WALD CONSISTENCY AND THE METHOD OF SIEVES IN REML ESTIMATION

By Jiming Jiang

Case Western Reserve University

We prove that for all unconfounded balanced mixed models of the analysis of variance, estimates of variance components parameters that maximize the (restricted) Gaussian likelihood are consistent and asymptotically normal—and this is true whether normality is assumed or not. For a general (nonnormal) mixed model, we show estimates of the variance components parameters that maximize the (restricted) Gaussian likelihood over a sequence of approximating parameter spaces (i.e., a sieve) constitute a consistent sequence of roots of the REML equations and the sequence is also asymptotically normal. The results do not require the rank p of the design matrix of fixed effects to be bounded. An example shows that, in some unbalanced cases, estimates that maximize the Gaussian likelihood over the full parameter space can be inconsistent, given the condition that ensures consistency of the sieve estimates.

1. Introduction. In many cases exploration of asymptotic properties of maximum likelihood (ML) estimates led to one of two types of consistency: Cramér (1946) or Wald (1949) types. The former establishes the consistency of some root of the likelihood equation(s) but usually gives no indication on how to identify such a root when the roots of the likelihood equation(s) are not unique. The latter, however, states that an estimate of the parameter vector that maximizes the likelihood function is consistent. Therefore it is not unusual that Wald consistency sometimes fails whereas the Cramér one holds [e.g., Le Cam (1979)].

Often, obstacles to Wald consistency arise on the boundary of the parameter space Θ. For this reason, the maximization is sometimes carried out over a sequence of subspaces which may belong to the interior of Θ and approaches Θ as sample size increases. Methods of this type have been studied in the literature, among them those that Grenander (1981) called *sieves*.

In this paper we consider asymptotic behaviors of REML—restricted or residual maximum likelihood—estimates in variance components estimation. The REML method was first proposed by W. A. Thompson (1962). Several authors have given overviews on REML, which can be found in Harville (1977), Khuri and Sahai (1985), Robinson (1987) and Searle, Casella and McCulloch (1992). A general mixed model can be written as

$$(1) \qquad y = X\beta + Z_1\alpha_1 + \cdots + Z_s\alpha_s + \varepsilon,$$

where y is an $N \times 1$ vector of observations, X is an $N \times p$ known matrix of full rank p, β is a $p \times 1$ vector of unknown constants (the fixed effects), Z_i is an $N \times m_i$ known matrix, α_i is a $m_i \times 1$ vector of i.i.d. random variables with mean 0 and variance σ_i^2 (the random effects), $i = 1, \ldots, s$; ε is an $N \times 1$ vector of i.i.d. random variables with mean 0 and variance σ_0^2 (the errors).

Asymptotic results of maximum likelihood type estimates for model (1) are few in number, with or without normality assumptions. Assuming normality, and the model having a standard ANOVA structure, Miller (1977) proved a result of Cramér consistency and asymptotic normality for the maximum likelihood estimates (MLE) of both the fixed effects and the variance components $\sigma_0^2, \ldots, \sigma_s^2$. Under conditions slightly stronger than those of Miller, Das (1979) obtained a similar result for the REML estimates. Also under the normality assumption, Cressie and Lahiri (1993) gave conditions under which Cramér consistency and asymptotic normality of the REML estimates in a general mixed model would hold, using a result of Sweeting (1980). Without assuming normality, in which case the REML estimates are defined as solutions of the REML equations derived under normality, Richardson and Welsh (1994) proved Cramér consistency and asymptotic normality for hierarchical (nested) mixed models. A common feature of the above results is that the rank p of the design matrix X of the fixed effects was held fixed. A more important and interesting question related to the (possible) superiority of REML over straight ML in estimating the variance components is how the REML estimates behave asymptotically with $p \to \infty$. In such situations a well-known example showing the inconsistency of the MLE is due to Neyman and Scott (1948). The question was answered recently by Jiang (1996), in which Cramér consistency and asymptotic normality of the REML estimates were proved under the assumption that the model is asymptotically identifiable and infinitely informative. The results do not require boundedness of p, normality, or any structure (such as ANOVA or nested design, etc.) for the model. As a consequence, Cramér consistency was shown to hold for all unconfounded balanced mixed models of the analysis of variance. In contrast to the preceding results, the only consistency result in variance components estimation that might be considered as Wald type was Hartley and Rao (1967) for MLE. As was noted by Rao (1977), a corrected version of Hartley–Rao's proof would establish the Wald consistency of the MLE, under the assumption that the number of observations falling into any particular level of any random factor stayed below a universal constant. To our knowledge, there has not been discussion on the method of sieves in variance components estimation.

From both theoretical and practical points of view, consistency of Wald type (nonsieve or sieve) is more gratifying than that of Cramér type, although the former often requires stronger assumptions. Since in nonnormal cases [e.g., Richardson and Welsh (1994), Jiang (1996)] the REML estimates are defined as a solution of the (normal) REML equations, a natural question regarding Wald type consistency is whether the solution that maximizes the (restricted) Gaussian likelihood is consistent. The aim of this paper is to establish the following.

1. Wald consistency holds for all unconfounded balanced mixed models of the analysis of variance. That is, estimates of the variance components parameters that maximize the restricted Gaussian likelihood are consistent.
2. The method of sieves works for all mixed models provided the models are asymptotically identifiable and infinitely informative globally. That is, there is a clear way of constructing the sieve which will produce consistent estimates.
3. In both (1) and (2) the estimates are also asymptotically normal.

Again the results do not require normality or boundedness of p or structures for the model. The Wald consistency in the balanced case is also without the assumption that the true parameter vector is an interior point of the parameter space. We also give an example showing that in some unbalanced cases there is sieve Wald consistency but no Wald consistency.

The restricted Gaussian likelihood is given in Section 2. Wald consistency in the balanced case is proved in Section 3. In Section 4 we discuss the method of sieves for the general mixed model (1). An example and a counterexample are given in Section 5. In Section 6 we make some remarks about the techniques we use and open problems.

2. Definitions, basic assumptions, and a simple lemma. There are two parametrizations of the variance components: $\theta_0 = \lambda = \sigma_0^2$, $\theta_i = \mu_i = \sigma_i^2/\sigma_0^2$, $1 \leq i \leq s$ [Hartley and Rao (1967)] and $\phi_i = \sigma_i^2$, $0 \leq i \leq s$. There is a one-to-one correspondence between the two sets of parameters. In this paper, estimates of the two sets of parameters are equivalent in the sense that consistency for one set of parameters implies that for the other.

The parameter spaces are $\Theta = \{\theta: \lambda > 0, \ \mu_i \geq 0, \ 1 \leq i \leq s\}$ and $\Phi = \{\phi: \sigma_0^2 > 0, \ \sigma_i^2 \geq 0, \ 1 \leq i \leq s\}$. The true parameter vectors will be denoted by $\theta_{(0)} = (\theta_{0i})_{0 \leq i \leq s} = (\lambda_0 \ \mu_0')' = (\lambda_0 \ (\mu_{0i})_{1 \leq i \leq s}')'$ and $\phi_{(0)} = (\phi_{0i})_{0 \leq i \leq s}$.

Given the data y, the restricted (or residual) Gaussian log-likelihood is the Gaussian log-likelihood based on $z = A'y$, where A is a $N \times (N - p)$ matrix such that

$$(2) \qquad \mathrm{rank}(A) = N - p, \quad A'X = 0.$$

Thus the restricted Gaussian log-likelihoods for estimating the two sets of parameters θ_i, $0 \leq i \leq s$ and ϕ_i, $0 \leq i \leq s$ are given, respectively, by

$$(3) \qquad L_N(\theta) = c - \left(\frac{N-p}{2}\right) \log \lambda - \frac{1}{2} \log |V(A, \mu)| - \frac{1}{2\lambda} z' V(A, \mu)^{-1} z,$$

and

$$(4) \qquad \tilde{L}_N(\phi) = c - \tfrac{1}{2} \log |U(A, \phi)| - \tfrac{1}{2} z' U(A, \phi)^{-1} z,$$

where $c = -((N - p)/2) \log 2\pi$, $V(A, \mu) = G_0 + \sum_{i=1}^s \mu_i G_i$, $U(A, \phi) = \sum_{i=0}^s \phi_i G_i = \lambda V(A, \mu)$ with $G_0 = A'A$, $G_i = A'Z_i Z_i' A$, $1 \leq i \leq s$. The

maximizers $\hat{\theta} = (\hat{\theta}_i)_{0 \le i \le s}$ and $\hat{\phi} = (\hat{\phi}_i)_{0 \le i \le s}$ of (3) and (4) do not depend on the choice of A so long as (2) is satisfied. Therefore we may assume, w.l.o.g., that

$$(5) \qquad\qquad A'A = I_{N-p}.$$

Note that the relation between $\hat{\theta}_i$, $0 \le i \le s$ and $\hat{\phi}_i$, $0 \le i \le s$ is the same as that for the parameters. The dependence of $V(A, \mu)$ and $U(A, \phi)$ on A is not important in this paper; therefore we will abbreviate these by $V(\mu)$ and $U(\phi)$, respectively.

Also w.l.o.g., we can focus on a sequence of designs indexed by N, since consistency and asymptotic normality hold iff they hold for any sequence with N increasing strictly monotonically. Thus, for example, p and m_i, $1 \le i \le s$ can be considered as functions of N, and y, X, Z_i's and so on as depending on N [e.g., Jiang (1996)].

Let $m_0 = N$, $\alpha_0 = \varepsilon$. The following assumptions A1 and A2 are made for model (1).

A1.   For each N, α_0, $\alpha_1, \ldots, \alpha_s$ are mutually independent.

A2.   For $0 \le i \le s$, the common distribution of $\alpha_{i1}, \ldots, \alpha_{im_i}$ may depend on N. However, it is required that

$$(6) \qquad\qquad \lim_{x \to \infty} \sup_N \max_{0 \le i \le s} E\alpha_{i1}^4 1_{(|\alpha_{i1}| > x)} = 0.$$

The basic idea of proving Wald consistency (nonsieve or sieve) is simple. Consider, for example, the log-likelihood (3). The difference $L_N(\theta) - L_N(\theta_{(0)})$ can be decomposed as

$$(7) \qquad L_N(\theta) - L_N(\theta_{(0)}) = e_N(\theta, \theta_{(0)}) + d_N(\theta, \theta_{(0)}),$$

where

$$
\begin{aligned}
e_N(\theta, \theta_{(0)}) &= E_{\theta_{(0)}}(L_N(\theta) - L_N(\theta_{(0)})) \\
(8) \qquad &= -\frac{1}{2}\left\{ (N - p)\log\frac{\lambda}{\lambda_0} + \log\frac{|V(\mu)|}{|V(\mu_0)|} \right. \\
&\qquad\qquad \left. + \frac{\lambda_0}{\lambda}\operatorname{tr}(V(\mu)^{-1}V(\mu_0)) - (N - p) \right\},
\end{aligned}
$$

$$
\begin{aligned}
d_N(\theta, \theta_{(0)}) &= L_N(\theta) - L_N(\theta_{(0)}) - E_{\theta_{(0)}}(L_N(\theta) - L_N(\theta_{(0)})) \\
(9) \qquad &= -\frac{1}{2}\left\{ z'\left[\frac{1}{\lambda}V(\mu)^{-1} - \frac{1}{\lambda_0}V(\mu_0)^{-1}\right]z \right. \\
&\qquad\qquad \left. - E_{\theta_{(0)}}z'\left[\frac{1}{\lambda}V(\mu)^{-1} - \frac{1}{\lambda_0}V(\mu_0)^{-1}\right]z \right\}.
\end{aligned}
$$

Essentially, what we are going to show is that, with probability $\to$ 1 and uniformly for θ outside a small neighborhood $N_\delta(\theta_{(0)}) = \{\theta: \|\theta - \theta_{(0)}\| < \delta\}$, the second term on the RHS of (7) is negligible compared with the first term, which is negative. Therefore the LHS of (7) will be less than 0 for $\theta \notin N_\delta(\theta_{(0)})$, and the rest of the argument is the same as Jiang (1996).

Let A, B, A_i, $1 \leq i \leq n$ be matrices. Define $|A|$ = the determinant of A, $\|A\| = (\lambda_{\max}(A'A))^{1/2}$ ($\lambda_{\max}$ denotes the largest eigenvalue), $\|A\|_R = (\operatorname{tr}(A'A))^{1/2}$, $\operatorname{Cor}(A_1, \ldots, A_n) = (\operatorname{cor}(A_i, A_j))$ if $A_i \neq 0$, $1 \leq i \leq n$, and 0 otherwise, where $\operatorname{cor}(A, B) = \operatorname{tr}(A'B)/\|A\|_R\|B\|_R$ if $A, B \neq 0$; $\operatorname{diag}(A_i)$ to be the block-diagonal matrix with A_i as its ith diagonal block. Define $b(\mu) = (I_N \ \sqrt{\mu_1}Z_1 \ \cdots \ \sqrt{\mu_s}Z_s)'$, $B(\mu) = AV(\mu)^{-1}A'$ [note that $B(\mu)$ indeed does not depend on A], $B_0(\mu) = b(\mu)B(\mu)b(\mu)'$, $B_i(\mu) = b(\mu)B(\mu)Z_iZ_i'B(\mu)b(\mu)'$, $i = 1, \ldots, s$.

Let $p_i(N)$, $0 \leq i \leq s$, be sequences of positive numbers. Denote $U_i(\phi) = U(\phi)^{-1/2}G_iU(\phi)^{-1/2}$, $0 \leq i \leq s$, $V_0(\theta) = (1/\lambda)I_{N-p}$, $V_i(\theta) = V(\mu)^{-1/2}G_iV(\mu)^{-1/2}$, $1 \leq i \leq s$; $I_{ij}^{(N)}(\theta) = \operatorname{tr}(V_i(\theta)V_j(\theta))/p_i(N)p_j(N)$,

$$K_{ij}^{(N)}(\theta) = \frac{1}{p_i(N)p_j(N)} \sum_{l=1}^{N+m} (EW_{Nl}^4 - 3)B_i(\mu)_{ll}B_j(\mu)_{ll}/\lambda^{1_{(i=0)}+1_{(j=0)}},$$

$$i, j = 0, 1, \ldots, s,$$

where $m = m_1 + \cdots + m_s$, B_{ll} denotes the lth diagonal element of B; $W_{Nl} = \varepsilon_l/\sqrt{\lambda_0}$, $1 \leq l \leq N$,

$$W_{Nl} = \alpha_{i\,l-N-\sum_{k<i}m_k}/\sqrt{\lambda_0\mu_{0i}}, \quad N + \sum_{k<i}m_k + 1 \leq l \leq N + \sum_{k\leq i}m_k, \qquad 1 \leq i \leq s,$$

or $\mathscr{W}_N = (W_{Nl})_{1\leq l\leq N+m} = (\varepsilon'/\sqrt{\lambda_0}, \ \alpha_1'/\sqrt{\lambda_0\mu_{01}} \ \cdots \ \alpha_s'/\sqrt{\lambda_0\mu_{0s}})'$, where $\alpha_i/\sqrt{\lambda_0\mu_{0i}}$ is understood as $(0\cdots0)'$ $(\alpha_{ij}/\sqrt{\lambda_0\mu_{0i}} \equiv 0$, $1 \leq j \leq m_i)$ if $\mu_{0i} = 0$. Let $I_N(\theta) = (I_{ij}^{(N)}(\theta))$, $K_N(\theta) = (K_{ij}^{(N)}(\theta))$, $J_N(\theta) = 2I_N(\theta) + K_N(\theta)$.

We say model (1) has positive variance components if $\theta_{(0)}$ is an interior point of Θ and is nondegenerate if

$$(10) \qquad\qquad \inf_N \min_{0\leq i\leq s} \operatorname{var}(\alpha_{i1}^2) > 0.$$

A sequence of estimates $\{(\hat\theta_{N0}, \ldots, \hat\theta_{Ns})'\}$ is called asymptotically normal if there are sequences of numbers $p_i(N) \to \infty$, $0 \leq i \leq s$ and matrices $\{M_N(\theta_{(0)})\}$ with $\limsup(\|M_N^{-1}(\theta_{(0)})\| \vee \|M_N(\theta_{(0)})\|) < \infty$ such that

$$(11) \qquad M_N(\theta_{(0)})(p_0(N)(\hat\theta_{N0} - \theta_{00}), \ldots, p_s(N)(\hat\theta_{Ns} - \theta_{0s}))' \xrightarrow{\mathscr{L}} N(0, I_{s+1}).$$

Similarly we define the asymptotic normality of estimates for the ϕ's.

The following lemma will be used in proofs of both Theorem 3.1 and Theorem 4.1 in the next two sections.

LEMMA 2.1. *Let $\mathscr{X} = (X_1, \ldots, X_n)'$ be a vector of independent random variables such that $EX_i = 0$, $EX_i^4 < \infty$, $1 \leq i \leq n$, and $A = (a_{ij})$ be an $n \times n$ symmetric matrix. Then*

$$\mathrm{var}(\mathscr{X}'A\mathscr{X}) = \sum_{i=1}^n a_{ii}^2 \, \mathrm{var}(X_i^2) + 2 \sum_{i \neq j} a_{ij}^2 \, \mathrm{var}(X_i)\,\mathrm{var}(X_j)$$

$$\leq 2 \vee \max_{1 \leq i \leq n} \left\{ \frac{\mathrm{var}(X_i^2)}{(\mathrm{var}(X_i))^2} \right\} \sum_{i,j=1}^n a_{ij}^2 \, \mathrm{var}(X_i)\,\mathrm{var}(X_j),$$

where $\mathrm{var}(X_i^2)/(\mathrm{var}(X_i))^2$ is understood as 0 if $\mathrm{var}(X_i) = 0$.

PROOF. We have

$$(12) \qquad \mathscr{X}'A\mathscr{X} - E\mathscr{X}'A\mathscr{X} = \sum_{i=1}^n \left\{ a_{ii}^2 (X_i^2 - EX_i^2) + 2 \left(\sum_{j<i} a_{ij} X_j \right) X_i \right\},$$

where $\sum_{j<i}(\cdot) = 0$ if $i = 1$. The summands in (12) form a sequence of martingale differences with respect to the σ-fields $\mathscr{F}_i = \sigma(X_1, \ldots, X_i)$, $1 \leq i \leq n$. The result then follows easily. $\square$

3. The balanced case: Wald consistency.

A balanced r-factor mixed model of the analysis of variance can be expressed (after possible reparametrization) in the following way [e.g., Searle, Casella and McCulloch (1992), Rao and Kleffe (1988)]:

$$(13) \qquad\qquad y = X\beta + \sum_{i \in S} Z_i \alpha_i + \varepsilon,$$

where $X = \bigotimes_{q=1}^{r+1} 1_{n_q}^{d_q}$ with $d = (d_1, \ldots, d_{r+1}) \in S_{r+1} = \{0, 1\}^{r+1}$, $Z_i = \bigotimes_{q=1}^{r+1} 1_{n_q}^{i_q}$ with $i = (i_1, \ldots, i_{r+1}) \in S \subset S_{r+1}$, $1_n^0 = I_n$, $1_n^1 = 1_n$ with I_n and 1_n being the n-dimensional identity matrix and vector of unit elements, respectively. [For examples, see Jiang (1996).] We assume, as usual, that factor $r + 1$ corresponds to "repetition within cells." As a consequence, in (13) one must have $d_{r+1} = 1$ and $i_{r+1} = 1$, $i \in S$. Also we can assume, w.l.o.g., that $n_q \geq 2$, $1 \leq q \leq r$ (since if $n_q = 1$, factor q is not really a factor and the model not really an r-factor one). The model is called unconfounded if (1) the fixed effects are not confounded with the random effects and errors, that is, $\mathrm{rank}(X, Z_i) > p, \forall\, i$ and $X \neq I_N$; (2) the random effects and errors are not confounded; that is, $I_N, Z_i Z_i', i \in S$ are linearly independent [e.g., Miller (1977)]. Note that under the balanced model (13) we have the expressions $m_i = \prod_{i_q=0} n_q$, $i \in S$. This allows us to extend the definition of m_i to all $i \in S_{r+1}$. In particular, $p = m_d = \prod_{d_q=0} n_q$, and $N = m_{(00\cdots0)} = \prod_{q=1}^{r+1} n_q$. Let $\bar{S} = \{(00\cdots0)\} \cup S$, $\mu_i = \mu_{0i} = 1$ if $i = (00\cdots0) \in S_{r+1}$.

In the balanced case we have the following nice representations for the $e_N(\theta, \theta_{(0)})$ and $d_N(\theta, \theta_{(0)})$ in (7).

LEMMA 3.1. *In the balanced case,*

$$
e_N(\theta, \theta_{(0)}) = -\tfrac{1}{2} S_{N,2}(\theta), \tag{14}
$$

$$
d_N(\theta, \theta_{(0)}) = \tfrac{1}{2} \sum_{l \not\geq d} (r_l(\theta) - 1) \xi_l, \tag{15}
$$

where for $u, v \in S, u \geq v$ iff $u_q \geq v_q, 1 \leq q \leq r+1$, and $u \not\geq v$ iff u is not $\geq v$; $r_l(\theta) = \lambda_0 C_l(\mu_0)/\lambda C_l(\mu)$ with $C_l(\mu) = \sum_{i \in \bar{S}} \mu_i (N/m_i) 1_{(i \leq l)}$; $S_{N,2}(\theta) = \sum_{l \not\geq d} (r_l(\theta) - 1 - \log r_l(\theta)) p_l^2$ with $p_l = (\prod_{l_q=0}(n_q - 1))^{1/2}$; and $\xi_l, \ l \not\geq d$ are random variables whose definition will be given later on in the proof.

PROOF. By the proof of Lemma 7.3 in Jiang (1996), there is an orthogonal matrix T such that $T'Z_i Z_i' T = \mathrm{diag}(\lambda_{ij})$, $T'XX'T = \mathrm{diag}(\lambda_{dj})$, where $\{\lambda_{i1}, \ldots, \lambda_{iN}\} = \{\prod_{q=1}^{r+1} \lambda_{iqk_q}, \ 1 \leq k_q \leq n_q, \ 1 \leq q \leq r+1\}$, $\{\lambda_{d1}, \ldots, \lambda_{dN}\} = \{\prod_{q=1}^{r+1} \lambda_{dqk_q}, \ 1 \leq k_q \leq n_q, \ 1 \leq q \leq r+1\}$ with $\lambda_{iqw} = 1 - i_q + n_q i_q \delta_{w,1}, \ i \in S \cup \{d\}$ ($\delta_{x,y} = 1$ if $x = y$ and $= 0$ otherwise). Let $A'T = B = (b_1, \ldots, b_N)$, then since [by (5)] $AA' = P_{X^\perp} = I_N - X(X'X)^{-1}X' = I_N - (p/N)XX'$, $B'B = \mathrm{diag}(\gamma_j)$, with $\gamma_j = 1 - (p/N)\lambda_{dj}, \ 1 \leq j \leq N$. It is easy to see that $\prod_{q=1}^{r+1}(1 - i_q + n_q i_q l_q) = (N/m_i) 1_{(i \leq l)}, \ i \in S \cup \{d\}, \ l \in S_{r+1}$. Hence

$$
\prod_{q=1}^{r+1} \lambda_{iqk_q} = \frac{N}{m_i} 1_{(i \leq \delta_{k,1})}, \ i \in S; \qquad \prod_{q=1}^{r+1} \lambda_{dqk_q} = \frac{N}{p} 1_{(d \leq \delta_{k,1})},
$$

where $\delta_{k,1} = (\delta_{k_1,1}, \ldots, \delta_{k_{r+1},1})$ for $k = (k_1, \ldots, k_{r+1})$. Thus

$$
\{\gamma_1, \ldots, \gamma_N\} = \{1_{(\delta_{k,1} \not\geq d)}, \ 1 \leq k_q \leq n_q, \ 1 \leq q \leq r+1\}.
$$

Now $G_i = A'Z_i Z_i' A = B \, \mathrm{diag}(\lambda_{ij}) B' = \sum_{j=1}^{N} \lambda_{ij} b_j b_j', \ i \in S$, thus

$$
V(\mu) = I_{N-p} + \sum_{t \in S} \mu_t G_t = BB' + \sum_{t \in S} \mu_t G_t
$$

$$
= \sum_{j=1}^{N} \left(1 + \sum_{t \in S} \mu_t \lambda_{tj}\right) b_j b_j' = \sum_{\gamma_j \neq 0} \left(1 + \sum_{t \in S} \mu_t \lambda_{tj}\right) b_j b_j', \tag{16}
$$

since $\gamma_j = 0 \Rightarrow b_j = 0$. Hence $V(\mu)^{-1} = \sum_{\gamma_j \neq 0}(1 + \sum_{t \in S} \mu_t \lambda_{tj})^{-1} b_j b_j'$ since $\gamma_j \neq 0 \Rightarrow \gamma_j = 1$. Let $\Gamma_+ = \{j: \gamma_j \neq 0\}$. Then $|\Gamma_+| = N - p$ (hereafter $|E|$ denotes the cardinality of a set E). It follows from (16) that the eigenvalues

of $V(\mu)$ are $\{1 + \sum_{t \in S} \mu_t \lambda_{tj}, \ j \in \Gamma_+\}$. Thus by (8),

$$
\begin{aligned}
e_N(\theta, \theta_{(0)}) &= -\frac{1}{2} \sum_{\gamma_j \neq 0} \left\{ \frac{\lambda_0(1 + \sum_{t \in S} \mu_{0t} \lambda_{tj})}{\lambda(1 + \sum_{t \in S} \mu_t \lambda_{tj})} - 1 - \log \frac{\lambda_0(1 + \sum_{t \in S} \mu_{0t} \lambda_{tj})}{\lambda(1 + \sum_{t \in S} \mu_t \lambda_{tj})} \right\} \\
&= -\frac{1}{2} \sum_{k_1=1}^{n_1} \cdots \sum_{k_{r+1}=1}^{n_{r+1}} 1_{(\delta_{k,1} \not\geq d)} \left\{ \frac{\lambda_0(1 + \sum_{t \in S} \mu_{0t}(N/m_t)1_{(t \leq \delta_{k,1})})}{\lambda(1 + \sum_{t \in S} \mu_t(N/m_t)1_{(t \leq \delta_{k,1})})} \right. \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \left. - 1 - \log(\cdots) \right\} \\
&= -\frac{1}{2} \sum_{l_1=0}^{1} \cdots \sum_{l_{r+1}=0}^{1} 1_{(l \not\geq d)} \left\{ \frac{\lambda_0(1 + \sum_{t \in S} \mu_{0t}(N/m_t)1_{(t \leq l)})}{\lambda(1 + \sum_{t \in S} \mu_t(N/m_t)1_{(t \leq l)})} \right. \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \left. - 1 - \log(\cdots) \right\} \prod_{l_q=0} (n_q - 1) \\
&= -\frac{1}{2} \sum_{l \not\geq d} \left\{ \frac{\lambda_0 C_l(\mu_0)}{\lambda C_l(\mu)} - 1 - \log \frac{\lambda_0 C_l(\mu_0)}{\lambda C_l(\mu)} \right\} \prod_{l_q=0} (n_q - 1) = -\frac{1}{2} S_{N,2}(\theta),
\end{aligned}
$$

using a formula given at the end of the proof of Lemma 7.3 in Jiang (1996) for the third equation; and by (9),

$$
\begin{aligned}
d_N(\theta, \theta_{(0)}) &= -\frac{1}{2} \sum_{\gamma_j \neq 0} \left[\lambda^{-1}\left(1 + \sum_{t \in S} \mu_t \lambda_{tj}\right)^{-1} - \lambda_0^{-1}\left(1 + \sum_{t \in S} \mu_{0t} \lambda_{tj}\right)^{-1} \right] \\
&\qquad\qquad \times [(b_j' z)^2 - E_{\theta_{(0)}}(b_j' z)^2] \\
&= \frac{1}{2} \sum_{k_1=1}^{n_1} \cdots \sum_{k_{r+1}=1}^{n_{r+1}} 1_{(\delta_{k,1} \not\geq d)} \left[\lambda^{-1}\left(1 + \sum_{t \in S} \mu_t \frac{N}{m_t} 1_{(t \leq \delta_{k,1})}\right)^{-1} \right. \\
&\qquad\qquad\qquad\qquad\qquad \left. - \lambda_0^{-1}\left(1 + \sum_{t \in S} \mu_{0t} \frac{N}{m_t} 1_{(t \leq \delta_{k,1})}\right)^{-1} \right] \\
&\qquad\qquad \times [(b_k' z)^2 - E_{\theta_{(0)}}(b_k' z)^2] \\
&= \frac{1}{2} \sum_{l_1=0}^{1} \cdots \sum_{l_{r+1}=0}^{1} \sum_{\delta_{k,1}=l} 1_{(\delta_{k,1} \not\geq d)} [\cdots - \cdots][(b_k' z)^2 - E_{\theta_{(0)}}(b_k' z)^2] \\
&= \frac{1}{2} \sum_{l \not\geq d} (r_l(\theta) - 1) \xi_l,
\end{aligned}
$$

where $\xi_l = (\lambda_0 C_l(\mu_0))^{-1} \sum_{\delta_{k,1}=l}[(b_k' z)^2 - E_{\theta_{(0)}}(b_k' z)^2]$, $l \not\geq d$. $\square$

Let $S_{N,1}(\theta) = \sum_{l \not\geq d}(r_l(\theta) - 1)^2 p_l^2$, $S_{N,3}(\theta) = \sum_{l \not\geq d}|r_l(\theta) - 1| p_l$. Define $S^+ = \{i \in S: \mu_{0i} > 0\}$, $\bar{S}^+ = \{(00\ldots0)\} \bigcup S^+$; $b_0 = (\min_{i \in S^+} \mu_{0i})^{-1} \sum_{i \in \bar{S}^+} \mu_{0i}$ if $S^+ \neq \varnothing$, and $b_0 = 1$ if $S^+ = \varnothing$; $\varepsilon_0 = (\lambda_0/3)(\sum_{i \in \bar{S}} \mu_{0i})^{-1}$; $b = 1 + [(\exp(\log 2/r) -$

$1)^{-1}]$ (here $[x]$ means the largest interger less than or equal to x); $s = |S|$; and $T_l = \{q : 1 \leq q \leq r+1, \ l_q = 1\}$.

The following lemma plays a key role in the proof of Wald consistency in the balanced case.

LEMMA 3.2. *Suppose model* (13) *is unconfounded. For any* $B > 0$,

$$(17) \qquad \inf_{\theta \notin \Theta_B} S_{N,1}(\theta) \geq (1/2)^{r+3} \left\{ \left(\min_{i \in \bar{S}} m_i \right) \wedge \left(\frac{\varepsilon_0 B}{3(\varepsilon_0 + s\lambda_0)} \right)^2 \right\},$$

$$(18) \qquad \inf_{\theta \notin \Theta_B} S_{N,2}(\theta) \geq (1/2)^{r+5} b_0^{-1} \left\{ \left(\min_{i \in \bar{S}} m_i \right) \wedge \left(\frac{\varepsilon_0 B}{3(\varepsilon_0 + s\lambda_0)} \right)^2 \right\},$$

$$(19) \qquad \begin{aligned} \sup_{\theta \notin \Theta_B} \frac{S_{N,3}(\theta)}{S_{N,2}(\theta)} &\leq \left(\frac{2^{r+3} b_0}{\inf_{\theta \notin \Theta_B} S_{N,2}(\theta)} \right)^{1/2} \\ &\quad + (2\sqrt{2})^{r+1} \left(\frac{2b_0 + 1}{1 - \log 2} \right) \left(\min_{i \in \bar{S}} m_i \right)^{-1/2}, \end{aligned}$$

where $\Theta_B = \{\theta \in \Theta : |\lambda - \lambda_0| \leq \varepsilon_0 B/\sqrt{N}, \ |\mu_l - \mu_{0l}| \leq b^{|T_l|} B/\sqrt{m_l}, \ l \in S\}$.

PROOF. Let $i_* = 0 = (0 \cdots 0)$ if $n_{r+1} \geq 2$; $i_* = (0 \cdots 01)$ if $n_{r+1} = 1$. It is easy to show that $i_* \not\geq d$, $i_* \not\geq i$, $i \in S$, and $p_{i_*}^2 \geq (\frac{1}{2})^{r+1} m_{i_*} = (\frac{1}{2})^{r+1} N$. So if $\lambda < (2/3)\lambda_0$ or $\lambda > 2\lambda_0$, then

$$(20) \qquad S_{N,1}(\theta) \geq (r_{i_*}(\theta) - 1)^2 p_{i_*}^2 = (\lambda_0/\lambda - 1)^2 p_{i_*}^2 \geq (1/2)^{r+3} N.$$

If $(2/3)\lambda_0 \leq \lambda \leq 2\lambda_0$ and $\mu_l > \lambda_0/\varepsilon_0$ for some $l \in S$, then $r_l(\theta) = (\lambda_0 D_l(\mu_0)/\lambda D_l(\mu)) < 1/2$, where $D_l(\mu) = \sum_{i \in \bar{S}} \mu_i (m_l/m_i) 1_{(i \leq l)}$. So $S_{N,1}(\theta) \geq (r_l(\theta) - 1)^2 p_l^2 \geq (1/2)^{r+2} m_l$.

Now suppose $(2/3)\lambda_0 \leq \lambda \leq 2\lambda_0$, $0 \leq \mu_l \leq \lambda_0/\varepsilon_0$, $l \in S$, and $\theta \notin \Theta_B$. If $|\lambda - \lambda_0| > \varepsilon_0 B/\sqrt{N}$, then as in (20) we have $S_{N,1}(\theta) \geq (1/2)^{r+3}(\varepsilon_0/\lambda_0)^2 B^2$. If $|\lambda - \lambda_0| \leq \varepsilon_0 B/\sqrt{N}$ and $|\mu_l - \mu_{0l}| > b^{|T_l|} B/\sqrt{m_l}$ for some $l \in S$, there is $l \in S$ such that $|\mu_l - \mu_{0l}| > b^{|T_l|} B/\sqrt{m_l}$ and $|\mu_i - \mu_{0i}| \leq b^{|T_i|} B/\sqrt{m_i}$, $i \in S$, $i \leq l$, $i \neq l$. Then

$$|D_l(\mu_0) - D_l(\mu)| \geq |\mu_{0l} - \mu_l| - \left| \sum_{i \in S} (\mu_{0i} - \mu_i)(m_l/m_i) 1_{(i \leq l, \ i \neq l)} \right| > B/\sqrt{m_l},$$

since

$$\begin{aligned} \sum_{i \in S} |\mu_{0i} - \mu_i|(m_l/m_i) 1_{(i \leq l, \ i \neq l)} &\leq (B/\sqrt{m_l}) \sum_{i \leq l, \ i \neq l} b^{|T_i|} \\ &= (B/\sqrt{m_l})((b+1)^{|T_l|} - b^{|T_l|}) \\ &\leq (B/\sqrt{m_l})(b^{|T_l|} - 1). \end{aligned}$$

Therefore,

$$|\lambda_0 D_l(\mu_0) - \lambda D_l(\mu)| \geq \lambda |D_l(\mu_0) - D_l(\mu)| - |\lambda - \lambda_0| D_l(\mu_0)$$
$$\geq (2/3)\lambda_0(B/\sqrt{m_l}) - \varepsilon_0(B/\sqrt{N}) \sum_{i \in \bar{S}} \mu_{0i}$$
$$\geq (\lambda_0/3)(B/\sqrt{m_l}).$$

Hence $S_{N,1}(\theta) \geq (r_l(\theta) - 1)^2 p_l^2 \geq (1/2)^{r+2}(\varepsilon_0 B/3(\varepsilon_0 + s\lambda_0))^2$.

To prove (18), let $l_* \in S_l^+ = \{i \in S^+ : i \leq l\}$ such that $m_{l_*} = \min_{i \in S_l^+} m_i$ if $S_l^+ \neq \varnothing$, and $l_* = i_*$ if $S_l^+ = \varnothing$. Then it is easy to show that $C_l(\mu_0)/C_{l_*}(\mu_0) \leq b_0$ and hence

$$(21) \qquad \frac{r_l(\theta)}{r_{l_*}(\theta)} = \frac{C_l(\mu_0)}{C_{l_*}(\mu_0)} \frac{C_{l_*}(\mu)}{C_l(\mu)} \leq b_0.$$

If $r_l(\theta) \geq 2b_0$ for some $l \not\geq d$, then $r_{l_*}(\theta) \geq 2$ by (21). Therefore,

$$S_{N,2}(\theta) \geq (r_{l_*}(\theta) - 1 - \log r_{l_*}(\theta)) p_{l_*}^2$$
$$(22)$$
$$\geq \left(\frac{1 - \log 2}{2}\right) r_{l_*}(\theta) p_{l_*}^2 \geq (1/2)^{r+1}(1 - \log 2) \min_{i \in \bar{S}} m_i,$$

using that $(x - 1 - \log x)/x \geq (1 - \log 2)/2, \ x \geq 2$.

If $r_l(\theta) < 2b_0$ for all $l \not\geq d$, then $S_{N,2}(\theta) \geq S_{N,1}(\theta)/4b_0$ by $x - 1 - \log x \geq (x - 1)^2/2L, \ 0 < x < L, \ L > 1$.

To prove (19), let $S_* = \{i_*\} \cup S$. Write

$$(23) \quad S_{N,3}(\theta) = \sum_{l \not\geq d, \, r_l(\theta) < 2b_0} |r_l(\theta) - 1| p_l + \sum_{l \not\geq d, \, r_l(\theta) \geq 2b_0} |r_l(\theta) - 1| p_l = S_1 + S_2.$$

For any $\varepsilon > 0$,

$$S_1 \leq \left(\sum_{l \not\geq d, \, r_l(\theta) < 2b_0} 1^2\right)^{1/2} \left(\sum_{l \not\geq d, \, r_l(\theta) < 2b_0} (r_l(\theta) - 1)^2 p_l^2\right)^{1/2}$$
$$(24) \qquad \leq (\sqrt{2})^{r-1}\left\{\varepsilon^{-1} + \varepsilon \sum_{l \not\geq d, \, r_l(\theta) < 2b_0} (r_l(\theta) - 1)^2 p_l^2\right\}$$
$$\leq (\sqrt{2})^{r-1}\left\{\frac{\varepsilon^{-1}}{S_{N,2}(\theta)} + 4b_0\varepsilon\right\} S_{N,2}(\theta)$$

by $x \leq (\varepsilon^{-1} + \varepsilon x^2)/2$ and that $S_{N,2}(\theta) \geq (4b_0)^{-1} \sum_{l \not\geq d, \, r_l(\theta) < 2b_0} (r_l(\theta) - 1)^2 p_l^2$. By (21), the facts that (according to the definition) l_* only takes values in

$S_* = \{i_*\} \cup S$ and $l_* \leq l$ (therefore $p_{l_*} \geq p_l$), and the fact that

$$S_{N,2}(\theta) \geq \left(\frac{1 - \log 2}{2}\right) \sum_{i \in S_*,\, r_i(\theta) \geq 2} r_i(\theta) p_i^2,$$

it follows that

$$
\begin{aligned}
S_2 &\leq (b_0 + 1/2) \sum_{l \not\geq d,\, r_{l_*}(\theta) \geq 2} r_{l_*}(\theta) p_l \\
&= (b_0 + 1/2) \sum_{i \in S_*,\, r_i(\theta) \geq 2} r_i(\theta) \sum_{l \not\geq d,\, l_* = i} p_l \\
&\leq 2^{r+1}(b_0 + 1/2) \sum_{i \in S_*,\, r_i(\theta) \geq 2} r_i(\theta) p_i \\
&\leq \frac{2^{r+1}(2b_0 + 1)}{(1 - \log 2) \min_{i \in S_*} p_i} S_{N,2}(\theta).
\end{aligned}
$$

(25)

Inequality (19) now follows by (23)–(25), taking $\varepsilon = (4b_0 \inf_{\theta \notin \Theta_B} S_{N,2}(\theta))^{-1/2}$ in (24), and the fact that

$$\min_{i \in \bar S_*} p_i \geq (\sqrt 2)^{-(r+1)}\left(\min_{i \in \bar S} m_i\right)^{1/2}. \qquad \square$$

Let $\hat\theta_N = (\hat\lambda_N, (\hat\mu_{Ni})'_{i \in S})' \in \Theta$ be the maximizer of (3) (one may define $\hat\theta_N$ as any fixed point in Θ when the maximum can not be reached).

THEOREM 3.1. *Let the balanced mixed model* (13) *be unconfounded. As* $N \to \infty$ *and* $m_i \to \infty$, $i \in S$, *we have the following*:

(i) $\{\hat\theta_N\}$ *is consistent and the sequence* $\{(\sqrt{N - p}(\hat\lambda_N - \lambda_0), (\sqrt{m_i}(\hat\mu_{Ni} - \mu_{0i}))'_{i \in S})'\}$ *is bounded in probability.*

(ii) *If, moreover, the model has positive variance components and is nondegenerate, then* $\{\hat\theta_N\}$ *is asymptotically normal with* $p_0(N) = \sqrt{N - p}$, $p_i(N) = \sqrt{m_i}$, $i \in S$ *and* $M_N(\theta_{(0)}) = J_N^{-1/2}(\theta_{(0)}) I_N(\theta_{(0)})$.

REMARK. Part (i) of Theorem 3.1 does not require that the model have positive variance components. In particular, when such requirement does hold, $\{\hat\theta_N\}$ constitutes (with probability $\to 1$) a consistent sequence of roots of the REML equations [e.g., Jiang (1996)], which are identical to the ANOVA estimates in the balanced case [e.g., Searle, Casella and McCulloch (1992)]. On the other hand, it is necessary for part (ii) of Theorem 3.1 that the model have positive variance components since otherwise the estimates, which are nonnegative, cannot be asymptotically normal.

PROOF. According to the definition in Section 2, it is always true (no matter whether $\mu_{0i} = 0$ or not) that $Z_i \alpha_i = \sqrt{\lambda_0 \mu_{0i}} Z_i(\alpha_i/\sqrt{\lambda_0 \mu_{0i}})$, $1 \leq i \leq s$, and

(26) $$z = A'y = \sqrt{\lambda_0} A'b(\mu_0)' \mathscr{W}_N.$$

Note that $\mathcal{W}_N$ is a vector of independent random variables satisfying $EW_{Nl} = 0$, $\mathrm{var}(W_{Nl}) \leq 1$ [$\mathrm{var}(W_{Nl}) = 1$ if the corresponding $\mu_{0i} \neq 0$], and $EW_{Nl}^4 < \infty$. By the definition of ξ_l at the end of the proof of Lemma 3.1, Lemma 2.1 and assumption A2 in Section 2, we see that with $B_N = b(\mu_0)A(\sum_{\delta_{k,1}=l} b_k b_k')A'b(\mu_0)'$,

$$
(27) \quad
\begin{aligned}
E_{\theta_{(0)}}\xi_l^2 = \mathrm{var}_{\theta_{(0)}}\left(\mathcal{W}_N' \frac{B_N}{C_l(\mu_0)} \mathcal{W}_N\right) &\leq \frac{C}{C_l^2(\mu_0)} \mathrm{tr}(B_N^2) \\
&= \frac{C}{C_l^2(\mu_0)} \mathrm{tr}\left(\left(\sum_{\delta_{k,1}=l} b_k b_k' V(\mu_0)\right)^2\right) = C p_l^2
\end{aligned}
$$

for some constant C. Thus, for any $M > 0$,

$$
(28) \qquad P_{\theta_{(0)}}(|\xi_l| > M p_l \text{ for some } l \not\geq d) \leq \frac{2^{r+1}C}{M^2}.
$$

For any $\varepsilon > 0$, choose $M > (2^{r+1}C\varepsilon^{-1})^{1/2}$. Then choose $B > 0$ and K such that by Lemma 3.2, $\sup_{\theta \notin \Theta_B}\{S_{N,3}(\theta)/S_{N,2}(\theta)\} < 1/(2M)$, $m_i > K$, $i \in \bar{S}$. It then follows by (7), (14), (15), (18) and (28) that

$$
P_{\theta_{(0)}}(L_N(\theta) < L_N(\theta_{(0)}) \; \forall \; \theta \notin \Theta_B) > 1 - \varepsilon,
$$

when $m_i > K$, $i \in \bar{S}$. The conclusion of (i) then follows by the definition of Θ_B. The proof for (ii) is the same as that of Theorem 4.1 of Jiang (1996). □

4. The general case: the method of sieves.

For convenience we consider in this section the parameters ϕ_i, $0 \leq i \leq s$.

Let δ_N, M_N be two sequences of positive numbers such that $\delta_N \to 0$, $M_N \to \infty$. Let $\hat{\phi}_N \in [\delta_N, M_N] = \{\phi \in \Phi: \delta_N \leq \phi_i \leq M_N, \; 0 \leq i \leq s\}$ be the maximizer of $\tilde{L}_N(\phi)$ of (4) over $[\delta_N, M_N]$. The above procedure is called the method of sieves [e.g., Grenander (1981)].

The following definitions are explained intuitively in Jiang (1996).

DEFINITION 4.1. Model (1) is called asymptotically identifiable under the invariant class (AI^2) at $\phi \in \Phi(\theta \in \Theta)$ if

$$
(29) \qquad \liminf_{N \to \infty} \lambda_{\min}(\mathrm{Cor}(V_0(\theta), \ldots, V_s(\theta))) > 0
$$

(see Section 2; $\lambda_{\min}$ means the smallest eigenvalue). Here the invariant class is the class of (location) invariant estimates

$$
(30) \quad \mathscr{I} = \{\text{estimates which are functions of } A'y \text{ with } A \text{ satisfying (2)}\}.
$$

DEFINITION 4.2. Model (1) is called infinitely informative under the invariant class (I^3) at $\phi \in \Phi(\theta \in \Theta)$ if

$$
(31) \qquad \lim_{N \to \infty} \|V_i(\theta)\|_R = \infty, \qquad 0 \leq i \leq s.
$$

The following notation is needed for this section: $\lambda_N(\theta) = \lambda_{\min}(\mathrm{Cor}(V_0(\theta), \ldots, V_s(\theta)))$, $\rho_N(a, b) = \inf_{\theta \in [a, b]} \lambda_N(\theta)$, $\tau_N(c, d) = \inf_{\phi \in [c, d]} \lambda_N(\theta)$, where $[a, b] = \{\theta: a \le \theta_i \le b, \ 0 \le i \le s\}$ and so on.

LEMMA 4.1. *For any* $x_i, \ 0 \le i \le s$ *and* $\phi \in \Phi$,

$$\left\| \sum_{i=0}^s x_i U_i(\phi) \right\|_R^2 \ge \left(\frac{\lambda_N(\theta)}{4(s+1)} \right) \left\{ x_0^2 \|V_0(\theta)\|_R^2 + \lambda^{-2} \sum_{i=1}^s x_i^2 \|V_i(\theta)\|_R^2 \right\}$$

$$\ge \left(\frac{\lambda_N(\theta)}{4M^2(s+1)} \right) \sum_{i=0}^s x_i^2 \|V_i(1)\|_R^2,$$

where $M = \max_{0 \le i \le s} \phi_i$.

PROOF. By the relation (see Section 2): $U_0(\phi) = V_0(\theta) - \sum_{i=1}^s \lambda^{-1} \mu_i V_i(\theta)$, $U_i(\phi) = \lambda^{-1} V_i(\theta), 1 \le i \le s$, we have

$$\left\| \sum_{i=0}^s x_i U_i(\phi) \right\|_R^2 = \mathrm{tr}\left(\left(\sum_{i=0}^s y_i V_i(\theta) \right)^2 \right) \ge \lambda_N(\theta) \sum_{i=0}^s y_i^2 \|V_i(\theta)\|_R^2,$$

where $y_0 = x_0$, $y_i = \lambda^{-1}(x_i - \mu_i x_0)$, $1 \le i \le s$. Let $B = \{1 \le i \le s: |x_i - \mu_i x_0| \ge (1/2)|x_i|\}$, then

$$\sum_{i=0}^s y_i^2 \|V_i(\theta)\|_R^2 \ge \frac{1}{s+1} x_0^2 \|V_0(\theta)\|_R^2 + \frac{1}{s+1} \sum_{i \notin B} x_0^2 \|V_0(\theta)\|_R^2 + \sum_{i \in B} y_i^2 \|V_i(\theta)\|_R^2$$

$$\ge \frac{1}{4(s+1)} \left\{ x_0^2 \|V_0(\theta)\|_R^2 + \lambda^{-2} \sum_{i=1}^s x_i^2 \|V_i(\theta)\|_R^2 \right\},$$

using the fact that $\|V_0(\theta)\|_R \ge \lambda^{-1} \mu_i \|V_i(\theta)\|_R$, $1 \le i \le s$. The second inequality is obvious. $\square$

LEMMA 4.2. *Suppose model* (1) *is* AI^2 *at all* $\theta \in [a, b]$ *with* $a > 0$, *then* $\liminf \rho_N(a, b) > 0$.

PROOF. By direct computation it can be shown that (see Section 2):

$$\left| \frac{\partial}{\partial \theta_k} \mathrm{cor}(V_i(\theta), V_j(\theta)) \right| \le 4\|V_k(\theta)\| \le 4a^{-1}, \qquad 0 \le i, j, \ k \le s.$$

The result thus follows by an argument of subsequences [e.g., Billingsley (1986), A10]. $\square$

Denote, respectively, the interiors of Θ and Φ by Θ° and Φ°.

COROLLARY 4.1. *Suppose model* (1) *is* AI^2 *at all* $\theta \in \Theta^\circ$, *then* $\liminf \tau_N(c, d) > 0$ *for all* $0 < c < d < \infty$.

Let model (1) be AI^2 at all $\theta \in \Theta^\circ$ and I^3 at $1 = (1 \cdots 1)'$. Define sequences $\{\delta_N\}$ and $\{M_N\}$ as follows.

Let $q_N = \min_{0 \le i \le s} \|V_i(1)\|$, a, b be positive numbers such that

$$(32) \qquad\qquad (a + b)(1 - b)^{-1} < 2(3s + 7)^{-1}.$$

Pick f, g, h such that

$$(33) \qquad\qquad 3a + 2b < f < (s + 1)^{-1}(2 - 4a - (s + 7)b),$$

$$(34) \qquad 0 < g < \min\left(\frac{f - 3a - 2b}{2},\ 1 - 2a - \frac{s + 7}{2}b - \frac{s + 1}{2}f\right),$$

$$(35) \qquad 0 < h < \min\left(f - 3a - 2b - 2g,\ 1 - 2a - \frac{s + 7}{2}b - \frac{s + 1}{2}f - g\right).$$

Now pick sequences $\{c_N\}$ and $\{d_N\}$ such that $0 \leftarrow \cdots \le c_2 \le c_1 \le 1 \le d_1 \le d_2 \le \cdots \rightarrow \infty$ and

$$(36) \qquad\qquad \liminf q_N^h \tau_N(c_N, d_N) > 0$$

(by Corollary 4.1, one certainly can do this). Finally, pick a "baseline region" $[\delta_B, M_B]$ such that $0 < \delta_B < 1 < M_B < \infty$ and let

$$\delta_N = \delta_B \wedge (c_N \vee q_N^{-a}), \qquad M_N = M_B \vee (d_N \wedge q_N^b).$$

NOTE. The selection of δ_B and M_B will not make a difference for the following asymptotic result. However, a reasonable choice of δ_B and M_B may be of practical convenience since, when N is not sufficiently large, the region $[c_N \vee q_N^{-a},\ d_N \wedge q_N^b]$ may appear to be "too narrow."

THEOREM 4.1. *Consider a general mixed model* (1) *having positive variance components.*

(i) *If the model is asymptotically identifiable and infinitely informative* (AI^4) *at all* $\theta \in \Theta^\circ$ *(or* $\phi \in \Phi^o$*), then* $\{\hat{\phi}_N\}$ *is consistent and the sequence* $\{(\|V_i(1)\|_R(\hat{\phi}_{Ni} - \phi_{0i}))_{0 \le i \le s}\}$ *is bounded in probability.*

(ii) *If, moreover, the model is nondegenerate, then* $\{\hat{\phi}_N\}$ *is asymptotically normal with* $p_0(N) = \sqrt{N - p}$, $p_i(N)$ *being any sequence* $\sim \|V_i(1)\|_R$, $1 \le i \le s$ *and* $M_N(\theta_{(0)}) = J_N^{-1/2}(\theta_{(0)})I_N(\theta_{(0)})$.

REMARK 1. Thus, with probability $\rightarrow 1$, $\{\hat{\phi}_N\}$ constitutes a consistent sequence of roots of the REML equations.

REMARK 2. The assumption that model (1) is AI^4 at all $\theta \in \Theta^\circ$ is equivalent to saying that it is AI^2 at all $\theta \in \Theta^\circ$ and I^3 at some $\theta^\circ \in \Theta^\circ$ and w.l.o.g., one can take $\theta^\circ = 1$.

PROOF. (i) Let $\Phi_N^* = \{\phi^* = (\phi_i^*)_{0 \le i \le s}: \phi_i^* = \delta_N + k_i \varepsilon_N, \ 0 \le i \le s$ for some integers $0 \le k_i < K_N, \ 0 \le i \le s\}$, where $K_N = [q_N^{b+f}], \ \varepsilon_N = (M_N - \delta_N)/K_N$. Let $\phi \in [\delta_N, M_N]$, then by (4) we have a decomposition similar to (7):

$$(37) \qquad \tilde{L}_N(\phi) - \tilde{L}_N(\phi_{(0)}) = \tilde{e}_N(\phi, \phi_{(0)}) + \tilde{d}_N(\phi, \phi_{(0)}).$$

By an expression similar to (8), we have

$$\tilde{e}_N(\phi, \phi_{(0)}) = E_{\phi_{(0)}}(\tilde{L}_N(\phi) - \tilde{L}_N(\phi_{(0)}))$$

$$= (1/2)\{\log |U(\phi_{(0)})^{1/2} U(\phi)^{-1} U(\phi_{(0)})^{1/2}|$$

$$+ \operatorname{tr}(I_{N-p} - U(\phi_{(0)})^{1/2} U(\phi)^{-1} U(\phi_{(0)})^{1/2})\}$$

$$= (1/2) \sum_{i=1}^{N-p} (\log \lambda_i + 1 - \lambda_i),$$

where $\lambda_1, \ldots, \lambda_{N-p}$ are the eigenvalues of $B(\phi) = U(\phi_{(0)})^{1/2} U(\phi)^{-1} U(\phi_{(0)})^{1/2}$. Since $\lambda_{\max}(B(\phi)) \le (1 \vee M_0) \delta_N^{-1}$, where $M_0 = \max_{0 \le i \le s} \phi_{0i}$, we have by $x - 1 - \log x \ge (x-1)^2/2L, \ 0 < x \le L, \ L \ge 1$ and Lemma 4.1 that

$$\tilde{e}_N(\phi, \phi_{(0)}) \le -\frac{\delta_N}{4(1 \vee M_0)} \operatorname{tr}((U(\phi_{(0)})^{1/2} U(\phi)^{-1} U(\phi_{(0)})^{1/2} - I_{N-p})^2)$$

$$(38) \qquad = -\frac{\delta_N}{4(1 \vee M_0)} \left\| \sum_{i=0}^{s} (\phi_i - \phi_{0i}) U_i(\phi) \right\|_R^2$$

$$\le -\frac{\delta_N \tau_N(c_N, d_N)}{16(s+1)(1 \vee M_0) M_N^2} \sum_{i=0}^{s} \|V_i(1)\|_R^2 (\phi_i - \phi_{0i})^2.$$

It is easy to verify the following identity which holds for any $\phi^* \in \Phi$:

$$U(\phi_{(0)})^{-1} - U(\phi)^{-1}$$

$$= \sum_{i=0}^{s} (\phi_i - \phi_{0i}) U(\phi_{(0)})^{-1} G_i U(\phi_{(0)})^{-1}$$

$$(39) \qquad - \sum_{i=0}^{s} \sum_{j=0}^{s} (\phi_i - \phi_{0i})(\phi_j - \phi_{0j}) U(\phi_{(0)})^{-1} G_i U(\phi^*)^{-1} G_j U(\phi_{(0)})^{-1}$$

$$+ \sum_{i=0}^{s} \sum_{j=0}^{s} (\phi_i - \phi_{0i})(\phi_j - \phi_{0j}) U(\phi_{(0)})^{-1} G_i U(\phi_{(0)})^{-1/2}$$

$$\times H^* U(\phi_{(0)})^{-1/2} G_j U(\phi_{(0)})^{-1},$$

where

$$H^* = (1/2) \sum_{k=0}^{s} (\phi_k - \phi_{*k}) U(\phi_{(0)})^{1/2}$$

$$\times (U(\phi)^{-1} G_k U(\phi^*)^{-1} + U(\phi^*)^{-1} G_k U(\phi)^{-1}) U(\phi_{(0)})^{1/2}.$$

In the following we pick $\phi^* \in \Phi_N^*$ in (39) such that $|\phi_{*i} - \phi_i| < \varepsilon_N,\ 0 \le i \le s.$ By an expression similar to (9) and by (39) we get

$$\tilde{d}_N(\phi, \phi_{(0)})$$

$$= \tilde{L}_N(\phi) - \tilde{L}_N(\phi_{(0)}) - E_{\phi_{(0)}}(\tilde{L}_N(\phi) - \tilde{L}_N(\phi_{(0)}))$$

$$= (1/2)\Bigg\{ \sum_{i=0}^s (\phi_i - \phi_{0i})(z'U(\phi_{(0)})^{-1}G_i U(\phi_{(0)})^{-1}z - E_{\phi_{(0)}}(\cdots))$$

$$- \sum_{i=0}^s \sum_{j=0}^s (\phi_i - \phi_{0i})(\phi_j - \phi_{0j})$$

(40)

$$\times \big(z'U(\phi_{(0)})^{-1}G_i U(\phi^*)^{-1}G_j U(\phi_{(0)})^{-1}z - E_{\phi_{(0)}}(\cdots)\big)$$

$$+ \sum_{i=0}^s \sum_{j=0}^s (\phi_i - \phi_{0i})(\phi_j - \phi_{0j}) z'U(\phi_{(0)})^{-1}G_i U(\phi_{(0)})^{-1/2}$$

$$\times H^* U(\phi_{(0)})^{-1/2}G_j U(\phi_{(0)})^{-1}z - E_{\phi_{(0)}}\bigg(\sum_{i=0}^s \sum_{j=0}^s \cdots\bigg)\Bigg\}$$

$$= (1/2)(I_1 - I_2 + I_3 - E_{\phi_{(0)}}I_3).$$

It follows from Lemma 2.1 and (26) that

$$\mathrm{var}_{\phi_{(0)}}(z'U(\phi_{(0)})^{-1}G_i U(\phi_{(0)})^{-1}z) \le C\|U_i(\phi_{(0)})\|_R^2 \le C\delta_0^{-2}\|V_i(1)\|_R^2,$$

where $C = 2 \vee (\mathrm{var}(\varepsilon_1^2)/(\mathrm{var}(\varepsilon_1))^2) \vee \max_{1 \le i \le s}(\mathrm{var}(\alpha_{i1}^2)/(\mathrm{var}(\alpha_{i1}))^2),\ \delta_0 = \min_{0 \le i \le s} \phi_{0i}.$ Thus with $L_1 = (3(s+1)C)^{1/2}q_N^{\bar{g}}$, we have on

$$E_1 = \big\{|z'U(\phi_{(0)})^{-1}G_i U(\phi_{(0)})^{-1}z - E_{\phi_{(0)}}\cdots| \le L_1 \delta_0^{-1}\|V_i(1)\|_R,\ 0 \le i \le s\big\}$$

that

(41)
$$|I_1| \le L_1(s+1)^{1/2}\delta_0^{-1}\bigg(\sum_{i=0}^s \|V_i(1)\|_R^2 (\phi_i - \phi_{0i})^2\bigg)^{1/2},$$

and

(42)
$$P_{\phi_{(0)}}(E_1^c) \le \sum_{i=0}^s \frac{\mathrm{var}_{\phi_{(0)}}(z'\cdots z)}{L_1^2 \delta_0^{-2}\|V_i(1)\|_R^2} \le (1/3)q_N^{-2g}.$$

Similarly,

$$\mathrm{var}_{\phi_{(0)}}(z'U(\phi_{(0)})^{-1}G_i U(\phi^*)^{-1}G_j U(\phi_{(0)})^{-1}z)$$

$$\le C\|U(\phi_{(0)})^{-1/2}G_i U(\phi^*)^{-1}G_j U(\phi_{(0)})^{-1/2}\|_R^2$$

$$\le C\|U_i(\phi_{(0)})\|\|U_j(\phi_{(0)})\|\|U_i(\phi^*)\|_R\|U_j(\phi^*)\|_R$$

$$\le C(\delta_0 \delta_N)^{-2}\|V_i(1)\|_R\|V_j(1)\|_R.$$

(Here is a brief derivation of the second inequality in the above:

$$\|U(\phi_{(0)})^{-1/2}G_i U(\phi^*)^{-1}G_j U(\phi_{(0)})^{-1/2}\|_R^2 = \|AB\|_R^2 \le \|A'A\|_R\|B'B\|_R,$$

where $A = U(\phi_{(0)})^{-1/2}G_iU(\phi^*)^{-1/2}$, $B = U(\phi^*)^{-1/2}G_jU(\phi_{(0)})^{-1/2}$. Now $A'A \le \lambda_{\max}(U_i(\phi_{(0)}))U_i(\phi^*)$. Using the fact that $A, B \ge 0$, $A \le B \Rightarrow \operatorname{tr}(A^2) \le \operatorname{tr}(B^2)$ [note that it is not true that $A, B \ge 0$, $A \le B \Rightarrow A^2 \le B^2$ (e.g., Chan and Kwong (1985))], we get $\|A'A\|_R \le \|U_i(\phi_{(0)})\|\|U_i(\phi^*)\|_R$, and so on.) Thus with $L_2 = (s+1)(3K_N^{s+1}C)^{1/2}q_N^g$ we have on

$$E_2 = \{|z'U(\phi_{(0)})^{-1}G_iU(\phi^*)^{-1}G_jU(\phi_{(0)})^{-1}z - E_{\phi_{(0)}}\cdots|$$
$$\le L_2(\delta_0\delta_N)^{-1}\|V_i(1)\|_R^{1/2}\|V_j(1)\|_R^{1/2}, \ 0 \le i, \ j \le s, \ \phi^* \in \Phi_N^*\}$$

that

$$
\begin{aligned}
(43) \quad |I_2| &\le L_2(\delta_0\delta_N)^{-1}\left(\sum_{i=0}^{s}\|V_i(1)\|_R^{1/2}|\phi_i - \phi_{0i}|\right)^2 \\
&\le L_2(s+1)^{3/2}(\delta_0\delta_N)^{-1}(M_0 \vee M_N)\left(\sum_{i=0}^{s}\|V_i(1)\|_R^2(\phi_i - \phi_{0i})^2\right)^{1/2},
\end{aligned}
$$

and

$$(44) \quad P_{\phi_{(0)}}(E_2^c) \le \sum_{\phi^* \in \Phi_N^*}\sum_{0 \le i, \, j \le s}\frac{\operatorname{var}_{\phi_{(0)}}(z'\cdots z)}{L_2^2(\delta_0\delta_N)^{-2}\|V_i(1)\|_R\|V_j(1)\|_R} \le (1/3)q_N^{-2g}.$$

Finally, $I_3 = U'H^*U$, where $U = \sum_{i=0}^{s}(\phi_i - \phi_{0i})U(\phi_{(0)})^{-1/2}G_iU(\phi_{(0)})^{-1}z$. We have

$$\|H^*\| \le (1/2)\sum_{k=0}^{s}|\phi_k - \phi_{*k}|\|U(\phi_{(0)})^{1/2}\cdots U(\phi_{(0)})^{1/2}\| \le (s+1)M_0\delta_N^{-2}\varepsilon_N,$$

$$|U|^2 \le (s+1)\sum_{i=0}^{s}(\phi_i - \phi_{0i})^2|U(\phi_{(0)})^{-1/2}G_iU(\phi_{(0)})^{-1}z|^2$$

and

$$E_{\phi_{(0)}}|U(\phi_{(0)})^{-1/2}G_iU(\phi_{(0)})^{-1}z|^2 \le \delta_0^{-2}\|V_i(1)\|_R^2.$$

Thus with $L_3 = (3(s+1))^{1/2}q_N^g$ we have on $E_3 = \{|U(\phi_{(0)})^{-1/2}G_iU(\phi_{(0)})^{-1}z| \le L_3\delta_0^{-1}\|V_i(1)\|_R, \ 0 \le i \le s\}$ that

$$(45) \quad |I_3| \le L_3^2(s+1)^2\delta_0^{-2}M_0\delta_N^{-2}\varepsilon_N\sum_{i=0}^{s}\|V_i(1)\|_R^2(\phi_i - \phi_{0i})^2$$

and

$$(46) \quad P_{\phi_{(0)}}(E_3^c) \le \sum_{i=0}^{s}\frac{E_{\phi_{(0)}}|\cdots z|^2}{L_3^2\delta_0^{-2}\|V_i(1)\|_R^2} \le (1/3)q_N^{-2g}.$$

It also follows that

$$(47) \quad E_{\phi_{(0)}}|I_3| \le (s+1)^2\delta_0^{-2}M_0\delta_N^{-2}\varepsilon_N\sum_{i=0}^{s}\|V_i(1)\|_R^2(\phi_i - \phi_{0i})^2.$$

Combining (37), (38), (40), (41), (43), (45) and (47), we have on $E = E_1 \cap E_2 \cap E_3$ that for $\phi \in [\delta_N, M_N]$,

$$\tilde{L}_N(\phi) - \tilde{L}_N(\phi_{(0)})$$

(48)
$$\leq Q_N^2 \left\{ -\frac{\delta_N \tau_N(c_N, d_N)}{16(s+1)(1 \vee M_0) M_N^2} + (1/2)(s+1)^2 \delta_0^{-2} M_0(L_3^2 + 1) \delta_N^{-2} \varepsilon_N \right.$$

$$\left. + (1/2)(s+1)^{1/2} \delta_0^{-1}(L_1 + (s+1)L_2 \delta_N^{-1}(M_0 \vee M_N)) Q_N^{-1} \right\},$$

where $Q_N = (\sum_{i=0}^s \|V_i(1)\|_R^2 (\phi_i - \phi_{0i})^2)^{1/2}$. And, combining (42), (44) and (46), we have

(49)
$$P_{\phi_{(0)}}(E^c) \leq q_N^{-2g}.$$

Note that by I^3 at 1 ($= (1 \cdots 1)'$), q_N tends to infinity, and so the probability of E tends to 1.

It remains to show that for any $\eta > 0$ there is $N_0 > 0$ such that for $N \geq N_0$ the $\{\cdots\}$ in (48) is less than 0 for all $\phi \in [\delta_N, M_N] \cap S_\eta(\phi_{(0)})^c$, where $S_\eta(\phi_{(0)}) = \{\phi \in \Phi \colon |\phi_i - \phi_{0i}| < \eta, \ 0 \leq i \leq s\}$. Since $\delta_N \to 0$, $M_N \to \infty$ as $N \to \infty$, this implies that $\{\hat{\phi}_N\}$ is consistent.

Since $\phi \in [\delta_N, M_N] \cap S_\eta(\phi_{(0)})^c$ implies $Q_N \geq \eta q_N$,

$$\{\cdots\} \text{ in (48)} \leq -\frac{q_N^h \tau_N(c_N, d_N)}{16(s+1)(1 \vee M_0)} q_N^{-a-2b-h} + C_1 q_N^{2a-f+2g}$$

$$+ C_2 \eta^{-1} q_N^{a+((s+3)/2)b+((s+1)/2)f+g-1},$$

where C_1, C_2 are constants. Also $\liminf q_N^h \tau_N(c_N, d_N) > 0$, and

$$-a - 2b - h > \max\left(2a - f + 2g, a + \frac{s+3}{2}b + \frac{s+1}{2}f + g - 1\right)$$

by the way we picked a, b, f, g, h. Thus $\{\cdots\}$ in (48) is less than 0 for large N uniformly for all $\phi \in [\delta_N, M_N] \cap S_\eta(\phi_{(0)})^c$.

To prove that $\{(\|V_i(1)\|_R(\hat{\phi}_{Ni} - \phi_{0i}))_{0 \leq i \leq s}\}$ is bounded in probability, replace δ_N, M_N, K_N, ε_N and q_N^g by $\delta = (1/2)\delta_0$, $M = (3/2)M_0$, K, $\varepsilon = (M - \delta)/K$ and η^{-1}, respectively. Then by the same argument we have with probability greater than or equal to $1 - \eta^2$ that for all $\phi \in [\delta, M]$,

(50)
$$\tilde{L}_N(\phi) - \tilde{L}_N(\phi_{(0)}) \leq Q_N^2 \left\{ -\frac{\delta \tau_N(\delta, M)}{16(s+1)M_0 M^2} + C_1 \eta^{-2} \delta^{-2} \varepsilon \right.$$

$$\left. + C_2 \eta^{-1} K^{(s+1)/2} \delta^{-1} M Q_N^{-1} \right\}.$$

The result follows from (50) and the consistency of $\{\hat{\phi}_N\}$.

The proof of (ii) is the same as that of Theorem 4.3 in Jiang (1996). $\quad\square$

5. Examples. The first example is used to illustrate the AI^4 condition.

EXAMPLE 5.1. Consider a two-way (unbalanced) nested model

(51) $$y_{ijk} = \beta_i + \alpha_{ij} + \varepsilon_{ijk},$$

where $i = 1, \ldots, p$, $j = 1, 2$, $k = 1, \ldots, n_i$, β, α and ε correspond to the fixed, random effects and errors, respectively. Assume, w.l.o.g., that $n_i \geq 1$, $1 \leq i \leq p$. Let $X = \text{diag}(1_{n_i} \otimes 1_2)$, $Z = \text{diag}(1_{n_i} \otimes I_2)$, then the model can be written as

(52) $$y = X\beta + Z\alpha + \varepsilon.$$

So $N = 2\sum_{i=1}^{p} n_i$. Direct calculation shows

$$H = Z'AA'Z = Z'(I_N - X(X'X)^{-1}X')Z = \text{diag}(n_i(I_2 - 2^{-1}J_2)),$$

where $J_2 = 1_2 1_2'$, and $Z'AV(\mu)^{-1}A'Z = H - H(\mu^{-1}I_{2p} + H)^{-1}H = \text{diag}(n_i(1 + \mu n_i)^{-1}(I_2 - 2^{-1}J_2))$. Thus

$$\text{tr}(V_1(\theta)) = \text{tr}(Z'AV(\mu)^{-1}A'Z) = \sum_{i=1}^{p} \frac{n_i}{1 + \mu n_i},$$

$$\text{tr}(V_1^2(\theta)) = \text{tr}((Z'AV(\mu)^{-1}A'Z)^2) = \sum_{i=1}^{p} \left(\frac{n_i}{1 + \mu n_i}\right)^2.$$

Let $q = \sum_{n_i > 1}(n_i - 1)$, $r = q/p$, then it is easy to show that

(53) $$\liminf \lambda_{\min}(\text{Cor}(V_0(\theta), V_1(\theta))) > 0$$
$$\text{for all } \theta \in \Theta^\circ = \{\theta = (\lambda, \mu) : \lambda > 0, \mu > 0\}$$

iff $\limsup p/N < 1/2$ iff $\liminf r > 0$. Since $q \geq |\{1 \leq i \leq p : n_i > 1\}|$, $\liminf r = 0$ would mean the model is asymptotically confounded.

Note $s = 1$ in this example, so (32)–(35) become $(a+b)(1-b)^{-1} < 0.2$, $3a + 2b < f < 1 - 2a - 4b$, $0 < g < (0.5f - 1.5a - b) \wedge (1 - 2a - 4b - f)$, $0 < h < (f - 3a - 2b - 2g) \wedge (1 - 2a - 4b - f - g)$.

The next example is a counterexample showing that AI^4 is not enough for Wald consistency.

EXAMPLE 5.2. Let $\lambda_1 = 0, \lambda_2, \ldots, \lambda_N$ be positive numbers such that

(54) $$\limsup \frac{\sum_{i=1}^{N} \lambda_i/(1 + \mu\lambda_i)}{N^{1/2}(\sum_{i=1}^{N}(\lambda_i/(1 + \mu\lambda_i))^2)^{1/2}} < 1,$$

and

(55) $$\lim \sum_{i=1}^{N} \left(\frac{\lambda_i}{1 + \mu\lambda_i}\right)^2 = \infty$$

for all $\mu \geq 0$ [e.g., $\lambda_2 = \cdots = \lambda_{[N/2]} = a$, $\lambda_{[N/2]+1} = \cdots = \lambda_N = b$, where $a, b > 0$, $a \neq b$].

Consider

(56) $$y_i = \sqrt{\lambda_i} \alpha_i + \varepsilon_i, \qquad i = 1, \ldots, N,$$

where α_i's are random effects, ε_i's are errors satisfying $P(\varepsilon_1 = 0) > 0$. Thus

$$y = Z\alpha + \varepsilon,$$

where $Z = \text{diag}(\sqrt{\lambda_i})$. Equations (54) and (55) imply $\liminf \lambda_{\min}(\text{Cor}(V_0(\theta), V_1(\theta))) > 0$ and $\lim \|V_i(\theta)\|_R = \infty$, $i = 0, 1$ for all $\theta \in \Theta = \{\theta = (\lambda, \mu): \lambda > 0, \mu \geq 0\}$, so the model is AI^4 not only at all $\theta \in \Theta^\circ$ but all $\theta \in \Theta$.

Assume $\phi_{(0)} \in \Phi^o$. It is easy to derive

(57)
$$\tilde{L}_N(\phi) - \tilde{L}_N(\phi_{(0)}) = \frac{1}{2}\left\{ \sum_{i=1}^N \log\left(\frac{\sigma_{00}^2 + \sigma_{01}^2 \lambda_i}{\sigma_0^2 + \sigma_1^2 \lambda_i} \right) \right.$$
$$\left. + \sum_{i=1}^N \left(1 - \frac{\sigma_{00}^2 + \sigma_{01}^2 \lambda_i}{\sigma_0^2 + \sigma_1^2 \lambda_i} \right) w_i^2 \right\},$$

where $w_i = (\varepsilon_i + \sqrt{\lambda_i}\alpha_i)/\sqrt{\sigma_{00}^2 + \sigma_{01}^2 \lambda_i}$, $i = 1, \ldots, N$.

Let $B = \{\phi = (\sigma_0^2, \sigma_1^2): \sigma_i^2 \geq \sigma_{0i}^2/2, \ i = 0, 1\}$, $\phi^* = (\sigma_{*0}^2, \sigma_{01}^2)$ with

$$0 < \sigma_{*0}^2 < \sigma_{00}^2 \exp\left\{ -3\delta^{-1}\left(N + (\sigma_{00}^2/\sigma_{01}^2)\left(\sum_{i=2}^N \lambda_i^{-1} \right) \right) - N \log 2 \right\},$$

where $\delta = P_{\phi_{(0)}}(\varepsilon_1 = 0)$. Define $\xi_N = (1/N)\sum_{i=1}^N w_i^2$, $\eta_N = \sum_{i=2}^N \lambda_i^{-1} w_i^2 / \sum_{i=2}^N \lambda_i^{-1}$, $E = \{\varepsilon_1 = 0, \ \xi_N \vee \eta_N \leq 3\delta^{-1}\}$. Then it follows from (57) that on E

$$\tilde{L}_N(\phi) \leq \tilde{L}_N(\phi_{(0)}) + (N/2)(\log 2 + \xi_N)$$

$$< \tilde{L}_N(\phi_{(0)}) + (1/2)\left\{ \log \frac{\sigma_{00}^2}{\sigma_{*0}^2} - \left(\frac{\sigma_{00}^2}{\sigma_{01}^2} \right)\left(\sum_{i=2}^N \lambda_i^{-1} \right) \eta_N \right\}$$

$$< \tilde{L}_N(\phi^*), \qquad \phi \in B.$$

Thus

$$P_{\phi_{(0)}}\left(\sup_{\phi \in B} \tilde{L}_N(\phi) < \tilde{L}_N(\phi^*) \right) \geq P_{\phi_{(0)}}(E) \geq \delta/3.$$

Therefore with probability greater than some positive number the maximum point $\hat{\phi}_N$ will not fall into B no matter how large N is. Note that in this example the REML estimates are the same as the MLE.

6. Concluding remarks.

6.1. In virtually all problems regarding Wald consistency, there are two sources of difficulties. The first is the arbitrariness of the parameters (here, θ or ϕ). The second is the randomness of some quantities involved in the likelihood function (here, $z = A'y$). Somehow, one does not want these two kinds of things to be tied up, and this turns out to be the technical origin of our proofs.

As we mentioned in Section 2, the basic idea is to prove the difference, say $L_N(\theta) - L_N(\theta_{(0)})$, is negative for θ outside a small neighborhood of $\theta_{(0)}$. Since the first term in the difference decomposition [e.g., (7)] is nonrandom, the focus is therefore on the second term (i.e., the d term).

In the balanced case, we are able to further decompose $d_N(\theta, \theta_{(0)})$ as a sum such that each summand is a product of two factors—the first depends only on θ and the second only on z, and the number of terms in the sum is *bounded*. This technique of separating the two kinds of difficulties mentioned above has been proved totally successful since it then becomes clear how to construct the "small neighborhood" (see the proof of Theorem 3.1).

Unfortunately, this nice decomposition of the d term no longer exists in the unbalanced case (thanks to a counterexample we are able to construct which helps clarify that). In the unbalanced situation we go back to the old idea of Wald (1949), that is, dividing the parameter space by small regions and approximating the likelihood within each small region by its value at, say, the center of the region. Note that once ϕ is fixed at some point ϕ^*, $\tilde{L}_N(\phi) = \tilde{L}_N(\phi^*)$ becomes a function of z alone. So this technique often does the job of separating ϕ and z except for one thing: *the number of such small regions has to be "well under control,"* which is impossible, as we have found, without using the method of sieves. Since we are dealing with some nonstandard situations, namely, quadratic forms of non-Gaussian random variables which can not be expressed as "i.i.d. sums," special techniques are needed to evaluate carefully the orders of different quantities and hence eventually construct the *sieve*. These include an expansion of some matrix-valued function and an inequality for the variance of a quadratic form.

Finally, the counterexample has made it clear that sieve-Wald consistency is all one can expect in a general unbalanced non-Gaussian situation.

6.2. In Section 4 a sequence of approximating spaces (i.e., a sieve) is constructed, and the consistency of the resulting estimates can be ensured under AI^4. Although the AI^4 condition seems minimal for a theorem of this generality, the convergence rate of the sieve to the full parameter space could be slow. The rate can be much improved if, for example, the true parameter vector is known to belong to a compact subspace of the interior of the parameter space, say, $\phi_0 \in [\delta, M]$, where $0 < \delta < M < \infty$ (e.g., $\delta = 10^{-6}$, $M = 10^6$). In such a case the actual parameter space is $\Phi_0 = [\delta, M]$, each approximating space can be taken as Φ_0 (i.e., $\delta_N = \delta_B = \delta$, $M_N = M_B = M$) and the corresponding sequence $\{\hat{\phi}_N\}$ is consistent [see (50) in the proof of Theorem 4.1].

6.3. However, the AI^4 condition is not sufficient for Wald consistency. Additional assumptions are needed, either on the structure of the model or on distributions of the random effects and errors. For example, it would be interesting to see whether Wald consistency holds when normality actually holds (in which case Gaussian likelihood is the true likelihood). Note that in Example 5.2 the distribution of the errors is not normal.

REFERENCES

BILLINGSLEY, P. (1986). *Probability and Measure*. Wiley, New York.

CHAN, N. N. and KWONG, M. K. (1985). Hermitian matrix inequalities and a conjecture. *Amer. Math. Monthly* **92** 533–541.

CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press.

CRESSIE, N. and LAHIRI, S. N. (1993). The asymptotic distribution of REML estimators. *J. Multivariate Anal.* **45** 217–233.

DAS, K. (1979). Asymptotic optimality of restricted maximum likelihood estimates for the mixed model. *Calcutta Statist. Assoc. Bull.* **28** 125–142.

GRENANDER, U. (1981). *Abstract Inference*. Wiley, New York.

HARTLEY, H. O. and RAO, J. N. K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika* **54** 93–108.

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

JIANG, J. (1996). REML estimation: asymptotic behavior and related topics. *Ann. Statist.* **24** 255–286.

KHURI, A. I. and SAHAI, H. (1985). Variance components analysis: a selective literature survey. *Internat. Statist. Rev.* **53** 279–300.

LE CAM, L. (1979). Maximum likelihood: an introduction. *Lecture Notes in Statist.* **18**. Univ. Maryland, College Park.

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

NEYMAN, J. and SCOTT, E. (1948). Consistent estimates based on partially consistent observations. *Econometrika* **16** 1–32.

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

RAO, J. N. K. (1977). Contribution to the discussion of "Maximum likelihood approaches to variance component estimation and to related problems" by D. A. Harville. *J. Amer. Statist. Assoc.* **72** 338–339.

RICHARDSON, A. M. and WELSH, A. H. (1994). Asymptotic properties of restricted maximum likelihood (REML) estimates for hierarchical mixed linear models. *Austral. J. Statist.* **36** 31–43.

ROBINSON, D. L. (1987). Estimation and use of variance components. *Statistician* **36** 3–14.

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

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

THOMPSON, JR., W. A. (1962). The problem of negative estimates of variance components. *Ann. Math. Statist.* **33** 273–289.

WALD, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20** 595–601.

DEPARTMENT OF STATISTICS
CASE WESTERN RESERVE UNIVERSITY
10900 EUCLID AVENUE
CLEVELAND, OHIO 44106-7054
E-MAIL: jiang@reml.cwru.edu