$$

where $z_{\alpha/2}$ is the upper $100(1-\alpha/2)\%$ point of $N(0, 1)$. Obviously, for this direct interval, the coverage probability is $1-\alpha$. However, when D_i is large as in the case of small area estimation, its length is too large to make any reasonable conclusion.

The choice $\hat{\theta}_i = x_i'\hat{\beta}$, where $\hat{\beta}$ is a consistent estimator of β, provides an interval based on the regression synthetic estimator of θ_i. Hall and Maiti (2006) considered this choice with $\hat{\tau}_i(\hat{\theta}_i) = \sqrt{\hat{A}}$, $\hat{A}$ being a consistent estimator of A, and obtained s_α using a parametric bootstrap method. This approach could be useful when y_i is missing for the ith area.

We call an interval empirical Bayes (EB) confidence interval if we choose an empirical Bayes estimator for $\hat{\theta}_i$. There has been a considerable interest in constructing empirical Bayes confidence intervals, starting from the work of Cox (1975) and Morris (1983a), because of good theoretical and empirical properties of empirical Bayes point estimators. Before introducing an empirical Bayes confidence interval, we introduce the Bayesian credible interval in the context of the Fay–Herriot model. When the hyperparameters β and A are known, the Bayesian credible interval of θ_i is obtained using the posterior distribution of $\theta_i : \theta_i | y_i; (\beta, A) \sim N[\hat{\theta}_i^B, \sigma_i(A)]$, where $\hat{\theta}_i^{\mathrm{B}} \equiv \hat{\theta}_i^B(\beta, A) = (1 - B_i)y_i + B_i x_i'\beta$, $B_i \equiv B_i(A) = \frac{D_i}{D_i + A}$, $\sigma_i(A) = \sqrt{\frac{AD_i}{A + D_i}}$ $(i = 1, \dots, m)$. Such a credible interval is given by

$$I_i^B(\beta, A) : \hat{\theta}_i^B(\beta, A) \pm z_{\alpha/2}\sigma_i(A).$$

The Bayesian credible interval cuts down the length of the direct confidence interval by $100 \times (1 - \sqrt{1 - B_i})\%$ while maintaining the exact coverage $1 - \alpha$ with respect to the joint distribution of y_i and θ_i. The maximum benefit from the Bayesian methodology is achieved when B_i is close to 1, that is, when the prior variance A is much smaller than the sampling variances D_i.

In practice, the hyperparameters are unknown. Cox (1975) initiated the idea of developing an one-sided empirical Bayes confidence interval for θ_i for a special case of the Fay–Herriot model with $p = 1$, $x_i'\beta = \beta$ and $D_i = D$ $(i = 1, \dots, m)$. The two-sided version of his confidence interval is given by

$$I_i^{\mathrm{Cox}}(\hat{\beta}, \hat{A}_{\mathrm{ANOVA}}) : \hat{\theta}_i^B(\hat{\beta}, \hat{A}_{\mathrm{ANOVA}}) \pm z_{\alpha/2}\sigma(\hat{A}_{\mathrm{ANOVA}}),$$

where $\hat{\theta}_i^B(\hat{\beta}, \hat{A}_{\mathrm{ANOVA}}) = (1 - \hat{B})y_i + \hat{B}\hat{\beta}$, an empirical Bayes estimator of θ_i; $\hat{\beta} = m^{-1}\sum_{i=1}^m y_i$ and $\hat{B} = D/(D + \hat{A}_{\mathrm{ANOVA}})$ with $\hat{A}_{\mathrm{ANOVA}} = \max\{(m - 1)^{-1}\sum_{i=1}^m (y_i - \hat{\beta})^2 - D, 0\}$. An extension of this ANOVA estimator for the Fay–Herriot model can be found in Prasad and Rao (1990).

Like the Bayesian credible interval, the length of the Cox interval is smaller than that of the direct interval. However, the Cox empirical Bayes confidence interval introduces a coverage error of the order $O(m^{-1})$, not accurate enough in most small area applications. In fact, Cox (1975) recognized the problem and considered a different α', motivated from a higher-order asymptotic expansion, in order to bring the coverage error down to $o(m^{-1})$. However, such an adjustment may cause the interval to be undefined when $\hat{A}_{\mathrm{ANOVA}} = 0$ and sacrifices an appealing feature of $I_i^{\mathrm{Cox}}(\hat{\mu}, \hat{A}_{\mathrm{ANOVA}})$, that is, the length of such interval may no longer be less than that of the direct method.

One may argue that Cox's method has an undercoverage problem because it does not incorporate uncertainty due to estimation of the regression coefficients β and prior variance A in measuring uncertainty of the empirical Bayes estimator of θ_i. Morris (1983a) used an improved measure of uncertainty for his empirical Bayes estimator that incorporates the additional uncertainty due to the estimation of the model parameters. Similar ideas can be found in Prasad and Rao (1990) for a more general model. However, Basu, Ghosh and Mukerjee (2003) showed that the coverage error of the empirical Bayes confidence interval proposed by Morris (1983a) remains $O(m^{-1})$. In the context of the Fay–Herriot model, Diao et al. (2014) examined the higher order asymptotic coverage of a class of empirical Bayes confidence intervals of the form: $\theta_i^{\mathrm{EB}} \pm z_{\alpha/2}\sqrt{\mathrm{mse}_i}$, where θ_i^{EB} is an empirical Bayes estimator of θ_i that uses a consistent estimator of A and mse_i is a second-order unbiased estimator of $\mathrm{MSE}(\theta_i^{\mathrm{EB}})$ given in Datta and Lahiri (2000). They showed that the coverage error for such an interval is $O(m^{-1})$. In a simulation study, Yoshimori (2014) observed poor finite sample performance of such empirical Bayes confidence intervals. Furthermore, it is not clear if the length of such confidence interval is always less than that of the direct method. Morris (1983b) considered a variation of his (1983a) empirical Bayes confidence interval where he used a hierarchical Bayes-type point estimator in place of the previously used empirical Bayes estimator and conjectured, with some evidence, that the coverage probability for his interval is at least $1-\alpha$. He also noted that the coverage probability tends to $1-\alpha$ as m goes to ∞ or D goes to zero. However, higher-order asymptotic properties of this confidence interval are unknown.

Using a Taylor series expansion, Basu, Ghosh and Mukerjee (2003) obtained expressions for the order $O(m^{-1})$ term of the coverage errors of the Morris' interval and another prediction interval proposed by Carlin and Louis [(1996), page 98], which were then used to calibrate the lengths of these empirical Bayes confidence intervals in order to reduce the coverage errors down to $o(m^{-1})$. However, it is not known if the lengths of their confidence intervals are always smaller than that of the direct method. Using a multilevel model, Nandram (1999) obtained an empirical Bayes confidence interval for a small area mean and showed that asymptotically it converges to the nominal coverage probability. However, he did not study the higher-order asymptotic properties of his interval.

Researchers considered improving the coverage property of the Cox-type empirical Bayes confidence interval by changing the normal percentile point $z_{\alpha/2}$. For the model used by Cox (1975), Laird and Louis (1987) proposed a prediction interval based on parametric bootstrap samples. However, the order of their coverage error has not been studied analytically. Datta et al. (2002) used a Taylor series approach similar to that of Basu, Ghosh and Mukerjee (2003) in order to calibrate the Cox-type empirical Bayes confidence interval for the general Fay–Herriot model. Using mathematical tools similar to Sasase and Sasase and Kubokawa (2005), Yoshimori (2014) extended the method of Datta et al. (2002) and Basu, Ghosh and Mukerjee (2003) when REML estimator of A is used.

For a general linear mixed model, Chatterjee, Lahiri and Li (2008) developed a parametric bootstrap empirical Bayes confidence interval for a general mixed effect and examined its higher order asymptotic properties. For the special case, this can be viewed as a Cox-type empirical Bayes confidence interval where $z_{\alpha/2}$ is replaced by percentile points obtained using a parametric bootstrap method. While the parametric bootstrap empirical Bayes confidence interval of Chatterjee, Lahiri and Li (2008) has good theoretical properties, one must apply caution in choosing B, the number of bootstrap replications, and the estimator of A. In two different simulation studies, Li and Lahiri (2010) and Yoshimori (2014) found that the parametric bootstrap empirical Bayes confidence interval did not perform well when REML method is used to estimate A. Li and Lahiri (2010) developed an adjusted REML estimator of A that works better than the REML in their simulation setting. Moreover, in absence of a sophisticated software, analysts with modest computing skills may find it a daunting task to evaluate parametric bootstrap confidence intervals in a large scale simulation experiment. The coverage errors of confidence intervals developed by Datta et al. (2002), Chatterjee, Lahiri and Li (2008) and Li and Lahiri (2010) are of the order $O(m^{-3/2})$. However, there is no analytical result that suggests the lengths of these confidence intervals are smaller than the length of the direct method.

In Section 2, we introduce a list of notation and regularity conditions used in the paper. In this paper, our goal is to find an empirical Bayes confidence interval of θ_i that (i) matches the coverage error properties of the best known empirical Bayes method such as the one proposed by Chatterjee, Lahiri and Li (2008), (ii) has length smaller than that of the direct method and (iii) does not rely on simulation-based heavy computation. In Section 3, we propose such a new interval method for the general Fay–Herriot model by replacing the ANOVA estimator of A in the Cox interval by a carefully devised adjusted residual maximum likelihood estimator of A. Lahiri and Li (2009) introduced a generalized (or adjusted) maximum likelihood method for estimating variance components in a general linear mixed model. Li and Lahiri (2010) and Yoshimori and Lahiri (2014) examined different adjustment factors for point estimation of the small area means in the context of the Fay–Herriot model. But none of the authors explored adjusted residual likelihood method for constructing small area confidence intervals. In Section 4, we compare our proposed confidence interval methods with the direct, different Cox-type EB confidence intervals and the parametric bootstrap empirical Bayes confidence interval method of Chatterjee, Lahiri and Li (2008) using a Monte Carlo simulation study. The proofs of all technical results presented in Section 3 are deferred to the Appendix.

2. A list of notation and regularity conditions. We use the following notation throughout the paper:

$y = (y_1, \dots, y_m)'$, a $m \times 1$ column vector of direct estimates;

$X' = (x_1, \dots, x_m)$, a $p \times m$ known matrix of rank p;
$q_i = x_i'(X'X)^{-1}x_i$, leverage of area i for level 2 model, $(i = 1, \dots, m)$;
$V = \mathrm{diag}(A + D_1, \dots, A + D_m)$, a $m \times m$ diagonal matrix;
$P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}$;
$L_{\mathrm{RE}}(A) = |X'V^{-1}X|^{-1/2}|V|^{-1/2}\exp(-\frac{1}{2}y'Py)$, the residual likelihood function of A;
$h_i(A)$ is a general area specific adjustment factor;
$L_{i;\mathrm{ad}}(A) = h_i(A) \times L_{\mathrm{RE}}(A)$, adjusted residual likelihood function of A with a general adjustment factor $h_i(A)$;
$\hat{A}_{h_i} = \arg\max_{A \in [0,\infty)} L_{i;\mathrm{ad}}(A)$, adjusted residual maximum likelihood estimator of A with respect to a general adjustment factor $h_i(A)$;
$l_{\mathrm{RE}}(A) = \log[L_{\mathrm{RE}}(A)]$;
$\tilde{l}_{i;\mathrm{ad}}(A) = \log h_i(A)$;
$l_{i;\mathrm{ad}}(A) = \log L_{i;\mathrm{ad}}(A)$;
$\tilde{l}_{i,\mathrm{ad}}^{(k)}(A) \equiv \frac{\partial^k \tilde{l}_{i,\mathrm{ad}}(A)}{\partial A^k}$, kth derivative of $\tilde{l}_{i,\mathrm{ad}}(A)$, $(k \geq 1)$;
$l_{i,\mathrm{ad}}^{(k)}(A) \equiv \frac{\partial^k l_{i,\mathrm{ad}}(A)}{\partial A^k}$, kth derivative of $l_{i,\mathrm{ad}}(A)$, $(k \geq 1)$;
$\hat{V} = \mathrm{diag}(\hat{A}_{h_1} + D_1, \dots, \hat{A}_{h_m} + D_m)$, $(i = 1, \dots, m)$;
$\tilde{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$, weighted least square estimator of β when A is known;
$\hat{\beta} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}y$, weighted least square estimator of β when $A + D_i$ is replaced by $\hat{A}_{h_i} + D_i$, $(i = 1, \dots, m)$;
$B_i = D_i/(\hat{A} + D_i)$, shrinkage factor for the i area, $(i = 1, \dots, m)$;
$\hat{B}_i \equiv \hat{B}_i(\hat{A}_{h_i}) = D_i/(\hat{A}_{h_i} + D_i)$, estimated shrinkage factor for the i area, $(i = 1, \dots, m)$;
$\hat{\theta}_i^{\mathrm{B}} \equiv \hat{\theta}_i^{B}(\beta, A) = (1 - B_i)y_i + B_i x_i'\beta$;
$\hat{\theta}_i^{\mathrm{EB}} \equiv \hat{\theta}_i^{\mathrm{EB}}(\hat{A}_{h_i}) \equiv \hat{\theta}_i^{\mathrm{B}}(\hat{\beta}, \hat{A}_{h_i}) = (1 - \hat{B}_i)y_i + \hat{B}_i x_i'\hat{\beta}$, empirical Bayes estimator of θ_i, $(i = 1, \dots, m)$;
$I_i^{\mathrm{Cox}}(\hat{\beta}, \hat{A}_{h_i}) \equiv I_i^{\mathrm{Cox}}(\hat{A}_{h_i}) : \hat{\theta}_i^{B}(\hat{\beta}, \hat{A}_{h_i}) \pm z_{\alpha/2}\sigma_i(\hat{A}_{h_i})$, Cox-type EB confidence interval of θ_i using adjusted REML $\hat{A}_{h_i}$, where $z = z_{\alpha/2}$ is the upper $100(1 - \alpha/2)\%$ point of the normal deviate.

We use the following regularity conditions in proving different results presented in this paper.

Regularity conditions:

R1: The logarithm of the adjustment term $\tilde{l}_{\mathrm{ad}}(A)$ [or $\tilde{l}_{i,\mathrm{ad}}(A)$] is free of y and is five times continuously differentiable with respect to A. Moreover, the gth power of the $|\tilde{l}_{\mathrm{ad}}^{(j)}(A)|$ [or $|\tilde{l}_{i,\mathrm{ad}}^{(j)}(A)|$] is bounded for $g > 0$ and $j = 1, 2, 3, 4, 5$;
R2: $\mathrm{rank}(X) = p$;
R3: The elements of X are uniformly bounded implying $\sup_{j \geq 1} q_j = O(m^{-1})$;
R4: $0 < \inf_{j \geq 1} D_j \leq \sup_{j \geq 1} D_j < \infty$, $A \in (0, \infty)$;
R5: $|\hat{A}_{h_i}| < C_+ m^{\lambda}$, where C_+ a generic positive constant and λ is small positive constant.

3. A new second-order efficient empirical Bayes confidence interval. We call an empirical Bayes interval of θ_i second-order efficient if the coverage error is of order $O(m^{-3/2})$ and length shorter than that of the direct confidence interval. The goal of this section is to produce such an interval that requires a fraction of computer time required by the recently proposed parametric bootstrap empirical Bayes confidence interval. Our idea is simple and involves replacement of the ANOVA estimator of A in the empirical Bayes interval proposed by Cox (1975) by a carefully devised adjusted residual maximum likelihood estimator of A.

Theorem 1 provides a higher-order asymptotic expansion of the confidence interval $I_i^{\mathrm{Cox}}(\hat{A}_{h_i})$. The theorem holds for any area $1 \le i \le m$, for large m.

THEOREM 1. *Under regularity conditions* R1–R5, *we have*

$$P\{\theta_i \in I_i^{\mathrm{Cox}}(\hat{A}_{h_i})\} = 1 - \alpha + z\phi(z)\frac{a_i + b_i[h_i(A)]}{m} + O(m^{-3/2}), \tag{3.1}$$

where

$$\begin{aligned} a_i = {} & -\frac{m}{\operatorname{tr}(V^{-2})}\left[\frac{4D_i}{A(A+D_i)^2} + \frac{(1+z^2)D_i^2}{2A^2(A+D_i)^2}\right] \\ & - \frac{mD_i}{A(A+D_i)} x_i' \operatorname{Var}(\tilde{\beta}) x_i, \end{aligned} \tag{3.2}$$

$$b_i \equiv b_i[h_i(A)] = \frac{2m}{\operatorname{tr}(V^{-2})}\frac{D_i}{A(A+D_i)} \times \tilde{l}_{i;\mathrm{ad}}^{(1)}. \tag{3.3}$$

We can produce higher order asymptotic expansion of the coverage probability of Cox-type EB confidence interval with any standard likelihood-based estimator of A available in the literature (e.g., residual maximum likelihood, profile maximum likelihood, different adjusted residual and profile maximum likelihood, etc.) simply by choosing an appropriate $h_i(A)$ [e.g., for REML, $h_i(A) = 1$] and using equation (3.1). We have verified that coverage errors for all these Cox-type EB confidence intervals are of order $O(m^{-1})$. We can, however, use equation (3.1) to reduce the coverage error to the order $O(m^{-3/2})$ by choosing $h_i(A)$ such that the order $O(m^{-1})$ term in the right-hand side of (3.1) vanishes. More specifically, we first obtain an expression for $h_i(A)$ by finding a solution to the following differential equation:

$$a_i + b_i[h_i(A)] = 0 \tag{3.4}$$

and then maximize the adjusted residual likelihood $L_{i;\mathrm{ad}}(A)$ with respect to $A \in [0, \infty)$ to obtain our adjusted residual maximum likelihood estimator of A, which is used to construct the desired Cox-type second-order efficient EB confidence interval for θ_i. Notice that we can produce two different new adjusted REML estimators of A by using generalized least square (GLS) and ordinary least square

(OLS) estimators of β in the EB estimator of θ_i. Let $h_{i;\text{gls}}(A)$ and $h_{i;\text{ols}}(A)$ denote the adjustment factors that are solutions of $h_i(A)$ in (3.4) with GLS and OLS estimators of β in $\hat{\theta}_i^{\text{EB}}$, respectively. We denote the corresponding adjusted residual maximum likelihood estimators of A by $\hat{A}_{i;\text{gls}}$ and $\hat{A}_{i;\text{ols}}$. Note that in general we cannot obtain $h_{i;\text{ols}}(A)$ as a special case of $h_{i;\text{gls}}(A)$ except for the balanced case $D_i = D, i = 1, \dots, m$ when the GLS and OLS estimators of β are identical. Consequently, $\hat{A}_{i;\text{gls}}$ is generally different from $\hat{A}_{i;\text{ols}}$ except for the balanced case when $\hat{A}_{i;\text{gls}} = \hat{A}_{i;\text{ols}} = \hat{A}_i$ (say).

Theorem 2 provides expressions for $h_{i;\text{gls}}(A)$ and $h_{i;\text{ols}}(A)$ and states the uniqueness of $\hat{A}_i$ for the balanced case. In Theorem 2 and elsewhere in the paper, C is a generic constant free of A.

THEOREM 2. (i) *The expressions for* $h_{i;\text{gls}}(A)$ *and* $h_{i;\text{ols}}(A)$ *are given by*

$$\begin{aligned} h_{i;\text{gls}}(A) &= C A^{(1/4)(1+z^2)}(A + D_i)^{(1/4)(7-z^2)} \\ &\quad \times \exp\left[\int \frac{1}{2}\operatorname{tr}(V^{-2})x_i'(X'V^{-1}X)^{-1}x_i \, dA\right], \end{aligned} \tag{3.5}$$

$$\begin{aligned} h_{i;\text{ols}}(A) &= C A^{(1/4)(1+z^2)}(A + D_i)^{(1/4)(7-z^2)}\left[\prod_{i=1}^{m}(A + D_i)\right]^{(1/2)q_i} \\ &\quad \times \exp\left[-\frac{1}{2}\operatorname{tr}(V^{-1})x_i'(X'X)^{-1}X'VX(X'X)^{-1}x_i\right]. \end{aligned} \tag{3.6}$$

(ii) *For the balanced case* $D_i = D$ $(i = 1, \dots, m)$, *we have*

$$h_{i;\text{gls}}(A) = h_{i;\text{ols}} = C A^{(1/4)(1+z^2)}(A + D)^{(1/4)(7-z^2)+(1/2)mq_i}. \tag{3.7}$$

In this balanced case, the $\hat{A}_i$ *is unique provided* $m > \frac{4+p}{1-q_i}$.

REMARK 1. Note that $h_{i;\text{gls}}(A)$ does not have a closed-form expression in A. But this is not an issue since finding a root of the corresponding likelihood equation remains simple in this case because the derivative of $\log[h_i(A)]$ has a closed-form expression. Just like the standard residual likelihood, our adjusted residual likelihood function could have multiple maxima in the general balanced case. We refer to Searle, Casella and McCulloch [(1992), Section 8.1] who suggested a way to search for the global maximum. In this connection, we refer to Gan and Jiang (1999) who proposed a method for testing for the global maximum. Moreover, in order to reduce the number of iterations, we suggest to use the simple ANOVA estimator of A proposed by Prasad and Rao (1990) as an initial value.

REMARK 2. In a real data analysis, one should check the condition $m > (4 + p)/(1 - h_i)$ for the existence of strictly positive estimates $\hat{A}_{i;\text{gls}}$ and $\hat{A}_{i;\text{gls}}$. Under the regularity conditions R2 and R3, the condition $m > (4+p)/(1-h_i)$ reduces to $m > m_0$, where m_0 is a fixed constant depending on p and the leverages q_i. Thus, for sufficiently large m, this condition does not pose any problem.

REMARK 3. One might be tempted to treat our adjustment factor $h_i(A)$ as a prior and conduct a regular hierarchical Bayesian analysis. But $h_i(A)$ may not be treated as a prior since in certain cases this leads to an improper posterior distribution of A. To illustrate our point, we consider the simple case: $D_i = D$ and $h_i(A) = h_{i;\mathrm{gls}}(A) = h_{i;\mathrm{ols}}(A)$, $i = 1, \dots, m$. Since

$$\begin{aligned} h_i(A) L_{\mathrm{RE}}(A) &= A^{(1+z^2)/4}(A+D)^{(7-z^2)/4+mq_i-m/2-p/2} \\ &\qquad \times \exp\left[-\frac{y'(I - X'(X'X)^{-1}X)y}{2(A+D)}\right] |X'X|^{-1/2} C \\ &\geq 0, \end{aligned}$$

under the regularity conditions, and $\exp[-\frac{y'(I-X'(X'X)^{-1}X)y}{2(A+D)}]$ and $A/(A+D)$ are increasing monotone functions of A, there exists $s < \infty$ such that

$$1 - \exp\left[-\frac{y'(I - X'(X'X)^{-1}X)y}{2(s+D)}\right] < \frac{1}{2}$$

and

$$1 - \frac{s}{s+D} < \frac{1}{2}.$$

Using the above results, we have

$$\int_0^\infty h_i(A) L_{\mathrm{RE}}\, dA \geq C \int_s^\infty (A+D)^{2+1/2[mq_i+p]-m/2}\, dA, \tag{3.8}$$

if $m > \frac{4+p}{1-q_i}$. Hence, if $-1 \leq 2 + 1/2[mq_i + p] - m/2 \leq 0$, the right-hand side of the above equation is infinite, even if $m > \frac{4+p}{1-q_i}$. Thus, in this case $h_i(A)$ cannot be treated as a prior since $\int_0^\infty h_i(A) L_{\mathrm{RE}}\, dA = \infty$ in case $-1 \leq 2 + 1/2[mq_i + p] - m/2 \leq 0$.

We now propose two empirical Bayes confidence intervals for θ_i:

$$I_i^{YL}(\hat{A}_{i;h}) : \hat{\theta}_i^{\mathrm{EB}}(\hat{A}_{i;h}) \pm z_{\alpha/2}\sigma_i(\hat{A}_{i;h}),$$

where $h = \mathrm{gls}, \mathrm{ols}$. Since $\sigma_i(\hat{A}_{i;h}) < \sqrt{D_i}$ ($h = \mathrm{gls}, \mathrm{ols}$), the length of our proposed Cox-type EB intervals, like the original Cox EB interval $I_i^{\mathrm{Cox}}(\hat{A}_{\mathrm{ANOVA}})$, are always shorter than that of the direct interval I_i^D. The following theorem compares the lengths of Cox EB confidence intervals of θ_i when A is estimated by $\hat{A}_{\mathrm{RE}}$, $\hat{A}_{i;\mathrm{gls}}$ and $\hat{A}_{i;\mathrm{ols}}$.

THEOREM 3. *Under the regularity conditions* R2–R4 *and* $m > (4+p)/(1-q_i)$, *we have*

$$\textit{Length of } I_i^{\mathrm{Cox}}(\hat{A}_{\mathrm{RE}}) \le \textit{Length of } I_i^{YL}(\hat{A}_{i;\mathrm{gls}}) \le \textit{Length of } I_i^{YL}(\hat{A}_{i;\mathrm{ols}}).$$

The following theorem provides the higher order asymptotic properties of a general class of adjusted residual maximum likelihood estimators of A.

THEOREM 4. *Under regularity conditions* R1–R5, *we have*:

(i) $E[\hat{A}_{h_i} - A] = \frac{2}{\mathrm{tr}(V^{-1})}\tilde{l}^{(1)}_{i,\mathrm{ad}}(A) + O(m^{-3/2})$,
(ii) $E(\hat{A}_{h_i} - A)^2 = \frac{2}{\mathrm{tr}(V^{-1})} + O(m^{-3/2})$.

COROLLARY TO THEOREM 4. *Under regularity conditions* R2–R5, *we have*:

(i) *Both* $\hat{A}_{i;\mathrm{gls}}$ *and* $\hat{A}_{i;\mathrm{gls}}$ *are strictly positive if* $m > \frac{4+p}{1-q_i}$,
(ii) $E[\hat{A}_{i;\mathrm{gls}} - A] = \frac{2}{\mathrm{tr}(V^{-2})}\tilde{l}^{(1)}_{i,\mathrm{ad};\mathrm{gls}}(A) + O(m^{-3/2})$,
(iii) $E[\hat{A}_{i;\mathrm{ols}} - A] = \frac{2}{\mathrm{tr}(V^{-2})}\tilde{l}^{(1)}_{i,\mathrm{ad};\mathrm{ols}}(A) + O(m^{-3/2})$,
(iv) $E(\hat{A}_{i;h} - A)^2 = \frac{2}{\mathrm{tr}(V^{-2})} + O(m^{-3/2})$,

where

$$\tilde{l}^{(1)}_{i;\mathrm{ad,gls}} = \frac{2}{A+D_i} + \frac{(1+z^2)D_i}{4A(A+D_i)} + \frac{1}{2}\,\mathrm{tr}(V^{-2})x_i'(X'V^{-1}X)^{-1}x_i,$$

$$\tilde{l}^{(1)}_{i;\mathrm{ad,ols}} = \frac{2}{A+D_i} + \frac{(1+z^2)D_i}{4A(A+D_i)} + \frac{1}{2}\,\mathrm{tr}(V^{-2})x_i'(X'X)^{-1}X'VX(X'X)^{-1}x_i.$$

REMARK 4. We reiterate that our true model variance is A, which is not area specific (i.e., it does not depend on i). However, unlike other likelihood based estimators of A, our theory driven proposed adjusted REML estimators $\hat{A}_{i;\mathrm{ols}}$ and $\hat{A}_{i;\mathrm{gls}}$ of A are area and confidence level specific. We would like to cite a similar situation that arises in the Bayesian small area inference. For the same two level model, flat prior distribution on A is widely accepted [see Morris and Tang (2011)]. However, in order to match the posterior variance with the classical MSE of EB with REML up to the order $O(m^{-1})$, Datta, Rao and Smith (2005) proposed a noncustomary prior for A that is area specific.

REMARK 5. The area and confidence level specific nature of our proposed estimators of a global parameter A naturally raises a concern that such proposed estimators may perform poorly when compared to rival estimators of A. To address this issue, first note that the consistency of the new adjusted REML estimators

$\hat{A}_{i;\text{ols}}$ and $\hat{A}_{i;\text{gls}}$ of A follows from part (iv) of the Corollary to Theorem 4. This is due to the fact that the leading term in the right-hand side tends to 0 as m tends to ∞, under the regularity conditions R2–R5. This result also implies that MSEs of the proposed estimators of A are identical, up to the order $O(m^{-1})$, to those of different likelihood based estimators of A such as REML, ML, different adjusted profile and residual maximum likelihood estimators of Li and Lahiri (2010) and Yoshimori and Lahiri (2014). Moreover, while such an area and confidence level specific adjustment causes the resulting proposed adjusted REML estimators to have more bias than that of REML, the biases remain negligible and are of order $O(m^{-1})$, same as the order of the bias of profile maximum likelihood or adjusted profile maximum likelihood estimators of A proposed by Li and Lahiri (2010) and Yoshimori and Lahiri (2014). Basically, we introduce this slight bias in $\hat{A}_{i;\text{ols}}$ and $\hat{A}_{i;\text{gls}}$ in order to achieve the desired low coverage error property while maintaining length always shorter than that of the corresponding direct confidence interval.

REMARK 6. Using the Corollary to Theorem 4 and the mathematical tools used in Li and Lahiri (2010), we obtain the following second-order approximation to the mean squared error (MSE) of $\hat{\theta}_i^{\text{EB}}(\hat{A}_{i;\text{gls}})$:

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

where $g_{1i}(A) = \frac{AD_i}{A+D_i}$, $g_{2i}(A) = \frac{D_i^2}{(A+D_i)^2}\operatorname{Var}(x_i'\hat{\beta}) = \frac{D_i^2}{(A+D_i)^2}x_i'(\sum_{j=1}^m \frac{x_j x_j'}{A+D_j})^{-1} \times x_i$, and $g_{3i}(A) = \frac{2D_i^2}{(A+D_i)^3}\{\sum_{j=1}^m \frac{1}{(A+D_j)^2}\}^{-1}$. Thus, in terms of MSE criterion, $\hat{\theta}_i^{\text{EB}}[\hat{A}_{i;\text{gls}}]$ is equally efficient, up to the order $O(m^{-1})$, as the empirical Bayes estimators of θ_i that use standard REML, PML and the adjusted PML and REML estimators of A proposed by Li and Lahiri (2010) and Yoshimori and Lahiri (2014).

We note that

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

where $g_{2i;\text{ols}}(A) = \frac{D_i^2}{(A+D_i)^2}x_i'(X'X)^{-1}X'VX(X'X)^{-1}x_i \geq \frac{D_i^2}{(A+D_i)^2} \times x_i'(X'V^{-1}X)^{-1}x_i$. Thus, in terms of higher order asymptotics $\hat{\theta}_i^{\text{EB}}(\hat{A}_{i;\text{ols}})$ is less efficient than $\hat{\theta}_i^{\text{EB}}(\hat{A}_{i;\text{gls}})$.

REMARK 7. We suggest the following second-order unbiased estimator of $\text{MSE}[\hat{\theta}_i^{\text{EB}}(\hat{A}_{i;\text{gls}})]$:

$$\text{mse}_i = g_{1i}(\hat{A}_{i;\text{gls}}) + g_{2i}(\hat{A}_{i;\text{gls}}) + 2g_{3i}(\hat{A}_{i;\text{gls}}) - \big[\hat{B}_i(\hat{A}_{i;\text{gls}})\big]^2\widehat{\text{Bias}}(\hat{A}_{i;\text{gls}}),$$

where $\hat{B}_i(\hat{A}_{i;\text{gls}}) = \frac{D_i}{D_i+\hat{A}_{i;\text{gls}}}$, and $\widehat{\text{Bias}}(\hat{A}_{i;\text{gls}}) = \frac{2}{\operatorname{tr}(V^{-2})}\tilde{l}_{i,\text{ad};\text{gls}}^{(1)}(\hat{A}_{i;\text{gls}})$. We provide expressions for the second-order MSE approximation and the second-order unbiased estimator of $\text{MSE}[\hat{\theta}_i^{\text{EB}}(\hat{A}_{i;\text{gls}})]$ for the benefit of researchers interested in such expressions. However, for the purpose of point estimation and the associated second-order unbiased MSE estimators, we recommend the estimators proposed by Yoshimori and Lahiri (2014). We recommend the use of $\hat{A}_{i;\mathrm{gls}}$ only for the construction of second-order efficient Cox-type EB confidence intervals.

4. A Monte Carlo simulation study. In this section, we design a Monte Carlo simulation study to compare finite sample performances of the following confidence intervals of θ_i for the Fay–Herriot model: direct, Cox-type EB using (i) REML estimator of A (Cox.RE), (ii) estimator of A proposed by Wang and Fuller (2003) (Cox.WF), (iii) estimator of A proposed by Li and Lahiri (2010) (Cox.LL), parametric bootstrap EB confidence interval of Chatterjee, Lahiri and Li (2008) using Li–Lahiri estimator of A (CLL.LL), our proposed Cox-type EB confidence intervals using GLS estimator of β (Cox.YL.GLS) and OLS estimator of β (Cox.YL.OLS). In Section 4.1, we consider a Fay–Herriot model with a common mean as in Datta, Rao and Smith (2005) and Chatterjee, Lahiri and Li (2008). In Section 4.2, we consider a Fay–Herriot model with one auxiliary variable in order to examine the effect of different leverage and sampling variance combinations on the coverage and average length of different confidence intervals of a small area mean.

4.1. *The Fay–Herriot model with a common mean.* Throughout this subsection, we assume a common mean $x_i'\beta = 0$, which is estimated using data as in other papers on small area estimation. Specifically, we generate $R = 10^4$ independent replicates $\{y_i, v_i, i = 1, \dots, m\}$ using the following Fay–Herriot model:

$$y_i = v_i + e_i,$$

where v_i and e_i are mutually independent with $v_i \overset{\text{i.i.d.}}{\sim} N(0, A)$, $e_i \overset{\text{ind}}{\sim} N(0, D_i)$, $i = 1, \dots, m$. We set $A = 1$. For the parametric bootstrap method, we consider $B = 6000$ bootstrap samples.

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 two patterns of the sampling variances: (a) $(0.7, 0.6, 0.5, 0.4, 0.3)$ and (b) $(4.0, 0.6, 0.5, 0.4, 0.1)$. Note that in pattern (a) all areas have sampling variances less than A. In contrast, in pattern (b), sampling variances of all but one area are less than A. The patterns (a) and (b) correspond to the sampling variance patterns (a) and (c) of Datta, Rao and Smith (2005).

The simulation results are displayed in Table 1. First note that while the direct method attains the nominal coverage most of the time it has the highest length compared to the other methods considered. The interval Cox.RE cuts down the length of the direct method considerably at the expense of undercoverage, which is more severe for pattern (b) than pattern (a). This could be due to the presence of three outlying areas (i.e., with respect to the sampling variances) in G_1. The intervals Cox.WF and Cox.LL improve on Cox.RE as both use strictly positive consistent estimators of A. Our new methods—Cox.YL.GLS and Cox.YL.OLS—and CLL.LL perform very well in terms of coverage although CLL.LL is showing a slight undercoverage. The CLL.LL method is slightly better than ours in terms of average length although we notice that in some simulation replications the length of the parametric bootstrap EB confidence interval is larger than that of the direct.

4.2. *Effect of leverage and sampling variance in a Fay–Herriot model with one auxiliary variable.* We generate $R = 10^4$ independent replicates $\{y_i, v_i, i = 1, \dots, m\}$ using the following Fay–Herriot model:

$$y_i = x_i\beta + v_i + e_i,$$

where v_i and e_i are mutually independent with $v_i \overset{\text{i.i.d.}}{\sim} N(0, A)$, $e_i \overset{\text{ind}}{\sim} N(0, D_i)$, $i = 1, \dots, m$. We set $A = 1$. For the parametric bootstrap method, we consider $B = 6000$ bootstrap samples.

In this subsection, we examine the effects of leverage and sampling variance on different confidence intervals for θ_i. We consider six different (leverage, sampling variance) patterns of the first area using leverages $(0.07, 0.22, 0.39)$ and sampling variances $D_1 = (1, 5, 10)$. For the remaining 14 areas, we assume equal small sampling variances $D_j = 0.01$, $j \geq 2$ and same leverage. Since the total leverage for all the areas must be 1, we obtain the common leverage for the other areas from the knowledge of leverage for the first area.

In Table 2, we report the coverages and average lengths for all the competing methods for the first area for all the six patterns. We do not report the results for the remaining 14 areas since they are similar, as expected, due to small sampling variances in those areas. The use of strictly positive consistent estimators of A such as WF and LL help bringing coverage of the Cox-type EB confidence interval closer to the nominal coverage of 95% than the one based on REML. For large sampling variances and leverages, the Cox-type EB confidence intervals based on REML, WF and LL methods have generally shorter length than ours or parametric bootstrap confidence interval but only at the expense of severe undercoverage. Our simulation results show that our proposed Cox.YL.GLS could perform better than Cox.YL.OLS and is very competitive to the more computer intensive CLL.LL method.

5. Concluding remarks. In this paper, we put forward a new simple approach for constructing second-order efficient empirical Bayes confidence interval for a small area mean using a carefully devised adjusted residual maximum likelihood estimator of the model variance in the well-known Cox empirical Bayes confidence interval. Our simulation results show that the proposed method performs much better than the direct or Cox EB confidence intervals with different standard likelihood

TABLE 1

Simulation results for Section 4.1: Simulated coverage and average length (in parenthesis) of different confidence intervals of small area means; nominal coverage is 95%

Pattern	G	Cox.WF		Cox.RE		Cox.LL		CLL.LL		Cox.YL.GLS		Cox.YL.OLS		Direct	
a	1	90.6	(2.4)	90.4	(2.4)	94.2	(2.6)	94.9	(2.7)	95.3	(2.8)	95.3	(2.8)	95.1	(3.3)
	2	91.2	(2.3)	90.8	(2.3)	94.3	(2.5)	94.9	(2.5)	95.3	(2.6)	95.3	(2.6)	94.9	(3.0)
	3	91.5	(2.1)	90.8	(2.1)	94.2	(2.3)	94.9	(2.4)	95.3	(2.4)	95.3	(2.4)	95.1	(2.8)
	4	91.8	(2.0)	91.2	(2.0)	94.3	(2.1)	94.9	(2.2)	95.2	(2.2)	95.3	(2.2)	95.2	(2.5)
	5	92.4	(1.8)	92.1	(1.8)	94.7	(1.9)	95.1	(1.9)	95.5	(2.0)	95.5	(2.0)	95.1	(2.1)
b	1	88.3	(3.3)	88.1	(3.3)	93.7	(3.8)	94.6	(4.0)	95.6	(4.3)	95.9	(4.3)	94.8	(7.8)
	2	90.9	(2.3)	90.0	(2.3)	93.9	(2.5)	94.7	(2.5)	95.2	(2.6)	95.3	(2.6)	94.9	(3.0)
	3	91.2	(2.1)	90.2	(2.1)	93.9	(2.3)	94.7	(2.4)	95.0	(2.5)	95.2	(2.5)	95.1	(2.8)
	4	91.7	(2.0)	90.9	(2.0)	94.4	(2.1)	95.0	(2.2)	95.3	(2.2)	95.4	(2.3)	95.0	(2.5)
	5	93.8	(1.1)	93.1	(1.1)	94.8	(1.2)	94.9	(1.2)	95.0	(1.2)	95.0	(1.2)	94.9	(1.2)

TABLE 2

Simulation results for Section 4.2: Simulated coverage and average length (in parenthesis) of different confidence intervals for the first small area mean for different combinations of leverage and sampling variance of the first area; nominal coverage is 95%

Leverage	D_1	Cox.WF		Cox.RE		Cox.LL		CLL.LL		Cox.YL.gls		Cox.YL.ols		Direct	
0.39	10	78.1	(3.2)	85.3	(3.6)	88.0	(3.9)	94.7	(5.0)	98.0	(6.9)	98.3	(8.1)	95.1	(12.4)
	5	81.5	(3.2)	86.6	(3.5)	89.1	(3.7)	95.0	(4.6)	97.0	(5.8)	97.3	(6.2)	94.9	(8.8)
	1	89.7	(2.7)	90.0	(2.7)	91.3	(2.8)	94.9	(3.2)	95.3	(3.4)	95.4	(3.4)	94.8	(3.9)
0.22	10	84.0	(3.4)	89.7	(3.7)	92.2	(3.9)	95.3	(4.5)	96.7	(5.0)	98.5	(5.7)	94.9	(12.4)
	5	85.8	(3.3)	89.9	(3.5)	91.9	(3.8)	95.0	(4.2)	96.0	(4.6)	97.1	(4.9)	95.0	(8.8)
	1	91.6	(2.7)	91.9	(2.7)	93.0	(2.8)	94.9	(3.0)	95.5	(3.2)	95.5	(3.2)	95.2	(3.9)
0.07	10	87.2	(3.5)	92.2	(3.7)	94.2	(3.9)	95.3	(4.1)	95.7	(4.2)	96.1	(4.3)	95.0	(12.4)
	5	89.2	(3.4)	92.7	(3.5)	94.4	(3.7)	95.5	(3.9)	95.8	(4.0)	95.9	(4.0)	94.8	(8.8)
	1	93.3	(2.7)	93.3	(2.7)	94.4	(2.8)	95.2	(2.9)	95.4	(3.0)	95.4	(3.0)	95.2	(3.9)

based estimators of the model variance. In our simulation, the parametric bootstrap empirical Bayes confidence interval also performs well and it generally produces intervals shorter than direct confidence intervals on the average. However, to the best of our knowledge, there is no analytical result that shows that the parametric bootstrap empirical Bayes confidence interval is always shorter than the direct interval. In fact, in our simulation we found cases where the length of parametric bootstrap empirical Bayes confidence interval is higher than that of the direct. In order to obtain good parametric bootstrap empirical Bayes confidence intervals, choices of the estimator of A and the bootstrap replication B appear to be important. To limit the computing time, we have considered a simple simulation setting with $m = 15$. During the course of our investigation, we feel the need for developing an efficient computer program that allows evaluation of parametric bootstrap empirical Bayes confidence intervals in a large scale simulation environment. Until the issues raised for the parametric bootstrap empirical Bayes confidence interval method are resolved, our proposed simple second-order efficient empirical Bayes confidence interval could serve as a promising method. The results presented in this paper is for the well-known Fay–Herriot model. It is not clear at this time how the results will extend to a general class of small area models—this will be a challenging topic for future research.

APPENDIX A

In this appendix, we provide an outline of proofs of different results presented in the paper. In order to facilitate the review, we supply a detailed proof of Theorem 4 in the supplementary material [Yoshimori and Lahiri (2014)].

PROOF OF THEOREM 1. For notational simplicity, we set $\hat{A}_{h_i} \equiv \hat{A}$ throughout the Appendix. Define

$$G_i(z, y) = z[\hat{\sigma}_i/\sigma_i - 1] + \{(B_i - \hat{B}_i)(y_i - x_i'\beta) + \hat{B}_i[x_i'(\hat{\beta} - \beta)]\}/\sigma_i.$$

Using calculations similar to the ones Chatterjee, Lahiri and Li (2008), we have

$$\begin{aligned} &P[\theta_i \le \hat{\theta}_i^{\mathrm{EB}}(\hat{A}) + z\hat{\sigma}_i] \\ &\quad = \Phi(z) + \phi(z)E\left[G_i(z, y) - \frac{z}{2}G_i^2(z, y)\right] \\ &\qquad + \frac{1}{2}E\left[\int_z^{z+G_i(z,y)} (z + G_i(z, y) - t)^2 (t^2 - 1)\phi(t)\, dt\right]. \end{aligned} \tag{A.1}$$

We shall first show that the third term of the right-hand side of (A.1) is of order $O(m^{-3/2})$. To this end, using

$$0 \le |z + G_i(z, y) - t| \le |G_i(z, y)| \quad \text{and} \quad (t^2 - 1)\phi(t) \le 2\phi(\sqrt{3}),$$

in $t \in (z, z + G_i(z, y))$, we have

$$\text{third term of (A.1)} \le \frac{1}{2} E\left[\int_z^{z+G_i(z,y)} \big(z + G_i(z, y) - t\big)^2 \big|(t^2 - 1)\phi(t)\big|\, dt\right]$$
$$\le C\phi(\sqrt{3}) E\big[G_i^3(z, y)\big].$$

Setting $\sigma_i^2 = S_i$ and using the Taylor series expansion, we have

$$\hat{\sigma}_i(\hat{S}_i) - \sigma_i(S_i) = \tfrac{1}{2} S_i^{-1/2}(\hat{S}_i - S_i) - \tfrac{1}{8} S_i^{-3/2}(\hat{S}_i - S_i)^2 + O_p\big(|\hat{S}_i - S_i|^3\big),$$

so that

$$\frac{\hat{\sigma}_i(\hat{S}_i)}{\sigma_i(S_i)} - 1 = \frac{1}{2S_i}(\hat{S}_i - S_i) - \frac{1}{8S_i^2}(\hat{S}_i - S_i)^2 + R_{A1}.$$

Using

$$\hat{B}_i - B_i = -(\hat{A} - A)\frac{D_i}{(A + D_i)^2} + (\hat{A} - A)^2 \frac{D_i}{(A + D_i)^3} + R_{A2},$$

$$\hat{\sigma}_i^2 - \sigma_i^2 = (\hat{A} - A)\frac{D_i^2}{(A + D_i)^2} - (\hat{A} - A)^2 \frac{D_i^2}{(A + D_i)^3} + R_{A3},$$

we can write $G_i(z, y) = G_{1i}(y) + G_{2i}(z, y)$, where

$$G_{1i}(y) = \frac{1}{\sqrt{m}} \hat{u}_{1i} + \frac{1}{m} \hat{u}_{2i} + R_{A4},$$

$$G_{2i}(z, y) = z\left[\frac{1}{\sqrt{m}} \hat{v}_{1i} + \frac{1}{m} \hat{v}_{2i}\right] + R_{A5},$$

with

$$\hat{u}_{1i} = \sqrt{m}\sigma_i^{-1}\left[B_i x_i'(\hat{\beta} - \beta) + (\hat{A} - A)\frac{D_i}{(A + D_i)^2}\big(y_i - x_i'\beta\big)\right],$$

$$\hat{u}_{2i} = m\sigma_i^{-1}\left[-(\hat{A} - A)^2 \frac{D_i}{(A + D_i)^3}\big(y_i - x_i'\beta\big)\right.$$
$$\left. + (\hat{A} - A)\frac{D_i}{(A + D_i)^2} B_i x_i'(\hat{\beta} - \beta)\right],$$

$$\hat{v}_{1i} = \sqrt{m}\frac{B_i^2}{2\sigma_i^2}(\hat{A} - A),$$

$$\hat{v}_{2i} = m\left[-\frac{1}{2\sigma_i^2}\frac{B_i^2}{A + D_i}(\hat{A} - A)^2 - \frac{1}{8\sigma_i^4}(\hat{A} - A)^2 B_i^4\right].$$

Using the fact that $E[|\hat{A} - A|^k] = O(m^{-3/2})$ for $k \ge 3$ [this can be proved using the mathematical tools used in Li and Lahiri (2010) and Das, Jiang and Rao

(2004)], we have, for $k = 1, 2, 3, 4, 5$ and large m,

$$E\big[|R_{Ak}|\big] \le C E\big[|\hat{A} - A|^3\big] = O\big(m^{-3/2}\big),$$

$$\big|\text{third term of (A.1)}\big| \le C\phi(\sqrt{3}) E\big[\big|G_i^3(z, y)\big|\big] \le C E\big[|\hat{A} - A|^3\big] = O\big(m^{-3/2}\big),$$

where C is a generic constant.

We also note that

$$E\big[G_i(z, y)\big] = m^{-1/2} E[\hat{u}_{1i} + z\hat{v}_{1i}] + m^{-1} E[\hat{u}_{2i} + z\hat{v}_{2i}] + O\big(m^{-3/2}\big),$$

so that

$$\text{the right-hand side of (A.1)} = \Phi(z) + \phi(z) E\bigg[G_i(z, y) - \frac{z}{2} G_i^2(z, y)\bigg] + O\big(m^{-3/2}\big).$$

Similarly,

$$\begin{aligned} P\big(\hat{\theta}_i^{\mathrm{EB}} - z\hat{\sigma}_i \le \theta_i\big) &= \Phi(-z) + \phi(-z) E\bigg[G_i(-z, y) + \frac{z}{2} G_i^2(-z, y)\bigg] \\ &\quad + O\big(m^{-3/2}\big), \end{aligned} \tag{A.2}$$

so that using

$$\begin{aligned} &G_i(z, y) - G_i(-z, y) - \frac{z}{2}\big[G_i^2(z, y) + G_i^2(-z, y)\big] \\ &\quad = 2G_{2i}(z, y) - \frac{z}{2}\big[G_{1i}^2(y) + G_{2i}^2(z, y)\big] \\ &\quad = \frac{2z}{\sqrt{m}} \hat{v}_{1i} + \frac{z}{m}\big\{2\hat{v}_{2i} - \hat{u}_{1i}^2 - z^2 \hat{v}_{1i}^2\big\} + R_{A6}, \end{aligned}$$

where $E[|R_{A6}|] = O(m^{-3/2})$ since $E[|\hat{A} - A|^k] = O(m^{-3/2})$ for $k \ge 3$.

We have

$$\begin{aligned} &P\big\{\theta_i \in I_{cox}(\hat{A})\big\} \\ &\quad = \Phi(z) - \Phi(-z) + \phi(z) E\big[G_i(z, y) - G_i(-z, y)\big] \\ &\qquad - \frac{z}{2}\phi(z) E\big[G_i^2(z, y) + G_i^2(-z, y)\big] + O\big(m^{-3/2}\big) \\ &\quad = 1 - \alpha + z\phi(z)\big\{m^{-1/2} E[2\hat{v}_{1i}] + m^{-1} E\big[2\hat{v}_{2i} - \hat{u}_{1i}^2 - z^2 \hat{v}_{1i}^2\big]\big\} \\ &\qquad + O\big(m^{-3/2}\big). \end{aligned}$$

Using Lemma 1, given below, and considerable algebra, we show that

$$a_i = E\big[2\hat{v}_{2i} - \hat{u}_{1i}^2 - z^2 \hat{v}_{1i}^2\big] \quad \text{and} \quad b_i = 2\sqrt{m} E[\hat{v}_{1i}].$$

This completes the proof of equation (3.1). $\square$

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

$$E\big[\hat{v}_{1i}^2(\hat{A})\big] = \frac{m}{\operatorname{tr}(V^{-2})}\frac{D_i^2}{2A^2(A+D_i)^2} + O\big(m^{-1/2}\big), \tag{A.3}$$

$$\begin{aligned} E\big[\hat{v}_{2i}(\hat{A})\big] &= -\frac{m}{\operatorname{tr}(V^{-2})}\left[\frac{D_i}{A(A+D_i)^2} + \frac{D_i^2}{4A^2(A+D_i)^2}\right] \\ &\quad + O\big(m^{-1/2}\big), \end{aligned} \tag{A.4}$$

$$\begin{aligned} E\big[\hat{u}_{1i}^2(\hat{A})\big] &= m\frac{D_i}{A(A+D_i)}\left[E\big[\{x_i'\tilde{\beta} - \beta)\}^2\big] + \frac{D_i}{A(A+D_i)^2}\frac{2}{\operatorname{tr}(V^{-2})}\right] \\ &\quad + O\big(m^{-1/2}\big), \end{aligned} \tag{A.5}$$

$$E\big[\hat{v}_{1i}(\hat{A})\big] = \frac{\sqrt{m}}{\operatorname{tr}(V^{-2})}\frac{D_i}{A(A+D_i)}\tilde{l}_{i;\mathrm{ad}}^{(1)} + O\big(m^{-1}\big). \tag{A.6}$$

PROOF OF THEOREM 2. First note that solution of $h_i(A)$ to the differential equation (3.4) depends on whether the OLS or GLS method is used to estimate β. Also note that the solution of $h_i(A)$ for the OLS case does not follow as a special case of GLS. Thus, we treat these two cases separately. The balanced case, that is, equation (3.7) follows from (3.5) or (3.6).

Case 1: *Unbalanced case: OLS* [*proof of equation* (3.6)]

From equation (3.4), we have

$$\tilde{l}_{i;\mathrm{ad}}^{(1)}(A) = \frac{2}{A+D_i} + \frac{(1+z^2)D_i}{4A(A+D_i)} + \frac{1}{2}x_i'(X'X)^{-1}X'VX(X'X)^{-1}x_i\operatorname{tr}(V^{-2}).$$

Therefore,

$$\begin{aligned} \tilde{l}_{i;\mathrm{ad}}(A) &= \int \tilde{l}_{i;\mathrm{ad}}^{(1)}\,dA \\ &= 2\log(A+D_i) + \frac{(1+z^2)}{4}\log\left(\frac{A}{A+D_i}\right) \\ &\quad + \frac{1}{2}x_i'(X'X)^{-1}X'JX(X'X)^{-1}x_i + C \\ &= 2\log(A+D_i) + \frac{(1+z^2)}{4}\log\left(\frac{A}{A+D_i}\right) \\ &\quad + \frac{1}{2}x_i'(X'X)^{-1}X'[-V\operatorname{tr}(V^{-1}) + \operatorname{tr}(V^{-1}) + C]X(X'X)^{-1}x_i + C \\ &= 2\log(A+D_i) + \frac{1+z^2}{4}\log\left(\frac{A}{A+D_i}\right) \end{aligned}$$

$$
\begin{aligned}
&- \frac{1}{2} x_i'(X'X)^{-1} X'VX(X'X)^{-1} x_i \operatorname{tr}(V^{-1}) \\
&+ \frac{1}{2} q_i \left[\sum_{i=1}^{m} \log(A + D_i) \right] + C.
\end{aligned}
$$

In addition,

$$
J = \operatorname{diag}\biggl(\int (A + D_1) \operatorname{tr}(V^{-2})\, dA, \ldots, \int (A + D_m) \operatorname{tr}(V^{-2})\, dA \biggr).
$$

Equation (3.6) follows noting that $h_i(A) = \exp[\tilde{l}_{i;\mathrm{ad}}(A)]$.

Case 2: *Unbalanced case: GLS* [*proof of* (3.5)]

Solving equation (3.4) for $\tilde{l}^{(1)}_{i;\mathrm{ad}}(A)$, we get

$$
\tilde{l}^{(1)}_{i;\mathrm{ad}}(A) = \frac{2}{A + D_i} + \frac{(1 + z^2) D_i}{4A(A + D_i)} + \frac{1}{2} x_i'(X'V^{-1}X)^{-1} x_i \operatorname{tr}(V^{-2}).
$$

Thus,

$$
\begin{aligned}
\tilde{l}_{i;\mathrm{ad}}(A) &= \int \tilde{l}^{(1)}_{i;\mathrm{ad}}\, dA \\
&= \int \frac{2}{A + D_i}\, dA + \int \frac{(1 + z^2) D_i}{4A(A + D_i)}\, dA \\
&\quad + \frac{1}{2} \int x_i'(X'V^{-1}X)^{-1} x_i \operatorname{tr}(V^{-2})\, dA \\
&= 2 \log(A + D_i) + \frac{1}{4}(1 + z^2) D_i \log \frac{A}{A + D_i} + \frac{1}{2} K + C, \qquad \text{say,}
\end{aligned}
$$

where $K = \int x_i'(X'V^{-1}X)^{-1} x_i \operatorname{tr}(V^{-2})\, dA$.

We now prove part (ii) of the theorem. To this end, note that the adjusted maximum residual likelihood estimator of A with the adjustment factor (3.7) is obtained as a solution of

$$
\begin{aligned}
&l^{(1)}_{\mathrm{RE}} + \tilde{l}^{(1)}_{i,\mathrm{ad}} = 0 \\
&\quad \Longleftrightarrow f(A) \equiv \{-2(m - p) + 8 + 2mq_i\} A^2 \\
&\qquad\qquad + \{2y'(I_m - X(X'X)^{-1}X')y - 2(m - p)D + 8D \\
&\qquad\qquad\qquad + (1 + z^2)D + 2mDq_i\} A \\
&\qquad\qquad + (1 + z^2) D^2 = 0.
\end{aligned}
$$

Therefore, under strict positiveness of the solution and $m > \frac{4+p}{1-q_i}$, $f(A)$ is a quadratic and concave function of A. Thus, due to $f(0) > 0$, there is a unique and strictly positive adjusted residual maximum likelihood estimator of A in the balanced case. $\square$

PROOF OF THEOREM 3. Note that the length of the Cox-type EB confidence interval of θ_i is given by $2\sigma(\hat{A}_i)$, where $\sigma(\hat{A}_i) = \sqrt{\frac{\hat{A}_i D_i}{\hat{A}_i + D_i}}$ and $\hat{A}_i$ is an estimator of A used to construct an empirical Bayes confidence interval for θ_i. We show that among the three intervals considered the length of the Cox EB confidence interval is the shortest when $\hat{A}_{\text{RE}}$ is used to estimate A, followed by $\hat{A}_{i,\text{gls}}$, and $\hat{A}_{i,\text{ols}}$. Since $\sigma(\hat{A}_i)$ is a monotonically increasing function of $\hat{A}_i$, it suffices to show that

$$\hat{A}_{\text{RE}} \le \hat{A}_{i,\text{gls}} \le \hat{A}_{i,\text{ols}}.$$

Note that

$$\begin{aligned} l_{\text{RE}}^{(1)}(\tilde{A}_{\text{RE}}) &= 0, \\ l_{\text{RE}}^{(1)}(\hat{A}_{i,\text{gls}}) + \tilde{l}_{i;\text{ad},\text{gls}}^{(1)}(\hat{A}_{i,\text{gls}}) &= 0, \\ l_{\text{RE}}^{(1)}(\hat{A}_{i,\text{ols}}) + \tilde{l}_{i;\text{ad},\text{ols}}^{(1)}(\hat{A}_{i,\text{ols}}) &= 0, \\ l_{\text{RE}}^{(2)}(\hat{A}) + \tilde{l}_{i;\text{ad}}^{(2)}(\hat{A}) &< 0, \end{aligned}$$

where $\hat{A} \in \{\tilde{A}_{\text{RE}}, \hat{A}_{i,\text{gls}}, \hat{A}_{i,\text{ols}}\}$ and $\tilde{A}_{\text{RE}}$ is a solution to the REML estimation equation. Hence, $\hat{A}_{\text{RE}}$ is always larger than $\hat{A}_{i,\text{gls}}$ or $\hat{A}_{i,\text{gls}}$ using the facts that $\hat{A}_{\text{RE}} = \max\{0, \tilde{A}_{\text{RE}}\}$ and $\hat{A}_{i,\text{gls}}$ or $\hat{A}_{i,\text{gls}}$ are strictly positive if $m > (4 + p)/(1 - q_i)$.

Finally, using that $0 < \tilde{l}_{i;\text{ad},\text{gls}}^{(1)} \le \tilde{l}_{i;\text{ad},\text{ols}}^{(1)}$ for $A \ge 0$, we have the result. $\square$

PROOF OF COROLLARY TO THEOREM 4. (i) Since for these two adjustment terms, $h_i(A)L_{\text{RE}}(A)|_{A=0} = 0$ and $h_i(A)L_{\text{RE}}(A) \ge 0$ for $A > 0$, it suffices to show that $\lim_{A\to\infty} h_i(A)L_{\text{RE}}(A) = 0$. For $h_i(A)$ given by (3.6),

$$(3.6) \le (A + D_i)^2 \Big(A + \sup_{i\ge1} D_i\Big)^{(1/2)mq_i} \le \Big(A + \sup_{i\ge1} D_i\Big)^{2+(1/2)mq_i}.$$

For (3.5), we have

$$\begin{aligned} (3.5) &\le (A + D_i)^2 \exp\left\{\frac{1}{2}\int \Big(A + \inf_{i\ge1} D_i\Big) q_i \operatorname{tr}\big(V^{-2}\big)\, dA\right\} \\ &\le (A + D_i)^2 \Big(A + \sup_{i\ge1} D_i\Big)^{(1/2)mq_i} \exp\left[-\frac{m}{2} q_i\right] \\ &\quad \times \exp\left[-\frac{1}{2}\inf_{i\ge1} D_i q_i \operatorname{tr}\big(V^{-1}\big)\right] \\ &\le \Big(A + \sup_{i\ge1} D_i\Big)^{2+(1/2)mq_i}. \end{aligned}$$

Using the fact $L_{\mathrm{RE}}(A) < C(A+\sup_{i\ge 1} D_i)^{p/2}|X'X|^{-1/2}(A+\inf_{i\ge 1} D_i)^{-m/2}$, we have

$$0 \le h_i(A)L_{\mathrm{RE}}(A) \le \Bigl(A+\sup_{i\ge 1} D_i\Bigr)^{2+(1/2)[mq_i+p]}\Bigl(A+\inf_{i\ge 1} D_i\Bigr)^{-m/2}|X'X|^{-1/2},$$

so that, under mild regularity conditions,

$$0 \le \lim_{A\to\infty} h_i(A)L_{\mathrm{RE}}(A) = \lim_{A\to\infty} A^{2+(1/2)[mq_i+p-m]}.$$

Thus, if $2+\frac{1}{2}[mq_i+p-m] < 0$, we have

$$\lim_{A\to\infty} h_i(A)L_{\mathrm{RE}}(A) = 0.$$

We first show that $\hat{A}_{i;\mathrm{gls}}$ and $\hat{A}_{i;\mathrm{gls}}$ satisfy the regularity conditions of Theorem 3. Since $0 < A < \infty$, we claim that $\tilde{l}^{k}_{i,\mathrm{ad}}(A) = O(1)$ $(k=1,2,3)$, for large m, for both the GLS and OLS estimators of β using the following facts.

For the GLS estimator,

$$\begin{aligned}
\tilde{l}^{(1)}_{i,\mathrm{ad}}(A) &= \biggl(2-\frac{(1+z^2)}{4}\biggr)\frac{1}{A+D_i}+\frac{(1+z^2)}{4A}+\frac{1}{2}\operatorname{tr}\bigl[V^{-2}\bigr]x_i'\bigl(X'V^{-1}X\bigr)^{-1}x_i,\\
\tilde{l}^{(2)}_{i,\mathrm{ad}}(A) &= -\biggl(2-\frac{(1+z^2)}{4}\biggr)\frac{1}{(A+D_i)^2}-\frac{(1+z^2)}{4A^2}\\
&\quad -\operatorname{tr}\bigl[V^{-3}\bigr]x_i'\bigl(X'V^{-1}X\bigr)^{-1}x_i\\
&\quad +\frac{1}{2}\operatorname{tr}\bigl[V^{-2}\bigr]x_i'\bigl(X'V^{-1}X\bigr)^{-1}X'V^{-2}X\bigl(X'V^{-1}X\bigr)^{-1}x_i,\\
\tilde{l}^{(3)}_{i,\mathrm{ad}}(A) &= \biggl(2-\frac{(1+z^2)}{4}\biggr)\frac{2}{(A+D_i)^3}+\frac{(1+z^2)}{2A^3}\\
&\quad +3\operatorname{tr}\bigl[V^{-4}\bigr]x_i'\bigl(X'V^{-1}X\bigr)^{-1}x_i\\
&\quad -2\operatorname{tr}\bigl[V^{-3}\bigr]x_i'\bigl(X'V^{-1}X\bigr)^{-1}X'V^{-2}X\bigl(X'V^{-1}X\bigr)^{-1}x_i\\
&\quad \times\operatorname{tr}\bigl[V^{-2}\bigr]\bigl[x_i'\bigl(X'V^{-1}X\bigr)^{-1}X'V^{-2}X\bigl(X'V^{-1}X\bigr)^{-1}\\
&\qquad\qquad \times X'V^{-2}X\bigl(X'V^{-1}X\bigr)^{-1}x_i\\
&\qquad\qquad -x_i'\bigl(X'V^{-1}X\bigr)^{-1}X'V^{-3}X\bigl(X'V^{-1}X\bigr)^{-1}x_i\bigr].
\end{aligned}$$

For the OLS estimator,

$$\begin{aligned}
\tilde{l}^{(1)}_{i,\mathrm{ad}}(A) &= \biggl(2-\frac{(1+z^2)}{4}\biggr)\frac{1}{A+D_i}+\frac{(1+z^2)}{4A}\\
&\quad +\frac{1}{2}\operatorname{tr}\bigl[V^{-2}\bigr]x_i'\bigl(X'X\bigr)^{-1}X'VX\bigl(X'X\bigr)^{-1}x_i,
\end{aligned}$$

$$\tilde{l}^{(2)}_{i,\mathrm{ad}}(A) = -\left(2 - \frac{(1+z^2)}{4}\right)\frac{1}{(A+D_i)^2} - \frac{(1+z^2)}{4A^2}$$
$$- \operatorname{tr}[V^{-3}]x_i'(X'X)^{-1}X'VX(X'X)^{-1}x_i + \frac{1}{2}\operatorname{tr}[V^{-2}]q_i,$$

$$\tilde{l}^{(3)}_{i,\mathrm{ad}}(A) = \left(2 - \frac{(1+z^2)}{4}\right)\frac{2}{(A+D_i)^3} + \frac{(1+z^2)}{2A^3}$$
$$+ 3\operatorname{tr}[V^{-4}]x_i'(X'X)^{-1}X'VX(X'X)^{-1}x_i - 2\operatorname{tr}[V^{-3}]q_i.$$

In addition, For GLS,

$$\tilde{l}^{(4)}_{i,\mathrm{ad}}(A) = -\left(12 - \frac{3(1+z^2)}{2}\right)\frac{1}{(A+D_i)^4} - \frac{3(1+z^2)}{2A^4} + \tilde{l}^{(4)}_{3,i,\mathrm{ad,gls}}(A).$$

For OLS,

$$\tilde{l}^{(4)}_{i,\mathrm{ad}}(A) = -\left(12 - \frac{3(1+z^2)}{2}\right)\frac{1}{(A+D_i)^4} - \frac{3(1+z^2)}{2A^4} + \tilde{l}^{(4)}_{3,i,\mathrm{ad,ols}}(A),$$

where

$$\begin{aligned}\tilde{l}^{(4)}_{3,i,\mathrm{ad,gls}}(A) = {} & -12\operatorname{tr}[V^{-5}]x_i'(X'V^{-1}X)^{-1}x_i \\
& + 6\operatorname{tr}[V^{-3}][x_i'(X'V^{-1}X)^{-1}X'V^{-3}X(X'V^{-1}X)^{-1}x_i \\
& \qquad - x_i'(X'V^{-1}X)^{-1}X'V^{-2}X(X'V^{-1}X)^{-1}X'V^{-2} \\
& \qquad \times X(X'V^{-1}X)^{-1}x_i] \\
& + 9\operatorname{tr}[V^{-4}]x_i'(X'V^{-1}X)^{-1}X'V^{-2}X(X'V^{-1}X)^{-1}x_i \\
& + \operatorname{tr}[V^{-2}][3x_i'(X'V^{-1}X)^{-1}X'V^{-4}X(X'V^{-1}X)^{-1}x_i \\
& \qquad - 4x_i'(X'V^{-1}X)^{-1}X'V^{-2}X(X'V^{-1}X)^{-1}X'V^{-3} \\
& \qquad \times X(X'V^{-1}X)^{-1}x_i \\
& \qquad - 4x_i'(X'V^{-1}X)^{-1}X'V^{-3}X(X'V^{-1}X)^{-1}X'V^{-2} \\
& \qquad \times X(X'V^{-1}X)^{-1}x_i \\
& \qquad + 3x_i'(X'V^{-1}X)^{-1}X'V^{-2}X(X'V^{-1}X)^{-1}X'V^{-2} \\
& \qquad \times X(X'V^{-1}X)^{-1} \\
& \qquad \times X'V^{-2}X(X'V^{-1}X)^{-1}x_i],\end{aligned}$$

$$\begin{aligned}\tilde{l}^{(4)}_{3,i,\mathrm{ad,ols}}(A) = {} & -12\operatorname{tr}[V^{-5}]x_i'(X'X)^{-1}X'VX(X'X)^{-1}x_i \\
& + 9\operatorname{tr}[V^{-4}]q_i.\end{aligned}$$

Using the above facts, we can prove that $|\tilde{l}^{(j)}_{i,\mathrm{ad,gls}}|$ and $|\tilde{l}^{(j)}_{i,\mathrm{ad,ols}}|$ are bounded for $j = 1, 2, 3, 4$ under the regularity conditions R2–R4. Similarly, we can show that the gth powers of $\sup_{A/2<A^*<2A} \frac{1}{m}|\tilde{l}^{(5)}_{i,\mathrm{ad};h}(A^*)|$ with $h = \mathrm{gls}, \mathrm{ols}$ are bounded for any fixed $g > 0$. Thus, the new area specific adjustment terms satisfy the regularity condition R1. Thus, an application of Theorem 4 leads to (ii)–(iv) of the Corollary to Theorem 4. $\square$

APPENDIX B: PROOF OF LEMMA 1

The proof of (A.5) is much more complex due to the dependence of $\hat{A}$ and y_i. We use the following lemma repeatedly for proving (A.5). For a proof of Lemma 2, see Srivastava and Tiwari (1976).

LEMMA 2. *Let $Z \sim N(0, \Sigma)$. Then for symmetric matrices Q, U and W,*

$$E[(Z'QZ)(Z'UZ)] = 2\,\mathrm{tr}(Q\Sigma U\Sigma) + \mathrm{tr}(Q\Sigma)\,\mathrm{tr}(U\Sigma),$$

$$\begin{aligned} E[(Z'QZ)(Z'UZ)(Z'WZ)] = {} & 8\,\mathrm{tr}(Q\Sigma U\Sigma W\Sigma) \\ & + 2\{\mathrm{tr}(Q\Sigma U\Sigma)\,\mathrm{tr}(W\Sigma) + \mathrm{tr}(Q\Sigma W\Sigma)\,\mathrm{tr}(U\Sigma) \\ & + \mathrm{tr}(U\Sigma W\Sigma)\,\mathrm{tr}(Q\Sigma)\} \\ & + \mathrm{tr}(Q\Sigma)\,\mathrm{tr}(U\Sigma)\,\mathrm{tr}(W\Sigma). \end{aligned}$$

The proof also needs the following lemma, which is immediate from Theorem 2.1 of Das, Jiang and Rao (2004).

LEMMA 3. *Assume the following regularity conditions:*

1. *$\tilde{l}_{i,\mathrm{ad}}(A)$, which is free of y, is four times continuously differentiable with respect to A,*
2. *the gth power of the following are bounded: $\frac{1}{\sqrt{m}}|\tilde{l}^{(1)}_{i,\mathrm{ad}}(A)|$, $\frac{1}{m}|\tilde{l}^{(2)}_{i,\mathrm{ad}}(A)|$, $\frac{1}{m}|\tilde{l}^{(3)}_{i,\mathrm{ad}}(A)|$, and $\frac{1}{m}\sup_{A/2<\tilde{A}<2A}|\tilde{l}^{(4)}_{i,\mathrm{ad}}(A)|_{A=\tilde{A}}|$ (fixed $g > 0$),*
3. *$A \in \Theta_0$, the interior of Θ, that is, $0 < A < \infty$.*

Then:

(i) *there is $\hat{A}_i$ such that for any $0 < \rho < 1$, there is a set Λ satisfying for large m and on Λ, $\hat{A} \in \Theta$, $l^{(1)}(A)|_{\hat{A}} = 0$, $\sqrt{m}|\hat{A}_i - A| < m^{(1-\rho)/2}$, and*

$$\hat{A}_i - A = I + II + III + r,$$

where $I = -E[l^{(2)}]^{-1}l^{(1)}$, $II = E[l^{(2)}]^{-2}l^{(2)}l^{(1)} - E[l^{(2)}]^{-1}l^{(1)}$, $III = -\frac{1}{2}E[l^{(2)}]^{-3}\{l^{(1)}\}^2 l^{(3)}$, and $r \le m^{-3\rho/2}u$ with $E[|u|^g]$ bounded;

(ii) *$P(\Lambda^c) \le m^{-\tau/2g}C$, where $\tau = 1/4 \wedge (1-\rho)$.*

First note that

$$E[\hat{u}_{1i}^2] = m\sigma_i^{-2}\left\{B_i^2 T_1 + 2B_i \frac{D_i}{(A+D_i)^2} T_2 + \frac{D_i^2}{(A+D_i)^4} T_3\right\},$$

where $T_1 = E[x_i'(\hat{\beta}-\beta)^2]$, $T_2 = E[(\hat{A}-A)x_i'(\hat{\beta}-\beta)(y_i - x_i'\beta)]$ and $T_3 = E[(\hat{A}-A)^2(y_i - x_i'\beta)^2]$. We now simplify these three terms.

We first prove that

$$E[T_1] = x_i' \operatorname{Var}(\tilde{\beta}) x_i + O(m^{-2}), \tag{B.1}$$

where $\operatorname{Var}(\tilde{\beta}) = (X'X)^{-1}X'VX(X'X)^{-1}$ if $\tilde{\beta}$ is the OLS estimator of β and $(X'V^{-1}X)^{-1}$ if $\tilde{\beta}$ is the GLS estimator of β.

Note that

$$\begin{aligned} E\big[\{x_i'(\hat{\beta}-\beta)\}^2\big] &= E\big[\{x_i'(\tilde{\beta}-\beta)\}^2\big] + E\big[\{x_i'(\hat{\beta}-\tilde{\beta})\}^2\big] \\ &= x_i' \operatorname{Var}(\tilde{\beta}) x_i + E\big[\{x_i'(\hat{\beta}-\tilde{\beta})\}^2\big], \end{aligned}$$

and we have the following facts:

$$E\big[\{x_i'(\hat{\beta}(\hat{A}_1,\ldots,\hat{A}_m) - \tilde{\beta}(A))\}^2\big] \le E\big[\{x_i'(\hat{\beta}(\hat{A}_U) - \tilde{\beta})\}^2\big], \tag{B.2}$$

where $\hat{A}_U = \arg\max_{\hat{A}_i} |x_i'(\hat{\beta}(\hat{A}_1,\ldots,\hat{A}_m) - \tilde{\beta}(A))|$.

We have $\frac{\partial \tilde{\beta}}{\partial A} = H(y - X\beta)$, where $H = 0$ for the OLS estimator of β and

$$H = (X'V^{-1}X)^{-1}X'V^{-2}X(X'V^{-1}X)^{-1}X'V^{-1} - (X'V^{-1}X)^{-1}X'V^{-2},$$

the GLS estimators of β.

Using the Taylor series expansion, we have

$$x_i'(\hat{\beta}(\hat{A}_U) - \tilde{\beta}) = (\hat{A}_U - A)x_i'Hy + r_1, \tag{B.3}$$

where $|r_1| = \frac{1}{2}(\hat{A}_U - A)^2 x_i' \frac{\partial H}{\partial A}|_{A=A^*} y$ with $A^* \in (A, \hat{A}_U)$ and

$$\begin{aligned} \frac{\partial H}{\partial A} &= 2(X'V^{-1}X)^{-1}X'V^{-2}(X(X'V^{-1}X)^{-1}X'V^{-1} - I) \\ &\quad \times V^{-1}(X(X'V^{-1}X)^{-1}X'V^{-1} - I). \end{aligned}$$

Let $H_s^{(1)}$ be the matrix with (i, j) components given by

$$\sup_{A/2<A^*<2A}\left\{\frac{\partial H}{\partial A}\bigg|_{A=A^*}\right\}_{(i,j)},$$

where $Q_{(i,j)}$ is (i, j) component of a matrix Q. Under the regularity conditions R3–R4, we can show that the components of $H_s^{(1)}$ are bounded and of order

$O(m^{-1})$ using an argument similar to that given in Proposition 3.2 of Das, Jiang and Rao (2004). Using the facts that $HX = 0$, $x_i' HVH' x_i = O(m^{-1})$, we have

$$\begin{aligned}
E\big[\{x_i'(\hat\beta - \tilde\beta)\}^2\big] &\le E\big[(\hat A_U - A)^2 (x_i' H y)(y' H' x_i)\big] \\
&\quad + 2E\big[(\hat A_U - A)^3 (x_i' H y)(y'[H_s^{(1)}]' x_i)\big] \\
&\quad + E\bigg[(\hat A_U - A)^4 \bigg(x_i' \frac{\partial H}{\partial A} y\bigg)(y'[H_s^{(1)}]' x_i)\bigg] \\
&\le E\big[(\hat A_U - A)^2\big] x_i' H V H' x_i + E\big[|\hat A_U - A|^3\big] x_i' H V [H_s^{(1)}]' x_i \\
&\quad + E\big[(\hat A_U - A)^4\big] x_i' H_s^{(1)} V [H_s^{(1)}]' x_i \\
&= O(m^{-2}).
\end{aligned}$$

Thus, this completes the proof of (B.1).

Next, we simplify $E[T_2]$. Let $l_{i;\mathrm{ad}}$ denote the adjusted residual log-likelihood function. Then $l_{i;\mathrm{ad}} = l_{\mathrm{RE}} + \tilde l_{i;\mathrm{ad}}$, where l_{RE} is the residual log-likelihood function and $\tilde l_{i;\mathrm{ad}} = \log h_i(A)$. Define $I_F = -1/E[\frac{\partial^2 l}{\partial A^2}]$. For notational simplicity, we set $l_{i;\mathrm{ad}} \equiv l_{\mathrm{ad}}$ and $\tilde l_{i;\mathrm{ad}} \equiv \tilde l_{\mathrm{ad}}$. Since $\tilde l_{\mathrm{ad}}$ is bounded and free from y, we obtain the following using Lemma 3,

$$\hat A - A = \frac{\partial l_{\mathrm{ad}}}{\partial A} I_F + r_{2.1} = l_{\mathrm{RE}}^{(1)} I_F + r_{2.2},$$

where $l_{\mathrm{RE}}^{(1)} = \frac{\partial l_{\mathrm{RE}}}{\partial A} = \frac{1}{2}[y' P^2 y - \mathrm{tr}(P)]$ and $E[|r_{2.2}|] = O(m^{-1})$ when ρ is taken as 3/4 in Lemma 3.

Since $\hat A$ is translation invariant and even function, we can substitute $\hat A(Z) - A$ for $\hat A(y) - A$, where $Z = y - X\beta \sim N(0, V)$. Thus,

$$\begin{aligned}
x_i'(\hat\beta - \beta) &= x_i' (X' \hat V^{-1} X)^{-1} X' \hat V^{-1} Z \\
&= \lambda_i' X (X' \hat V^{-1} X)^{-1} X' \hat V^{-1} Z \\
&= \lambda_i' X (X' V^{-1} X)^{-1} X' V^{-1} Z + r_{1.2} Z,
\end{aligned}$$

where λ_i denotes a $m \times 1$ vector with i component 1 and the rest 0 and $r_{1.2} Z \le (\hat A_U - A) x_i' H Z + r_1$.

Hence,

$$\begin{aligned}
E[T_2] &\le E\big[(l_{\mathrm{RE}}^{(1)} I_F + r_{2.2})\{\lambda_i' X (X' V^{-1} X)^{-1} X' V^{-1} Z + r_{1.2} Z\}(\lambda_i' Z)\big] \\
&= I_F \{E[l_{\mathrm{RE}}^{(1)} Z' E_i X (X' V^{-1} X)^{-1} X' V^{-1} Z]\} + E\big[(\hat A_U - A) r_{1.2} Z (\lambda_i' Z)\big] \\
&\quad + E\big[r_{2.2} Z' E_i X (X' V^{-1} X)^{-1} X' V^{-1} Z\big] \\
&= I_F T_{2.1} + T_{2.2} + T_{2.3},
\end{aligned}$$

where E_i denotes a $m \times m$ matrix with the (i, i) component one and rest zeroes.

Using Lemma 2 and the following facts:

(i) $PVP = P$,

(ii) $\mathrm{tr}[C_i V]$ and $\mathrm{tr}[P^2 V C_i V]$ are of order $O(m^{-1})$, under the regularity conditions,

we have

$$T_{2.1} = \tfrac{1}{2}\{E[(Z'P^2Z)(Z'C_iZ)] - \mathrm{tr}[P]E[Z'C_iZ]\}$$
$$= \mathrm{tr}[P^2VC_iV] + \tfrac{1}{2}\mathrm{tr}[P^2V]\mathrm{tr}[C_iV] - \tfrac{1}{2}\mathrm{tr}[P]\mathrm{tr}[C_iV]$$
$$= O(m^{-1}),$$

where $C_i = E_i X(X'V^{-1}X)^{-1}X'V^{-1}$.

Using $\lambda_i' \frac{\partial H}{\partial A} = O(m^{-1})$, we have

$$T_{2.2} = E[(\hat{A}_U - A) r_{1.2} Z(\lambda_i' Z)]$$
$$= E[(\hat{A}_U - A)^2(\lambda_i' XHZ)(\lambda_i' Z)] + E[(\hat{A}_U - A) r_1 (\lambda_i' Z)]$$
$$\leq E[(\hat{A}_U - A)^2]E[Z'H'X'E_iZ] + E\left[(\hat{A}_U - A)^3\left(\lambda_i' X \frac{\partial H}{\partial A}\bigg|_{A=A} Z\right)(\lambda_i' Z)\right]$$
$$= O(m^{-2}).$$

Using $E[|r_{2.2}|] = O(m^{-1})$,

$$T_{2.3} = E[r_{2.2}Z'C_iZ] \leq E[|r_{2.2}|]\mathrm{tr}[C_iV] = O(m^{-2}).$$

Therefore,

$$E[T_2] \leq O(m^{-2}).$$

Hence, using the above results and $E[T_2] \geq O(m^{-2})$ with same calculation, we have

$$E[T_2] = O(m^{-2}). \tag{B.4}$$

Since I_F is of order $O(m^{-1})$, we have

$$E[T_3] = E[(\hat{A} - A)^2(y_i - x_i'\beta)^2]$$
$$= E[(I_F l_{\mathrm{RE}}^{(1)} + r_{2.2})^2 \lambda_i' ZZ' \lambda_i]$$
$$= I_F^2\{\tfrac{1}{4}E[(Z'P^2Z)(Z'P^2Z)(Z'E_iZ)] - \tfrac{1}{2}E[(Z'P^2Z)(Z'E_iZ)]\mathrm{tr}[P]$$
$$+ \tfrac{1}{4}E[Z'E_iZ]\mathrm{tr}[P]^2\}$$
$$+ I_F E[r_{2.2}(Z'P^2Z - \mathrm{tr}[P])Z'E_iZ] + E[r_{2.2}^2 Z'E_iZ]$$
$$\leq I_F^2\Gamma + I_F E[|r_{2.2}|]\{2\,\mathrm{tr}[P^2VE_iV] + \mathrm{tr}[P^2V]\mathrm{tr}[E_iV] - \mathrm{tr}[P]\mathrm{tr}[E_iV]\}$$
$$+ E[r_{2.2}^2]\mathrm{tr}[E_iV]$$
$$= \Gamma I_F^2 + O(m^{-2}).$$

Using Lemma 2 and the following facts:

(i) $PVP = P$,
(ii) $\operatorname{tr}(E_i V) = (A + D_i)$, and
(iii) $|\operatorname{tr}(P^k) - \operatorname{tr}(V^{-k})| = O(1)$, for $k \geq 1$,

we have

$$\begin{aligned}
\Gamma &= \{\tfrac{1}{4}E[(Z'P^2Z)(Z'P^2Z)(Z'E_iZ)] - \tfrac{1}{2}E[(Z'P^2Z)(Z'E_iZ)]\operatorname{tr}[P] \\
&\qquad + \tfrac{1}{4}E[Z'E_iZ]\operatorname{tr}[P]^2\} \\
&= \tfrac{1}{4}[8\operatorname{tr}(P^2VP^2E_iV) + 2\{\operatorname{tr}(P^2VP^2V)\operatorname{tr}(E_iV) + 2\operatorname{tr}(P^2VE_iV)\operatorname{tr}(P^2V)\} \\
&\qquad + \operatorname{tr}(P^2V)^2\operatorname{tr}(E_iV)] \\
&\quad - \operatorname{tr}(P^2VE_iV)\operatorname{tr}(P) - \tfrac{1}{2}\operatorname{tr}(P^2V)\operatorname{tr}(E_iV)\operatorname{tr}(P) + \tfrac{1}{4}\operatorname{tr}(P)^2\operatorname{tr}(E_iV) \\
&= 2\operatorname{tr}(P^3VE_iV) + \tfrac{1}{2}\operatorname{tr}(P^2)\operatorname{tr}(E_iV) = \tfrac{1}{2}\operatorname{tr}(P^2)\operatorname{tr}(E_iV) + O(1) \\
&= \tfrac{1}{2}\operatorname{tr}(V^{-2})(A + D_i) + O(1).
\end{aligned}$$

Hence,

$$E[T_3] = I_F^2 \frac{1}{2}\operatorname{tr}(V^{-2})(A + D_i) + O(m^{-2}) = \frac{2(A + D_i)}{\operatorname{tr}(V^{-2})} + O(m^{-2}). \tag{B.5}$$

Thus, we can show (A.5) using (B.1), (B.4) and (B.5).

Acknowledgments. M. Yoshimori conducted this research while visiting the University of Maryland, College Park, USA, as a research scholar under the supervison of the second author. The authors thank Professor Yutaka Kano for reading an earlier draft of the paper and making constructive comments. We are also grateful to three referees, an Associate Editor and Professor Runze Li, Coeditor, for making a number of constructive suggestions, which led to a significant improvement of our paper.

SUPPLEMENTARY MATERIAL

Supplemental proof (DOI: 10.1214/14-AOS1219SUPP; .pdf). We provide a proof of Theorem 4.

REFERENCES

BASU, R., GHOSH, J. K. and MUKERJEE, R. (2003). Empirical Bayes prediction intervals in a normal regression model: Higher order asymptotics. *Statist. Probab. Lett.* **63** 197–203. MR1986689

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. U.S. Census Bureau, Washington, DC.

CARLIN, B. P. and LOUIS, T. A. (1996). *Bayes and Empirical Bayes Methods for Data Analysis. Monographs on Statistics and Applied Probability* **69**. Chapman & Hall, London. MR1427749

CARTER, G. M. and ROLPH, J. F. (1974). Empirical Bayes methods applied to estimating fire alarm probabilities. *J. Amer. Statist. Assoc.* **69** 880–885.

CHATTERJEE, S., LAHIRI, P. and LI, H. (2008). Parametric bootstrap approximation to the distribution of EBLUP and related prediction intervals in linear mixed models. *Ann. Statist.* **36** 1221–1245. MR2418655

COX, D. R. (1975). Prediction intervals and empirical Bayes confidence intervals. In *Perspectives in Probability and Statistics* (*papers in Honour of M. S. Bartlett on the Occasion of His* 65*th Birthday*) J. Gani, ed.) 47–55. Applied Probability Trust, Univ. Sheffield, Sheffield. MR0403046

DAS, K., JIANG, J. and RAO, J. N. K. (2004). Mean squared error of empirical predictor. *Ann. Statist.* **32** 818–840. MR2060179

DATTA, G. S. and LAHIRI, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statist. Sinica* **10** 613–627. MR1769758

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. MR2158619

DATTA, G. S., GHOSH, M., SMITH, D. D. and LAHIRI, P. (2002). On an asymptotic theory of conditional and unconditional coverage probabilities of empirical Bayes confidence intervals. *Scand. J. Stat.* **29** 139–152. MR1894387

DIAO, L., SMITH, D. D., DATTA, G. S., MAITI, T. and OPSOMER, J. D. (2014). Accurate confidence interval estimation of small area parameters under the Fay–Herriot model. *Scand. J. Stat.* **41** 497–515.

EFRON, B. and MORRIS, C. N. (1975). Data analysis using Stein's estimator and its generalizations. *J. Amer. Statist. Assoc.* **70** 311–319.

FAY, R. E. III and HERRIOT, R. A. (1979). Estimates of income for small places: An application of James–Stein procedures to census data. *J. Amer. Statist. Assoc.* **74** 269–277. MR0548019

GAN, L. and JIANG, J. (1999). A test for global maximum. *J. Amer. Statist. Assoc.* **94** 847–854. MR1723335

HALL, P. and MAITI, T. (2006). On parametric bootstrap methods for small area prediction. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 221–238. MR2188983

LAHIRI, P. and LI, H. (2009). 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*. Available at http://www.fcsm.gov/events/papers2009.html.

LAIRD, N. M. and LOUIS, T. A. (1987). Empirical Bayes confidence intervals based on bootstrap samples. *J. Amer. Statist. Assoc.* **82** 739–757. MR0909979

LI, H. and LAHIRI, P. (2010). An adjusted maximum likelihood method for solving small area estimation problems. *J. Multivariate Anal.* **101** 882–892. MR2584906

MORRIS, C. N. (1983a). Parametric empirical Bayes inference: Theory and applications. *J. Amer. Statist. Assoc.* **78** 47–65. MR0696849

MORRIS, C. N. (1983b). Parametric empirical Bayes confidence intervals. In *Scientific Inference, Data Analysis, and Robustness* (*Madison, Wis.*, 1981) (G. E. P. Box, T. Leonard and C. F. J. Wu, eds.). *Publ. Math. Res. Center Univ. Wisconsin* **48** 25–50. Academic Press, Orlando, FL. MR0772762

MORRIS, C. N. and TANG, R. (2011). Estimating random effects via adjustment for density maximization. *Statistical Sci.* **26** 271–287. MR2858514

NANDRAM, B. (1999). An empirical Bayes prediction interval for the finite population mean of a small area. *Statist. Sinica* **9** 325–343. MR1707843

PRASAD, N. G. N. and RAO, J. N. K. (1990). The estimation of the mean squared error of small-area estimators. *J. Amer. Statist. Assoc.* **85** 163–171. MR1137362

RAO, J. N. K. (2003). *Small Area Estimation*. Wiley, Hoboken, NJ. MR1953089

SASASE, Y. and KUBOKAWA, T. (2005). Asymptotic correction of empirical Bayes confidence intervals and its application to small area estimation (in Japanese). *J. Japan Statist. Soc.* **35** 27–54.

SEARLE, S. R., CASELLA, G. and MCCULLOCH, C. E. (1992). *Variance Components*. Wiley, New York. MR1190470

SRIVASTAVA, V. K. and TIWARI, R. (1976). Evaluation of expectations of products of stochastic matrices. *Scand. J. Stat.* **3** 135–138. MR0518922

WANG, J. and FULLER, W. A. (2003). The mean squared error of small area predictors constructed with estimated area variances. *J. Amer. Statist. Assoc.* **98** 716–723. MR2011685

YOSHIMORI, M. (2014). Numerical comparison between different prediction intervals using EBLUP under the Fay–Herriot model. *Comm. Statist. Simulation Comput*. To appear.

YOSHIMORI, M. and LAHIRI, P. (2014). Supplement to "A second-order efficient empirical Bayes confidence interval." DOI:10.1214/14-AOS1219SUPP.

YOSHIMORI, M. and LAHIRI, P. (2014). A new adjusted maximum likelihood method for the Fay–Herriot small area model. *J. Multivariate Anal.* **124** 281–294. MR3147326

DEPARTMENT OF MEDICAL INNOVATION
OSAKA UNIVERSITY HOSPITAL
2-2 YAMADAOKA
SUITA, OSAKA 5650871
JAPAN
E-MAIL: masayo@sigmath.es.osaka-u.ac.jp

THE JOINT PROGRAM IN SURVEY METHODOLOGY
UNIVERSITY OF MARYLAND
1218 LEFRAK HALL
COLLEGE PARK, MARYLAND 20742
USA
E-MAIL: plahiri@survey.umd.edu