JOURNAL OF MULTIVARIATE ANALYSIS 45, 217-233 (1993)

The Asymptotic Distribution of REML Estimators

NOEL CRESSIE AND SOUMENDRA NATH LAHIRI

Iowa State University

Communicated by the Editors

Restricted maximum likelihood (REML) estimation is a method employed to estimate variance-covariance parameters from data that follow a Gaussian linear model. In applications, it has either been conjectured or assumed that REML estimators are asymptotically Gaussian with zero mean and variance matrix equal to the inverse of the restricted information matrix. In this article, we give conditions under which the conjecture is true and apply our results to variance-components models. An important application of variance components is to census undercount; a simulation is carried out to verify REML's properties for a typical census undercount model.

1. INTRODUCTION

Consider the Gaussian general linear model,

$$\mathbf{Y} \sim N_n(X\boldsymbol{\beta}, \Sigma(\boldsymbol{\theta})), \tag{1.1}$$

where $\mathbf{Y}$ is an $n \times 1$ data vector $(Y_1, ..., Y_n)'$, X is an $n \times p$ matrix of explanatory variables, $\boldsymbol{\beta} \equiv (\beta_1, ..., \beta_p)'$ is a $p \times 1$ vector of unknown large-scale effects, and $\Sigma(\boldsymbol{\theta})$ is an $n \times n$ positive-definite variance matrix which is known up to a $k \times 1$ vector of small-scale effects $\boldsymbol{\theta} \equiv (\theta_1, ..., \theta_k)' \in \Theta$ open in $\mathbb{R}^k$. Define $\boldsymbol{\varepsilon} \equiv \mathbf{Y} - X\boldsymbol{\beta}$, the $n \times 1$ vector of errors.

For $\boldsymbol{\theta}$ known, estimation of $\boldsymbol{\beta}$ is straightforward. Assuming only that $E(\mathbf{Y}) = X\boldsymbol{\beta}$ and $\text{var}(\mathbf{Y}) = \Sigma(\boldsymbol{\theta})$,

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) \equiv (X'\Sigma(\boldsymbol{\theta})^{-1} X)^{-1} X'\Sigma(\boldsymbol{\theta})^{-1} \mathbf{Y} \tag{1.2}$$

is the best linear unbiased estimator of $\boldsymbol{\beta}$, in the sense that for any linear unbiased estimator $\bar{\boldsymbol{\beta}}$, $\text{var}(\bar{\boldsymbol{\beta}}) - \text{var}(\hat{\boldsymbol{\beta}})$ is nonnegative-definite [1]. More

Received March 18, 1992; revised August 5, 1992.

AMS 1980 subject classifications: primary 62F12, secondary 62F20.

Key words and phrases: agricultural field trials, animal breeding, asymptotic efficiency, asymptotic normality, census undercount, restricted maximum likelihood, variance-components models.

0047-259X/93 $5.00

realistically, $\boldsymbol{\theta}$ is unknown and has to be estimated; substitution of that estimator into (1.2) then yields an *estimated* generalized least squares estimator of $\boldsymbol{\beta}$.

Note that under model (1.1), the negative loglikelihood of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ is

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}) \equiv (n/2)\log(2\pi) + \tfrac{1}{2}\log(|\Sigma(\boldsymbol{\theta})|) + \tfrac{1}{2}(\mathbf{Y} - X\boldsymbol{\beta})' \Sigma(\boldsymbol{\theta})^{-1} (\mathbf{Y} - X\boldsymbol{\beta}), \tag{1.3}$$

where for any matrix A, $|A|$ denotes its determinant. Minimization of this function yields the maximum likelihood (m.l.) estimates $\hat{\boldsymbol{\beta}}_{ml}$ and $\hat{\boldsymbol{\theta}}_{ml}$. The estimating equations, obtained from the profile likelihood of $\boldsymbol{\theta}$, are

$$\operatorname{tr}\{\Sigma(\boldsymbol{\theta})^{-1} \Sigma_i(\boldsymbol{\theta})\} + \mathbf{Y}'\{\partial\Pi(\boldsymbol{\theta})/\partial\theta_i\}\mathbf{Y} = 0; \qquad i = 1, ..., k, \tag{1.4}$$

where $\Sigma_i(\boldsymbol{\theta}) \equiv \partial\Sigma(\boldsymbol{\theta})/\partial\theta_i$ and

$$\Pi(\boldsymbol{\theta}) \equiv \Sigma(\boldsymbol{\theta})^{-1} - \Sigma(\boldsymbol{\theta})^{-1} X(X'\Sigma(\boldsymbol{\theta})^{-1} X)^{-1} X'\Sigma(\boldsymbol{\theta})^{-1}. \tag{1.5}$$

Under appropriate regularity conditions [16, 9], $\hat{\boldsymbol{\theta}}_{ml}$ is asymptotically unbiased and

$$J(\boldsymbol{\theta})^{1/2}(\hat{\boldsymbol{\theta}}_{ml} - \boldsymbol{\theta}) \xrightarrow{d} N_k(\mathbf{0}, I), \tag{1.6}$$

where the (i, j)th element of the information matrix $J(\boldsymbol{\theta})$ is

$$(J(\boldsymbol{\theta}))_{ij} = \tfrac{1}{2}\operatorname{tr}\{\Sigma(\boldsymbol{\theta})^{-1} \Sigma_i(\boldsymbol{\theta})\, \Sigma(\boldsymbol{\theta})^{-1} \Sigma_j(\boldsymbol{\theta})\}. \tag{1.7}$$

Although they are asymptotically fully efficient, dissatisfaction with m.l. estimators has come from their finite-sample properties. Simulation studies by *inter alia* Swallow and Monahan [15], Mardia and Marshall [9], and Zimmerman [18, Section 2.4] all demonstrate that m.l. estimators can be badly biased when n is small. Depending on the context, this bias can have serious consequences, for example, in census undercount (e.g., [2]).

An alternative method of estimating $\boldsymbol{\theta}$ is considered in this article. Section 2 gives the definition and basic properties of restricted maximum likelihood (REML) estimation, due to Patterson and Thompson [11, 12]. Section 3, which contains the most important results of the article, gives general conditions under which the normalized REML estimator is asymptotically a zero-mean Gaussian random vector. Section 4 applies these results to variance-components models, and Section 5 compares, via simulation, exact and asymptotic distributions of the REML estimator in a census-undercount model. Section 6 contains proofs of the results given in Sections 3 and 4.

2. Restricted Maximum Likelihood

The method of restricted maximum likelihood (REML) estimation of variance-matrix parameters $\boldsymbol{\theta}$, developed originally by Patterson and Thompson [11, 12], applies maximum likelihood to error contrasts rather than to the data themselves. (Rao [13] calls this method MML, marginal maximum likelihood, in the context of estimation of variance components. Recently, some authors have also called it residual maximum likelihood, although they have retained the abbreviation REML.)

A linear combination $\boldsymbol{\gamma}'\mathbf{Y}$ is called an error contrast if $E(\boldsymbol{\gamma}'\mathbf{Y})=0$, for all $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. Let $\mathbf{U} \equiv \Gamma'\mathbf{Y}$ represent a vector of $n-p$ linearly independent error contrasts; i.e., the $(n-p)$ columns of Γ are linearly independent and $\Gamma'X=0$. Under the Gaussian assumption (1.1), $\mathbf{U} \sim N_{n-p}(\mathbf{0}, \Gamma'\Sigma(\boldsymbol{\theta})\,\Gamma)$, which does not depend on $\boldsymbol{\beta}$. Thus, when Γ does not depend on $\boldsymbol{\theta}$, the negative loglikelihood function based on $\mathbf{U}$ is

$$
\begin{aligned}
L_U(\boldsymbol{\theta}) = ((n-p)/2)\log(2\pi) + \tfrac{1}{2}\log(|\Gamma'\Sigma(\boldsymbol{\theta})\,\Gamma|) \\
+ \tfrac{1}{2}\mathbf{U}'(\Gamma'\Sigma(\boldsymbol{\theta})\,\Gamma)^{-1}\mathbf{U}.
\end{aligned} \tag{2.1}
$$

If another set of $(n-p)$ linearly independent contrasts were used to define $\mathbf{U}$, the new negative log likelihood function would differ from $L_U(\boldsymbol{\theta})$ only by an additive constant (Harville [5]). Harville also shows that for a Γ that satisfies $\Gamma\Gamma' = I - X(X'X)^{-1}X'$ (and $\Gamma'\Gamma = I$),

$$
\begin{aligned}
L_U(\boldsymbol{\theta}) = ((n-p)/2)\log(2\pi) - \tfrac{1}{2}\log(|X'X|) + \tfrac{1}{2}\log(|\Sigma(\boldsymbol{\theta})|) \\
+ \tfrac{1}{2}\log(|X'\Sigma(\boldsymbol{\theta})^{-1}X|) + \tfrac{1}{2}\mathbf{Y}'\Pi(\boldsymbol{\theta})\,\mathbf{Y},
\end{aligned} \tag{2.2}
$$

where $\Pi(\boldsymbol{\theta})$ is given by (1.5). A REML estimator of $\boldsymbol{\theta}$, denoted $\hat{\boldsymbol{\theta}}_{rl}$, is obtained by minimizing (2.2) with respect to $\boldsymbol{\theta}$. The distinction between REML and m.l. estimation becomes important when p is large relative to n.

The REML method was originally proposed to estimate *variance-component* parameters: Numerical algorithms [6] and robust adaptations [4] have been developed in this context, although distribution theory is lacking. REML can also be used to estimate *spatial-dependence* parameters: Kitanidis [8] and Zimmerman [18] give computational details for producing an iterative minimization of (2.2) in this case.

Harville [5] provides a Bayesian justification for REML by assuming a noninformative prior for $\boldsymbol{\beta}$, which is statistically independent of $\boldsymbol{\theta}$, and showing that the marginal posterior density of $\boldsymbol{\theta}$ is proportional to (2.2) multiplied by the prior for $\boldsymbol{\theta}$. When that prior is noninformative, REML estimates correspond to marginal MAP (maximum *a posteriori*) estimates. Thus, in the situation where noninformative prior distributions for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are independent, REML can be seen as a compromise between m.l. and

Bayes estimation with squared error loss: REML averages the full likelihood over $\boldsymbol{\beta}$ but then maximizes the resulting (restricted or marginal) likelihood over $\boldsymbol{\theta}$, whereas the Bayesian approach yields $\int \boldsymbol{\theta} \exp\{-L_U(\boldsymbol{\theta})\}\, d\boldsymbol{\theta}$. Although the Bayesian interpretation of REML helps to explain its properties, $\hat{\boldsymbol{\theta}}_{rl}$ also has the obvious frequentist interpretation of being an m.l. estimator based on restricted information.

Minimization of (2.2) with respect to $\boldsymbol{\theta}$ can proceed by any of the gradient algorithms (e.g., [6]). The estimating equations, obtained by differentiating (2.2) with respect to $\boldsymbol{\theta}$ and setting the result equal to zero, are easily seen to be

$$\operatorname{tr}\{\Pi(\boldsymbol{\theta})\,\Sigma_i(\boldsymbol{\theta})\} + \mathbf{Y}'\{\partial \Pi(\boldsymbol{\theta})/\partial\theta_i\}\,\mathbf{Y} = 0; \qquad i = 1, ..., k. \tag{2.3}$$

A comparison of (2.3) and (1.4) shows remarkable similarities between the REML and m.l. estimating equations. It is not difficult to establish that, $E[\mathbf{Y}'\{\partial\Pi(\boldsymbol{\theta})/\partial\theta_i\}\,\mathbf{Y}] = -\operatorname{tr}\{\Pi(\boldsymbol{\theta})\,\Sigma_i(\boldsymbol{\theta})\}$, and hence the m.l. estimating equations (1.4) are biased but the REML estimating equations are exactly unbiased. This goes some way towards explaining the REML estimator's superior bias properties in small samples.

3. Asymptotic Distribution of $\hat{\boldsymbol{\theta}}_{rl}$

In contrast to $\hat{\boldsymbol{\theta}}_{ml}$, asymptotic properties for the REML estimator $\hat{\boldsymbol{\theta}}_{rl}$ have received little attention in the literature. Mardia and Marshall [9] studied the consistency and asymptotic normality (CAN) of $\hat{\boldsymbol{\theta}}_{ml}$ under the general linear model (1.1) and also gave results for a spatial linear model. Miller [10] obtained like results for $\hat{\boldsymbol{\theta}}_{ml}$ assuming a variance-components structure for $\Sigma(\boldsymbol{\theta})$. However, nothing seems to be known about the CAN properties of $\hat{\boldsymbol{\theta}}_{rl}$. In this article, we show that, under appropriate regularity conditions, $\hat{\boldsymbol{\theta}}_{rl}$ is asymptotically normal with mean $\boldsymbol{\theta}$ and dispersion matrix $((E_{\boldsymbol{\theta}}(\partial^2 L_U(\boldsymbol{\theta})/\partial\theta_i\partial\theta_j)))^{-1}$. The main tool used in the proof is a general result of Sweeting [16] on CAN of m.l. estimators, adapted here for REML estimators.

For easy reference later on, we shall state a version of Sweeting's [16] result that is most useful for the present problem. Let $\mathscr{I}_n(\boldsymbol{\theta})$ denote the matrix of second-order partial derivatives of the negative log likelihood function $L_U(\cdot)$, defined by (2.1). For a matrix $M \equiv (\boldsymbol{\theta}_1^0, ..., \boldsymbol{\theta}_k^0)_{k\times k}$, $\boldsymbol{\theta}_i^0 \in \Theta$, write $\mathscr{I}_n(M)$ for the matrix with (i, j)th element $(\partial^2 L_U(\boldsymbol{\theta})/\partial\theta_i\,\partial\theta_j)|_{\boldsymbol{\theta}=\boldsymbol{\theta}_i^0}$, $1 \leqslant i, j \leqslant k$. For any $l \times m$ matrix B, let $\|B\| = \{\operatorname{tr}(B'B)\}^{1/2}$, $\|B\|_s = \sup\{\|B\mathbf{t}\|: \|\mathbf{t}\| = 1, \mathbf{t} \in \mathbb{R}^m\}$ and $(B)_{ij} =$ the (i, j)th element of B. Also write $((a_{ij}))$ for a matrix whose (i, j)th element is a_{ij}. Let $\xrightarrow[u]{}$ denote uniform convergence of nonrandom functions over compact subsets of Θ, and for

l-dimensional random vectors $\mathbf{V}_n$, $\mathbf{V}$, write $\mathbf{V}_n \underset{u}{\Rightarrow} \mathbf{V}$ if $E_{\boldsymbol\theta} f(\mathbf{V}_n) \underset{u}{\to} E_{\boldsymbol\theta} f(\mathbf{V})$ for all bounded uniformly continuous functions $f\colon \mathbb{R}^l \to \mathbb{R}$. Here, and in what is to follow, it is understood that all the limits are taken as $n \to \infty$. Note that, by the well-known Helly–Bray theorem, $\mathbf{V}_n \underset{u}{\Rightarrow} \mathbf{V}$ implies $\mathbf{V}_n$ converges in distribution to $\mathbf{V}$.

The following result can be easily deduced from Theorems 1 and 2 of Sweeting [16]:

THEOREM 3.1. *Assume that*

(C.1) $\Sigma(\boldsymbol\theta)$ *is twice continuously differentiable on* Θ,

(C.2) *there exist nonrandom,* $k \times k$ *matrices* $W(\boldsymbol\theta)$ *and* $\{B_n(\boldsymbol\theta)\colon n \geqslant 1\}$, *continuous in* $\boldsymbol\theta$, *such that* $\|B_n(\boldsymbol\theta)^{-1}\| \underset{u}{\to} 0$ *and* $B_n(\boldsymbol\theta)^{-1} \mathcal{I}_n(\boldsymbol\theta)(B_n(\boldsymbol\theta)^{-1})' \underset{u}{\Rightarrow} W(\boldsymbol\theta)$, *where* $W(\boldsymbol\theta)$ *is positive-definite (p.d.) for all* $\boldsymbol\theta \in \Theta$,

(C.3) *for all* $c > 0$, $\eta > 0$,

(i) $\sup\{\|B_n(\boldsymbol\theta)^{-1} B_n(\boldsymbol\theta^0) - I\|\colon \ \|(B_n(\boldsymbol\theta))'(\boldsymbol\theta - \boldsymbol\theta^0)\| \leqslant c; \ \ \boldsymbol\theta, \boldsymbol\theta^0 \in \Theta\} \underset{u}{\to} 0$;

(ii) *for* $M \equiv (\boldsymbol\theta_1^0, ..., \boldsymbol\theta_k^0)$, $P_{\boldsymbol\theta}(\sup\{\|B_n(\boldsymbol\theta)^{-1}(\mathcal{I}_n(M) - \mathcal{I}_n(\boldsymbol\theta))(B_n(\boldsymbol\theta)^{-1})'\|\colon \|(B_n(\boldsymbol\theta))'(\boldsymbol\theta - \boldsymbol\theta_i^0)\| < c, 1 \leqslant i \leqslant k\} > \eta) \underset{u}{\to} 0$. *Let* $\hat{\boldsymbol\theta}_n$ *be the* REML *estimator based on the first* n *observations. Then,*

$$(B_n(\boldsymbol\theta))'(\hat{\boldsymbol\theta}_n - \boldsymbol\theta) \underset{u}{\Rightarrow} N_k(\mathbf{0}, W(\boldsymbol\theta)^{-1}).$$

In applying Theorem 3.1, one needs to check the validity of assumptions (C.1), (C.2), and (C.3). Conditions (C.1) and (C.3)(ii) require smoothness of the dispersion matrix $\Sigma(\boldsymbol\theta)$ as a function of $\boldsymbol\theta$ and can be checked directly for a given model. Condition (C.3)(i) depends on the choice of the normalizing matrices $\{B_n(\boldsymbol\theta)\}_{n \geqslant 1}$; typically, such matrices can be chosen with enough smoothness to guarantee (C.3)(i). Consequently, we shall concentrate on the verification of (C.2) and, for the rest of this section, assume

ASSUMPTION (A.1). *Conditions* (C.1) *and* (C.3) *of Theorem* 3.1 *hold.*

This assumption is verified in Section 4 for variance-components models. Note that if $\{B_n(\boldsymbol\theta)\}_{n \geqslant 1}$ and $W(\boldsymbol\theta)$ satisfy the requirements of Theorem 3.1, then for any $k \times k$ nonsingular matrix $B(\boldsymbol\theta)$, continuous in $\boldsymbol\theta$, $\bar{B}_n(\boldsymbol\theta) \equiv B_n(\boldsymbol\theta) B(\boldsymbol\theta)^{-1}$ and $\bar{W}(\boldsymbol\theta) \equiv B(\boldsymbol\theta)\, W(\boldsymbol\theta)\, B(\boldsymbol\theta)'$ also satisfy conditions (C.2) and (C.3). Consequently one may choose $B_n(\boldsymbol\theta)$ suitably to simplify verification of these conditions.

Before we formulate Theorem 3.2, we need to state two technical assumptions that will be seen to be sufficient for condition (C.2). Define $C_n(\boldsymbol{\theta}) \equiv (B_n(\boldsymbol{\theta}))^{-2}$, $c_{ij}(\boldsymbol{\theta}) \equiv (C_n(\boldsymbol{\theta}))_{ij}$, and $\Sigma_{ij}(\boldsymbol{\theta}) \equiv \partial^2 \Sigma(\boldsymbol{\theta})/\partial\theta_i\, \partial\theta_j$, $1 \leq i, j \leq k$.

ASSUMPTION (A.2). *As* $n \to \infty$, $\sum_{i=1}^k \sum_{j=1}^k \sum_{l=1}^k \sum_{p=1}^k c_{il} c_{jp} \{\mathrm{tr}(\Pi(\boldsymbol{\theta})\, \Sigma_{lj}(\boldsymbol{\theta})\, \Pi(\boldsymbol{\theta})\, \Sigma_{pi}(\boldsymbol{\theta})) + \sum_{r=1}^n (\Sigma(\boldsymbol{\theta})^{1/2}\, \Pi(\boldsymbol{\theta})\, \Sigma_{lj}(\boldsymbol{\theta})\, \Pi(\boldsymbol{\theta})\, \Sigma(\boldsymbol{\theta})^{1/2})_{rr} \cdot (\Sigma(\boldsymbol{\theta})^{1/2}\, \Pi(\boldsymbol{\theta})\, \Sigma_{pi}(\boldsymbol{\theta})\, \Pi(\boldsymbol{\theta})\, \Sigma(\boldsymbol{\theta})^{1/2})_{rr}\} \xrightarrow[u]{} 0$.

ASSUMPTION (A.3). *As* $n \to \infty$, $\sum_{i=1}^k \sum_{j=1}^k \sum_{l=1}^k \sum_{p=1}^k c_{il} c_{jp} \{\mathrm{tr}(\Pi(\boldsymbol{\theta})\, \Sigma_l(\boldsymbol{\theta})\, \Pi(\boldsymbol{\theta})\, \Sigma_j(\boldsymbol{\theta})\, \Pi(\boldsymbol{\theta})\, \Sigma_p(\boldsymbol{\theta})\, \Pi(\boldsymbol{\theta})\, \Sigma_i(\boldsymbol{\theta})) + \sum_{r=1}^n (\Sigma(\boldsymbol{\theta})^{1/2}\, \Pi(\boldsymbol{\theta})\, \Sigma_l(\boldsymbol{\theta})\, \Pi(\boldsymbol{\theta})\, \Sigma_j(\boldsymbol{\theta})\, \Pi(\boldsymbol{\theta})\, \Sigma(\boldsymbol{\theta})^{1/2})_{rr} \cdot (\Sigma(\boldsymbol{\theta})^{1/2}\, \Pi(\boldsymbol{\theta})\, \Sigma_p(\boldsymbol{\theta})\, \Pi(\boldsymbol{\theta})\, \Sigma_i(\boldsymbol{\theta})\, \Pi(\boldsymbol{\theta})\, \Sigma(\boldsymbol{\theta})^{1/2})_{rr}\} \xrightarrow[u]{} 0$.

THEOREM 3.2. *Assume that there exist nonrandom,* p.d. *matrices* $W(\boldsymbol{\theta})$ *and* $\{B_n(\boldsymbol{\theta})\}_{n \geq 1}$, *continuous in* $\boldsymbol{\theta}$, *such that* $\|B_n(\boldsymbol{\theta})^{-1}\| \xrightarrow[u]{} 0$ *and* $B_n(\boldsymbol{\theta})^{-1} E_{\boldsymbol{\theta}} \mathscr{I}_n(\boldsymbol{\theta})\, B_n(\boldsymbol{\theta})^{-1} \xrightarrow[u]{} W(\boldsymbol{\theta})$. *Then, under Assumptions* (A.1), (A.2), *and* (A.3), $\hat{\boldsymbol{\theta}}_n \equiv \hat{\boldsymbol{\theta}}_{rl}$ *satisfies*

$$B_n(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \underset{u}{\Longrightarrow} N_k(\mathbf{0}, W(\boldsymbol{\theta})^{-1}).$$

COROLLARY 3.1. *Under the conditions of Theorem* 3.2,

$$\mathbf{V}_n \equiv [E_{\boldsymbol{\theta}} \mathscr{I}_n(\boldsymbol{\theta})]^{1/2}\, (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \underset{u}{\Longrightarrow} N_k(\mathbf{0}, I).$$

For constructing confidence sets, Corollary 3.1 is not very useful since the normalizing matrix $[E_{\boldsymbol{\theta}} \mathscr{I}_n(\boldsymbol{\theta})]^{1/2}$ depends on the unknown parameter $\boldsymbol{\theta}$. The next result allows for a data-dependent choice of the normalizing factor.

COROLLARY 3.2. *Define* $R_n(\boldsymbol{\theta}) \equiv [E_{\boldsymbol{\theta}} \mathscr{I}_n(\boldsymbol{\theta})]^{1/2}$. *Suppose that there exist nonrandom matrices* $\tilde{W}(\boldsymbol{\theta})$ *and* $\{B_n(\boldsymbol{\theta}): n \geq 1\}$, *continuous in* $\boldsymbol{\theta}$, *such that* $B_n(\boldsymbol{\theta})$ *is* p.d., $\tilde{W}(\boldsymbol{\theta})$ *is nonsingular and*

$$R_n(\boldsymbol{\theta})\, B_n(\boldsymbol{\theta})^{-1} \xrightarrow[u]{} \tilde{W}(\boldsymbol{\theta}). \tag{3.1}$$

Then, under Assumptions (A.1), (A.2), *and* (A.3),

$$R_n(\hat{\boldsymbol{\theta}}_n)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \underset{u}{\Longrightarrow} N_k(\mathbf{0}, I).$$

Verification of Assumptions (A.2) and (A.3) can be quite lengthy if one wants to use the natural choice of normalizing matrices, viz., $B_n(\boldsymbol{\theta}) = [E_{\boldsymbol{\theta}} \mathscr{I}_n(\boldsymbol{\theta})]^{1/2}$. However, the amount of computation can be significantly reduced under some special choices of $B_n(\boldsymbol{\theta})$. In Theorem 3.3 below we choose $B_n(\boldsymbol{\theta})$ suitably and formulate some easy-to-check sufficient conditions for Assumptions (A.2) and (A.3) to hold. Define the matrix $Q_n(\boldsymbol{\theta})$ by

$$(Q_n(\boldsymbol{\theta}))_{ij} \equiv \mathrm{tr}(\Pi(\boldsymbol{\theta})\, \Sigma_i(\boldsymbol{\theta})\, \Pi(\boldsymbol{\theta})\, \Sigma_j(\boldsymbol{\theta}))/(\|\Pi(\boldsymbol{\theta})\, \Sigma_i(\boldsymbol{\theta})\|\ \|\Pi(\boldsymbol{\theta})\, \Sigma_j(\boldsymbol{\theta})\|), \tag{3.2}$$

$1 \leqslant i, j \leqslant k$. Also, let $|\lambda_{1n}(\boldsymbol{\theta})| \leqslant \cdots \leqslant |\lambda_{nn}(\boldsymbol{\theta})|$ denote the absolute eigenvalues of $\Sigma(\boldsymbol{\theta})$, $|\lambda^i_{1n}(\boldsymbol{\theta})| \leqslant \cdots \leqslant |\lambda^i_{nn}(\boldsymbol{\theta})|$ denote those of $\Sigma_i(\boldsymbol{\theta})$, and $|\lambda^{ij}_{1n}(\boldsymbol{\theta})| \leqslant \cdots \leqslant |\lambda^{ij}_{nn}(\boldsymbol{\theta})|$ denote those of $\Sigma_{ij}(\boldsymbol{\theta})$, $1 \leqslant i, j \leqslant k$. For convenience, we usually suppress dependence on $\boldsymbol{\theta}$ and n unless clarity demands it.

THEOREM 3.3. *Assume that Assumption* (A.1) *holds and that there exists a* p.d. *matrix* $W(\boldsymbol{\theta})$, *continuous in* $\boldsymbol{\theta}$, *such that* $Q_n(\boldsymbol{\theta}) \underset{u}{\to} W(\boldsymbol{\theta})$, *where* $Q_n(\cdot)$ *is defined by* (3.2). *Assume further that* $\sum_{i=1}^k \|\Pi(\boldsymbol{\theta})\,\Sigma_i(\boldsymbol{\theta})\|^{-1} \underset{u}{\to} 0$. *If there exists a sequence* $\{r_n\}_{n \geqslant 1}$ *with* $1 \leqslant r_n < n - p$ *for all* $n \geqslant 1$, *such that*

$$(\lambda_n/\lambda_1)^4 \{n/(n-r_n-p)^2\} \left\{ \sum_{i=1}^k (\lambda^i_n/\lambda^i_{r_n})^2 \right\}^2 \underset{u}{\to} 0, \tag{3.3}$$

$$(\lambda_n^2/\lambda_1)^2 \{n/(n-r_n-p)^2\}^2 \left\{ \sum_{i=1}^k \sum_{j=1}^k (\lambda^{ij}_n)^2 (\lambda^i_{r_n}\lambda^j_{r_n})^{-2} \right\} \underset{u}{\to} 0, \tag{3.4}$$

then $\hat{\boldsymbol{\theta}}_n \equiv \hat{\boldsymbol{\theta}}_{rl}$ *satisfies*

$$[E_{\boldsymbol{\theta}}(\mathcal{I}_n(\boldsymbol{\theta}))]^{1/2} (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \underset{u}{\Rightarrow} N(\mathbf{0}, I).$$

Remark 3.2. Each of the following conditions implies (3.3) and (3.4):

(S.1) There exist $0 < \eta < 1$ and a sequence of integers $\{r_n\}$ such that for all $1 \leqslant i, j \leqslant k$, (a) $\limsup_{n \to \infty} n^{-1} r_n < 1 - \eta$, (b) $\limsup_{n \to \infty} |\lambda^i_n|\,|\lambda^i_{r_n}|^{-1} < \eta^{-1}$, (c) $\lambda_n^4 \lambda_1^{-4} = o(n)$, and (d) $\lambda_n^4 (\lambda^{ij}_n)^4 (\lambda^i_{r_n}\lambda^j_{r_n})^{-4} = o(n)$.

(S.2) There exists a sequence of integers $\{r_n\}$ such that $r_n = o(n)$, $\liminf_{n \to \infty} |\lambda^i_{r_n}| > 0$, $\liminf_{n \to \infty} \lambda_1 > 0$, and $\lambda_n^4(|\lambda^i_n|^4 + |\lambda^i_n|^2) = o(n)$, for all $1 \leqslant i, j \leqslant n$.

COROLLARY 3.3. *Let* $B_n(\boldsymbol{\theta}) \equiv \operatorname{diag}(\|\Pi(\boldsymbol{\theta})\,\Sigma_1(\boldsymbol{\theta})\|, \ldots, \|\Pi(\boldsymbol{\theta})\,\Sigma_k(\boldsymbol{\theta})\|)$ *and recall that* $R_n(\boldsymbol{\theta}) = [E_{\boldsymbol{\theta}} \mathcal{I}_n(\boldsymbol{\theta})]^{1/2}$. *Assume that* (3.1) *holds for some nonsingular matrix* $\tilde{W}(\boldsymbol{\theta})$, *continuous in* $\boldsymbol{\theta}$. *Then, under assumptions* (3.3) *and* (3.4),

$$R_n(\hat{\boldsymbol{\theta}}_n)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \underset{u}{\Rightarrow} N_k(\mathbf{0}, I).$$

4. VARIANCE-COMPONENTS MODELS

In this section, we specialize the original model (1.1) to

$$\Sigma(\boldsymbol{\theta}) = \sigma_0^2 A_0 + \sigma_1^2 A_1 + \cdots + \sigma_k^2 A_k, \tag{4.1}$$

where $\sigma_0^2 \equiv 1$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)' \equiv (\sigma_1^2, \ldots, \sigma_k^2)' \in (0, \infty) \times \cdots \times (0, \infty) \equiv \Theta$, $A_0, A_1, \ldots, A_k$ are known nonnegative-definite (n.n.d.) matrices, and A_1 is p.d. Asymptotic properties of $\hat{\boldsymbol\theta}_{ml}$ in variance-components models have been investigated by Miller [10]. Although $\hat{\boldsymbol\theta}_{rl}$ is preferred to $\hat{\boldsymbol\theta}_{ml}$ in many applications, asymptotic results for REML estimators have not been available. For the variance-components problem, Rao and Kleffe [14, p. 236] establish the equivalence of REML and a form of iterated MINQ. They go on to show that if the starting value of the iteration is consistent, then so is the iterated estimator (their Section 10.5). Further, they consider the efficiency of REML estimators *assuming* that the estimators are asymptotically normal (their Section 10.6). The following theorem gives sufficient conditions under which REML estimators of $\boldsymbol\theta = (\sigma_1^2, ..., \sigma_k^2)$ are asymptotically normal.

THEOREM 4.1. *Assume that the model* (1.1), (4.1) *holds, and that*

(1) *there exists a p.d. matrix* $W(\boldsymbol\theta)$, *continuous in* $\boldsymbol\theta$, *such that* $\operatorname{tr}(\Pi(\boldsymbol\theta) A_i \Pi(\boldsymbol\theta) A_j)/\|\Pi(\boldsymbol\theta) A_i\| \, \|\Pi(\boldsymbol\theta) A_j\| \underset{u}{\to} (W(\boldsymbol\theta))_{ij}$ *for all* $1 \leqslant i, j \leqslant k$, *and*

(2) *there exist* $\eta \in (0, 1)$ *and a sequence* $\{r_n\}$ *of positive integers such that for large* n, $r_n < (1-\eta)\, n$; $(\sum_{i=0}^k \alpha_{in}^8)(\sum_{i=1}^k \alpha_{ir_n}^{-8}) = o(n)$; *and* $\alpha_{1n} > \eta$, *where* $\alpha_{i1} \leqslant \cdots \leqslant \alpha_{in}$ *are the eigenvalues of* A_i, $0 \leqslant i \leqslant k$. *Then,* $\hat{\boldsymbol\theta}_n \equiv \hat{\boldsymbol\theta}_{rl}$ *satisfies*

$$[E_{\boldsymbol\theta} \mathscr{I}_n(\boldsymbol\theta)]^{1/2} (\hat{\boldsymbol\theta}_n - \boldsymbol\theta) \underset{u}{\Longrightarrow} N_k(\mathbf{0}, I).$$

COROLLARY 4.1. *Let* $B_n(\boldsymbol\theta) \equiv \operatorname{diag}(\|\Pi(\boldsymbol\theta) A_1\|, ..., \|\Pi(\boldsymbol\theta) A_k\|)$ *and recall that* $R_n(\boldsymbol\theta) \equiv [E_{\boldsymbol\theta} \mathscr{I}_n(\boldsymbol\theta)]^{1/2}$. *If* (3.1) *holds for some nonsingular matrix* $\tilde W(\boldsymbol\theta)$, *continuous in* $\boldsymbol\theta$, *and condition* (2) *of Theorem* 4.1 *holds, then*

$$R_n(\hat{\boldsymbol\theta}_n)(\hat{\boldsymbol\theta}_n - \boldsymbol\theta) \underset{u}{\Longrightarrow} N_k(\mathbf{0}, I).$$

Remark 4.1. There are no further conditions beyond (1) and (2) needed to prove Theorem 4.1. In particular, Assumption (A.1) of Section 3 is not in effect in this section. The growth condition $(\sum_{i=0}^k \alpha_{in}^8)(\sum_{i=1}^k \alpha_{ir_n}^{-8}) = o(n)$, on the eigenvalues of $A_0, A_1, ..., A_k$, is used for verifying conditions (C.2) and (C.3). In many applications (e.g., [7]), $\{\alpha_{in}\}_{n \geqslant 1}$, $0 \leqslant i \leqslant k$, are bounded and $\alpha_{ir_n} > \eta > 0$ for all $1 \leqslant i \leqslant k$, so that this condition automatically holds. However, there are some simple examples (e.g., the two-way random-effects model), where the component matrices $\{A_i\}$ have too many zero eigenvalues and condition (2) of Theorem 4.1 fails. In such cases, the more general result given by Theorem 3.2 may be applied directly.

5. Application of REML to Census Undercount in the U.S.A.

Although a census attempts to carry out a complete enumeration of the population, for various reasons the final tallies are inaccurate. Suppose the United States is divided into $i = 1, ..., n$ areas (e.g., states, including Washington, D.C.). In the ith area, let T_i be the true (unknown) count and C_i be the census count. Then the undercount (in percentage) is defined as, $U_i \equiv \{(T_i - C_i)/T_i\}$ 100. Adjustment of the census count C_i, by the adjustment factor $F_i \equiv T_i/C_i$, yields the true count, $T_i = F_i C_i$. Clearly, F_i and U_i are monotonic increasing functions of each other. In what is to follow, $\mathbf{F} \equiv (F_1, ..., F_n)'$ will be predicted based on further information obtained from a post-enumeration survey (PES). The PES samples several hundred thousand households from which capture–recapture estimates are computed to obtain "raw" adjustment factors $\mathbf{Y}$.

Assume, given $\mathbf{F}$, that $\mathbf{Y} \sim N_n(\mathbf{F}, \Delta)$, where the $n \times n$ variance matrix Δ is known from sampling considerations. The vector $\mathbf{F}$ is unknown and to be predicted from the data $\mathbf{Y}$. Assume further that $\mathbf{F} \sim N_n(X\boldsymbol{\beta}, \tau^2 D)$, where X is an $n \times p$ matrix of explanatory variables, $\boldsymbol{\beta}$ is a $p \times l$ vector of (unknown) coefficients of the linear model, τ^2 is an unknown variance parameter, and D is a known $n \times n$ variance matrix. Then the model (1.1) holds with $\boldsymbol{\theta} = \tau^2$ and $\Sigma(\tau^2) = \Delta + \tau^2 D$.

Based on 1980 PES data at the state level, Cressie [2] models $D = \mathrm{diag}(1/C_1, ..., 1/C_{51})$ and $X = [\mathbf{X}_0 \mathbf{X}_1 \mathbf{X}_2]$, where $\mathbf{X}_0$ is the 51×1 vector of 1's, $\mathbf{X}_1$ is the vector of percent minority, and $\mathbf{X}_2$ is the vector of percentage of people over 25 who have not graduated from high school. (These and other explanatory variables are discussed by Ericksen, Kadane, and Tukey [3].)

If τ^2 is known, then the best linear unbiased predictor (BLUP) of $\mathbf{F}$ is (e.g., [2]):

$$\hat{\mathbf{p}}(\mathbf{Y}; \tau^2) = \{I - \Delta \Pi(\tau^2)\} \mathbf{Y}, \tag{5.1}$$

where $\Pi(\tau^2)$ is given by (1.5). Typically, τ^2 is unknown, so that it has to be estimated from the data $\mathbf{Y}$. Two possibilities are maximum likelihood (m.l.) and restricted maximum likelihood (REML). Inference (asymptotic) on τ^2 follows from results presented in the previous sections. For example, if conditions (1) and (2) of Theorem 4.1 hold, then $\hat{\tau}^2_{rl}$ is CAN. In this case, $k = 1$, so that condition (1) holds trivially with $W(\boldsymbol{\theta}) \equiv 1$. Condition (2) says that, after removing from consideration areas with large populations, the maximum of census count ratios should grow slower than $n^{1/8}$. Since $n = 51$ is fixed, the condition cannot be verified, but it will be seen from simulation that Theorem 4.1 offers a reasonable approximation.

To check the asymptotic distribution theory of $\hat{\tau}^2_{ml}$ and $\hat{\tau}^2_{rl}$, a simulation was carried out using realistic parameter values (obtained from various estimation methods applied to the 1980 Census and PES): $\beta_0 = 1.0330$, $\beta_1 = 0.000712$, $\beta_2 = -0.000110$, $\tau^2 = 95.00$. The simulation of

$$\mathbf{Y} \sim N_n(X\boldsymbol{\beta}, \Delta + \tau^2 D) \tag{5.2}$$

was performed 500 times and, each time, the estimates $\hat{\tau}^2_{ml}$ and $\hat{\tau}^2_{rl}$ were

```
a  0 000000000000001155556667
   1 0001223566667889
   2 000112356677899
   3 00111223445555779999
   4 001111111222233344445555566667777888888899999
   5 000012223333333445556666667788888899999999
   6 00000111111122222222233333444455666777777788889999
   7 0001111111122222334444445555666677777888888999
   8 001112222222333333333344455566777777788889999
   9 000011112222222333334555677777788
  10 0000111111123333444444566777777888899
  11 000111222222344444456667899
  12 000111122222333333667777888899
  13 1223345556677999
  14 000122233444566679 9
  15 0000122233445589 99
  16 157899
  17 001122233589
  18 2568
  19 145
  20 7
  21 2
  22 88

b  0 00000001234777799
   1 0012334567789
   2 012225556888899
   3 1122344445568889
   4 00133444455555666777888888899
   5 0000112223333333444445566777788888999
   6 0001112222222233444444444566667777888899999 9
   7 000000111111222222222333344445556777788888899999
   8 000000011112223334555555666667777778889
   9 000111222222222333333444455555566689999999
  10 0000000001112233344455556677788889 9
  11 0001122233344455666677788888899
  12 000111111222333444455567789
  13 0001333345555555556788
  14 00011123444456677789
  15 0011122234456678 8
  16 0011122223355557999
  17 011235556
  18 00112566777779
  19 117
  20 013478
  21 123
  22 7
  23 6
  24
  25 02
```

FIG. 1. Stem-and-leaf plots of estimated variance component τ^2, based on 500 simulations of (5.2): (a) maximum likelihood; (b) restricted maximum likelihood [2].

computed using the Gauss–Newton iterative method. (Whenever a negative value was obtained, the estimate was set equal to zero.) The stem-and-leaf plots of the two estimates are presented in Figs. 1a and b. Note the larger number of zeros for $\hat{\tau}^2_{ml}$ (Fig. 1a).

The means ($\bar{X}$) and the standard deviations (S) of the distributions shown in Fig. 1 are, for $\hat{\tau}^2_{ml}$: $\bar{X}=83.56$, $S=45.65$, and, for $\hat{\tau}^2_{rl}$: $\bar{X}=94.27$, $S=49.17$. The means can be compared to the true value of $\tau^2=95.00$. The bias in $\hat{\tau}^2_{ml}$ is apparent, whereas $\hat{\tau}^2_{rl}$ has very little bias. In terms of standard deviations, $\hat{\tau}^2_{ml}$ appears to have an advantage over $\hat{\tau}^2_{rl}$. (In fact, in terms of mean squared error, this advantage dominates.)

A direct consequence of negative bias in the estimation of τ^2 (such as for m.l. estimation) is to shrink the data $\mathbf{Y}$ too far towards the model $X\hat{\boldsymbol{\beta}}$, something the undercount research staff at the U.S. Census Bureau are very wary of. Since a single set of predictors will be used for adjustment purposes, bias in estimation of τ^2 is of primary importance. Thus, $\hat{\tau}^2_{rl}$ may be preferred to $\hat{\tau}^2_{ml}$, although final mean-squared prediction errors may be larger.

Formulas for asymptotic standard errors can be compared to the sample standard deviations obtained from the simulation experiment. Upon substituting $\tau^2=95.00$ into (1.7) (which is $E\{\partial^2 L(\boldsymbol{\beta}, \tau^2)/\partial(\tau^2)^2\}$), we obtain $\{\mathrm{var}(\hat{\tau}^2_{ml})\}^{1/2} \simeq 48.73$, which should be compared to $S=45.65$. The discrepancy may be partly due to replacing negative variance estimates with zero in the simulation. A better result is obtained with REML estimation; substituting $\tau^2=95.00$ into $E\{\partial^2 L_U(\tau^2)/\partial(\tau^2)^2\}$ yields $\{\mathrm{var}(\hat{\tau}^2_{rl})\}^{1/2} \simeq 50.14$, which should be compared to $S=49.17$.

6. Proofs

In this section, we give the proofs of the results stated in Sections 3 and 4. Since Theorem 3.1 can be deduced directly from Theorems 1 and 2 of Sweeting [16], we omit its proof.

Proof of Theorem 3.2. In view of Theorem 3.1 and assumption (A.1), it is enough to verify condition (C.2) only. Define $W_n(\boldsymbol{\theta}) \equiv E_{\boldsymbol{\theta}} \mathscr{I}_n(\boldsymbol{\theta})$. Note that by Lemma 2.3 of Zimmerman [17],

$$\begin{aligned}(\mathscr{I}_n(\boldsymbol{\theta}))_{ij} = (L''_U(\boldsymbol{\theta}))_{ij} &= \mathrm{tr}\{\Pi(\boldsymbol{\theta})(\Sigma_{ij}(\boldsymbol{\theta}) - \Sigma_i(\boldsymbol{\theta})\,\Pi(\boldsymbol{\theta})\,\Sigma_j(\boldsymbol{\theta}))\}/2 \\ &\quad - \mathbf{Y}'\Pi(\boldsymbol{\theta})(\Sigma_{ij}(\boldsymbol{\theta}) - 2\Sigma_i(\boldsymbol{\theta})\,\Pi(\boldsymbol{\theta})\,\Sigma_j(\boldsymbol{\theta}))\,\Pi(\boldsymbol{\theta})\,\mathbf{Y}/2 \qquad (6.1)\end{aligned}$$

and

$$(W_n(\boldsymbol{\theta}))_{ij} = (E_{\boldsymbol{\theta}} L''_U(\boldsymbol{\theta}))_{ij} = \mathrm{tr}\{\Pi(\boldsymbol{\theta})\,\Sigma_i(\boldsymbol{\theta})\,\Pi(\boldsymbol{\theta})\,\Sigma_j(\boldsymbol{\theta})\}/2,$$

for all $1 \leqslant i, j \leqslant k$. Since $\Pi(\boldsymbol{\theta})\,\mathbf{Y} = \Pi(\boldsymbol{\theta})\,\boldsymbol{\varepsilon}$, one obtains (from (6.1))

$$B_n(\boldsymbol{\theta})^{-1}\,\mathcal{I}_n(\boldsymbol{\theta})\,B_n(\boldsymbol{\theta})^{-1} = B_{1n}(\boldsymbol{\theta}) + B_{2n}(\boldsymbol{\theta}) + B_{3n}(\boldsymbol{\theta}),$$

where

$$-2B_{1n}(\boldsymbol{\theta}) \equiv B_n(\boldsymbol{\theta})^{-1}((\boldsymbol{\varepsilon}'\Pi(\boldsymbol{\theta})\,\Sigma_{ij}(\boldsymbol{\theta})\,\Pi(\boldsymbol{\theta})\,\boldsymbol{\varepsilon} - \operatorname{tr}\{\Pi(\boldsymbol{\theta})\,\Sigma_{ij}(\boldsymbol{\theta})\}))\,B_n(\boldsymbol{\theta})^{-1}$$

$$B_{2n}(\boldsymbol{\theta}) \equiv B_n(\boldsymbol{\theta})^{-1}\,((\boldsymbol{\varepsilon}'\Pi(\boldsymbol{\theta})\,\Sigma_i(\boldsymbol{\theta})\,\Pi(\boldsymbol{\theta})\,\Sigma_j(\boldsymbol{\theta})\,\Pi(\boldsymbol{\theta})\,\boldsymbol{\varepsilon} - \operatorname{tr}\{\Pi(\boldsymbol{\theta})\,\Sigma_i(\boldsymbol{\theta})\,\Pi(\boldsymbol{\theta})\,\Sigma_j(\boldsymbol{\theta})\}))\,B_n(\boldsymbol{\theta})^{-1}$$

and

$$B_{3n}(\boldsymbol{\theta}) \equiv B_n(\boldsymbol{\theta})^{-1}\,W_n(\boldsymbol{\theta})\,B_n(\boldsymbol{\theta})^{-1}.$$

By assumption, $B_{3n}(\boldsymbol{\theta}) \underset{u}{\longrightarrow} W(\boldsymbol{\theta})$. Hence, by Lemma 4.2 of Sweeting [16] and Chebyshev's inequality, it is enough to show that, for $i = 1, 2$,

$$E_{\boldsymbol{\theta}}\,\|B_{in}(\boldsymbol{\theta})\|^2 \underset{u}{\longrightarrow} 0. \tag{6.2}$$

In the rest of the proof, where necessary, we suppress the dependence on $\boldsymbol{\theta}$ and write $\Pi(\boldsymbol{\theta}) = \Pi$, $c_{ij}(\boldsymbol{\theta}) = c_{ij}$, $\Sigma_i(\boldsymbol{\theta}) = \Sigma_i$, and so forth.

Define $J_n \equiv ((\boldsymbol{\varepsilon}'\Pi\Sigma_{ij}\Pi\boldsymbol{\varepsilon}))$ and $\tilde{Q}_n \equiv ((\operatorname{tr}\{\Pi\Sigma_{ij}\}))$. Since $\Pi\Sigma\Pi = \Pi$ and $E(J_n) = \tilde{Q}_n$,

$$\begin{aligned} 4E_{\boldsymbol{\theta}}\,\|B_{1n}(\boldsymbol{\theta})\|^2 &= E\,\|B_n^{-1}(J_n - \tilde{Q}_n)\,B_n^{-1}\|^2 \\ &= \sum_{i=1}^{k}\sum_{j=1}^{k} \operatorname{cov}((C_nJ_n)_{ij}, (C_nJ_n)_{ji}), \end{aligned}$$

which, after some algebra, is equal to the l.h.s. of assumption (A.2). Similarly, one can show that $E_{\boldsymbol{\theta}}\,\|B_{2n}(\boldsymbol{\theta})\|^2$ is equal to the l.h.s. of assumption (A.3). Hence, by (A.2), (A.3), and (6.2), condition (C.2) holds. This completes the proof of Theorem 3.2. ∎

Proof of Corollary 3.1. Let $\phi_{n,\boldsymbol{\theta}}(\mathbf{t}) \equiv E_{\boldsymbol{\theta}} \exp(i\mathbf{t}'B_n(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}))$ and $\phi(\mathbf{t}) \equiv \exp(-\|\mathbf{t}\|^2/2)$. Since $B_n(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \underset{u}{\Longrightarrow} N_k(\mathbf{0}, W(\boldsymbol{\theta})^{-1})$, for any $M > 0$,

$$\sup_{\|\mathbf{t}\| < M} |\phi_{n,\boldsymbol{\theta}}(W(\boldsymbol{\theta})^{1/2}\,\mathbf{t}) - \phi(\mathbf{t})| \underset{u}{\longrightarrow} 0. \tag{6.3}$$

Note that, by assumption, $B_n(\boldsymbol{\theta})^{-1}\,W_n(\boldsymbol{\theta})\,B_n(\boldsymbol{\theta})^{-1} \underset{u}{\longrightarrow} W(\boldsymbol{\theta})$. Hence,

$$\|W_n(\boldsymbol{\theta})^{1/2}\,B_n(\boldsymbol{\theta})^{-1}\,W(\boldsymbol{\theta})^{-1/2}\|^2 \underset{u}{\longrightarrow} \operatorname{tr}(I) = k. \tag{6.4}$$

For $\mathbf{t} \in \mathbb{R}^k$, define $\mathbf{t}_n(\boldsymbol{\theta}) \equiv W(\boldsymbol{\theta})^{-1/2} B_n(\boldsymbol{\theta})^{-1} W_n(\boldsymbol{\theta})^{1/2} \mathbf{t}$. Then, by (6.4),

$$\begin{aligned} | \, \|\mathbf{t}_n(\boldsymbol{\theta})\|^2 - \|\mathbf{t}\|^2 | &= |\mathbf{t}_n(\boldsymbol{\theta})' (I - W(\boldsymbol{\theta})^{1/2} B_n(\boldsymbol{\theta}) W_n(\boldsymbol{\theta})^{-1} B_n(\boldsymbol{\theta}) W(\boldsymbol{\theta})^{1/2}) \mathbf{t}_n(\boldsymbol{\theta})| \\ &\leqslant \|\mathbf{t}_n(\boldsymbol{\theta})\|^2 \, \|W(\boldsymbol{\theta})\| \, \|W(\boldsymbol{\theta})^{-1} \\ &\quad - B_n(\boldsymbol{\theta}) W_n(\boldsymbol{\theta})^{-1} B_n(\boldsymbol{\theta})\| \underset{u}{\to} 0. \end{aligned} \tag{6.5}$$

Hence, by (6.3) and (6.5), it follows that for any $\mathbf{t} \in \mathbb{R}^k$,

$$\begin{aligned} &|\phi_{n,\boldsymbol{\theta}}(B_n(\boldsymbol{\theta})^{-1} W_n(\boldsymbol{\theta})^{1/2} \mathbf{t}) - \phi(\mathbf{t})| \\ &\quad \leqslant |\phi_{n,\boldsymbol{\theta}}(W(\boldsymbol{\theta})^{1/2} \mathbf{t}_n) - \phi(\mathbf{t}_n)| + |\phi(\mathbf{t}_n) - \phi(\mathbf{t})| \underset{u}{\to} 0. \quad \blacksquare \end{aligned}$$

Proof of Corollary 3.2. By the assumption on $R_n(\boldsymbol{\theta})$, continuity of $\tilde{W}(\boldsymbol{\theta})$, (C.3)(i), and Theorem 3.2, we obtain $R_n(\hat{\boldsymbol{\theta}}_n) B_n(\hat{\boldsymbol{\theta}}_n)^{-1} \tilde{W}(\hat{\boldsymbol{\theta}}_n)^{-1} \underset{u}{\to} Y$, in probability $(P_{\boldsymbol{\theta}})$. Corollary 3.2 then follows from Theorem 3.2 and Slutsky's Theorem. $\blacksquare$

Proof of Theorem 3.3. Without loss of generality, assume (see [6]) that $\Gamma'\Gamma = I_{n-p}$. As in the proof of Theorem 3.2, where necessary we shall suppress the dependence on $\boldsymbol{\theta}$ and n. Fix $1 \leqslant i \leqslant k$. Note that there exists an orthogonal matrix O_i such that $\Sigma_i = O_i \operatorname{diag}(\lambda_1^i, ..., \lambda_n^i) O_i'$. Write $O_i'\Gamma \equiv [\boldsymbol{\xi}_1, ..., \boldsymbol{\xi}_{n-p}]_{n \times (n-p)}$. Then, there exists $\{\boldsymbol{\xi}_s : n-p < s \leqslant n\}$ such that $\{\boldsymbol{\xi}_1, ..., \boldsymbol{\xi}_n\}$ forms an orthogonal basis of $\mathbb{R}^n$. Write $\boldsymbol{\xi}_s \equiv (\xi_{s1}, ..., \xi_{sn})'$, $1 \leqslant s \leqslant n$. Hence, for any $1 \leqslant r < n-p$,

$$\begin{aligned} \operatorname{tr}\{\Gamma'\Sigma_i\Sigma_i\Gamma\} &= \operatorname{tr}\{\Gamma' O_i \operatorname{diag}(\lambda_1^i, ..., \lambda_n^i)^2 O_i'\Gamma\} = \sum_{v=1}^{n} \left(\sum_{s=1}^{n-p} \xi_{sv}^2 \right) (\lambda_v^i)^2 \\ &\geqslant (\lambda_r^i)^2 \left[\sum_{v=r}^{n} \sum_{s=1}^{n-p} \xi_{sv}^2 \right] \\ &= (\lambda_r^i)^2 \sum_{v=r}^{n} \left(\sum_{s=1}^{n} \xi_{sv}^2 - \sum_{s>n-p} \xi_{sv}^2 \right) \\ &\geqslant (\lambda_r^i)^2 \left[(n-r+1) - \sum_{s>n-p} \left(\sum_{v=1}^{n} \xi_{sv}^2 \right) \right] \\ &\geqslant (\lambda_r^i)^2 \, [(n-r-p)]. \end{aligned}$$

Next, note that the smallest eigenvalue of $(\Gamma'\Sigma\Gamma)^{-2}$ is $\geqslant \lambda_n^{-2}$. Hence, for any $1 \leqslant r < n-p$, and for all $1 \leqslant i \leqslant k$,

$$\|\Pi\Sigma_i\|^2 \geqslant \lambda_n^{-2} \operatorname{tr}\{\Sigma_i\Gamma\Gamma'\Sigma_i\} \geqslant \lambda_n^{-2}(\lambda_r^i)^2 \, (n-r-p). \tag{6.6}$$

Now choose $B_n(\boldsymbol{\theta}) = \text{diag}(\|\Pi\Sigma_1\|, ..., \|\Pi\Sigma_k\|)$. Then, $c_{ij} = 0$ if $i \neq j$ and $c_{ii} = \|\Pi\Sigma_i\|^{-2}$, for $1 \leqslant i, j \leqslant k$. Hence by (6.6), we have

$$E_{\boldsymbol{\theta}} \|B_{2n}(\boldsymbol{\theta})\|^2 \leqslant 2 \sum_{i=1}^{k} \sum_{j=1}^{k} c_{ii}c_{jj}(\text{tr}(\Pi\Sigma_i \Pi\Sigma_j \Pi\Sigma_j \Pi\Sigma_i)) = \sum_{i=1}^{k} \sum_{j=1}^{k} c_{ii}c_{jj}$$
$$\times \text{tr}\{(\Pi^{1/2}\Sigma_i \Pi^{1/2})(\Pi^{1/2}\Sigma_j \Pi\Sigma_j \Pi^{1/2})(\Pi^{1/2}\Sigma_i \Pi^{1/2})\}$$
$$\leqslant 2n(n-r-p)^{-2} (\lambda_n/\lambda_1)^4 \left\{\sum_{i=1}^{k} (\lambda_n^i/\lambda_r^i)^2\right\}^2. \tag{6.7}$$

By similar arguments,

$$4E_{\boldsymbol{\theta}} \|B_{1n}(\boldsymbol{\theta})\|^2 \leqslant 2n(n-r-p)^{-2} \lambda_n^4 \lambda_1^{-2} \left\{\sum_{j=1}^{k} \sum_{i=1}^{k} (\lambda_n^{ij})^2 (\lambda_r^i \lambda_r^j)^{-2}\right\}. \tag{6.8}$$

Theorem 3.3 now follows from (3.3), (3.4), (6.7), (6.8), and Theorem 3.2. ∎

Proof of Corollary 3.3. Similar to the proof of Corollary 3.2.

Proof of Theorem 4.1. By (4.1), $\Sigma(\boldsymbol{\theta})$ has partial derivatives of all orders on Θ and $\Sigma_i(\boldsymbol{\theta}) = A_i$, $\Sigma_{ij}(\boldsymbol{\theta}) = 0$, for all $1 \leqslant i, j \leqslant k, \boldsymbol{\theta} \in \Theta$, implying condition (C.1) of Theorem 3.1. Note that for A p.d. and B n.n.d., $A^{-1} - (A+B)^{-1}$ is n.n.d. Hence, for $1 \leqslant i \leqslant k$, $\Pi = \Gamma(\Gamma'\Sigma\Gamma)^{-1}\Gamma'$ implies

$$\|A_i^{1/2}\Pi A_i^{1/2}\|_s^2 = \sup\{\mathbf{t}'A_i^{1/2}\Pi A_i \Pi A_i^{1/2}\mathbf{t}: \|\mathbf{t}\| = 1\}$$
$$\leqslant \sigma_i^{-2} \sup\{\mathbf{t}'A_i^{1/2}\Pi\Sigma\Pi A_i^{1/2}\mathbf{t}: \|\mathbf{t}\| = 1\}$$
$$= \sigma_i^{-2} \sup\{\mathbf{t}'A_i^{1/2}\Pi A_i^{1/2}\mathbf{t}: \|\mathbf{t}\| = 1\} \leqslant \sigma_i^{-4}. \tag{6.9}$$

Since Π and $A_0, A_1, ..., A_k$ are n.n.d., it follows (cf. (6.7), (6.9)) that

$$\text{tr}\{\Pi A_i \Pi A_j \Pi A_j \Pi A_i\} \leqslant \|A_j^{1/2}\Pi A_j^{1/2}\|_s \, \text{tr}\{\Pi^{1/2}A_i \Pi A_j \Pi A_i \Pi^{1/2}\}$$
$$\leqslant \sigma_j^{-2} \|A_j^{1/2}\Pi A_j^{1/2}\|_s \, \text{tr}\{A_i \Pi A_i \Pi\} \leqslant n/(\sigma_j^4 \sigma_i^4). \tag{6.10}$$

Now use (6.9) and (6.10) to verify conditions (C.2) and (C.3) of Theorem 3.1.

Verification of (C.2). Let $B_n(\boldsymbol{\theta}) = \text{diag}(\|\Pi(\boldsymbol{\theta}) A_1\|, ..., \|\Pi(\boldsymbol{\theta}) A_k\|)$. Clearly, it is enough to show that (6.2) holds. Since $\lambda_n \leqslant \sum_{i=0}^{k} \sigma_i^2 \alpha_{in}$, by (6.10) and the inequalities leading to (6.7), we obtain

$$E_{\boldsymbol{\theta}} \|B_{2n}(\boldsymbol{\theta})\|^2 \leqslant 2n\lambda_n^4 \left(\sum_{i=1}^{k} \sigma_i^{-4} \alpha_{ir_n}^{-2}\right)^2 \Big/ (n - r_n - p)^2 \underset{u}{\rightarrow} 0.$$

Also, under (4.1), $B_{1n}(\boldsymbol{\theta}) \equiv 0$ for all $\boldsymbol{\theta}$. Therefore, condition (C.2) holds.

Verification of (C.3)(i). Note that condition (C.3)(i) holds if and only if for all $c>0$, $\sup\{|\,\|\Pi(\boldsymbol{\theta})\,A_i\|^{-2}\,\|\Pi(\mathbf{t})\,A_i\|^2-1|: \mathbf{t}\in\Theta_n(c)\}\underset{u}{\rightarrow}0$, where

$$\Theta_n(c)\equiv\left\{\mathbf{t}\equiv(t_1,\,....,\,t_k)'\in\Theta\colon \sum_{i=1}^{k}(t_i-\theta_i)^2\,\|\Pi(\boldsymbol{\theta})\,A_i\|^2<c^2\right\}. \quad (6.11)$$

Fix $1\leqslant i\leqslant k$ and $\mathbf{t}\in\Theta_n(c)$. Define $D_1\equiv\Gamma'\Sigma(\mathbf{t})\,\Gamma$ and $D_2\equiv\Gamma'\Sigma(\boldsymbol{\theta})\,\Gamma$. Let $K_1, K_2, \ldots$ denote positive constants, not depending on n. Then,

$$\begin{aligned}
&|\,\|\Pi(\mathbf{t})\,A_i\|^2-\|\Pi(\boldsymbol{\theta})\,A_i\|^2|\\
&=\left|\sum_{j=1}^{k}(\theta_j-t_j)\operatorname{tr}\{A_i\Gamma D_2^{-1}(\Gamma'A_j\Gamma D_1^{-1}+D_2^{-1}\Gamma'A_j\Gamma)\,D_1^{-1}\Gamma'A_i\}\right|\\
&\leqslant K_1\sum_{j=1}^{k}\|\Pi(\boldsymbol{\theta})\,A_j\|^{-1}\operatorname{tr}\{A_i(\Pi(\boldsymbol{\theta})\,A_j\Pi(\mathbf{t})^2\\
&\quad+\Pi(\boldsymbol{\theta})^2\,A_j\Pi(\mathbf{t}))\,A_i\}. \quad (6.12)
\end{aligned}$$

Note that, $\|\Pi(\boldsymbol{\theta})\|\leqslant(\theta_1\alpha_{1n})^{-1}$, for all $\boldsymbol{\theta}\in\Theta$. Hence, by (6.6), (6.9), (6.11), and (6.12), we obtain, uniformly in $\mathbf{t}\in\Theta_n(c)$,

$$\begin{aligned}
&\sum_{i=1}^{k}|\,\|\Pi(\mathbf{t})\,A_i\|^2\,\|\Pi(\boldsymbol{\theta})\,A_i\|^{-2}-1|\\
&\quad\leqslant K_2 n^{1/2}\sum_{i=1}^{k}(\|\Pi(\mathbf{t})\|_s^{3/2}\,\|\Pi(\mathbf{t})^{1/2}\,A_i^{1/2}\|_s\,\|A_i\|_s^{3/2}\\
&\quad\quad+\|\Pi(\mathbf{t})\|^{1/2}\,\|\Pi(\mathbf{t})^{1/2}\,A_i^{1/2}\|_s\,\|A_i\|_s^{3/2}\,\|\Pi(\boldsymbol{\theta})\|_s)(\|\Pi(\boldsymbol{\theta})\,A_i\|)^{-2}\\
&\quad\leqslant K_3 n^{1/2}(n-r_n-p)^{-1}\left(\sum_{i=1}^{k}\theta_i^{-1/2}\alpha_{in}^{3/2}\alpha_{ir_n}^{-2}\right)\\
&\quad\quad\times(\theta_1\alpha_{1n})^{-3/2}\left(\sum_{j=0}^{k}\theta_j\alpha_{jn}\right)^2\underset{u}{\rightarrow}0,
\end{aligned}$$

for every $\boldsymbol{\theta}\in\Theta$, where recall $\theta_0\equiv 1$. Hence, (C.3)(i) holds.

Verification of C.3(ii). Let $Z_1\equiv\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'-\Sigma(\boldsymbol{\theta})$, $U_1\equiv\Sigma(\boldsymbol{\theta})^{-1}\,\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'-I$, and $b_{ij}(\mathbf{t})=\operatorname{tr}\{Z_1(\Pi(\boldsymbol{\theta})\,A_i\Pi(\boldsymbol{\theta})\,A_j\Pi(\boldsymbol{\theta})-\Pi(\mathbf{t})\,A_i\Pi(\mathbf{t})\,A_j\Pi(\mathbf{t}))\}$, $\mathbf{t}\in\Theta$, $1\leqslant i, j\leqslant k$. Note that for $1\leqslant i, j\leqslant k$, the (i, j)th element of $[\mathscr{I}_n(M)-\mathscr{I}_n(\boldsymbol{\theta})]$ is

$$\begin{aligned}
&b_{ij}(\boldsymbol{\theta}_i^0)+\operatorname{tr}\{A_i(\Pi(\boldsymbol{\theta}_i^0)\,A_j\Pi(\boldsymbol{\theta}_i^0)-\Pi(\boldsymbol{\theta})\,A_j\Pi(\boldsymbol{\theta}))\}/2\\
&\quad+\operatorname{tr}\{\Sigma(\boldsymbol{\theta})(\Pi(\boldsymbol{\theta})\,A_i\Pi(\boldsymbol{\theta})\,A_j\Pi(\boldsymbol{\theta})\\
&\quad-\Pi(\boldsymbol{\theta}_i^0)\,A_i\Pi(\boldsymbol{\theta}_i^0)\,A_j\Pi(\boldsymbol{\theta}_i^0))\}. \quad (6.13)
\end{aligned}$$

First, we show that for all $c > 0$,

$$\sup\{\|B_n(\boldsymbol{\theta})^{-1}((b_{ij}(\boldsymbol{\theta}_i^0)))\, B_n(\boldsymbol{\theta})^{-1}\| : \boldsymbol{\theta}_i^0 \in \Theta_n(c),\ 1 \leqslant i \leqslant k\} \underset{n}{\longrightarrow} 0, \qquad (6.14)$$

in probability ($P_{\boldsymbol{\theta}}$). By Taylor's expansion,

$$b_{ij}(\mathbf{t}) = (\mathbf{t} - \boldsymbol{\theta})'\, \partial b_{ij}(\boldsymbol{\theta})/\partial \boldsymbol{\theta} + R_{ij}(\mathbf{t}), \qquad (6.15)$$

where, uniformly over $\mathbf{t} \in \Theta_n(c)$,

$$\begin{aligned} |R_{ij}(\mathbf{t})| &\leqslant K_4\, \|\mathbf{t} - \boldsymbol{\theta}\|^2\, \|U_1\| \\ &\quad \times \sum_{l=1}^{k} \sum_{m=1}^{k} \sup\{\|\Sigma(\mathbf{t})^{1/2}\, \Pi(\mathbf{t})\, A_{i_1} \Pi(\mathbf{t})\, A_{i_2} \Pi(\mathbf{t})\, A_{i_3} \Pi(\mathbf{t}) \\ &\quad \times A_{i_4} \Pi(\mathbf{t})\, \Sigma(\mathbf{t})^{1/2}\| : \{i_1, i_2, i_3, i_4\} \subseteq \{i, j, l, m\}\} \\ &\leqslant K_5\, \|\mathbf{t} - \boldsymbol{\theta}\|^2\, \|U_1\|\, n^{1/2} \max\{\theta_i^{-4} : 1 \leqslant i \leqslant k\}. \end{aligned}$$

As in the proof of $E_{\boldsymbol{\theta}} \|B_{2n}(\boldsymbol{\theta})\|^2 \underset{n}{\longrightarrow} 0$, above, one obtains,

$$E_{\boldsymbol{\theta}}\, \|U_1\|^2 = n(n+1);$$

$$E_{\boldsymbol{\theta}}([\partial b_{ij}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}]_l)^2\, [c_{ii}(\boldsymbol{\theta})\, c_{jj}(\boldsymbol{\theta})\, c_{ll}(\boldsymbol{\theta})]^{1/2} \underset{n}{\longrightarrow} 0. \qquad (6.16)$$

Now, using condition (2) and relations (6.11), (6.15), and (6.16), one can complete the proof of (6.14).

Next, applying the mean value theorem to the last two terms of (6.13), one can show that, uniformly in $\mathbf{t} \in \Theta_n(c)$,

$$\begin{aligned} &|\operatorname{tr}\{A_i(\Pi(\mathbf{t})\, A_j \Pi(\mathbf{t}) - \Pi(\boldsymbol{\theta})\, A_j \Pi(\boldsymbol{\theta}))\}| \\ &\qquad \leqslant K_6 n\, \|\mathbf{t} - \boldsymbol{\theta}\| \left(\sum_{l=1}^{k} (\theta_i \theta_j \theta_l)^{-1} \right); \\ &|\operatorname{tr}\{A_l(\Pi(\boldsymbol{\theta})\, A_i \Pi(\boldsymbol{\theta})\, A_j \Pi(\boldsymbol{\theta}) - \Pi(\mathbf{t})\, A_i \Pi(\mathbf{t})\, A_j \Pi(\mathbf{t}))\}| \\ &\qquad \leqslant K_7 n\, \|\mathbf{t} - \boldsymbol{\theta}\| \left(\sum_{m=1}^{k} (\theta_i \theta_j \theta_l \theta_m)^{-1} \right). \end{aligned}$$

Now (C.3)(ii) follows from (6.11), (6.13), and (6.14).

Acknowledgments

Cressie's research was partially supported by the National Science Foundation under Grant DMS-9001862 and by Joint Statistical Agreement JSA 90-41 between the U.S. Bureau of the Census and Iowa State University. The findings and opinions expressed in this article are of the authors alone and do not necessarily reflect those of the Census Bureau.

REFERENCES

[1] AITKEN, A. C. (1935). On least squares and linear combinations of observations. *Proc. Roy. Soc. Edinburgh Sect. A* **55** 42–48.

[2] CRESSIE, N. (1992). REML estimation in empirical Bays smoothing of census undercount. *Survey Methodol.* **18** 75–94.

[3] ERICKSON, E. P., KADANE, J. B., AND TUKEY, J. W. (1989). Adjusting the 1980 Census of population and housing. *J. Amer. Statist. Assoc.* **84** 927–944.

[4] FELLNER, W. H. (1986). Robust estimation of variance components. *Technometrics* **28** 51–60.

[5] HARVILLE, D. A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika* **61** 383–385.

[6] HARVILLE, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *J. Amer. Statist. Assoc.* **72** 320–340.

[7] HENDERSON, C. R. (1984). *Applications of Linear Models in Animal Breeding.* University of Guelph, Guelph, Ontario.

[8] KITANIDIS, P. K. (1983). Statistical estimation of polynomial generalized covariance functions and hydrologic applications. *Water Resources Res.* **19** 909–921.

[9] MARDIA, K. V., AND MARSHALL, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71** 135–146.

[10] MILLER, J. J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *Ann. Statist.* **5** 746–762.

[11] PATTERSON, H. D., AND THOMPSON, R. (1971). Recovery of interblock information when block sizes are unequal. *Biometrika* **58** 545–554.

[12] PATTERSON, H. D., AND THOMPSON, R. (1974). Maximum likelihood estimation of components of variance. In *Proceedings, 8th International Biometric Conference.* Biometric Society, Washington, DC.

[13] RAO, C. R. (1979). MINQE theory and its relation to ML and MML estimation of variance components. *Sankhyā B* **41** 138–153.

[14] RAO, C. R., AND KLEFFE, J. (1988). *Estimation of Variance Components and Applications.* North-Holland, Amsterdam.

[15] SWALLOW, W. H., AND MONAHAN, J. F. (1984). Monte Carlo comparison of ANOVA, MIVQUE, REML, and ML estimators of variance components. *Technometrics* **26** 47–57.

[16] SWEETING, T. J. (1980). Uniform asymptotic normality of the maximum likelihood estimator. *Ann. Statistics* **8** 1375–1381.

[17] ZIMMERMAN, D. L. (1986). *A Random Field Approach to Spatial Experiments.* Unpublished Ph.D. dissertation, Department of Statistics, Iowa State University.

[18] ZIMMERMAN, D. L. (1989). Computationally efficient restricted maximum likelihood estimation of generalized covariance functions. *Math. Geol.* **21** 655–672.