

A UNIFIED MEASURE OF UNCERTAINTY OF ESTIMATED BEST LINEAR UNBIASED PREDICTORS IN SMALL AREA ESTIMATION PROBLEMS

G. S. Datta and P. Lahiri

University of Georgia and University of Nebraska-Lincoln

Abstract: We obtain a second order approximation to the mean squared error (MSE), and its estimate, of the empirical or estimated best linear unbiased predictor (EBLUP) of a mixed effect in a general mixed linear normal model. This covers many important small area models in the literature. Unlike previous research in this area, we provide a unified theory of measuring uncertainty of an EBLUP for a complex small area model where the variance components are estimated by various standard methods including restricted or residual maximum likelihood (REML) and maximum likelihood (ML). It turns out that the MSE approximations for the REML and the ML methods are exactly the same in the second order asymptotic sense. However, the second order accurate estimator of this MSE based on the former method requires less bias correction than the one based on the latter method. This is due to a result in the paper which shows that the bias of the REML estimators of variance components is of lower order than that of the ML estimators. A simulation is undertaken to compare different methods of estimating the variance components and to study the properties of various estimators of the MSE of the mixed effect. In our context it is interesting to note that the residual likelihood is same as the conditional profile likelihood (CPL) of Cox and Reid (1987). Thus, this paper addresses an important open problem raised by Cox and Reid (1987) in small area prediction using the CPL method.

Key words and phrases: Borrow strength, conditional profile likelihood, maximum likelihood, mean squared error, mixed linear model, residual maximum likelihood, random effect.

1. Introduction

Research on small area estimation has received considerable attention in recent years due to growing demand for reliable small area statistics by various federal and local government agencies (e.g., the U.S. Census Bureau, U.S. Bureau of Labor Statistics, Statistics Canada, Australian Bureau of Statistics, and the Central Statistical Office of U.K.). A small area (domain) usually refers to a subgroup of a population from which samples are drawn. The subgroup may be a geographical region (e.g., county or municipality) or a group obtained by cross-classification of demographic factors such as age, race or sex. The importance

of reliable small area statistics cannot be over-emphasized as these are needed in regional planning and fund allocation in many federal and local government programs.

Most surveys provide very little information on a particular small area of interest since surveys are generally designed to produce statistics for larger populations. Thus, direct design-based estimators (see Cochran (1977)) are unreliable since only a few samples are available from the particular small area of interest. The main idea to improve on a design-based survey estimator is to use relevant supplementary information, usually available from various administrative records, in conjunction with the sample survey data. See Ghosh and Rao (1994), Holt, Smith and Tomberlin (1979), Rao (1986), among others.

The estimated best linear unbiased prediction (EBLUP) method has been widely used to produce small area statistics. In this approach, using Henderson's (1975) method, first the best linear unbiased predictor (BLUP) of the mixed effect of interest is produced using a normal mixed linear model. This predictor usually contains unknown variance components which are then replaced by their estimates using a standard variance component estimation procedure. The resulting predictor is known as an EBLUP of the mixed effect. Cressie (1992) and Dick (1995), among others, prefer the residual maximum likelihood method (REML) to the maximum likelihood (ML) method in estimating variance components in complex small area models. We prove in Theorem A.3 that the order of bias of the REML estimator is lower than that of the ML estimator, which justifies the preference of the former estimator over the latter. In Remark 1 in Section 5 we connect the residual likelihood as defined in Patterson and Thompson (1971) to the conditional profile likelihood (CPL) of Cox and Reid (1987) in our context.

A naive measure of uncertainty of an EBLUP is the mean squared error (MSE) of the corresponding BLUP. Prasad and Rao (1990) noted that this measure could lead to severe underestimation of the true uncertainty of an EBLUP since it does not incorporate the variability due to the estimation of the variance components. Using analysis of variance (ANOVA) estimates of the variance components, Prasad and Rao (1990) approximated the MSE of an EBLUP for a mixed linear normal model. In their approximation, they neglected terms of order $o(m^{-1})$ where m is the number of small areas. To the same order of approximation they also derived an estimator of their MSE approximation. The work of Singh, Stukel and Pfeffermann (1998) considers similar approximation to the MSE.

Prasad and Rao (1990) justified their approximation only for the ANOVA estimates of variance components. However, the ML and the REML methods are also widely used to produce EBLUP in small area estimation. Thus, there is a

need to advance their work when the variance components are estimated by these other methods. Also, Prasad and Rao (1990) *rigorously* justified the accuracy of their MSE only for the Fay-Herriot model.

In Section 2, we consider the general normal mixed linear model used by Prasad and Rao (1990). The BLUP and EBLUP of a mixed effect are presented in Section 3. In Section 4, a second order approximation to the MSE of an EBLUP is provided under certain regularity conditions. This approximation is valid for a number of variance component estimation methods including the ML and the REML. In Section 5, an estimator of the MSE of the proposed EBLUP is given for the general model considered in Section 2. The bias of this estimator is $o(m^{-1})$ and its small-sample accuracy is investigated in Section 6 through simulation for a special case of the model proposed in Section 2. Since the REML estimator is identical to the maximum CPL estimator, we have addressed an important problem raised by Cox and Reid (1987) in the context of small area prediction using the CPL method. In the Appendix, expressions for asymptotic biases of the ML and the maximum CPL estimators are provided in general setting. These expressions are used to derive MSE approximations and their estimators, accurate up to $o(m^{-1})$, of EBLUP's of mixed effects.

2. Notation and the Model

Consider the following general normal mixed linear model in small area estimation:

$$Y_i = X_i\beta + Z_iv_i + e_i, \quad i = 1, \dots, m, \quad (1)$$

where X_i ($n_i \times p$) and Z_i ($n_i \times b_i$) are known matrices, v_i and e_i are independently distributed with $v_i \stackrel{ind}{\sim} N_{b_i}(0, G_i)$ and $e_i \stackrel{ind}{\sim} N_{n_i}(0, R_i)$, $i = 1, \dots, m$. We assume that $G_i = G_i(\psi)$ ($b_i \times b_i$) and $R_i = R_i(\psi)$ ($n_i \times n_i$) possibly depend on $\psi = (\psi_1, \dots, \psi_q)'$, a $q \times 1$ vector of fixed variance components. Using the notations given in Prasad and Rao (1990), write $Y = col_{1 \leq i \leq m} Y_i$, $e = col_{1 \leq i \leq m} e_i$, $X = col_{1 \leq i \leq m} (X_i)$, $Z = diag_{1 \leq i \leq m} (Z_i)$, $G(\psi) = diag_{1 \leq i \leq m} G_i$, $v = col_{1 \leq i \leq m} v_i$ and $R(\psi) = diag_{1 \leq i \leq m} R_i$. We assume that X has full column rank p . Let $\Sigma(\psi) = R(\psi) + ZG(\psi)Z'$, the variance-covariance matrix of Y . With these notations we can write (1) as

$$Y = X\beta + Zv + e, \quad (2)$$

where v and e are independently distributed with $v \sim N_b(0, G)$, $e \sim N_n(0, R)$, $n = \sum_{i=1}^m n_i$ and $b = \sum_{i=1}^m b_i$. This model covers the following two important small area models.

Fay-Herriot Model (see Fay and Herriot (1979)):

$$Y_i = x_i'\beta + v_i + e_i, \quad i = 1, \dots, m,$$

where v_i 's and e_i 's are independent with $v_i \stackrel{iid}{\sim} N(0, A)$ and $e_i \stackrel{iid}{\sim} N(0, D_i)$, D_i ($i = 1, \dots, m$) being known. Here, $n_i = b_i = 1$, $Z_i = 1$, $\psi = A$, $R_i(\psi) = D_i$ and $G_i(\psi) = A$ ($i = 1, \dots, m$).

Nested Error Regression Model (see Battese, Harter and Fuller (1988)):

$$Y_{ij} = x'_{ij}\beta + v_i + e_{ij}, \quad i = 1, \dots, m; \quad j = 1, \dots, n_i, \quad (3)$$

where v_i and e_{ij} are independently distributed with $v_i \stackrel{iid}{\sim} N(0, \sigma_v^2)$ and $e_{ij} \stackrel{iid}{\sim} N(0, \sigma_e^2)$, $i = 1, \dots, m$; $j = 1, \dots, n_i$. In this case, $Y_i = (Y_{i1}, \dots, Y_{in_i})'$, $X_i = \text{col}_{1 \leq j \leq n_i} x'_{ij}$, $Z_i = 1_{n_i}$, $e_i = (e_{i1}, \dots, e_{in_i})'$, $\psi = (\sigma_e^2, \sigma_v^2)'$, $R_i(\psi) = \sigma_e^2 I_{n_i}$, $G_i(\psi) = \sigma_v^2$ ($i = 1, \dots, m$) where I_{n_i} ($n_i \times n_i$) is an identity matrix and 1_{n_i} ($n_i \times 1$) is a vector of ones.

3. Best Linear Unbiased Predictor (BLUP) and Estimated BLUP

In this paper we are interested in predicting a general mixed effect $\theta = h'\beta + \lambda'v$, where h and λ are known vectors of order $p \times 1$ and $b \times 1$ respectively. When ψ is known, the BLUP of θ is given by $\hat{\theta}(\psi) = h'\hat{\beta}(\psi) + s'(\psi)[Y - X\hat{\beta}(\psi)]$, where $s(\psi) = \Sigma^{-1}(\psi)ZG(\psi)\lambda$ and $\hat{\beta}(\psi) = [X'\Sigma^{-1}(\psi)X]^{-1}[X'\Sigma^{-1}(\psi)Y]$.

In practice ψ is unknown and is estimated from the data. Let $\hat{\psi}$ be an estimator of ψ . In this paper, we consider $\hat{\psi}$ which satisfy condition (f) given in Section 4. Then an EBLUP of θ is $\hat{\theta}(\hat{\psi})$, obtained from $\hat{\theta}(\psi)$ with ψ replaced by $\hat{\psi}$.

4. Second Order Approximation to MSE of $\hat{\theta}(\hat{\psi})$

The MSE of $\hat{\theta}(\hat{\psi})$ is $\text{MSE}[\hat{\theta}(\hat{\psi})] = E[\hat{\theta}(\hat{\psi}) - \theta]^2$, where E denotes the expectation with respect to model (1). Under certain regularity conditions, Prasad and Rao (1990) provided a second order approximation to $\text{MSE}[\hat{\theta}(\hat{\psi})]$. They obtained the approximation when ψ is estimated by $\hat{\psi}^{PR} = (\hat{\psi}_1^{PR}, \dots, \hat{\psi}_q^{PR})'$, where $\hat{\psi}_d^{PR} = \max\{0, Y'Q_d Y\}$ with $Q_d = \text{diag}_{1 \leq i \leq m} [O(m^{-1})]_{n_i \times n_i} + [O(m^{-2})]_{n \times n}$, $[O(m^{-r})]_{k \times k}$ being a $k \times k$ matrix with all elements $O(m^{-r})$. Note that the REML and the ML estimators of ψ are generally not of the form $\hat{\psi}^{PR}$. While $\hat{\psi}^{PR}$ has closed-form expression, the REML and the ML estimators of ψ generally do not have closed-form expressions. Thus the Prasad-Rao theory is not useful for our purposes when $\hat{\psi}$ is the REML or the ML estimator of ψ .

We obtain an approximation to $\text{MSE}[\hat{\theta}(\hat{\psi})]$ where $\hat{\psi}$ may be the REML or the ML estimate of ψ . In our approximation we assume m is large and neglect all terms of order $o(m^{-1})$. The following regularity conditions, referred to as RC later on, will be assumed:

- (a) The elements of X and Z are uniformly bounded such that $X'\Sigma^{-1}(\psi)X = [O(m)]_{p \times p}$;

- (b) $\sup_{i \geq 1} n_i < \infty$ and $\sup_{i \geq 1} b_i < \infty$;
 (c) $h - X's(\psi) = [O(1)]_{p \times 1}$;
 (d) $\frac{\partial}{\partial \psi_d} [X's(\psi)] = [O(1)]_{p \times 1}$ for $d = 1, \dots, q$;
 (e) $R_i(\psi) = \sum_{j=0}^q \psi_j D_{ij} D'_{ij}$ and $G_i(\psi) = \sum_{j=0}^q \psi_j F_{ij} F'_{ij}$, where $\psi_0 = 1$, D_{ij} and F_{ij} ($i = 1, \dots, m, j = 0, \dots, q$) are known matrices of order $n_i \times b_i$ and $b_i \times b_i$ respectively, and the elements are uniformly bounded known constants such that $R_i(\psi)$ and $G_i(\psi)$ ($i = 1, \dots, m$) are all positive definite matrices. (In special cases, some of D_{ij} and F_{ij} may be null matrices.)
 (f) $\hat{\psi}$ is an estimator of ψ which satisfies (i) $\hat{\psi} - \psi = O_p(m^{-1/2})$, (ii) $\hat{\psi} - \tilde{\psi} = O_p(m^{-1})$, (iii) $\hat{\psi}(-Y) = \hat{\psi}(Y)$, and (iv) $\hat{\psi}(Y + Xh) = \hat{\psi}(Y)$ for any $h \in R^p$ and for all Y , where $\tilde{\psi}$ is the ML estimator of ψ .

Conditions (a)-(d) are in Prasad and Rao (1990) though (c) and (d) were not explicitly stated. Condition (e) reduces (1) to a generalization of the usual variance components model (see Rao (1973)). Under (e), we can write (1) as $Y_i = X_i \beta + \sum_{j=0}^q E_{ij} \xi_{ij}$, where $E_{ij} = [Z_i F_{ij}, D_{ij}]$, $\xi_{ij} \stackrel{ind}{\sim} N(0, \psi_j I_{2b_i})$, ($i = 1, \dots, m; j = 0, \dots, q$). For our model, unlike the usual variance components model, we do not restrict matrices E_{ij} to 0 and 1 and entries nor $E'_{ij} E_{ij}$ to be diagonal for all $i = 1, \dots, m; j = 0, \dots, q$. Condition (f) is generally satisfied by the ML and the REML estimators of variance components.

Define $g_1(\psi) = \lambda' G(\psi) \lambda - s'(\psi) Z G(\psi) \lambda$, $g_2(\psi) = [h - X's(\psi)]' (X' \Sigma^{-1}(\psi) X)^{-1} [h - X's(\psi)]$, $L_d(\psi) = \frac{\partial}{\partial \psi_d} s(\psi)$ ($d = 1, \dots, q$), $L(\psi) = \text{col}_{1 \leq d \leq q} L'_d(\psi)$ and $g_3(\psi) = \text{tr}[L(\psi) \Sigma(\psi) L'(\psi) \text{Var}(\hat{\psi})]$, where $\text{Var}(\hat{\psi})$ denotes the asymptotic variance of $\hat{\psi}$.

It is shown in Theorem A.1 in the Appendix that, under regularity conditions RC,

$$MSE[\hat{\theta}(\hat{\psi})] \doteq g_1(\psi) + g_2(\psi) + g_3(\psi), \quad (4)$$

where \doteq means that neglected terms are of order $o(m^{-1})$. A naive approximation $g_1(\psi) + g_2(\psi)$ may be a serious underestimate of $MSE[\hat{\theta}(\hat{\psi})]$ since $g_3(\psi)$ is $O(m^{-1})$, the order of $g_2(\psi)$. For this reason, we shall not ignore any $O(m^{-1})$ term throughout the paper.

We now write down the formula (4) when ψ is estimated by $\hat{\psi}_{ML}$ and $\hat{\psi}_{REML}$, the ML and the REML estimates, respectively. Let $\eta = (\beta', \psi')'$ and $l_i(\eta)$ denote the log-likelihood of η based on Y_i ($i = 1, \dots, m$). Let $H(\eta) = ((\frac{\partial^2 \bar{l}}{\partial \eta_\alpha \partial \eta_\gamma}))$, where $\bar{l} = m^{-1} \sum_{i=1}^m l_i(\eta)$ and $I_\eta(\psi) = -E[H(\eta)]$, the average information matrix for η . It can be checked that, in the sense of Cox and Reid (1987) and Huzurbazar (1950), β and ψ are orthogonal, i.e., $I_\eta(\psi) = \text{diag}(I_\beta(\psi), I_\psi(\psi))$, where $I_\beta(\psi)$ and $I_\psi(\psi)$ are the information matrices for β and ψ respectively. It follows from properties of ML estimators that $\text{Var}(\hat{\psi}_{ML}) \doteq \text{Var}(\hat{\psi}_{REML}) \doteq m^{-1} I_\psi^{-1}(\psi)$ so that $g_3(\psi) \doteq m^{-1} \text{tr}[L(\psi) \Sigma(\psi) L'(\psi) I_\psi^{-1}(\psi)] = g_3^*(\psi)$, say, where $\hat{\psi}_{ML}$ and

$\hat{\psi}_{REML}$ are the ML and the REML estimators, respectively. Thus (4) reduces to

$$MSE[\hat{\theta}(\hat{\psi})] \doteq g_1(\psi) + g_2(\psi) + g_3^*(\psi). \quad (5)$$

It is interesting to note that formula (5) is valid for both the ML and the REML methods of estimating the variance components. We now spell out formula (5) for the Fay-Herriot and nested error regression models for estimating the small area mean. For the Fay-Herriot model let $\hat{\theta}_i(\hat{A})$ denote an EBLUP of $\theta_i = x_i'\beta + v_i$ where \hat{A} satisfies regularity condition (f). Then $g_{1i}(A) = AD_i/(A + D_i)$, $g_{2i}(A) = D_i^2(A + D_i)^{-2}x_i'[\sum_{u=1}^m(A + D_u)^{-1}x_u x_u']^{-1}x_i$ (as in Prasad and Rao (1990)) and $g_{3i}^*(A) = 2D_i^2(A + D_i)^{-3}[\sum_{u=1}^m(A + D_u)^{-2}]^{-1}$.

For the nested error regression model, let $\hat{\psi} = (\hat{\sigma}_v^2, \hat{\sigma}_e^2)'$ be an estimator of $\psi = (\sigma_v^2, \sigma_e^2)'$ and let $\hat{\theta}_i(\hat{\psi})$ be an EBLUP of $\theta_i = \bar{X}_i'\beta + v_i$ based on $\hat{\psi}$, where \bar{X}_i is the known population mean of the covariate vector x_{ij} for the i th small area ($i = 1, \dots, m$). Then $g_{1i}(\psi) = (1 - \gamma_i)\sigma_v^2$, $g_{2i}(\psi) = (\bar{X}_i - \gamma_i\bar{x}_i)'(X'\Sigma^{-1}(\psi)X)^{-1}(\bar{X}_i - \gamma_i\bar{x}_i)$, where $\gamma_i = \sigma_v^2(\sigma_v^2 + \sigma_e^2 n_i^{-1})^{-1}$ and \bar{x}_i is the sample mean vector of the covariate X 's for the i th small area (as in Prasad and Rao (1990)). Also, $g_{3i}(\psi) = n_i^{-2}(\sigma_v^2 + n_i^{-1}\sigma_e^2)^{-3}[\sigma_e^4 I^{vv} + \sigma_v^4 I^{ee} - 2\sigma_e^2\sigma_v^2 I^{ve}]$, $I^{vv} = 2a^{-1}\sum_{u=1}^m[(n_u - 1)\sigma_e^{-4} + w_u^{-2}]$, $I^{ee} = 2a^{-1}\sum_{u=1}^m n_u^2 w_u^{-2}$, $I^{ve} = -2a^{-1}\sum_{u=1}^m n_u w_u^{-2}$, $a = [\sum_{u=1}^m n_u^2 w_u^{-2}] [\sum_{u=1}^m \{(n_u - 1)\sigma_e^{-4} + w_u^{-2}\}] - (\sum_{u=1}^m n_u w_u^{-2})^2$, $w_i = \sigma_e^2 + n_i\sigma_v^2$ and I^{vv} , I^{ve} and I^{ee} are the elements of I_ψ^{-1} .

Assuming a superpopulation model of the form (3) for the N_i population units in the i th area, it can be shown that the EBLUP of $\bar{Y}_i = N_i^{-1}\sum_{j=1}^{N_i} Y_{ij}$, the i th finite population mean, is given by $\hat{\theta}_i^F(\hat{\psi}) = f_i\bar{y}_i + (1 - f_i)\hat{\theta}_i^*(\hat{\psi})$ under the nested error regression model, where $f_i = n_i/N_i$, $\hat{\theta}_i^*(\hat{\psi})$ is $\hat{\theta}_i(\hat{\psi})$ except that \bar{X}_i is replaced by \bar{X}_i^* , the mean of the x_{ij} for the $N_i - n_i$ non-sampled units, and F stands for finite populations (see (3.6) of Prasad and Rao (1990)). From (4.12) of Prasad and Rao (1990), it follows that $MSE[\hat{\theta}_i^F(\hat{\psi})] = (1 - f_i)^2[MSE(\hat{\theta}_i^*(\hat{\psi})) + N_i^{-1}(1 - f_i)^{-1}\sigma_e^2]$ where $MSE(\hat{\theta}_i^*(\hat{\psi}))$ is $MSE(\hat{\theta}_i(\hat{\psi}))$, except that \bar{X}_i in g_{2i} will be replaced by \bar{X}_i^* .

5. Estimator of the MSE of EBLUP

Let $b_{\hat{\psi}}(\psi)$ be the bias of $\hat{\psi}$, i.e., $E(\hat{\psi}) - \psi$, up to the order $o(m^{-1})$. Let $\nabla g_1(\psi) = (\frac{\partial}{\partial\psi_1}g_1(\psi), \dots, \frac{\partial}{\partial\psi_q}g_1(\psi))'$, where

$$\frac{\partial}{\partial\psi_d}g_1(\psi) = \lambda'F_dF_d'\lambda - [\frac{\partial s(\psi)}{\partial\psi_d}]'ZG(\psi)\lambda - s'(\psi)Z\frac{\partial G(\psi)}{\partial\psi_d}\lambda, \quad (6)$$

and $F_d = \text{diag}_{1 \leq i \leq m} F_{id}$, $d = 1, \dots, q$.

Using (4), we get an estimator of $MSE[\hat{\theta}(\hat{\psi})]$ whose bias is of order $o(m^{-1})$. It is shown in Theorem A.2 that

$$E[g_1(\hat{\psi})] \doteq g_1(\psi) + b'_{\hat{\psi}}(\psi)\nabla g_1(\psi) - g_3(\psi), \quad (7)$$

$$E[g_2(\hat{\psi})] \doteq g_2(\psi), \quad (8)$$

$$E[g_3(\hat{\psi})] \doteq g_3(\psi). \quad (9)$$

It now follows from (7) and (8) that the naive estimator $mse^N\{\hat{\theta}(\hat{\psi})\} = g_1(\hat{\psi}) + g_2(\hat{\psi})$ has a bias of order $O(m^{-1})$ in estimating $MSE[\hat{\theta}(\hat{\psi})]$. Using (4) and (7)-(9), we get

$$E[g_1(\hat{\psi}) + g_2(\hat{\psi}) + 2g_3(\hat{\psi}) - b'_{\hat{\psi}}(\hat{\psi})\nabla g_1(\hat{\psi})] \doteq MSE[\hat{\theta}(\hat{\psi})].$$

Thus, $mse[\hat{\theta}(\hat{\psi})] = g_1(\hat{\psi}) + g_2(\hat{\psi}) + 2g_3(\hat{\psi}) - b'_{\hat{\psi}}(\hat{\psi})\nabla g_1(\hat{\psi})$ is an estimator of $MSE[\hat{\theta}(\hat{\psi})]$ such that $E\{mse[\hat{\theta}(\hat{\psi})]\} \doteq MSE[\hat{\theta}(\hat{\psi})]$.

Remark 1. It follows from Remark A.3 in the Appendix that for the REML estimator of ψ , $\hat{\psi}_{REML}$, $b_{\hat{\psi}_{REML}}(\psi) \doteq 0$, and it follows from Theorem A.3 in the Appendix that for the ML estimator of ψ , $\hat{\psi}_{ML} = \tilde{\psi}$, $b_{\hat{\psi}_{ML}}(\psi) \doteq \frac{1}{2}m^{-1}I_{\psi}^{-1}(\psi) col_{1 \leq d \leq q} tr[I_{\beta}^{-1}(\psi) \frac{\partial}{\partial \psi_d} I_{\beta}(\psi)]$. Thus, the method based on the REML estimators requires less bias correction than the method based on the ML estimators. It can be checked that for the Fay-Herriot model the bias of \hat{A}_{ML} is $b_{\hat{A}_{ML}}(A) = -tr\{[\sum_{u=1}^m (A + D_u)^{-1} x_u x_u']^{-1} [\sum_{u=1}^m (A + D_u)^{-2} x_u x_u']\} / \sum_{u=1}^m (A + D_u)^{-2}$. Since $\nabla g_1(A) = (\frac{D_i}{A+D_i})^2 > 0$ and the bias of \hat{A}_{ML} is negative, using the Prasad-Rao MSE estimate for the ML estimate of A would result in underestimation of the approximate MSE up to $o(m^{-1})$.

6. A Simulation Study

In this section we conduct a simulation study for a special case of the Fay-Herriot model without covariates. In this case, $x'_i \beta = \mu$ (i.e., $x_i \equiv 1$). We investigate the finite sample performances of $\hat{\theta}_i(\hat{A})$ when \hat{A} , the estimator of the variance component A , is obtained by ANOVA, REML (\hat{A}_{REML}) and ML (\hat{A}_{ML}) methods. Since the ANOVA estimator could produce negative values, we need to truncate it at an arbitrary nonnegative real number whenever it yields a negative value. In this simulation, we consider three different ANOVA estimators, \hat{A}_{AOV0} , \hat{A}_{AOV1} and \hat{A}_{AOV2} , which correspond to the truncation points 0, 0.0001 and 0.01, respectively. The finite sample accuracy of the proposed estimators of MSE are also investigated.

Since the MSE is translation invariant, we set $\mu = 0$ without loss of generality. We selected $m = 30$, $A = 1$ and considered two different D_i ($i = 1, \dots, 30$)

patterns (see Table 1). Note that the D_i 's are more dispersed in pattern (b) than in pattern (a). These patterns were also used by Lahiri and Rao (1995) in their robustness study of the Prasad-Rao MSE approximation under a non-normality assumption of the small area effect in Fay-Herriot model.

Table 1. Values of D_i 's for the simulation study.

i	1-6	7-12	13-18	19-24	25-30
Pattern (a)	0.7	0.6	0.5	0.4	0.3
Pattern (b)	4.0	0.6	0.5	0.4	0.1

Using IMSL FORTRAN library, we generated 10,000 independent sets of (v_i, e_i) , $i = 1, \dots, 30$, for each case with specified parameters. We thus generate 10,000 independent data sets Y_i , $i = 1, \dots, 30$, using $Y_i = v_i + e_i$, $i = 1, \dots, 30$. Simulated $MSE[\hat{\theta}_i(\hat{A})]$, $E[mse(\hat{\theta}_i(\hat{A}))]$ and $E[mse^N\{\hat{\theta}_i(\hat{A})\}]$ were then computed using these 10,000 data sets for $\hat{A} = \hat{A}_{AOV0}, \hat{A}_{AOV1}, \hat{A}_{AOV2}, \hat{A}_{REML}$ and \hat{A}_{ML} and averaged over the small areas with the same value of D_i . To obtain \hat{A}_{ML} , we used the EM algorithm. We obtained \hat{A}_{REML} using the relation given in part (b) of Theorem (A.3) with A (i.e., ψ) in the second term in the right hand side of part (b) replaced by \hat{A}_{ML} . (This approximation to \hat{A}_{REML} reduced computational burden. While the true REML can be obtained by maximizing (A.13), the above approximation is quite accurate.) The ANOVA estimator of A (i.e., \hat{A}_{AOV0}) took negative values 4 times for pattern (a) and 328 times for pattern (b).

Table 2 reports the simulated MSE's of the small area estimators $\hat{\theta}_i(\hat{A})$ for $\hat{A} = \hat{A}_{AOV0}, \hat{A}_{AOV1}, \hat{A}_{AOV2}, \hat{A}_{REML}$, and \hat{A}_{ML} . There is not much difference in efficiency between the REML and the ML methods, see our comment following equation (5) in Section 4. Although the efficiency of the ANOVA method is comparable with those of the REML and the ML for pattern (a), the REML and the ML methods are much more efficient than the ANOVA method for pattern (b).

Table 2. Simulated $MSE \times 100$ of $\hat{\theta}_i(\hat{A})$ for $\hat{A} = \hat{A}_{AOV0}, \hat{A}_{AOV1}, \hat{A}_{AOV2}, \hat{A}_{REML}$ and \hat{A}_{ML} , averaged over areas with the same values of D_i .

D_i	Pattern (a)					Pattern (b)				
	0.7	0.6	0.5	0.4	0.3	4.0	0.6	0.5	0.4	0.1
AOV0	43.5	39.3	35.6	29.9	24.1	89.5	43.2	39.3	34.1	12.3
AOV1	43.6	39.8	35.1	30.1	24.0	89.1	43.5	40.0	34.2	11.9
AOV2	43.2	39.6	35.0	30.1	24.3	90.0	44.0	38.9	33.7	11.9
REML	43.5	39.3	35.6	29.8	23.9	85.6	39.3	35.1	30.1	9.3
ML	43.6	39.4	35.7	29.9	24.0	85.6	39.4	35.2	30.2	9.3

Percent relative biases (averaged over small areas with the same values of D_i) of the naive MSE estimators and the proposed MSE estimators of $\hat{\theta}_i(\hat{A})$, when $\hat{A} = \hat{A}_{AOV0}, \hat{A}_{AOV1}, \hat{A}_{AOV2}, \hat{A}_{REML}, \hat{A}_{ML}$, are reported in Table 3. We

denote the naive estimators by AOV0N, AOV1N, AOV2N, REMLN and MLN, when A is estimated by ANOVA (truncating the negative estimate at 0, 0.0001 and 0.01), REML and ML methods, respectively. Note that $MSE(\hat{\theta}_i(\hat{A}_{AOV0}))$ is simply the Prasad-Rao approximation. The MSE estimators based on the ML and the REML are less sensitive to the D_i patterns than the Prasad-Rao approximation, i.e., the one based on the ANOVA method. For pattern (a), percent relative biases of all the MSE estimators except the naive estimators are negligible. However, for pattern (b), the Prasad-Rao MSE estimator tends to overestimate the true MSE quite a bit. Note that the naive estimator based on the ML method has more negative bias than the one based on the REML method. This is due to the fact that bias of \hat{A}_{ML} is of order $O(m^{-1})$ but that of \hat{A}_{REML} is of order $o(m^{-1})$.

Table 3. Percent relative bias of estimators of MSE, averaged over areas with the same values of D_i .

D_i	Pattern (a)					Pattern (b)				
	0.7	0.6	0.5	0.4	0.3	4.0	0.6	0.5	0.4	0.1
AOV0	-0.08	0.34	-1.64	-0.07	-0.06	0.11	12.72	14.16	18.43	143.71
AOV0N	-7.24	-6.78	-8.47	-6.60	-6.13	-12.51	-21.04	-22.74	-23.83	-30.77
AOV1	0.10	-0.44	0.31	-0.11	0.42	0.83	12.18	14.92	18.16	145.10
AOV1N	-7.01	-7.43	-6.57	-6.67	-5.68	-11.85	-21.37	-22.17	-23.90	-28.84
AOV2	0.71	-0.02	0.23	-0.44	-0.82	-11.07	10.82	14.96	19.29	118.52
AOV2N	-6.46	-7.07	-6.68	-7.01	-6.88	-12.66	-22.22	-21.97	-22.86	-28.54
REML	-0.70	-0.23	-2.05	-0.19	0.09	-1.12	0.06	-0.71	-1.01	-0.02
REMLN	-7.19	-6.57	-8.22	-6.20	-5.44	-5.28	-6.34	-6.89	-6.85	-3.08
ML	-0.41	0.12	-1.75	0.00	0.46	-1.06	0.39	-0.38	0.00	0.36
MLN	-9.38	-8.70	-10.13	-7.92	-6.95	-9.14	-8.59	-8.96	-8.71	-3.87

Remark 2. A referee asked why the percent relative biases of the MSE estimators reported in Table 3 corresponding to $D_i = 0.6, 0.5$ and 0.4 , the values common between pattern (a) and pattern (b), are quite different in one pattern from the other for the ANOVA estimator of A . It was further asked why the same behavior was not obtained for ML and REML estimators of A . At first sight, this difference in relative biases of Prasad-Rao MSE estimates between pattern (a) and pattern (b) appears puzzling. The reason that the relative biases of the MSE estimates based on either ML or REML estimate of A show the same behavior for both patterns is that the third term, namely, g_{3i} , which is more influential than the second term, remains stable for ML/REML estimates of A . (Recall that in our simulations for the Fay-Herriot model, this term for ML/REML estimation of variance component is $g_{3i}^*(A) = 2D_i^2(A + D_i)^{-3} \{\sum_{u=1}^m (A + D_u)^{-2}\}^{-1}$, whereas for ANOVA estimates of A , it is given by $g_{3i}^{AO} = 2D_i^2(A + D_i)^{-3} \sum_{u=1}^m (A + D_u)^2 / m^2$.)

The first one involves the harmonic mean of $(A + D_u)^2$ values, whereas the second one involves the arithmetic mean.) Based on the values of D_i and $A = 1$ that are used in simulations, we calculate the following table.

Table 4. Relative contribution of the estimation error in estimating the regression coefficient to the estimation error in estimating the variance component.

i	Pattern (a)					Pattern (b)				
	1-6	7-12	13-18	19-24	25-30	1-6	7-12	13-18	19-24	25-30
D_i	0.7	0.6	0.5	0.4	0.3	4.0	0.6	0.5	0.4	0.1
g_{2i}	0.008	0.007	0.006	0.004	0.003	0.035	0.008	0.006	0.004	0.0004
g_{3i}^{AO}	0.015	0.013	0.011	0.009	0.006	0.056	0.039	0.033	0.026	0.004
g_{3i}^*	0.014	0.012	0.011	0.009	0.006	0.019	0.013	0.011	0.009	0.001
g_{3i}^{AO}/g_{2i}	1.87	1.85	1.83	2.25	2.00	1.60	4.88	5.50	6.25	10.00
g_{3i}^*/g_{2i}	1.75	1.71	1.83	2.25	2.00	0.54	1.62	1.83	2.25	2.50

From the table we see that while g_{3i}^*/g_{2i} remains fairly stable for both patterns, fluctuation of g_{3i}^{AO}/g_{2i} is substantial. This is the main reason that we do not get similar relative biases for the MSE estimates for the small areas with the D_i values that are common between the two patterns, when the variance component A is estimated by ANOVA method. Furthermore, while our algorithm converges to some positive value for ML/REML estimates of A for both the patterns, the ANOVA method produces more negative estimates of A for the second pattern.

Acknowledgements

G. S. Datta's research was supported in part by an NSF grant and an ASA/NSF/BLS/Census fellowship. P. Lahiri's research was supported in part by NSF grants SES-9511202 and SBR-9705574. The views expressed here are those of authors, and do not necessarily reflect those of the Census Bureau and the Bureau of the Labor Statistics. The authors are thankful to Professor J. K. Ghosh for many useful discussions and to Professor T. Maiti for help in computing. Thanks are also due to an associate editor and a referee for many useful comments, leading to a better manuscript.

Appendix

Some notation is used here. Let $t = p + q$. Unless noted otherwise the subscripts α, γ and δ will range from 1 to t , a from 1 to p and d, e, f from 1 to q . We use $((x_{ad}))_{a=1, \dots, p; d=1, \dots, q} \equiv ((x_{ad}))_{a,d}$ to denote a $p \times q$ matrix. Define $S_1(\eta) = [S'_{1\beta}(\eta), S'_{1\psi}(\eta)]'$, where $S_{1\beta}(\eta) = \text{col}_{1 \leq a \leq p} \frac{\partial}{\partial \beta_a} \bar{l}(\eta)$ and $S_{1\psi}(\eta) = \text{col}_{1 \leq d \leq q} \frac{\partial}{\partial \psi_d} \bar{l}(\eta)$.

Theorem A.1. Under (1) and the regularity conditions RC,

$$MSE[\hat{\theta}(\hat{\psi})] \doteq g_1(\psi) + g_2(\psi) + g_3(\psi), \tag{4}$$

where the neglected terms are of order $o(m^{-1})$.

To prove Theorem A.1 use the following lemma. For a proof of the lemma, see Srivastava and Tiwari (1976).

Lemma A.1. Let $U \sim N(0, \Sigma)$. Then for symmetric matrices A, B and C ,

$$\begin{aligned} & E[(U'AU)(U'BU)(U'CU)] \\ &= 8tr(A\Sigma B\Sigma C\Sigma) + 2\{tr(A\Sigma B\Sigma)tr(C\Sigma) + tr(A\Sigma C\Sigma)tr(B\Sigma) \\ &+ tr(B\Sigma C\Sigma)tr(A\Sigma)\} + tr(A\Sigma)tr(B\Sigma)tr(C\Sigma). \end{aligned}$$

Proof of Theorem A.1. Kackar and Harville (1984) showed that $MSE[\hat{\theta}(\hat{\psi})] = MSE[\hat{\theta}(\psi)] + E[\{\hat{\theta}(\hat{\psi}) - \hat{\theta}(\psi)\}^2]$. It is straightforward to show that $MSE[\hat{\theta}(\psi)] = g_1(\psi) + g_2(\psi)$. Now $\hat{\psi} - \tilde{\psi} = O_p(m^{-1})$ implies that $E[\{\hat{\theta}(\hat{\psi}) - \hat{\theta}(\psi)\}^2] \doteq E[\{\hat{\theta}(\tilde{\psi}) - \hat{\theta}(\psi)\}^2]$. Thus it is enough to show that $E[\{\hat{\theta}(\tilde{\psi}) - \hat{\theta}(\psi)\}^2] \doteq g_3(\psi)$. Writing $\nabla\hat{\theta}(\psi) = [\frac{\partial}{\partial\psi_1}\hat{\theta}(\psi), \dots, \frac{\partial}{\partial\psi_q}\hat{\theta}(\psi)]'$, and using $\tilde{\psi} - \psi = O_p(m^{-1/2})$, $\frac{\partial^2}{\partial\psi_d\partial\psi_e}\hat{\theta}(\psi)|_{\psi=\tilde{\psi}^*} = O_p(1)$ when $\|\tilde{\psi}^* - \psi\| \leq \|\tilde{\psi} - \psi\|$, by a Taylor expansion of $\hat{\theta}(\tilde{\psi})$ around ψ , we get

$$\hat{\theta}(\tilde{\psi}) - \hat{\theta}(\psi) = (\tilde{\psi} - \psi)' \nabla\hat{\theta}(\psi) + O_p(m^{-1}). \tag{A.1}$$

Now using $\frac{\partial}{\partial\psi_d}\hat{\theta}(\psi) = \sum_{a=1}^p \frac{\partial}{\partial\beta_a}\hat{\theta}^*(\beta, \psi)|_{\beta=\hat{\beta}(\psi)} \frac{\partial}{\partial\psi_d}\hat{\beta}_a(\psi) + \frac{\partial}{\partial\psi_d}\hat{\theta}^*(\beta, \psi)|_{\beta=\hat{\beta}(\psi)}$, where $\hat{\theta}^*(\beta, \psi) = h'\beta + s'(\psi)(Y - X\beta)$, condition (f), and the fact that $\frac{\partial}{\partial\psi_d}\hat{\beta}_a(\psi) = O_p(m^{-1/2})$ (see Cox and Reid (1987)), we get from (A.1) that

$$\hat{\theta}(\tilde{\psi}) - \hat{\theta}(\psi) = (\tilde{\psi} - \psi)' \nabla\hat{\theta}^*(\psi) + O_p(m^{-1}), \tag{A.2}$$

where $\nabla\hat{\theta}^*(\psi) = [\frac{\partial}{\partial\psi_1}\hat{\theta}^*(\beta, \psi)|_{\beta=\hat{\beta}(\psi)}, \dots, \frac{\partial}{\partial\psi_q}\hat{\theta}^*(\beta, \psi)|_{\beta=\hat{\beta}(\psi)}]'$ $= L(\psi)(Y - X\hat{\beta}(\psi))$.

Using conditions (c), (d), the fact that $\hat{\beta}(\psi) - \beta = O_p(m^{-1/2})$ and (A.2), we get

$$E[\{\hat{\theta}(\tilde{\psi}) - \hat{\theta}(\psi)\}^2] \doteq E[\{(\tilde{\psi} - \psi)' L(\psi)(Y - X\beta)\}^2]. \tag{A.3}$$

Using the Taylor series expansion of the likelihood equation $S_1(\hat{\eta}) = 0$ and the orthogonality of β and ψ , it follows that

$$\begin{aligned} \tilde{\psi} - \psi &= I_{\psi}^{-1}(\psi)S_{1\psi}(\hat{\eta}) + O_p(m^{-1}) \\ &= \frac{1}{2}m^{-1}I_{\psi}^{-1}(\psi)col_{1 \leq d \leq q}[-tr\{\Sigma^{-1}(\psi)\frac{\partial}{\partial\psi_d}\Sigma(\psi)\} + u'A_d(\psi)u] + O_p(m^{-1}), \end{aligned} \tag{A.4}$$

where $u = Y - X\beta$ and $A_d(\psi) = \Sigma^{-1}(\psi)\{\frac{\partial}{\partial\psi_d}\Sigma(\psi)\}\Sigma^{-1}(\psi)$. Using (A.3) and (A.4), we get

$$E[\{\hat{\theta}(\tilde{\psi}) - \hat{\theta}(\psi)\}^2] \doteq \frac{1}{4m^2}E\left[\sum_{d=1}^q\sum_{e=1}^q\{u'A_d(\psi)u - \nu_d\}\{u'A_e(\psi)u - \nu_e\}u'\{C_e(\psi)C'_d(\psi)\}u\right], \quad (\text{A.5})$$

where $C(\psi) = \text{col}_{1 \leq d \leq q} C'_d(\psi) = I_\psi^{-1}(\psi)L(\psi)$ and $\nu_d(\psi) = E[u'A_d(\psi)u]$, $d = 1, \dots, q$.

Note that $u'C_e(\psi)C'_d(\psi)u = u'O_{de}(\psi)u$, where $O_{de}(\psi) = \frac{1}{2}[C_d(\psi)C'_e(\psi) + C_e(\psi)C'_d(\psi)]$ is symmetric, and for any arbitrary matrix M ,

$$\text{tr}[O_{de}M] = \frac{1}{2}[C'_e(\psi)MC_d(\psi) + C'_d(\psi)MC_e(\psi)].$$

Now applying Lemma A.1, we get from (A.5) that

$$E[\{\hat{\theta}(\tilde{\psi}) - \hat{\theta}(\psi)\}^2] \doteq \frac{1}{4}m^{-2}\text{tr}[\text{Var}\{\text{col}_{1 \leq d \leq q}u'A_d(\psi)u\}\text{Var}\{C(\psi)u\}]. \quad (\text{A.6})$$

Thus $g_3(\psi) = \text{tr}[L(\psi)\Sigma(\psi)L'(\psi)\text{Var}(\tilde{\psi})]$.

Remark A.1. It follows from (A.4) and $C(\psi) = I_\psi^{-1}(\psi)L(\psi)$ that $g_3(\psi) = \text{tr}[\text{Var}(I_\psi(\psi)\tilde{\psi})\text{Var}\{I_\psi^{-1}(\psi)L(\psi)u\}] = \text{tr}[\text{Var}(\tilde{\psi})\text{Var}(L(\psi)u)]$.

Remark A.2. Let $F(\psi)$ denote the first term in the right hand side of (A.4). Define $W(\psi) = [\{F(\psi)\}'\nabla\hat{\theta}^*(\psi)]^2$. Then, $[\hat{\theta}(\tilde{\psi}) - \hat{\theta}(\psi)]^2 = W(\psi) + o_p(m^{-1})$. We interpret \doteq in (A.6) in the following sense: the expression in the right hand side of (A.6) is based on the limiting distribution of $[\hat{\theta}(\tilde{\psi}) - \hat{\theta}(\psi)]^2$ up to $o(m^{-1})$, and is calculated based on the distribution of $W(\psi)$. This interpretation is an adaptation of Ghosh (1994, p.8 and p.19). In his monograph, Ghosh (1994) obtained higher order asymptotic expansion of mean, variance and risk function of an ML estimator. These expansions were not obtained as the limits of the moments of the ML estimator, rather they were calculated using the limiting distribution (say, Edgeworth expansion) of the ML estimator. For more discussion, see Sections 2.1 and 2.7 of the monograph.

Theorem A.2. Under (1) and regularity conditions RC, we have

- (a) $E[g_1(\hat{\psi}) - b'_{\hat{\psi}}(\hat{\psi})\nabla g_1(\hat{\psi}) + g_3(\hat{\psi})] \doteq g_1(\psi)$,
- (b) $E[g_2(\hat{\psi})] \doteq g_2(\psi)$,
- (c) $E[g_3(\hat{\psi})] \doteq g_3(\psi)$,

where neglected terms are of order $o(m^{-1})$.

Proof. (a) Since $G(\psi)$ and $\Sigma(\psi)$ are linear functions of ψ , we get (A.8) from (A.7).

$$L_d(\psi) = \frac{\partial}{\partial \psi_d} s(\psi) = -\Sigma^{-1}(\psi) \frac{\partial \Sigma(\psi)}{\partial \psi_d} \Sigma^{-1}(\psi) ZG(\psi)\lambda + \Sigma^{-1}(\psi) Z \frac{\partial G(\psi)}{\partial \psi_d} \lambda, \tag{A.7}$$

$$\begin{aligned} \frac{\partial^2 s(\psi)}{\partial \psi_e \partial \psi_d} &= \Sigma^{-1}(\psi) \frac{\partial \Sigma(\psi)}{\partial \psi_e} \Sigma^{-1}(\psi) \frac{\partial \Sigma(\psi)}{\partial \psi_d} \Sigma^{-1}(\psi) ZG(\psi)\lambda \\ &\quad + \Sigma^{-1}(\psi) \frac{\partial \Sigma(\psi)}{\partial \psi_d} \Sigma^{-1}(\psi) \frac{\partial \Sigma(\psi)}{\partial \psi_e} \Sigma^{-1}(\psi) ZG(\psi)\lambda \\ &\quad - \Sigma^{-1}(\psi) \frac{\partial \Sigma(\psi)}{\partial \psi_d} \Sigma^{-1}(\psi) Z \frac{\partial G(\psi)}{\partial \psi_e} \lambda \\ &\quad - \Sigma^{-1}(\psi) \frac{\partial \Sigma(\psi)}{\partial \psi_e} \Sigma^{-1}(\psi) Z \frac{\partial G(\psi)}{\partial \psi_d} \lambda. \end{aligned} \tag{A.8}$$

Using (6), (A.7) and (A.8), we get for $d, e = 1, \dots, q$,

$$\begin{aligned} \frac{\partial^2 g_1(\psi)}{\partial \psi_e \partial \psi_d} &= -2s'(\psi) \frac{\partial \Sigma(\psi)}{\partial \psi_d} \Sigma^{-1}(\psi) \frac{\partial \Sigma(\psi)}{\partial \psi_e} s(\psi) \\ &\quad + 2s'(\psi) \frac{\partial \Sigma(\psi)}{\partial \psi_d} \Sigma^{-1}(\psi) \frac{\partial}{\partial \psi_e} (ZG(\psi)\lambda) \\ &\quad + 2s'(\psi) \frac{\partial \Sigma(\psi)}{\partial \psi_e} \Sigma^{-1}(\psi) \frac{\partial}{\partial \psi_d} (ZG(\psi)\lambda) \\ &\quad - 2 \frac{\partial}{\partial \psi_d} (ZG(\psi)\lambda)' \Sigma^{-1}(\psi) \frac{\partial}{\partial \psi_e} (ZG(\psi)\lambda). \end{aligned} \tag{A.9}$$

Using the definition of $s(\psi)$ it follows that $L_d(\psi) = -\Sigma^{-1}(\psi) \frac{\partial \Sigma(\psi)}{\partial \psi_d} s(\psi) + \Sigma^{-1}(\psi) \frac{\partial}{\partial \psi_d} [ZG(\psi)\lambda]$. Then with (A.9), one can show that

$$\text{Cov}(L'_d(\psi)Y, L'_e(\psi)Y) = -\frac{1}{2} \frac{\partial^2 g_1(\psi)}{\partial \psi_e \partial \psi_d}. \tag{A.10}$$

Writing $H_{g_1}(\psi) = ((\frac{\partial^2 g_1(\psi)}{\partial \psi_d \partial \psi_e}))$, the Hessian matrix of g_1 , from (A.10) and Remark A.1,

$$g_3(\psi) = -\frac{1}{2} \text{tr}[H_{g_1}(\psi)\text{Var}(\hat{\psi})]. \tag{A.11}$$

Using $\hat{\psi} - \psi = O_p(m^{-1/2})$ and $\frac{\partial^3}{\partial \psi_d \partial \psi_e \partial \psi_f} g_1(\psi)|_{\psi=\hat{\psi}^*} = O_p(1)$, when $\|\hat{\psi}^* - \psi\| \leq \|\hat{\psi} - \psi\|$ the Taylor series expansion of $g_1(\hat{\psi})$ around ψ gives

$$g_1(\hat{\psi}) \doteq g_1(\psi) + (\hat{\psi} - \psi)' \nabla g_1(\psi) + \frac{1}{2} \text{tr}[(\hat{\psi} - \psi)(\hat{\psi} - \psi)' H_{g_1}(\psi)] + o_p(m^{-1}) \tag{A.12}$$

To obtain (a) use (7), (A.11) and (A.12).

(b) Using $\hat{\psi} - \psi = O_p(m^{-1/2})$ and $\frac{\partial^2}{\partial \psi_d \partial \psi_e} g_2(\psi)|_{\psi=\hat{\psi}} = O_p(m^{-1})$, where $\|\hat{\psi}^* - \psi\| \leq \|\hat{\psi} - \psi\|$, we obtain (b) by an application of Taylor's series expansion of $g_2(\hat{\psi})$ around ψ .

Proof of (c) is similar to that of (b).

Theorem A.3. Under (1) and regularity conditions RC, we have

$$(a) E(\tilde{\psi} - \psi) \doteq \frac{1}{2} m^{-1} I_{\psi}^{-1}(\psi) \text{col}_{1 \leq d \leq q} \text{tr} [I_{\beta}^{-1}(\psi) \frac{\partial}{\partial \psi_d} I_{\beta}(\psi)],$$

$$(b) \hat{\psi}_{CP} \doteq \tilde{\psi} - \frac{1}{2} m^{-1} I_{\psi}^{-1}(\psi) \text{col}_{1 \leq d \leq q} [\text{tr} I_{\beta}^{-1}(\psi) \frac{\partial}{\partial \psi_d} I_{\beta}(\psi)],$$

$$(c) E(\hat{\psi}_{CP} - \psi) \doteq 0.$$

Proof. Part (a) follows from Peers and Iqbal (1985, p.554) after rewriting their notations in ours. For part (b) note that, when β and ψ are orthogonal, the average conditional profile log-likelihood of ψ in Cox and Reid (1987, Equation 10) is given by

$$\bar{l}_{CP}(\psi) = \bar{l}(\psi, \tilde{\beta}(\psi)) - \frac{1}{2m} \ln | I_{\beta}(\psi, \tilde{\beta}(\psi)) |, \quad (\text{A.13})$$

where $\tilde{\beta}(\psi)$ is the ML estimator of β for given ψ and $\bar{l}(\psi, \beta)$ is the average log-likelihood for $(\beta', \psi)'$. With the orthogonality of ψ and β one can show from (A.13) that

$$\hat{\psi}_{CP} \doteq \tilde{\psi} - \frac{1}{2} m^{-1} I_{\psi}^{-1}(\psi) \text{col}_{1 \leq d \leq q} [\text{tr} I_{\beta}^{-1}(\psi) \frac{\partial}{\partial \psi_d} I_{\beta}(\psi)],$$

where $\hat{\psi}_{CP}$ is the maximum CPL estimator of ψ . The details are omitted to save space. Part (c) follows from part (a) and part (b).

Remark A.3. It follows from Harville (1974, Eqn. 2) and Cox and Reid (1987, Eqn. 10) that for the normal mixed linear model given by (2), the residual likelihood for ψ is proportional to the CPL. It then follows from Theorem A.3 that the expression for the asymptotic bias for the REML estimator of ψ , $\hat{\psi}_{REML}$, is $b_{\hat{\psi}_{REML}}(\psi) \doteq 0$.

For recent results on the asymptotic distribution of the REML of variance components one may refer to the work of Jiang (1996).

References

- Battese, G. E., Harter, R. M. and Fuller, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc.* **80**, 28-36.
- Cochran, W. (1977). *Sampling Techniques*. 3rd edition. Wiley, New York.
- Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *J. Roy. Statist. Soc. Ser. B* **49**, 1-39 (with discussion).

- Cressie, N. (1992). REML estimation in empirical Bayes smoothing of census undercount. *Survey Methodology* **18**, 75-94.
- Dick, P. (1995). Modeling net undercoverage in the 1991 census. *Survey Methodology* **21**, 45-54.
- Fay, R. E. and Herriot, R. A. (1979). Estimates of income for small places: an application of James-Stein procedure to census data. *J. Amer. Statist. Assoc.* **74**, 269-277.
- Ghosh, J. K. (1994). *Higher Order Asymptotics*. Institute of Mathematical Statistics, Hayward, California.
- Ghosh, M. and Rao, J. N. K. (1994). Small area estimation: an appraisal. *Statist. Sci.* **9**, 55-93 (with discussion).
- Harville, D. A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika* **61**, 383-385.
- Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31**, 423-447.
- Holt, D., Smith, T. M. F. and Tomberlin, T. J. (1979). A model-based approach to estimation for small subgroups of a population. *J. Amer. Statist. Assoc.* **74**, 405-410.
- Huzurbazar, V. S. (1950). Probability distributions and orthogonal parameters. *Proc. Camb. Phil. Soc.* **46**, 281-284.
- Jiang, J. (1996). REML estimation: asymptotic behavior and related topics. *Ann. Statist.* **24**, 255-286.
- Lahiri, P. and Rao, J. N. K. (1995). Robust estimation of mean squared error of small area estimators. *J. Amer. Statist. Assoc.* **90**, 758-766.
- Kackar, R. N. and Harville, D. A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *J. Amer. Statist. Assoc.* **79**, 853-862.
- Patterson, H. D. and Thompson, R. (1971). Recovery of interblock information when block sizes are unequal. *Biometrika* **58**, 545-554.
- Peers, H. W. and Iqbal, M. (1985). Asymptotic expansions for confidence limits in the presence of nuisance parameters, with applications. *J. Roy. Statist. Soc. Ser. B* **47**, 547-554.
- Prasad, N. G. N. and Rao, J. N. K. (1990). The estimation of mean squared errors of small area estimators. *J. Amer. Statist. Assoc.* **85**, 163-171.
- Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. 2nd edition. Wiley, New York.
- Rao, J. N. K. (1986). Synthetic estimators, SPREE and the best model based predictors. *Proceedings of the Conference on Survey Methods in Agriculture*, 1-16., U.S. Dept. of Agriculture, Washington, D.C.
- Singh, A. C., Stukel, D. M. and Pfeiffermann, D. (1998). Bayesian versus frequentist measures of error in small area estimation. *J. Roy. Statist. Soc. Ser. B* **60**, 377-396.
- Srivastava, V. K. and Tiwari, R. (1976). Evaluation of expectations of products of stochastic matrices. *Scand. J. Statist.* **3**, 135-138.

Department of Statistics, University of Georgia, Athens, GA 30602-1952, U.S.A.

E-mail: gauri@stat.uga.edu

Department of Mathematics and Statistics, University of Nebraska-Lincoln, Lincoln, NE 68588-0323, U.S.A.

E-mail: plahiri@unlinfo.unl.edu

(Received September 1997; accepted May 1999)