The Annals of Statistics
1996, Vol. 24, No. 1, 255–286

REML ESTIMATION: ASYMPTOTIC BEHAVIOR AND RELATED TOPICS

By Jiming Jiang

University of California, Berkeley

The restricted maximum likelihood (REML) estimates of dispersion parameters (variance components) in a general (non-normal) mixed model are defined as solutions of the REML equations. In this paper, we show the REML estimates are consistent if the model is asymptotically identifiable and infinitely informative under the (location) invariant class, and are asymptotically normal (A.N.) if in addition the model is asymptotically nondegenerate. The result does not require normality or boundedness of the rank p of design matrix of fixed effects. Moreover, we give a necessary and sufficient condition for asymptotic normality of Gaussian maximum likelihood estimates (MLE) in non-normal cases. As an application, we show for all unconfounded balanced mixed models of the analysis of variance the REML (ANOVA) estimates are consistent; and are also A.N. provided the models are nondegenerate; the MLE are consistent (A.N.) if and only if certain constraints on p are satisfied.

1. Introduction. The restricted or residual maximum likelihood (REML) method was proposed by Thompson (1962) as a way of estimating dispersion parameters associated with linear models. Several authors have given overviews on REML, which will be given in the sequel.

Although the REML method has been used and studied over the past 30 years, questions remain on how good REML is compared with other estimates. Some of the questions are related to the asymptotic behavior of the REML estimates, especially when the rank p of design matrix of fixed effects tends to infinity. In such cases it is well known, by the Neyman–Scott example [Neyman and Scott (1948)], that the maximum likelihood estimates (MLE) can be inconsistent. What can we say about the REML estimates? And under what condition will the MLE be consistent (asymptotically normal)? Furthermore, can the REML estimates obtained under normality still perform well asymptotically in nonnormal cases? In particular, is it true that for balanced data the ANOVA estimates, which agree with solutions of REML equations under normality, are always consistent even if normality does not hold and $p \to \infty$? These questions, along with others, will be investigated in this paper.

The REML method was put on a broad basis for unbalanced data by Patterson and Thompson (1971). Surveys of REML can be found in articles of

Received May 1994; revised May 1995

AMS 1991 *subject classification*. 62F12.

Key words and phrases. Mixed models, restricted maximum likelihood, MLE, ANOVA, consistency, asymptotic normality.

Harville (1977), Khuri and Sahai (1985) and Robinson (1987) and in a recent book by Searle, Casella and McCulloch (1992). Different derivations of the REML show it may also be regarded as a method of marginal likelihood [Harville (1974) and Verbyla (1990)] or modified profile likelihood [Barndorff-Nielsen (1983)]. Other areas in which REML has been used include the following: estimating smoothing parameters in penalized estimation [Wahba (1990); see Speed (1991) for discussion]; the estimation of parameters in ARMA processes and other time series in the presence of fixed effects [Cooper and Thompson (1977) and Azzalini (1984)]; REML estimation in spatial models [Green (1985) and Gleeson and Cullis (1987)]; the analysis of longitudinal data [Laird and Ware (1982)]; and REML estimation in empirical Bayes smoothing of the census undercount [Cressie (1992)].

Consider a general mixed model

$$y = X\beta + Z_1\alpha_1 + \cdots + Z_s\alpha_s + \varepsilon, \tag{1}$$

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 an $m_i \times 1$ vector of i.i.d. random variables with mean 0 and variance σ_i^2 (the random effects), $i = 1, \dots, s$; and 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 for the mixed model (1) are few in number, with or without normality assumptions. Assuming normality and assuming that the model has a standard ANOVA structure with the number $p = \text{rank}(X)$ fixed, Miller (1977) considered the MLE for both fixed effects and variance components $\sigma_0^2, \sigma_1^2, \dots, \sigma_s^2$. He formulated a set of conditions under which the consistency and asymptotic normality of a sequence of solutions of the likelihood equations were proved. He also noted that normalizing sequences of different orders of magnitude might be required for estimates of different parameters. Under conditions slightly stronger than those of Miller (in particular, with normality and p fixed), Das (1979) obtained a similar result for the REML estimates and found that in his situation the REML estimates and the MLE are in some sense equivalent. In a quite different direction, Speed (1986) proved that in the balanced case with $p = 1$ the usual ANOVA estimates of variance components are consistent without assuming normality. Also without normality, Westfall (1986) obtained asymptotic normality of the ANOVA estimates of variance components for unbalanced mixed models with a nested structure; Brown (1976) proved asymptotic normality of C. R. Rao's MINQUE, and the so-called I-MINQUE [e.g., Rao and Kleffe (1988), Section 9.1] under replicated error structure [e.g., Anderson (1973)]. Recently, asymptotic behavior of the REML estimates was discussed by Cressie and Lahiri (1993) and by Richardson and Welsh (1994). Normality was assumed in the first paper but not in the second, although the second was restricted to hierarchical (nested) models. However, p was held fixed in both studies. It should be pointed out that when p is fixed or bounded, the REML estimates and the MLE for the variance components are equivalent in the sense that their suitably normalized difference converges to zero in probability (and hence there would be no essential difference asymptotically between the two estimates). It follows that the boundedness of p is a serious restriction, and an important and interesting question regarding the (possible) superiority of REML over straight ML in estimating the variance components is how do the REML estimates behave asymptotically when $p \to \infty$.

Our main goal is to establish a general theorem on the asymptotic behavior of the REML estimates for the dispersion parameters (variance components) in model (1) without any assumption on the structure of the model (such as balancedness, nestedness or ANOVA design), boundedness of p and normality. This may sound confusing because the REML estimates are usually, if not always, defined under normality. One obvious approach is to use the true likelihood of $A'y$, where A is an $N \times (N-p)$ matrix of full rank with $A'X = 0$. However, such likelihood may not have a simple closed form. Furthermore, the estimates thus obtained may depend on A, which will not occur with the normal likelihood. An alternative is to treat the REML estimates as a kind of M-estimates [e.g., Huber (1981)], that is, solutions of the REML equations, taking into account the nonnegativity constraints. We employed the second definition. Similarly we treat the Gaussian MLE in the nonnormal cases.

Before we give the general result, we want to take a look at a simple case of (1), the balanced case. For balanced data the REML solutions are identical to the ANOVA estimates, and this is true whether normality is assumed or not [e.g., Searle, Casella and McCulloch (1992), page 253]. It is well known that for balanced data the ANOVA estimates are best quadratic unbiased estimates (best unbiased estimates under normality). However, it is not clear whether in all balanced cases the ANOVA estimates are consistent (asymptotically normal) even without normality and possibly with $p \to \infty$, although one would expect such a result. Results covering special cases are available [see Speed (1986) and Westfall (1986)]. It is well known that, with $p \to \infty$, the MLE for variance components can be inconsistent, but it is also not clear under exactly what conditions the MLE will be consistent (asymptotically normal). In particular, is it true that under normality and with p fixed for all balanced mixed ANOVA models the MLE are asymptotically normal and efficient in the sense of attaining the Cramér–Rao lower bound? Such results are, of course, expectable, and simple examples have been discussed in, for example, Hartley and Rao (1967) and Miller (1977). We give complete answers to these questions. For all balanced mixed models of the analysis of variance, the REML (ANOVA) estimates are consistent, provided the models are not confounded and variance components are positive; they are also asymptotically normal if, furthermore, the models are nondegenerate. The MLE are consistent (asymptotically normal) if and only if certain constraints on p are satisfied. In particular, the answer to the last question on efficiency is positive.

The general result for model (1) turns out to be the following. The REML estimates are consistent if the model is asymptotically identifiable and infinitely informative under the invariant class (AI^4); they are also asymptotically normal if, furthermore, the model is asymptotically nondegenerate (AND). We will make it clear what AI^4 and AND mean. A necessary and sufficient condition for asymptotic normality of the MLE is also given.

The proof of the above result is based on a central limit theorem for quadratic forms of random variables, which seems to be of interest itself in limit theorems.

In Section 2 we define our estimates. Main results are given and explained in Section 4. Section 5 develops some central limit theorems for the quadratic forms. Comments and remarks are made in Section 6 and proofs given in Section 7.

2. Definitions of the estimates. Following Hartley and Rao (1967) (but using different notation), we consider the following parameters of variance components: $\lambda = \sigma_0^2$, $\mu_i = \sigma_i^2/\sigma_0^2$, $i = 1, \dots, s$. There is a 1–1 correspondence between the two sets of parameters λ, μ_i, $1 \le i \le s$ and σ_i^2, $0 \le i \le s$, and all results we obtain in this paper for the first set of parameters have analogues for the second set. Therefore we will focus on the first set of parameters. The parameter space is

$$\Theta = \{\theta\colon \lambda > 0,\ \mu_i \ge 0,\ i = 1, \dots, s\}, \tag{2}$$

where $\theta = (\lambda, \mu_1, \dots, \mu_s)'$. Basic results such as deriving the REML and the maximum likelihood (ML) equations can be found in Searle, Casella and McCulloch (1992), Section 6. The REML equations under normality are equivalent to

$$z'V(A, \mu)^{-1} z = \lambda(N - p), \tag{3}$$

$$z'V(A, \mu)^{-1} A'Z_i Z_i' A V(A, \mu)^{-1} z = \lambda \operatorname{tr}\left(Z_i' A V(A, \mu)^{-1} A' Z_i\right), \quad 1 \le i \le s, \tag{4}$$

where $z = A'y$, A is an $N \times (N - p)$ matrix with

$$\operatorname{rank}(A) = N - p, \qquad A'X = 0, \tag{5}$$

and

$$V(A, \mu) = A'A + \sum_{i=1}^{s} \mu_i A'Z_i Z_i' A.$$

Note that (3) and (4) do not depend on the choice of A so long as (5) is satisfied.

In general (without assuming normality), the REML estimates for λ, μ_i, $i = 1, \dots, s$, are defined as solutions of (3) and (4) that belong to Θ, whenever such solutions exist.

Remark 2.1. Although the REML equations (3) and (4) are derived by assuming $y \sim N(X\beta, \lambda V_\mu)$, where

$$V_\mu = I_N + \sum_{i=1}^{s} \mu_i Z_i Z_i',$$

normal likelihood is not the only one that can lead to the REML equations. For example, suppose in (1) that y has a multivariate-t distribution with degree of freedom k, $y \sim t_N(X\beta, \lambda V_\mu, k)$, where $Y \sim t_n(\mu, \Sigma, k)$ if Y has density

$$p(y) = \frac{\Gamma((n+k)/2)}{(\pi k)^{n/2}\Gamma(k/2)} \det(\Sigma)^{-1/2}\left[1 + \frac{1}{k}(y-\mu)'\Sigma^{-1}(y-\mu)\right]^{-(n+k)/2}.$$

Such distributions have been used in multiple linear regression [e.g., Zellner (1976)]. It can be shown that under the multivariate-t distribution, the likelihood of $A'y$ again leads to (3) and (4).

Similarly, the MLE for both the fixed effects and λ, μ_i, $i = 1, \dots, s$, are defined as solutions of the ML equations under normality with the same constraints on the λ and μ_i's. The ML equations are equivalent to

$$(X'V_\mu^{-1}X)\beta = X'V_\mu^{-1}y, \tag{6}$$

$$z'V(A,\mu)^{-1}z = \lambda N, \tag{7}$$

$$z'V(A,\mu)^{-1}A'Z_iZ_i'AV(A,\mu)^{-1}z = \lambda \operatorname{tr}(Z_i'V_\mu^{-1}Z_i), \qquad 1 \le i \le s. \tag{8}$$

In this paper our interest is the MLE for the parameters of variance components, namely, the λ and μ_i's. So in the following the term MLE will refer to the MLE for the λ and μ_i's.

It is seen from (6)–(8) that the MLE belong to the class of (location) invariant estimates (invariant class)

$$\mathscr{I} = \{\text{estimates which are function of } A'y \text{ with } A \text{ satisfying (5)}\} \tag{9}$$

[see Rao and Kleffe (1988), Section 4.4]. Other estimates that belong to $\mathscr{I}$ include the ANOVA estimates for variance components in a mixed model [Henderson's method III, (e.g., Searle, Casella and McCulloch (1992), Section 5)] and some of the MINQE's [e.g., Rao and Kleffe (1988), Sections 5 and 9.1]. Note that the REML estimates are the MLE based on $A'y$. From this point of view, the REML method seems to lose no information in estimating the parameters of variance components, and there is reason to expect the REML to behave well asymptotically, as will be seen next.

3. Notation. Let $A, B, A_1, \dots, A_s$ be matrices, and let $a_1, \dots, a_s$ be numbers. Define

$$\|A\| = \lambda_{\max}^{1/2}(A'A), \qquad \|A\|_R = \operatorname{tr}^{1/2}(A'A),$$

$$\langle A, B\rangle_R = \operatorname{tr}(A'B), \qquad \operatorname{cor}(A,B) = \frac{\langle A, B\rangle_R}{\|A\|_R\|B\|_R}, \qquad A, B \ne 0,$$

$\operatorname{Cor}(A_1, \dots, A_s) = ((\operatorname{cor}(A_i, A_j))$ if $A_1, \dots, A_s \ne 0$, and is 0 otherwise; A_{ll} is the lth diagonal element of A; $\operatorname{diag}(a_i)$ is the diagonal matrix with diagonal elements a_i, $i = 1, \dots, s$; I_n and 1_n are the n-dimensional identity matrix and vector of all 1's, respectively. Let $\theta_0 = (\lambda_0, \mu_0')'$ be the true parameter vector, and let $p_i(N)$, $i = 0, \dots, s$, be sequences of positive numbers; write $b(\mu) = (I_N\ \sqrt{\mu_1}Z_1 \cdots \sqrt{\mu_s}Z_s)'$,

$$V(\mu) = AV(A,\mu)^{-1}A', \qquad V_0(\mu) = b(\mu)V(\mu)b(\mu)',$$
$$V_i(\mu) = b(\mu)V(\mu)Z_iZ_i'V(\mu)b(\mu)', \qquad i = 1,\dots,s,$$
$$U_0 = \frac{I_N}{\lambda_0}, \qquad U_i = V_{\mu_0}^{-1/2}Z_iZ_i'V_{\mu_0}^{-1/2},$$
$$V_0 = \frac{I_{N-p}}{\lambda_0}, \qquad V_i = V(A,\mu_0)^{-1/2}A'Z_iZ_i'AV(A,\mu_0)^{-1/2}, \qquad i = 1,\dots,s,$$
$$I_{ij}^{(N)} = \frac{\operatorname{tr}(V_iV_j)}{p_i(N)p_j(N)}, \qquad I_{ij}^{*(N)} = \frac{\operatorname{tr}(U_iU_j)}{p_i(N)p_j(N)},$$
$$K_{ij}^{(N)} = \frac{1}{p_i(N)p_j(N)}\sum_{l=1}^{N+m}(EW_{Nl}^4 - 3)V_i(\mu_0)_{ll}V_j(\mu_0)_{ll}/\lambda_0^{1_{(i=0)}+1_{(j=0)}},$$
$$i, j = 0, 1, \dots, s,$$

where $m = m_1 + \cdots + m_s$,

$$W_{Nl} = \begin{cases} \dfrac{\varepsilon_l}{\sqrt{\lambda_0}}, & 1 \le l \le N, \\[2ex] \dfrac{\alpha_{il-N-\sum_{k<i}m_k}}{\sqrt{\lambda_0\mu_{0i}}}, & N + \sum_{k<i}m_k + 1 \le l \le N + \sum_{k\le i}m_k,\ 1 \le i \le s. \end{cases}$$

Define $I_N(\theta_0) = (I_{ij}^{(N)})$, $I_N^*(\theta_0) = (I_{ij}^{*(N)})$, $K_N(\theta_0) = (K_{ij}^{(N)})$ and $J_N(\theta_0) = 2I_N(\theta_0) + K_N(\theta_0)$.

The abbreviation w.p. $\to 1$ refers to "with probability tending to 1"; the abbreviation v.c., to "variance components."

4. Main results. First we note that in considering consistency and asymptotic normality of our estimates as $N \to \infty$, each m_i can be, w.l.o.g., considered as a function of N. Since such results hold iff they hold for each sequence with N increasing strictly monotonically, in which case the m_i's can readily be regarded as functions of N. (Note that the y, X, Z_i's, A, etc., also depend on N.) The following assumptions A1 and A2 are made for model (1). Let $\alpha_0 = \varepsilon$ and $m_0 = N$.

A1. For each N, $\alpha_0, \alpha_1, \dots, \alpha_s$ are mutually independent;

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

$$\lim_{x\to\infty}\sup_N \max_{0\le i\le s} E\alpha_{i1}^4 1_{(|\alpha_{i1}|>x)} = 0. \tag{10}$$

NOTE. If the common distribution of $\alpha_{i1}, \dots, \alpha_{im_i}$ is assumed not to depend on N, $0 \le i \le s$, as is usually the case, then (10) is equivalent to the existence of the fourth moments of α_{i1}, $0 \le i \le s$. In general, (10) implies $\sup_N \max_{0 \le i \le s} E\alpha_{i1}^4 < \infty$ but not the converse.

DEFINITION 4.1. We say model (1) has *positive v.c.* (p.v.c.) if the true parameter vector of v.c. is an interior point of Θ; and it is *nondegenerate* (ND) if

$$\inf_N \min_{0 \le i \le s} \operatorname{var}(\alpha_{i1}^2) > 0. \tag{11}$$

A sequence of estimates $\{(\hat{\lambda}_N, \hat{\mu}_{N1}, \dots, \hat{\mu}_{Ns})'\}$ is called asymptotically normal (AN) if there are sequences of positive numbers $p_i(N) \to \infty$, $0 \le i \le s$, and a sequence of matrices $\{M_N(\theta_0)\}$ such that $\limsup(\|M_N^{-1}(\theta_0)\| \vee \|M_N(\theta_0)\|) < \infty$ and

$$\begin{aligned} M_N(\theta_0)\big(p_0(N)(\hat{\lambda}_N - \lambda_0),& \\ &p_1(N)(\hat{\mu}_{N1} - \mu_{01}), \dots, p_s(N)(\hat{\mu}_{Ns} - \mu_{0s})\big)' \\ \to_{\mathscr{L}} N(0, I_{s+1}).& \end{aligned} \tag{12}$$

Two sequences $\{p_N\}$ and $\{q_N\}$ are called equivalent, denoted by $p_N \sim q_N$, if $0 < \liminf(p_N/q_N) \le \limsup(p_N/q_N) < \infty$.

4.1. *The balanced case.* A general balanced r-factor mixed model (of the analysis of variance) can be expressed (after possible reparametrization) in the form (1), where X and the Z_i's are Kronecker products [e.g., Searle, Casella and McCulloch (1992), Section 4.6; Rao and Kleffe (1988), pages 172–173]. By introducing indexes in $S_{r+1} = \{0, 1\}^{r+1}$, this can be written as

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

where $X = 1_{n_1}^{d_1} \otimes \cdots \otimes 1_{n_{r+1}}^{d_{r+1}} = \otimes_{q=1}^{r+1} 1_{n_q}^{d_q}$, with $d = (d_1, \dots, d_{r+1}) \in S_{r+1}$, $Z_i = \otimes_{q=1}^{r+1} 1_{n_q}^{i_q}$ with $i = (i_1, \dots, i_{r+1}) \in S \subset S_{r+1}$, $1_n^0 = I_n$ and $1_n^1 = 1_n$. Hence $N = \prod_{q=1}^{r+1} n_q$, $p = \prod_{d_q = 0} n_q$ and $m_i = \prod_{i_q = 0} n_q$, $i \in S$ ($\prod_{\varnothing} \cdot \equiv 1$).

EXAMPLE 4.1. $y_{ijk} = \mu + a_i + b_j + c_{ij} + \varepsilon_{ijk}$, $1 \le i \le I$, $1 \le j \le J$, $1 \le k \le K$, where a, b and c are random effects with c corresponding to the interaction (between factors associated with a and b). The model can be written as

$$\begin{aligned} y = {} & (1_I \otimes 1_J \otimes 1_K)\mu + (I_I \otimes 1_J \otimes 1_K)a + (1_I \otimes I_J \otimes 1_K)b \\ & + (I_I \otimes I_J \otimes 1_K)c + \varepsilon. \end{aligned}$$

Thus $r = 2$, $n_1 = I$, $n_2 = J$, $n_3 = K$; $N = IJK$; $d = (1\ 1\ 1)$, $p = 1$; $S = \{(0\ 1\ 1), (1\ 0\ 1), (0\ 0\ 1)\}$, $m_{(0\ 1\ 1)} = I$, $m_{(1\ 0\ 1)} = J$, $m_{(0\ 0\ 1)} = IJ$. This model was discussed by Miller (1977), where he showed that under normality and $I, J \to \infty$ (which implies $N \to \infty$ and $m_i \to \infty$, $i \in S$) the MLE were AN.

EXAMPLE 4.2. $y_{ijkl} = \alpha_i^{(1)} + \alpha_{ij}^{(2)} + \alpha_{ik}^{(3)} + \beta_k^{(1)} + \beta_l^{(2)} + \varepsilon_{ijkl}$, $1 \le i \le a$, $1 \le j \le b$, $1 \le k \le c$, $1 \le l \le d$, where $\beta^{(1)}$ and $\beta^{(2)}$ are fixed main effects, $\alpha^{(1)}$, $\alpha^{(2)}$ and $\alpha^{(3)}$ are random effects corresponding to a random main effect, a nested random factor and a fixed-by-random interaction. After reparametrization, namely, letting $\beta_{kl} = \beta_k^{(1)} + \beta_l^{(2)}$, $\beta = (\beta_{kl})$, the model can be written as (13).

EXAMPLE 4.3 (Neyman–Scott problem). $y_{ij} = \mu_i + \varepsilon_{ij}$, $1 \le i \le n$, $j = 1, 2$. This corresponds to (13) with $S = \emptyset$, $X = I_n \otimes 1_2$, $p = n$. It was shown by Neyman and Scott (1948) that as $p \to \infty$ the MLE for σ_ε^2 is inconsistent. However, the REML estimates are known to be AN [Hammerstrom (1978)].

EXAMPLE 4.4 (Random model). When $p = 1$, (13) is called a balanced random (effects) model. Speed (1986) proved the consistency of the ANOVA estimates in such a model without assuming normality.

EXAMPLE 4.5 (Nested design). A balanced nested or hierarchical model is (13) with $\{d\} \cup S$ being a completely ordered subset of S_{r+1} ($u \le v$ iff $u_q \le v_q$, $1 \le q \le r+1$ gives a partial order in S_{r+1}). Westfall (1986) showed that under certain conditions the ANOVA estimates are AN. The result did not require normality or balancedness, although p was assumed to satisfy $p/N \to 0$ (therefore it did not cover Example 4.3).

The above examples are special cases of two general theorems which we will state in the sequel.

DEFINITION 4.2. A general mixed ANOVA model (not necessarily balanced) is called unconfounded if (i) the fixed effects are not confounded with the random effects and errors [i.e., $\operatorname{rank}(X, Z_i) > p$, $\forall i$ and $X \ne I_N$] and (ii) the random effects and errors are not confounded [i.e., the matrices I_N and $Z_i Z_i'$, $i \in S$, are linearly independent [e.g., Miller (1977)].

THEOREM 4.1. *Let the balanced model* (13) *be unconfounded and have p.v.c. As $N \to \infty$ and $m_i \to \infty$, $i \in S$, the following hold*:

(i) *There exist w.p. $\to 1$ REML estimates $\hat{\lambda}_N$ and $\hat{\mu}_{Ni}$, $i \in S$, which are 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 is ND, then the REML estimates in* (i) *are AN with $p_0(N) = \sqrt{N-p}$ and $p_i(N) = \sqrt{m_i}$, $i \in S$, and $M_N(\theta_0) = J_N^{-1/2}(\theta_0) I_N(\theta_0)$.*

REMARK 4.1. The conclusions are also true for the ANOVA estimates [e.g. Searle, Casella and McCulloch (1992), page 253].

REMARK 4.2. There is no restriction on p in Theorem 4.1. For example, in Example 4.3, $N \to \infty$ iff $p \to \infty$.

Let $u, v \in S_{r+1}$; define $u \vee v = (u_1 \vee v_1, \dots, u_{r+1} \vee v_{r+1})$, $S_u = \{v \in S: v \le u\}$, $m_u = \prod_{u_q = 0} n_q$ and $m_{u,S} = \min_{v \in S_u} m_v$ if $S_u \ne \emptyset$ and 1 if $S_u = \emptyset$.

THEOREM 4.2. *Let the balanced model* (13) *be unconfounded and have p.v.c. As $N \to \infty$ and $m_i \to \infty$, $i \in S$, the following hold*:

(i) *There exist w.p.* $\to 1$ *MLE which are consistent if and only if*

$$\frac{p}{N} \to 0, \qquad \frac{m_{i \vee d} m_{i \vee d, S}}{m_i^2} \to 0, \qquad i \in S. \tag{14}$$

(ii) *If*, *moreover*, *the model is ND*, *then there exist w.p.* $\to 1$ *MLE which are AN if and only if*

$$p_0(N) \sim \sqrt{N - p}, \qquad p_i(N) \sim \sqrt{m_i}, \qquad i \in S, \tag{15}$$

and

$$\frac{p}{\sqrt{N}} \to 0, \qquad \frac{m_{i \vee d} m_{i \vee d, S}}{m_i^{3/2}} \to 0, \qquad i \in S. \tag{16}$$

When (16) *is satisfied*, *the MLE are AN with the same $p_i(N)$, $i \in \{0\} \cup S$, and $M_N(\theta_0)$ as for the REML estimates*.

4.2. *The general case*. The assumption of a mixed ANOVA model not being confounded is a natural requirement for the v.c. to be "identifiable." More generally we have the following.

DEFINITION 4.3. A v.c. model

$$Y = (Y_1, \dots, Y_N)' = X\beta + \varepsilon, \tag{17}$$

where $E\varepsilon = 0$ and $\mathrm{Var}(\varepsilon) = \Sigma(\theta) = \theta_1 \Sigma_1 + \cdots + \theta_r \Sigma_r$, is said to be identifiable of its v.c. (ID) if the matrices $\Sigma_1, \dots, \Sigma_r$ are linearly independent.

Note that our definition of identifiability is equivalent to requiring that every parameter θ_i, $1 \le i \le r$, be identifiable in the sense of Rao and Kleffe [(1988), Section 4.2]. Let A be a matrix. Then

$$A'Y = A'X\beta + A'\varepsilon \tag{18}$$

is again a v.c. model like (17).

DEFINITION 4.4. Model (17) is said to be identifiable of its v.c. under the invariant class (IDI) if model (18) is ID for some $N \times (N - p)$ matrix A [$p = \mathrm{rank}(X)$] such that (5).

It is clear that model (17) is IDI iff (18) is ID for every A satisfying (5) iff $A'\Sigma_1 A, \dots, A'\Sigma_r A$ are linearly independent for every A (or some A) satisfying (5).

Now consider the general mixed model (1).

LEMMA 4.1. *Model* (1) *is IDI iff* $\lambda_{\min}(\mathrm{Cor}(V_0, V_1, \dots, V_s)) > 0$.

Note $\mathrm{Cor}(V_0, V_1, \dots, V_s)$ does not depend on the choice of A so long as (5).

In considering the asymptotic behavior of our estimates, we need model (1) to be IDI in the asymptotic sense. Lemma 4.1 inspires the following definition.

DEFINITION 4.5. We say model (1) is asymptotically identifiable (of its v.c.) under the invariant class, abbreviated by AI^2, at θ_0 if $\liminf \lambda_{\min}(\mathrm{Cor}(V_0, V_1, \dots, V_s)) > 0$.

We now take another look at the property AI^2. We now return to (12). The feature of this definition is that different normalizing sequences (NS) are used for estimates of different parameters. The necessity of this was noted by Miller (1977). Harville (1977) described Miller's NS as "the effective number of levels for the ith random factor ($i = 1, \dots, c$)." Searle, Casella and McCulloch [(1992), page 240] questioned how in general the NS should be chosen and asked "what is meant by 'sample size tending to infinity'." We have seen that in the balanced case there is virtually no other choice of NS (see Theorem 4.2). Now we consider the problem from another point of view.

Let $\hat{\theta}_N \in \mathscr{I}$ in (9) and let it satisfy (12). The asymptotic covariance matrix of $\hat{\theta}_N$ is

$$V_{\hat{\theta}_N} = \mathrm{diag}\left(\frac{1}{p_i(N)}\right)\left(M_N(\theta_0)' M_N(\theta_0)\right)^{-1} \mathrm{diag}\left(\frac{1}{p_i(N)}\right).$$

If we want our estimates to be efficient in some sense, we would like to see $V_{\hat{\theta}_N}$ to be not too far from the Cramér–Rao lower bound $I^{(N)}(\theta_0)^{-1}$, where

$$I^{(N)}(\theta_0) = -\left(E_{\theta_0}\left\{\frac{\partial^2 L_N}{\partial\theta_i\,\partial\theta_j}\bigg|_{\theta_0}\right\}\right)$$

(L_N is the log-likelihood of $A'y$); that is, there exist bounds $\delta, M > 0$ such that $\delta I^{(N)}(\theta_0) \le V_{\hat{\theta}_N}^{-1} \le M I^{(N)}(\theta_0)$, which, under normality, leads to the following requirement on the NS $p_i(N)s$:

$$0 < \liminf \lambda_{\min}(I_N(\theta_0)) \le \limsup \lambda_{\max}(I_N(\theta_0)) < \infty, \tag{19}$$

where $I_N(\theta_0)$ is as in Section 3 [see Miller (1977), Assumption 3.5].

That (19) is closely related to the AI^2 is seen in the following lemma.

LEMMA 4.2. *The following are equivalent*:

(i) *There are sequences of positive numbers* $p_i(N) \to \infty$, $0 \le i \le s$, *such that* (19);

(ii) $\|V_i\|_R \to \infty$, $0 \le i \le s$, *and the model is* AI^2 *at* θ_0.

In fact, whenever (i) *holds, we must have* $p_i(N) \sim \|V_i\|_R$, $0 \le i \le s$.

The quantities $\|V_i\|_R$ can be interpreted intuitively. Under normality,

$$\frac{1}{2}\|V_i\|_R^2 = -E_{\theta_0}\left[\frac{\partial^2 L_N}{\partial \theta_i^2}\bigg|_{\theta_0}\right],$$

which is the information that $A'y$ contains about the true parameter θ_{0i}, $0 \le i \le s$. This leads to the following definition.

DEFINITION 4.6. Model (1) is called infinitely informative (about is v.c.) under the invariant class at θ_0 if $\lim\|V_i\|_R = \infty$, $0 \le i \le s$.

The main theorem is now stated as follows.

THEOREM 4.3. *Consider a general mixed model* (1) *having p.v.c.*

(i) *If the model is asymptotically identifiable and infinitely informative under the invariant class at* θ_0, *then there exist w.p.*$\to 1$ *REML estimates* $\hat{\lambda}_N$ *and* $\hat{\mu}_{Ni}$, $1 \le i \le s$, *which are consistent, and the sequence*

$$\left\{\left(\sqrt{N-p}\,(\hat{\lambda}_N - \lambda_0), \|V_1\|_R(\hat{\mu}_{N1} - \mu_{01}), \dots, \|V_s\|_R(\hat{\mu}_{Ns} - \mu_{0s})\right)'\right\}$$

is bounded in probability.

(ii) *If, moreover, the model is ND, then the REML estimates in* (i) *are AN with* $p_0(N) = \sqrt{N-p}$, $p_i(N)$ *being any sequence* $\sim \|V_i\|_R$, $1 \le i \le s$, *and* $M_N(\theta_0) = J_N^{-1/2}(\theta_0) I_N(\theta_0)$.

ABBREVIATION. We use AI^4 for "asymptotically identifiable and infinitely informative under the invariant class."

NOTE. A necessary and sufficient condition for AI^4 is given by Lemma 4.2(i). In particular, all balanced mixed models (13) are AI^4, provided the models are unconfounded, have p.v.c., and $N \to \infty$, $m_i \to \infty$, $i \in S$ (see the proof of Theorem 4.1).

THEOREM 4.4. *Consider a general mixed model* (1) *having p.v.c.*

(i) *For the MLE to exist w.p.*$\to 1$ *and be consistent, it is necessary that*

$$\frac{p}{N} \to 0, \qquad \frac{\mathrm{tr}(C_i(\mu_0))}{m^*} \to 0, \qquad 1 \le i \le s, \tag{20}$$

where $C_i(\mu_0) = Z_i'(V_{\mu_0}^{-1} - V(\mu_0))Z_i$, $m^* = \max_{1 \le i \le s} m_i$.

(ii) *If, moreover, the model is ND, then the following are equivalent*:

(a) *There exist w.p.* $\to 1$ *MLE* $\hat{\lambda}_N^*$ *and* $\hat{\mu}_{Ni}^*$, $1 \le i \le s$, *which are AN with* $p_i(N)$, $0 \le i \le s$, *satisfying*

$$0 < \liminf \lambda_{\min}(I_N^*(\theta_0)) \le \limsup \lambda_{\max}(I_N^*(\theta_0)) < \infty; \tag{21}$$

(b) *The model is* AI^4 *at* θ_0 *and*

$$\frac{p}{\sqrt{N}} \to 0, \qquad \frac{\operatorname{tr}(C_i(\mu_0))}{\|V_i\|_R} \to 0, \qquad 1 \le i \le s. \tag{22}$$

In either case, the MLE and the REML estimates in Theorem 4.3 *are equivalent in the sense that they are AN for the same* $p_i(N)$, $0 \le i \le s$, *and* $M_N(\theta_0)$ *as in Theorem* 4.3(ii), *and*

$$\left(\sqrt{N-p}\left(\hat{\lambda}_N^* - \hat{\lambda}_N\right), p_1(N)(\hat{\mu}_{N1}^* - \hat{\mu}_{N1}), \dots, p_s(N)(\hat{\mu}_{Ns}^* - \hat{\mu}_{Ns})\right)' \to 0 \tag{23}$$

in probability.

Condition (21) is implied, for example, by Miller [(1977), Assumption 3.5], which also shows the dependence of Miller's NS on θ_0, and the relation between the two sets of parameters. See also Das (1979).

The assumption ND in Theorems 4.3(ii) and 4.4(ii) can be weakened to the following (24) called asymptotically nondegenerate (AND)

$$\liminf \lambda_{\min}(J_N(\theta_0)) > 0, \tag{24}$$

where $J_N(\theta_0)$ is given in Section 3 with $p_i(N) = \|V_i\|_R$, $0 \le i \le s$.

It can also been shown that under a condition weaker than (22) the MLE exist w.p. $\to 1$ and are consistent.

5. Some central limit theorems for quadratic forms. The proof of our main theorem is based on a central limit theorem for quadratic forms of random variables (r.v.'s). For each n, let $X_{n1}, \dots, X_{nk_n}$ be independent with mean 0, and let $A_n = (a_{nij})_{1 \le i, j \le k_n}$ be symmetric. There have been studies on the central (noncentral) limit theorems of the quadratic form $\mathscr{X}_n' A_n \mathscr{X}_n$, where $\mathscr{X}_n = (X_{n1}, \dots, X_{nk_n})'$. Some of the results are either for special kind of r.v.'s [e.g., Guttorp and Lockhart (1988)] or for A_n with a special structure [e.g., Fox and Taqqu (1985)], or with the assumption that $a_{ii} = 0$, $1 \le i \le k_n$ [e.g., de Jong (1987)].

A general theorem was given in Schmidt and Thrum (1981) and was extended by Rao and Kleffe [(1988), Theorem 2.5.2]. However, as was pointed out by Rao and Kleffe [(1988), page 51], "the application of (the theorem) might be limited as it is essentially based on the assumption that the off diagonal blocks of A_n tend to zero." Such results could be used for models with replicated error structure [e.g., Anderson (1973) and Brown (1976)], but not for general model (1).

We will state two theorems. The first removes the unpleasant restriction noted by Rao and Kleffe. The second extends the first. The results can be extended to the vector case considered by Schmidt and Thrum (1981) and Rao and Kleffe (1988). Extension to the case where X_{ni} are martingale differences is also possible. We begin with some simple examples.

EXAMPLE 5.1. If $X_{n1}, \dots, X_{nk_n}$ are $N(0,1)$ distributed, then a necessary and sufficient condition for

$$\frac{\mathscr{X}_n' A_n \mathscr{X}_n - E\mathscr{X}_n' A_n \mathscr{X}_n}{[\mathrm{var}(\mathscr{X}_n' A_n \mathscr{X}_n)]^{1/2}} \to_{\mathscr{L}} N(0,1) \tag{25}$$

is that

$$\frac{\lambda_{\max}(A_n^2)}{\mathrm{tr}(A_n^2)} \to 0. \tag{26}$$

EXAMPLE 5.2. Let $A_n = I_n$, $P(X_{ni} = -1) = P(X_{ni} = 1) = 1/2 - 1/n$, $P(X_{ni} = -\sqrt{2}) = P(X_{ni} = \sqrt{2}) = 1/(2n)$ and $P(X_{ni} = 0) = 1/n$, $1 \le i \le n$. By the Lindeberg–Feller theorem it is easy to show that (25) does not hold, although (26) is satisfied.

The situation in Example 5.2 is extreme because the random variables are "asymptotically degenerate." Such cases must be excluded if one attempts to generalize the result of Example 5.1. Let $A_n^0 = A_n - \mathrm{diag}(a_{nii})$, $\mathscr{A}_n = \{1 \le i \le k_n,\ a_{nii} \ne 0\}$.

THEOREM 5.1. *Suppose*

$$\inf_n \Big(\min_{1 \le i \le k_n} \mathrm{var}(X_{ni}) \Big) \wedge \Big(\min_{i \in \mathscr{A}_n} \mathrm{var}(X_{ni}^2) \Big) > 0, \tag{27}$$

$$\sup_n \Big(\max_{1 \le i \le k_n} EX_{ni}^2 1_{(|X_{ni}| > x)} \Big) \vee \Big(\max_{i \in \mathscr{A}_n} EX_{ni}^4 1_{(|X_{ni}| > x)} \Big) \to 0, \tag{28}$$

as $x \to \infty$. *Then* (26) *implies* (25).

Let $\{L_{ni},\ 1 \le i \le k_n,\ n \ge 1\}$ be numbers; define

$$\gamma_{ni}^{(1)} = EX_{ni}^4 1_{(|X_{ni}| \le L_{ni})}, \qquad \gamma_{ni}^{(2)} = E(X_{ni}^2 - 1)^4 1_{(|X_{ni}| \le L_{ni})},$$

$$\delta_{ni}^{(1)} = EX_{ni}^2 1_{(|X_{ni}| > L_{ni})}, \qquad \delta_{ni}^{(2)} = E(X_{ni}^2 - 1)^2 1_{(|X_{ni}| > L_{ni})};$$

$$\gamma_{nij} = \begin{cases} \gamma_{ni}^{(1)} \gamma_{nj}^{(1)}, & \text{if } i \ne j, \\ \gamma_{ni}^{(2)}, & \text{if } i = j, \end{cases}$$

$$\delta_{nij} = \begin{cases} \frac{1}{2}\big(\delta_{ni}^{(1)} + \delta_{nj}^{(1)}\big), & \text{if } i \ne j, \\ \delta_{ni}^{(2)}, & \text{if } i = j \in \mathscr{A}_n \\ 0, & \text{otherwise.} \end{cases}$$

THEOREM 5.2. *Suppose $EX_{ni}^2 = 1$, $1 \le i \le k_n$, and there are numbers $\{L_{ni}, 1 \le i \le k_n, n \ge 1\}$ such that*

$$\frac{1}{\sigma_n^2} \sum_{i,j=1}^{k_n} a_{nij}^2 \delta_{nij} \to 0, \tag{29}$$

$$\frac{1}{\sigma_n^4} \left[\sum_{i,j=1}^{k_n} a_{nij}^4 \gamma_{nij} + \sum_{i=1}^{k_n} \left(\sum_{j \ne i} a_{nij}^2 \right)^2 \gamma_{ni}^{(1)} \right] \to 0, \tag{30}$$

where $\sigma_n^2 = var(\mathscr{X}_n' A_n \mathscr{X}_n)$. Then (25) *is true provided*

$$\frac{\lambda_{\max}\left(\left(A_n^0 \right)^2 \right)}{\sigma_n^2} \to 0. \tag{31}$$

In particular, with $a_{nii} = 0$, $1 \le i \le k_n$, we get Theorem 5.2 in de Jong (1987) under slightly weaker assumption.

The following lemma plays a crucial role in the proof of Theorem 5.2 and hence of Theorem 5.1.

LEMMA 5.1 (A lemma of linear algebra). *Let $B = (b_{ij} 1_{(i > j)})$ be a lower triangular matrix. Then*

$$\mathrm{tr}\left((B'B)^2 \right) \le 2 \lambda_{\max}\left((B' + B)^2 \right) \mathrm{tr}(B'B). \tag{32}$$

6. Discussion.

6.1. As in many maximum-likelihood-related problems the solution or root of the REML (ML) equations sometimes presents difficulties. Theorems 4.1 and 4.3 ensure the existence of a consistent sequence of roots (CSR) of the REML equations and asymptotic normality of any sequence of roots such that $\{(p_0(N)(\hat{\lambda}_N - \lambda_0), p_1(N)(\hat{\mu}_{N1} - \mu_{01}), \dots, p_s(N)(\hat{\mu}_{Ns} - \mu_{0s}))'\}$ is bounded in probability (see the proofs of the theorems). However, the theorems do not provide a way of identifying such a sequence when the roots are not unique. In other words, the established theorems are results of Cramér type [e.g., Miller (1977)].

Some methods were proposed in the literature to overcome this difficulty [e.g., Lehmann (1983)]. These methods basically require that some sequence of consistent (but not necessarily AN) estimates be available. A candidate of such estimates in our cases is Rao's MINQE, asymptotic properties of which are discussed in Rao and Kleffe [(1988), Section 10].

In some cases, the uniqueness of the roots can be ensured. For example, under certain conditions the ANOVA estimates are uniquely defined [e.g., Westfall (1986)]. Since in the balanced case solutions of the REML equations are identical to the ANOVA estimates, these conditions also guarantee the uniqueness of the REML estimates in the balanced case. Necessary and sufficient conditions for existence of (unique) explicit solution of the ML equations in a balanced mixed model of the analysis of variance are given by Szatrowski and Miller (1980). General sufficient conditions for the uniqueness of solutions of ML equations can be found in Mäkeläinen, Schmidt and Styan (1981).

The identification of a CSR is a problem of both theoretical and practical interests not only in mixed model analysis but also in a wide range of areas where M-estimates [Huber (1981)] are involved. Questions can always be asked such as whether solutions that maximize the Gaussian likelihood form a CSR, although the Gaussian likelihood is not necessarily the true likelihood. Note that all the REML estimates proved to exist in this paper are actually at least local maxima of the Gaussian likelihood of $A'y$.

Finally, Theorems 4.2 and 4.4 give necessary and sufficient conditions for existence of a CSR of the ML equations (and asymptotic normality of such a sequence). When these conditions are violated, no sequence of roots of the ML equations can be consistent (AN).

6.2. From Theorems 4.3 and 4.4 we see the asymptotic covariance matrix of both $\hat{\theta}_N = (\hat{\lambda}_N, \hat{\mu}_{N1}, \dots, \hat{\mu}_{Ns})'$ and $\hat{\theta}_N^* = (\hat{\lambda}_N^*, \hat{\mu}_{N1}^*, \dots, \hat{\mu}_{Ns}^*)'$ is $\tilde{V}(\theta_0) = \tilde{I}_N(\theta_0)^{-1}\tilde{J}_N(\theta_0)\tilde{I}_N(\theta_0)^{-1}$, where $\tilde{I}_N(\theta_0) = (\mathrm{tr}(V_iV_j))$, $\tilde{J}_N(\theta_0) = 2\tilde{I}_N(\theta_0) + \tilde{K}_N(\theta_0)$ with $\tilde{K}_N(\theta_0) = (\sum_{l=1}^{N+m}(EW_{Nl}^4 - 3)V_i(\mu_0)_{ll}V_j(\mu_0)_{ll}/\lambda_0^{1_{(i=1)}+1_{(j=0)}})$. Thus one can construct approximate confidence intervals for the parameters of variance components. It is seen that under normality [in which case $\tilde{V}(\theta_0) = 2\tilde{I}_N(\theta_0)^{-1}$, which is the inverse of the restricted information matrix] and the condition that $p/N \to 0$, $\mathrm{tr}(C_i(\mu_0))/\|V_i\|_R^2 \to 0$, $1 \le i \le s$ [which implies $2\tilde{I}_N(\theta_0)^{-1} \sim 2\tilde{I}_N^*(\theta_0)^{-1}$, the inverse of the (unrestricted) information matrix; see, e.g., Searle, Casella and McCulloch (1992), Section 6], the REML estimates are efficient in the sense of attaining asymptotically the Cramér–Rao lower bound [i.e., Miller (1977)]. By Theorem 4.4(ii) and similar discussion as for (19), the MLE are efficient in the same sense if and only if (22) holds. In particular, with p fixed for all balanced mixed models of the analysis of variance, both the REML estimates and the MLE are efficient.

However, efficiency in the non-i.i.d. case, especially in the presence of a large number of nuisance parameters, ought to be defined in a stricter sense [see Bickel (1993) and Pfanzagl (1993)]. Further work is needed before a conclusion is made about whether the REML estimates are the asymptotically best.

6.3. In all theorems in this paper, we assume the model has p.v.c. [e.g., Miller (1977)]. It can be shown that even without this assumption but with the assumption $\sup_N \max_{1 \le i \le s} \lambda_{\max}(Z_i'V(\mu_0)Z_i) < \infty$, a sequence of solutions to the REML equations can still be consistent and AN. However, the solutions are not guaranteed to fall into Θ asymptotically and therefore not the REML estimates by our definition.

7. Proofs

PROOF OF LEMMA 5.1. For any $1 \le i \le n$, let

$$A = B' + B = \left(\begin{array}{cccc|cccc} 0 & b_{21} & \cdots & b_{i1} & b_{i+11} & \cdots & \cdots & b_{n1} \\ b_{21} & \ddots & & \vdots & \vdots & & & \vdots \\ \vdots & & \ddots & \vdots & \vdots & & & \vdots \\ b_{i1} & \cdots & \cdots & 0 & b_{i+1i} & \cdots & \cdots & b_{ni} \\ \hline b_{i+11} & \cdots & \cdots & b_{i+1i} & 0 & \cdots & \cdots & b_{ni+1} \\ \vdots & & & \vdots & \vdots & \ddots & & \vdots \\ \vdots & & & \vdots & \vdots & & \ddots & b_{nn-1} \\ b_{n1} & \cdots & \cdots & b_{ni} & b_{ni+1} & \cdots & b_{nn-1} & 0 \end{array}\right)$$

$$= \begin{pmatrix} A_i & \tilde{A}_i' \\ \tilde{A}_i' & A_i^* \end{pmatrix},$$

$a = (0 * ')'$ and $b = (\tilde{*}\,'0)'$, where $* = (b_{i+1i}, \dots, b_{ni})'$, $\tilde{*} = (\sum_{k>i} b_{ki} b_{k1}, \dots, \sum_{k>i} b_{ki} b_{ki})'$. Then it is easy to check that

$$|\tilde{*}|^2 = a'Ab \le \|A\|\,|a|\,|b| = \|A\|\,|*|\,|\tilde{*}|$$
$$\Rightarrow \quad \sum_{j \le i} \Bigl(\sum_{k>i} b_{ki} b_{kj} \Bigr)^2 \le \|A\|^2 \sum_{k>i} b_{ki}^2.$$

Thus

$$\operatorname{tr}\bigl((B'B)^2\bigr) = \sum_{i,j} \Bigl(\sum_k b_{ki} b_{kj} 1_{(k > i \vee j)} \Bigr)^2 \le 2 \sum_i \sum_{j \le i} \Bigl(\sum_{k>i} b_{ki} b_{kj} \Bigr)^2$$
$$\le 2\|A\|^2 \sum_i \sum_{k>i} b_{ki}^2 = 2\lambda_{\max}(A^2)\operatorname{tr}(B'B). \qquad \square$$

PROOF OF THEOREM 5.2. We have

$$\frac{1}{\sigma_n}(\mathscr{X}_n' A_n \mathscr{X}_n - E\mathscr{X}_n' A_n \mathscr{X}_n) = \sum_{i=1}^{k_n} \xi_{ni} + \sum_{i=1}^{k_n} \eta_{ni},$$

where

$$\xi_{ni} = \frac{1}{\sigma_n}\Biggl[a_{nii} U_{ni} + 2\Bigl(\sum_{j<i} a_{nij} u_{nj} \Bigr) u_{ni} \Biggr],$$

$$\eta_{ni} = \frac{1}{\sigma_n}\Biggl[a_{nii} V_{ni} + 2\Bigl(\sum_{j<i} a_{nij} v_{nj} \Bigr) u_{ni} + 2\Bigl(\sum_{j<i} a_{nij} X_{nj} \Bigr) v_{ni} \Biggr],$$

$$U_{ni} = (X_{ni}^2 - 1) 1_{(|X_{ni}| \le L_{ni})} - E(X_{ni}^2 - 1) 1_{(|X_{ni}| \le L_{ni})}, \qquad V_{ni} = X_{ni}^2 - 1 - U_{ni},$$
$$u_{ni} = X_{ni} 1_{(|X_{ni}| \le L_{ni})} - E X_{ni} 1_{(|X_{ni}| \le L_{ni})}, \qquad v_{ni} = X_{ni} - u_{ni}.$$

Condition (29) implies $\sum_{i=1}^{k_n}\eta_{ni} \to_{L^2} 0$. So it is enough to check that the array of martingale difference ξ_{ni}, $1 \le i \le k_n$, satisfies conditions (3.18)–(3.20) of Hall and Heyde [(1980), Theorem 3.2] (see the remarks after that theorem):

$$\frac{1}{\sigma_n^4} E\left\{ \max_{1\le i\le k_n} (a_{nii}U_{ni})^4 \right\} \le \frac{1}{\sigma_n^4} \sum_{i=1}^{k_n} a_{nii}^4 EU_{ni}^4 \to 0,$$

by (30); and, by Rosenthal's inequality [see Hall and Heyde (1980)],

$$E\left(\sum_{j<i} a_{nij}u_{nj} \right)^4 \le c\left\{ \left(\sum_{j<i} a_{nij}^2 \right)^2 + \sum_{j<i} a_{nij}^4 Eu_{nj}^4 \right\}$$

for some constant c. So, by a similar argument and (30),

$$\frac{1}{\sigma_n^4} E\left\{ \max_{1\le i\le k_n} \left[\left(\sum_{j<i} a_{nij}u_{nj} \right) u_{ni} \right]^4 \right\} \to 0.$$

Thus $\max_{1\le i\le k_n} |\xi_{ni}|$ is bounded in L^2 and goes to 0 in probability. We have the following:

$$\sum_{i=1}^{k_n} \xi_{ni}^2 = \sum_{i=1}^{3} U_i + \sum_{i=1}^{3} V_i,$$

where

$$U_1 = \sigma_n^{-2} \sum_{i=1}^{k_n} a_{nii}^2 \left(U_{ni}^2 - EU_{ni}^2\right),$$

$$U_2 = 4\sigma_n^{-2} \sum_{i=1}^{k_n} a_{nii} \left(\sum_{j<i} a_{nij}u_{nj} \right) (U_{ni}u_{ni} - EU_{ni}u_{ni}),$$

$$U_3 = 4\sigma_n^{-2} \sum_{i=1}^{k_n} \left(\sum_{j<i} a_{nij}u_{nj} \right)^2 \left(u_{ni}^2 - Eu_{ni}^2\right),$$

$$V_1 = \sigma_n^{-2} \sum_{i=1}^{k_n} a_{nii}^2 EU_{ni}^2,$$

$$V_2 = 4\sigma_n^{-2} \sum_{i=1}^{k_n} a_{nii} \left(\sum_{j<i} a_{nij}u_{nj} \right) EU_{ni}u_{ni},$$

$$V_3 = 4\sigma_n^{-2} \sum_{i=1}^{k_n} \left(\sum_{j<i} a_{nij}u_{nj} \right)^2 Eu_{ni}^2.$$

It follows from (30) that $U_i \to_{L^2} 0$, $i = 1, 2, 3$ (using Rosenthal's inequality for U_3).

Also, $V_1 = \sigma_n^{-2}\sum_{i\in\mathscr{A}_n} a_{nii}^2 \operatorname{var}(X_{ni}^2) + o(1)$ by (29) and the fact that

$$\sigma_n^2 = \sum_{i\in\mathscr{A}_n} a_{nii}^2 \operatorname{var}\left(X_{ni}^2\right) + 2\sum_{i\ne j} a_{nij}^2. \tag{33}$$

Let $B_n = (a_{nij}1_{(i>j)})$, $l_n = (a_{nii}EU_{ni}u_{ni})$ and $\lambda_n = 16\sigma_n^{-4}$. Then, by Lemma 5.1, (31) and (33),

$$\begin{aligned}
EV_2^2 &= \lambda_n E\left[\sum_{j=1}^{k_n}\left(\sum_{i>j} a_{nij}a_{nii}EU_{ni}u_{ni}\right)u_{nj}\right]^2 \\
&= \lambda_n \sum_{j=1}^{k_n}\left(\sum_{i>j} a_{nij}a_{nii}EU_{ni}u_{ni}\right)^2 Eu_{nj}^2 \\
&\le \lambda_n |B_n' l_n|^2 \le \lambda_n \lambda_{\max}(B_n B_n')|l_n|^2 \le \lambda_n \left[\operatorname{tr}\left((B_n'B_n)^2\right)\right]^{1/2}|l_n|^2 \\
&\le \frac{16}{\sqrt{2}}\left[\frac{\lambda_{\max}\left((A_n^0)^2\right)}{\sigma_n^2}\right]^{1/2}\sigma_n^{-2}\sum_{i\in\mathscr{A}_n} a_{nii}^2 \operatorname{var}(X_{ni}^2) \\
&\le \frac{16}{\sqrt{2}}\left[\frac{\lambda_{\max}\left((A_n^0)^2\right)}{\sigma_n^2}\right]^{1/2} \to 0.
\end{aligned}$$

Also

$$\begin{aligned}
(29) \text{ and } (33) \quad \Rightarrow \quad V_3 &= 4\sigma_n^{-2}\sum_{i=1}^{k_n}\left(\sum_{j<i} a_{nij}u_{nj}\right)^2 + o_p(1) \\
&= \mathscr{U}_n' C_n \mathscr{U}_n + o_p(1),
\end{aligned}$$

where $\mathscr{U}_n = (u_{ni})$ and $C_n = 4\sigma_n^{-2}B_n'B_n$. It follows from (30) that

$$\sum_{i=1}^{k_n} c_{nii}^2 \operatorname{var}(u_{ni}^2) = \lambda_n \sum_{i=1}^{k_n}\left(\sum_{j>i} a_{nij}^2\right)^2 \operatorname{var}(u_{ni}^2) \to 0;$$

and, by Lemma 5.1, (31) and (33),

$$\sum_{j<i} c_{nij}^2 Eu_{ni}^2 Eu_{nj}^2 \le \frac{1}{2}\sum_{i\ne j} c_{nij}^2 \le \frac{\lambda_n}{2}\operatorname{tr}\left((B_n'B_n)^2\right) \le \frac{4\lambda_{\max}\left((A_n^0)^2\right)}{\sigma_n^2} \to 0.$$

Thus, by Lemma 7.1 in the following, we have $\mathscr{U}_n'C_n\mathscr{U}_n - E\mathscr{U}_n'C_n\mathscr{U}_n \to_{L^2} 0$.

Finally,

$$(29) \text{ and } (33) \quad \Rightarrow \quad E\mathscr{U}_n'C_n\mathscr{U}_n = 2\sigma_n^{-2}\sum_{i\ne j} a_{nij}^2 + o(1).$$

Thus we conclude, using (33), that $\sum_{i=1}^{k_n}\xi_{ni}^2 = 1 + o_p(1)$. □

LEMMA 7.1. *Let $X_{n1},\dots,X_{nk_n}$ be independent with $EX_{ni} = 0$, $EX_{ni}^2 = \sigma_{ni}^2$, and $EX_{ni}^4 < \infty$, and let A_n be symmetric. Then $\mathscr{X}_n'A_n\mathscr{X}_n - E\mathscr{X}_n'A_n\mathscr{X}_n \to_{L^2} 0$ provided $\sum_{i=1}^{k_n} a_{nii}^2 \operatorname{var}(X_{ni}^2) \to 0$ and $\sum_{j<i} a_{nij}^2\sigma_{ni}^2\sigma_{nj}^2 \to 0$.*

COROLLARY 7.1. *If, in Lemma 7.1,* $\sup_{n,i} EX_{ni}^4 < \infty$, *then* $\mathscr{X}_n' A_n \mathscr{X}_n - E\mathscr{X}_n' A_n \mathscr{X} \to_{L^2} 0$ *provided* $\operatorname{tr}(A_n^2) \to 0$.

PROOF OF THEOREM 5.1. First we assume $EX_{ni}^2 = 1$. Then

$$(27), (28) \text{ and } (33) \quad \Rightarrow \quad \sigma_n^2 \sim \operatorname{tr}(A_n^2).$$

The result now follows by letting $L_{ni} = L_n = (\sigma_n^2/\lambda_{\max}(A_n^2))^{\delta}$, $1 \le i \le k_n$, with $0 < \delta < \frac{1}{4}$ being fixed, and checking the conditions of Theorem 5.2. The general case is proved by making the transformation $\tilde{X}_{ni} = X_{ni}/\sigma_{ni}$ ($\sigma_{ni}^2 = EX_{ni}^2$), $\tilde{A}_n = \operatorname{diag}(\sigma_{ni}) A_n \operatorname{diag}(\sigma_{ni})$. □

The proof of Theorem 4.3 requires the following lemma.

LEMMA 7.2. *Let* $L_N = L_N(\theta, \mathscr{Y}_n)$, *where* $\theta = (\theta_1, \dots, \theta_s)' \in \Theta C R^s$ *and* $\mathscr{Y}_N$, $N \ge 1$, *are random vectors, be continuously differentiable w.r.t.* θ. *Suppose there are sequences of positive numbers* $p_i(N)$ *and* $q_i(N)$, $1 \le i \le s$, *such that* $p_i(N) \to \infty$, $p_i(N) q_i(N) \to \infty$, $1 \le i \le s$,

$$\left(\frac{1}{p_i(N) p_j(N)} \frac{\partial^2 L_N}{\partial \theta_i \, \partial \theta_j} \bigg|_{\theta_0} \right) = I_N(\theta_0) + o_p(1) \quad \textit{with } \liminf \lambda_{\min}(I_N(\theta_0)) > 0,$$

and

$$\frac{1}{p_i(N) p_j(N) p_k(N)} \sup_{\theta \in \Theta_N} \left| \frac{\partial^3 L_N}{\partial \theta_i \, \partial \theta_j \, \partial \theta_k} \right| \to_P 0, \qquad 1 \le i, j, k \le s,$$

where $\Theta_N = \{\theta\colon |\theta_i - \theta_{0i}| < q_i(N),\ 1 \le i \le s\}$. *Let*

$$A_N(\theta_0) = \left(\frac{1}{p_i(N)} \frac{\partial L_N}{\partial \theta_i} \bigg|_{\theta_0} \right), p(N)(\theta - \theta_0) = (p_i(N)(\theta_i - \theta_{0i})).$$

(i) *If* $\{|A_N(\theta_0)|\}$ *is bounded in probability, then w.p.* $\to 1$ *the equations* $\partial L_N / \partial \theta_i = 0$, $i = 1, \dots, s$, *have a solution* $\hat{\theta}_N = (\hat{\theta}_{Ni})$ *such that* $\{|p(N)(\hat{\theta}_N - \theta_0)|\}$ *is bounded in probability; therefore* $\hat{\theta}_N$ *is consistent.*

(ii) *If, furthermore, there is a sequence of symmetric matrices* $\{J_N(\theta_0)\}$ *such that* $\liminf \lambda_{\min}(J_N(\theta_0)) > 0$ *and* $J_N^{-1/2}(\theta_0) A_N(\theta_0) \to_{\mathscr{L}} N(0, I_s)$, *then* $J_N^{-1/2}(\theta_0) I_N(\theta_0) p(N)(\hat{\theta}_N - \theta_0) \to_{\mathscr{L}} N0, I_s)$.

The proof of (i) is similar to Weiss (1971). Namely, let θ_N be defined by $p(N)(\theta_N - \theta_0) = -\tilde{I}_N^{-1}(\theta_0) A_N(\theta_0)$, where

$$\tilde{I}_N(\theta_0) = \left(\frac{1}{p_i(N) p_j(N)} \frac{\partial^2 L_N}{\partial \theta_i \, \partial \theta_j} \bigg|_{\theta_0} \right),$$

one has $L_N(\theta) - L_N(\theta_N) = \frac{1}{2}(p(N)(\theta - \theta_N))' \tilde{I}_N(\theta_0)(p(N)(\theta - \theta_N)) + R_N(\theta, \theta_0) - R_N(\theta_N, \theta_0)$, and the R's are uniformly ignorable compared with the first term as θ varies on the ellipsoid $\{\theta\colon |p(N)(\theta - \theta_N)| = 1\}$. Part (ii) follows from the relation $-A_N(\theta_0) = I_N(\theta_0) p(N)(\hat{\theta}_N - \theta_0) + o_p(1)$.

It is sometime convenient to denote $\lambda, \mu_1, \dots, \mu_s$ by $\theta_0, \theta_1, \dots, \theta_s$. To avoid confusion, we use from now on the symbol $\underline{\theta}_0$ for the true parameter vector. Let $G_i = A'Z_iZ_i'A$, $1 \le i \le s$, $H_\mu = V(A, \mu)^{-1}$.

PROOF OF THEOREM 4.3. (i) By Lemma 4.2 there are sequences of positive numbers $p_i(N) \to \infty$, $0 \le i \le s$, such that (19) holds, and we can assume, w.l.o.g., that $p_0(N) = \sqrt{N-p}$.

See notation in Sections 2 and 3. Let $L_N = (N-p)\log\lambda + \log|H_\mu^{-1}| + (1/\lambda)z'H_\mu z$, where $z = A'y = \sqrt{\lambda_0}A'b(\mu_0)'\mathscr{W}_N$ with $\mathscr{W}_N = (W_{Nl})_{1\le l\le N+m}$. By Corollary 7.1 we have

$$\frac{1}{p_0^2(N)}\frac{\partial^2 L_N}{\partial\lambda^2}\bigg|_{\underline{\theta}_0} = \frac{1}{\lambda_0^2} + \frac{2}{\lambda_0^2}[\mathscr{W}_N' B_N \mathscr{W}_N - \operatorname{tr}(B_N)] = \frac{1}{\lambda_0^2} + o_p(1)$$

with $B_N = V_0(\mu_0)/(N-p)$. Similarly,

$$\frac{1}{p_i(N)p_0(N)}\frac{\partial^2 L_N}{\partial\mu_i\,\partial\lambda}\bigg|_{\underline{\theta}_0} = I_{i0}^{(N)} + o_p(1),$$

$$\frac{1}{p_i(N)p_j(N)}\frac{\partial^2 L_N}{\partial\mu_i\,\partial\mu_j}\bigg|_{\underline{\theta}_0} = I_{ij}^{(N)} + o_p(1), \qquad 1 \le i, j \le s.$$

Also

$$\frac{\partial^3 L_N}{\partial\lambda^3} = \frac{2}{\lambda^3}(N-p) - \frac{6}{\lambda^4}z'H_\mu z,$$

$$\frac{\partial^3 L_N}{\partial\lambda^2\,\partial\mu_i} = -\frac{2}{\lambda^3}z'H_\mu G_i H_\mu z,$$

$$\frac{\partial^3 L_N}{\partial\lambda\,\partial\mu_i\,\partial\mu_j} = -\frac{2}{\lambda^2}z'H_\mu G_i H_\mu G_j H_\mu z,$$

$$\frac{\partial^3 L_N}{\partial\mu_i\,\partial\mu_j\,\partial\mu_k} = T_{ijk} + T_{ikj} - \lambda^{-1}(S_{ijk} + S_{jki} + S_{kij} + S_{ikj} + S_{kji} + S_{jik}),$$

with $T_{ijk} = \operatorname{tr}(H_\mu G_i H_\mu G_j H_\mu G_k)$, $S_{ijk} = z'H_\mu G_i H_\mu G_j H_\mu G_k H_\mu z$. For $1 \le i \le s$, let λ_{ij}, $1 \le j \le m_i$, be the eigenvalues of $Z_i'P_{X^\perp}Z_i$. Then [e.g., Chan and Kwong (1985), Lemma 3] we have

$$\|H_{\mu_0}^{1/2}A'Z_i\|^2 = \lambda_{\max}(Z_i'V(\mu_0)Z_i) \le \lambda_{\max}\Big(Z_i'A(A'A + \mu_{0i}G_i)^{-1}A'Z_i\Big)$$

(34)
$$= \max_{1\le j\le m_i}\frac{\lambda_{ij}}{1+\mu_{0i}\lambda_{ij}} \le \mu_{0i}^{-1}.$$

Now let

$$q_0(N) = \frac{\lambda_0}{2}, \qquad q_i(N) = \frac{(\min_{1 \le v \le s} p_v(N))^{1/2}}{p_i(N)}, \qquad 1 \le i \le s,$$

and

$$\mathscr{M}_N = \{\mu : |\mu_i - \mu_{0i}| < q_i(N), 1 \le i \le s\},$$
$$\Theta_N = \{\theta = (\lambda, \mu')' : |\lambda - \lambda_0| < q_0(N), \mu \in \mathscr{M}_N\}.$$

Then for large N, $\frac{1}{2}H_{\mu_0} \le H_\mu \le 2H_{\mu_0}$, $\mu \in \mathscr{M}_N$. Thus by (34) and the identity

$$H_\mu = H_{\mu_0} + \sum_{l=0}^{s} (\mu_{0l} - \mu_l) H_{\mu_0} G_l H_\mu,$$

it can be shown that, for any $b \in R^{N-p}$,

$$|b' H_\mu^{1/2} G_i H_\mu z| \le |b| \left(\sqrt{2} |H_{\mu_0}^{1/2} G_i H_{\mu_0} z| + \sum_{l=1}^{s} 2\mu_{0i}^{-1} q_l(N) |H_\mu^{1/2} G_l H_\mu z| \right),$$

which implies

$$\sup_{\mu \in \mathscr{M}_N} |H_\mu^{1/2} G_i H_\mu z| \le \sqrt{2} |H_{\mu_0}^{1/2} G_i H_{\mu_0} z| + \sum_{l=1}^{s} 2\mu_{0i}^{-1} q_l(N) \sup_{\mu \in \mathscr{M}_N} |H_\mu^{1/2} G_l H_\mu z|,$$
$$1 \le i \le s.$$

It follows by solving the inequalities that there exist N_0 and $\{e_{il}(N)\}$, $1 \le i$, $l \le s$, not depending on z such that $\{e_{ij}(N)\}$ are bounded and, for $N \ge N_0$,

$$(35) \quad \begin{aligned} \sup_{\mu \in \mathscr{M}_N} |H_\mu^{1/2} G_i H_\mu z| &\le e_{ii}(N) |H_{\mu_0}^{1/2} G_i H_{\mu_0} z| \\ &\quad + \sum_{l \ne i} e_{il}(N) q_l(N) |H_{\mu_0}^{1/2} G_l H_{\mu_0} z|, \qquad 1 \le i \le s \end{aligned}$$

for all z. Thus it is not hard to conclude that

$$(36) \quad \frac{1}{p_i(N) p_j(N) p_k(N)} \sup_{\theta \in \Theta_N} \left| \frac{\partial^3 L_N}{\partial \theta_i \, \partial \theta_j \, \partial \theta_k} \right| \to_{L^1} 0, \qquad 0 \le i, j, k \le s.$$

Finally, an analogue of Corollary 7.1 which says for any symmetric $B_N, \mathscr{W}_N' B_N \mathscr{W}_N - \mathrm{tr}(B_N)$ is L^2 bounded provided $\mathrm{tr}(B_N^2)$ is bounded implies

$$\left\{ A_N(\underline{\theta}_0)_i = \frac{1}{p_i(N)} \left. \frac{\partial L_N}{\partial \theta_i} \right|_{\underline{\theta}_0} \right\}, \qquad 0 \le i \le s,$$

are bounded in L^2.

The result now follows from Lemma 7.2(i) and Lemma 4.2.

(ii) First we note that any such sequences $p_i(N)$, $0 \le i \le s$, can play the same role as those in the proof of (i), and

$$\liminf \lambda_{\min}(I_N(\underline{\theta}_0)) > 0 \text{ plus ND} \quad \Rightarrow \quad \liminf \lambda_{\min}(J_N(\underline{\theta}_0)) > 0.$$

For any $a \in R^{s+1} \setminus \{0\}$, let $b_N = J_N^{-1/2}(\underline{\theta}_0)a$. Then

$$a' J_N^{-1/2}(\underline{\theta}_0) A_N(\underline{\theta}_0) = -[\mathscr{W}_N' B_N \mathscr{W}_N - E\mathscr{W}_N' B_N \mathscr{W}_N],$$

where $B_N = B_N(b_N)$ with

$$B_N(b) = \frac{b_0}{\lambda_0 p_0(N)} V_0(\mu_0) + \sum_{i=1}^{s} \frac{b_i}{P_i(N)} V_i(\mu_0).$$

Since, for any $b \in R^{s+1}$,

$$b' J_N(\underline{\theta}_0) b = \operatorname{var}(\mathscr{W}_N' B_N(b) \mathscr{W}_N) \le M \operatorname{tr}\big(B_N^2(b)\big),$$

by (33), where $M = 2 \vee \sup_N \max_{1 \le l \le N+m} \operatorname{var}(W_{Nl}^2) < \infty$ by Assumption A2, we have

$$\operatorname{tr}\big(B_N^2\big) \ge \frac{1}{M} \operatorname{var}(\mathscr{W}_N' B_N \mathscr{W}_N) = \frac{1}{M} b_N' J_N(\underline{\theta}_0) b_N = \frac{|a|^2}{M},$$

and

$$\lambda_{\max}\big(B_N^2\big) \le \left(\lambda_0^{-2} p_0^{-2}(N) + \sum_{i=1}^{s} \mu_{0i}^{-2} p_i^{-2}(N)\right) \frac{|a|^2}{\lambda_{\min}\big(J_N(\underline{\theta}_0)\big)},$$

by (34). The result now follows from Theorem 5.1 and Lemma 7.2(ii). It is seen from the proof that the ND condition can be weaked to (24). □

PROOF OF THEOREM 4.4. (i) The consistency of the MLE $\theta_N^* = (\hat{\lambda}_N^*, \hat{\mu}_{N1}^*, \ldots, \hat{\mu}_{Ns}^*)'$ implies $p < N$ at least for large N (otherwise the model is not IDI). Let $L_N^* = N \log \lambda + \log|V_\mu| + (1/\lambda) z' H_\mu z$. Simple relations exist between derivatives of L_N^* and L_N (see the proof of Theorem 4.3). For example,

$$\frac{\partial L_N^*}{\partial \lambda} = \frac{p}{\lambda} + \frac{\partial L_N}{\partial \lambda}, \qquad \frac{\partial L_N^*}{\partial \mu_i} = \operatorname{tr}\big(C_i(\mu)\big) + \frac{\partial L_N}{\partial \mu_i}, \qquad \frac{\partial^2 L_N^*}{\partial \mu_i\, \partial \lambda} = \frac{\partial^2 L_N}{\partial \mu_i\, \partial \lambda},$$

$$\frac{\partial^2 L_N^*}{\partial \mu_i\, \partial \mu_j} = \operatorname{tr}\Big(\big(Z_i' V(\mu) Z_j\big)'\big(Z_i' V(\mu) Z_j\big)\Big) - \operatorname{tr}\Big(\big(Z_i' V_\mu^{-1} Z_j\big)'\big(Z_i' V_\mu^{-1} Z_j\big)\Big)$$
$$+ \frac{\partial^2 L_N}{\partial \mu_i\, \partial \mu_j}, \quad \text{etc.}$$

Let $q_i(N) \to 0$, $1 \le i \le s$, be such that

$$P_{\underline{\theta}_0}\left(|\hat{\lambda}_N^* - \lambda_0| < \frac{\lambda_0}{2}, |\hat{\mu}_{Ni}^* - \mu_{0i}| < q_i(N), 1 \le i \le s\right) \to 1.$$

With such $q_i(N)$, the same inequality as (35) can be established. Similarly, we obtain

$$\sup_{\mu \in \mathscr{M}_N} |Z_i' A H_\mu z| \le e_{ii}(N) |Z_i' A H_{\mu_0} z| + \sum_{l \ne i} e_{il}(N) q_l(N) |Z_l' A H_{\mu_0} z|,$$
$$1 \le i \le s,\ N \ge N_0,$$

for all z. Using those inequalities and the fact that, by (34),

$$\|Z_i' A H_{\mu_0} z\|_2^2 = \lambda_0 \operatorname{tr}(Z_i' V(\mu_0) Z_i) \le \frac{\lambda_0}{\mu_{0_i}} m_i \wedge (N-p),$$

it is seen that

$$\frac{1}{N} \sup_{\theta \in \Theta_N} \left| \frac{\partial^2 L_N}{\partial \lambda^2} \right|, \qquad \frac{1}{N \wedge m^*} \sup_{\theta \in \Theta_N} \left| \frac{\partial^2 L_N}{\partial \mu_i \, \partial \lambda} \right| \quad \text{and} \quad \frac{1}{m^*} \sup_{\theta \in \Theta_N} \left| \frac{\partial^2 L_N}{\partial \mu_i \, \partial \mu_j} \right|, \qquad 1 \le i, j \le s,$$

are bounded in L^2.

By the relations between derivatives of L_n^* and L_n it is easy to see that the above assertion of L^2 boundedness also holds with L_N replaced by L_N^*. Note that we have, in addition to (34), $\lambda_{\max}(Z_i' V_{\mu_0}^{-1} Z_i) \le \mu_{0_i}^{-1}$. Then, by Taylor expansions,

$$0 = \frac{1}{N} \left(\frac{\partial L_N^*}{\partial \lambda} \bigg|_{\underline{\theta}_0} + \frac{\partial^2 L_N^*}{\partial \lambda^2} \bigg|_{\theta_{N,0}^*} \left(\hat{\lambda}_N^* - \lambda_0 \right) + \sum_{j=1}^{s} \frac{\partial^2 L_N^*}{\partial \mu_j \, \partial \lambda} \bigg|_{\theta_{N,0}^*} \left(\hat{\mu}_{Nj}^* - \mu_{0j} \right) \right),$$

$$0 = \frac{1}{m^*} \left(\frac{\partial L_N^*}{\partial \mu_i} \bigg|_{\underline{\theta}_0} + \frac{\partial^2 L_N^*}{\partial \lambda \, \partial \mu_i} \bigg|_{\theta_{N,i}^*} \left(\hat{\lambda}_N^* - \lambda_0 \right) + \sum_{j=1}^{s} \frac{\partial^2 L_N^*}{\partial \mu_j \, \partial \mu_i} \bigg|_{\theta_{N,i}^*} \left(\hat{\mu}_{Nj}^* - \mu_{0j} \right) \right),$$

where $\theta_{N,i}^*$ is between $\underline{\theta}_0$ and $\hat{\theta}_N^*$, $0 \le i \le s$, we see that

$$\frac{1}{N} \frac{\partial L_N^*}{\partial \lambda} \bigg|_{\underline{\theta}_0} \to_{P_{\theta_0}} 0 \quad \text{and} \quad \frac{1}{m^*} \frac{\partial L_N^*}{\partial \mu_i} \bigg|_{\underline{\theta}_0} \to_{P_{\theta_0}} 0, \qquad 1 \le i \le s.$$

Thus, by the last part of the proof of Theorem 4.3(i) (note that $\|V_i\|_R \le \mu_{0_i}^{-1} \sqrt{m_i}$, $1 \le i \le s$), it is easy to conclude that

$$\frac{1}{N} E_{\underline{\theta}_0} \left(\frac{\partial L_N^*}{\partial \lambda} \bigg|_{\underline{\theta}_0} \right) \to 0, \qquad \frac{1}{m^*} E_{\underline{\theta}_0} \left(\frac{\partial L_N^*}{\partial \mu_i} \bigg|_{\underline{\theta}_0} \right) \to 0, \qquad 1 \le i \le s.$$

The result thus follows.

(ii) [(a) $\Rightarrow$ (b)] Condition (21) implies, as in Lemma 4.2, that $p_i(N) \sim \|U_i\|_R$, $0 \le i \le s$. By the proof of Theorem 4.3 and the relations between derivatives of L_N^* and L_N, we have (36) with L_N replaced by L_N^* [same $q_i(N)$'s as in (36)] and that

$$\tilde{I}_N^*(\underline{\theta}_0) = \left(\frac{1}{p_i(N) p_j(N)} \frac{\partial^2 L_N^*}{\partial \theta_i \, \partial \theta_j} \bigg|_{\underline{\theta}_0} \right) = O(1) + o_p(1).$$

It follows by Taylor-expanding $\partial L_N^*/\partial\theta_i|_{\hat{\theta}_N^*}$ at $\underline{\theta}_0$ that

$$-A_N^*(\underline{\theta}_0) = -\left(\frac{1}{p_i(N)}\frac{\partial L_N^*}{\partial\theta_i}\bigg|_{\underline{\theta}_0}\right) = \big(O(1) + o_p(1)\big)p(N)\big(\hat{\theta}_N^* - \underline{\theta}_0\big),$$

which is bounded in probability. On the other hand, by the proof of Theorem 4.3,

$$A_N(\underline{\theta}_0) = \left(\frac{1}{p_i(N)}\frac{\partial L_N}{\partial\theta_i}\bigg|_{\underline{\theta}_0}\right)$$

is bounded in L^2 [note that the r.h.s. of (21) implies the r.h.s. of (19)]; thus $A_N^*(\underline{\theta}_0) - A_N(\underline{\theta}_0)$ is bounded in probability and hence in L^2 since it is nonrandom. So $A_N^*(\underline{\theta}_0)$ is bounded in L^2. Thus $E_{\underline{\theta}_0}A_N^*(\underline{\theta}_0) \to 0$ by Corollary 8.1.7 in Chow and Teicher (1978) and by an argument of subsequences (a.o.s.) [note that $O(1)$ is not random], which implies $I_N^*(\underline{\theta}_0) - I_N(\underline{\theta}_0) \to 0$ and hence AI^4, and (22) by Lemma 4.2.

[(b) ⇒ (a)] AI^4 implies the existence of $p_i(N) \to \infty$, $0 \le i \le s$, such that (19) holds; so $p_i(N) \sim \|V_i\|_R$, $0 \le i \le s$, by Lemma 4.2, and hence (21) by (22). That the MLE are AN with such $p_i(N)$'s follows from (22), the relations between derivatives of L_N^* and L_N, the proof of Theorems 4.3, and Lemma 7.2.

Finally, the equivalence of the REML estimates and the MLE is easy to prove, given the equivalence of (a) and (b) [use Taylor expansion for both estimates and (22)]. □

To prove Theorem 4.1 we need the following.

Lemma 7.3. *For the balanced mixed model* (13),

$$\mathrm{tr}(H_\mu G_i) = \left(\frac{N}{m_i}\right)\sum_{u \ge i} C_{u,\mu}^{-1}\prod_{u_q=0}(n_q - 1)1_{(u \not\ge d)},$$

$$\mathrm{tr}(H_\mu G_i H_\mu G_j) = \left(\frac{N}{m_i}\right)\left(\frac{N}{m_j}\right)\sum_{u \ge i \vee j} C_{u,\mu}^{-2}\prod_{u_q=0}(n_q - 1)1_{(u \not\ge d)},$$

$i, j \in S$, *where* $C_{u,\mu} = 1 + \sum_{v \in S}\mu_v m_v^{-1}N1_{(v \le u)}$, *and* $u \not\ge v$ *means* u *is not* $\ge v$.

Proof. Choose A such that $A'A = I_{N-p}$, so that $AA' = P_{X^\perp}$. Let T_q be $(n_q \times n_q)$ orthogonal such that $T_q' J_{n_q} T_q = \mathrm{diag}(n_q, 0, \dots, 0)$, where $J_{n_q} = 1_{n_q}1_{n_q}'$, $1 \le q \le r+1$; $T = \otimes_{q=1}^{r+1}T_q$, $B = A'T = (b_1 \cdots b_N)$. Then $T'Z_iZ_i'T = \mathrm{diag}(\lambda_{ik})$, $T'XX'T = \mathrm{diag}(\lambda_{dk})$, where $\{\lambda_{i1}, \dots, \lambda_{iN}\} = \{\prod_{q=1}^{r+1}\lambda_{iqk_q},\ 1 \le$

$k_q \le n_q,\ 1 \le q \le r+1\}$ with $\lambda_{iqw} = 1 - i_q + n_q i_q \delta_{w,1},\ i \in S \cup \{d\}$; and $T'AA'T = \operatorname{diag}(\gamma_k)$ with $\gamma_k = 1 - (p/N)\lambda_{dk}$, $1 \le k \le N$. It follows that

$$H_\mu G_i = \sum_{k=1}^{N} \lambda_{ik}\Big(1 + \sum_{t \in S} \mu_t \lambda_{tk}\Big)^{-1} b_k b_k'.$$

So

$$\begin{aligned}
\operatorname{tr}(H_\mu G_i) &= \sum_{k=1}^{N} \gamma_k \Big(1 + \sum_{t\in S} \mu_t \lambda_{tk}\Big)^{-1} \lambda_{ik} \\
&= \sum_{k_1=1}^{n_1} \cdots \sum_{k_{r+1}=1}^{n_{r+1}} \Big(1 - \frac{p}{N} \prod_{q=1}^{r+1} \lambda_{dqk_q}\Big)\Big(1 + \sum_{t\in S} \mu_t \prod_{q=1}^{r+1} \lambda_{tqk_q}\Big)^{-1} \prod_{q=1}^{r+1} \lambda_{iqk_q} \\
&= \frac{N}{m_i} \sum_{l \ge i} C_{l,\mu}^{-1} \prod_{l_q=0} (n_q - 1) 1_{(l \not\ge d)},
\end{aligned}$$

using the formula that, for any functions $f(x_1, \dots, x_{r+1})$ and $g_q(x)$, $1 \le q \le r+1$,

$$\begin{aligned}
&\sum_{k_1=1}^{n_1} \cdots \sum_{k_{r+1}=1}^{n_{r+1}} f(\delta_{k_1,1}, \dots, \delta_{k_{r+1},1}) \prod_{q=1}^{r+1} g_q(\delta_{k_q,1}) \\
&\qquad = \sum_{l_1=0}^{1} \cdots \sum_{l_{r+1}=0}^{1} f(l_1, \dots, l_{r+1}) \prod_{q=1}^{r+1} (n_q - 1)^{1-l_q} g_q(l_q).
\end{aligned}$$

Similarly we obtain the equations for $\operatorname{tr}(H_\mu G_i H_\mu G_j)$. □

PROOF OF THEOREM 4.1. We can assume w.l.o.g. that $n_q \ge 2$, $1 \le q \le r$ (since if $n_q = 1$, factor q is not really a factor and the model is not really an r-factor model). The nonconfounding assumption implies $d \not\le i$, $d \ne 0$, $i \ne 0$, $i \in S$ and $n_{r+1} \ge 2$ whenever $(0 \cdots 0\ 1) \in \{d\} \cup S$.

(i) By Theorem 4.3 and Lemma 4.2 it is enough to show (19) with $p_0(N) = \sqrt{N-p}$, $p_i(N) = \sqrt{m_i}$, $i \in S$. The r.h.s. of (19) is obvious by (34), so we can focus on the l.h.s.

First we assume the following limits exist as $N \to \infty$: $f_q = \lim n_q^{-1}$, $1 \le q \le r+1$, and $c_{iu} = \lim c_{iu}^{(N)}$, $i \in \bar{S} = \{0\} \cup S$, $u \not\ge d$, where $c_{iu}^{(N)} = (m_i C_u)^{-1} N 1_{(i \le u)}$, $C_u = C_{u,\mu_0}$. Then we have

$$\begin{aligned}
a_u^{(N)} &= \Big[(N-p)^{-1} \prod_{u_q=0} (n_q - 1)\Big]^{1/2} \to a_u \\
&= \Big[\Big(1 - \prod_{d_q \ne 0} f_q\Big)^{-1}\Big(\prod_{u_q \ne 0} f_q\Big)\Big(\prod_{u_q=0} (1 - f_q)\Big)\Big]^{1/2},
\end{aligned}$$

with $\sum_{u \ngeq d} a_u^2 = \lim \sum_{u \ngeq d} (a_u^{(N)})^2 = 1$; $b_{iu}^{(N)} = a_{iu}^{(N)} c_{iu}^{(N)} \to b_{iu} = a_{iu} c_{iu} 1_{(i \le u)}$, $i \in \overline{S}$, $u \ngeq d$. So, by Lemma 7.3,

$$\gamma_i^{(N)} = \frac{\operatorname{tr}(H_{\mu_0} G_i)}{p_0(N) p_i(N)} = \sum_{u \ngeq d} a_u^{(N)} b_{iu}^{(N)} \to \sum_{u \ngeq d} a_u b_{iu} = \gamma_i,$$

$$\gamma_{ij}^{(N)} = \frac{\operatorname{tr}(H_{\mu_0} G_i H_{\mu_0} G_j)}{p_i(N) p_j(N)} = \sum_{u \ngeq d} b_{iu}^{(N)} b_{ju}^{(N)} \to \sum_{u \ngeq d} b_{iu} b_{ju} = \gamma_{ij}, \qquad i, j \in S.$$

Therefore $I_N(\underline{\theta}_0) \to I(\underline{\theta}_0)$, where $I(\underline{\theta}_0)_{00} = 1/\lambda_0^2$, $I(\underline{\theta}_0)_{0i} = I(\underline{\theta}_0)_{i0} = \gamma_i$, $I(\underline{\theta}_0)_{ij} = \gamma_{ij}$, $i, j \in S$.

It is easy to see that, for any $x = (x_0\ (x_i)'_{i \in S})'$, $x' I(\underline{\theta}_0) x = \sum_{u \ngeq d} [a_u(x_0/\lambda_0) + \sum_{i \in S} b_{iu} x_i]^2$. So $x' I(\underline{\theta}_0) x = 0 \ \Rightarrow\ a_u(x_0/\lambda_0) + \sum_{i \in S} b_{iu} x_i = 0$, $u \ngeq d$. Let $i_* = 0$ if $f_{r+1} < 1$ and $(0 \cdots 0\ 1)$ if $f_{r+1} = 1$, then it is easy to see that $i_* \le i$, $i_* \ngeq d$, $i_* \ngeq i$, $i \in S$ and $a_{i_*} \ge (1/2)^{(r+1)/2} > 0$. So we have by the above equation that $a_{i_*}(x_0/\lambda_0) = 0 \ \Rightarrow\ x_0 = 0$. Thus $\sum_{i \in S} b_{iu} x_i = 0$, $u \in S$, which implies $x_u = 0$, $u \in S_i$ note $b_{uu} \ge (1/2)^{r/2}(1 + \sum_{v \in S_u} \mu_v)^{-1} > 0$, $u \in S$.

In general, since $\{1/n_q\}$'s and $\{c_{iu}^{(N)}\}$'s are bounded, the result follows from an a.o.s.

(ii) This follows from Theorem 4.3 and Lemma 4.2. □

The following simple lemma tells a basic idea for the proof of Theorem 4.2(i).

LEMMA 7.4. *Let $f(x) = f(x_1, \dots, x_s)$ be a differentiable function, let S be a closed convex set in R^s and let x_0 be a point in the interior of S. Suppose for each point x on the boundary of S there is a set of index $\{i_1, \dots, i_r\}$ depending on x such that*

$$\sum_{j=1}^{r} (x_{i_j} - x_{0 i_j}) \frac{\partial f}{\partial x_{i_j}}(x) > 0.$$

Then $f(x)$ attains its minimum value over S at an interior point $x^* \in S$, therefore $(\partial f/\partial x_i)(x^*) = 0$, $i = 1, \dots, s$.

By Lemma 7.3, it is easy to show the following.

LEMMA 7.5. *Assuming that $\mu_{0i} > 0$, $i \in S$, then, for $i \in S$,*

$$\frac{1}{m_i} \operatorname{tr}(C_i(\mu_0)) \to 0 \quad \text{iff} \quad \frac{m_{i \vee d} m_{i \vee d, S}}{m_i^2} \to 0,$$

$$\frac{1}{\sqrt{m_i}} \operatorname{tr}(C_i(\mu_0)) \to 0 \quad \text{iff} \quad \frac{m_{i \vee d} m_{i \vee d, S}}{m_i^{3/2}} \to 0.$$

PROOF OF THEOREM 4.2. First we note in the balanced case $Z_iZ_i'P_{X^\perp}Z_jZ_j' = Z_jZ_j'P_{X^\perp}Z_iZ_i'$, $\forall i, j$, which implies that, for any $z = A'y$ and i, j, k, $z'H_\mu G_i H_\mu z$, $z'H_\mu G_i H_\mu G_j H_\mu z$ and $z'H_\mu G_i H_\mu G_j H_\mu G_k H_\mu z$ are nonnegative and decreasing functions of μ. As a direct consequence, now it is easy to show that

$$\sup_{\theta \in \Theta_0} \left| \frac{\partial^2 L_N^*}{\partial \theta_i \, \partial \theta_j} \right| = m_i \wedge m_j O_p(1),$$
$$\sup_{\theta \in \Theta_0} \left| \frac{\partial^3 L_N^*}{\partial \theta_i \, \partial \theta_j \, \partial \theta_k} \right| = m_i \wedge m_j \wedge m_k O_p(1),$$

$i, j, k \in \overline{S} = \{0\} \cup S$, where $\Theta_0 = \{\lambda \geq \lambda_0/2,\ (1/2)\mu_{0i} \leq \mu_i \leq (3/2)\mu_{0i},\ i \in S\}$.

(i) The "only if" part not follows as in the proof of Theorem 4.4(i), using the above result for the second derivatives.

We now show the "if" part. Without loss of generality, we can assume the limits $\lim(m_i/m_j)$, $i, j \in \overline{S}$, exist as $N \to \infty$, which can be positive numbers, 0 or ∞, since the general case is then dealt with by an a.o.s. Divide m_i, $i \in \overline{S}$, by groups

$$\{m_u,\ 0 \leq u \leq s_1\}, \{m_u,\ s_1 + 1 \leq u \leq s_2\}, \dots, \{m_u,\ s_{c-1} + 1 \leq u \leq s_c\}, \tag{37}$$

where $m_0 = N$, $s_c = s = |S|$, such that within each $\{\cdots\}$ the m's are of same order, and the $\{\cdots\}$'s are of decreasing order. Note now the indexes of the m's are integers. With such ordering the parameters are ordered correspondingly as $\theta_0, \theta_1, \dots, \theta_s$ with $\theta_0 = \lambda$. Also, we have the partition of matrix

$$H^* = \left(\left. \frac{\partial^2 L_N^*}{\partial \theta_u \, \partial \theta_v} \right|_{\underline{\theta}_0} \right)_{0 \leq u, v \leq s} = (I_{ab})_{1 \leq a, b \leq c},$$

where

$$I_{ab} = \left(\left. \frac{\partial^2 L_N^*}{\partial \theta_u \, \partial \theta_v} \right|_{\underline{\theta}_0} \right)_{s_{a-1}+1 \leq u \leq s_a,\ s_{b-1}+1 \leq v \leq s_b}, \qquad s_0 = -1.$$

Condition (14) implies $\tilde{I}_N^*(\underline{\theta}_0) = \tilde{I}_N(\underline{\theta}_0) + o(1)$, where $\tilde{I}_N^*(\underline{\theta}_0) = \mathrm{diag}(m_u^{-1/2}) H^* \,\mathrm{diag}(m_u^{-1/2})$, $\tilde{I}_N(\underline{\theta}_0)$ is $\tilde{I}_N^*(\underline{\theta}_0)$ with L_N^* replaced by L_N. Thus by the proof of Theorem 4.1, there exists $\delta > 0$ such that

$$P\Big(\lambda_{\min}\big(\tilde{I}_N^*(\underline{\theta}_0)\big) \geq \delta\Big) \to 1. \tag{38}$$

Let the order of the a-th group in (37) be $m^{(a)}$, that is, there are positive numbers p_a and q_a such that $p_a m^{(a)} \le m_u \le q_a m^{(a)}$, $s_{a-1}+1 \le u \le s_a$, $1 \le a \le c$. Let $\underline{\theta}_0 = (\theta_{0u})$. Define, for any $0 < \varepsilon < 2^{-1}(1 \wedge \min_{0 \le u \le s} \theta_{0u})$,

$$S_\varepsilon = \left\{ \sum_{u=s_{a-1}+1}^{s_a} (\theta_u - \theta_{0u})^2 \le \varepsilon^{2(c-a+1)}, 1 \le a \le c \right\},$$

$$\partial S_{\varepsilon, a} = \left\{ \sum_{u=s_{a-1}+1}^{s_a} (\theta_u - \theta_{0u})^2 = \varepsilon^{2(c-a+1)}, \right.$$

$$\left. \sum_{u=s_{b-1}+1}^{s_b} (\theta_u - \theta_{0u})^2 \le \varepsilon^{2(c-b+1)}, b \ne a \right\}, \quad 1 \le a \le c;$$

$$\psi_{a,\theta} = \sum_{u=s_{a-1}+1}^{s_a} \frac{\partial L_N^*}{\partial \theta_u}\bigg|_{\underline{\theta}_0} (\theta_u - \theta_{0u}),$$

$$\psi_{a,\theta}^{(v)} = \sum_{u=s_{a-1}+1}^{s_a} \frac{\partial^2 L_N^*}{\partial \theta_u \, \partial \theta_v}\bigg|_{\underline{\theta}_0} (\theta_u - \theta_{0u})(\theta_v - \theta_{0v}),$$

$$\psi_{a,\theta,\tilde{\theta}}^{(v,w)} = \sum_{u=s_{a-1}+1}^{s_a} \frac{\partial^3 L_N^*}{\partial \theta_u \, \partial \theta_v \, \partial \theta_w}\bigg|_{\tilde{\theta}_u} (\theta_u - \theta_{0u})(\theta_v - \theta_{0v})(\theta_w - \theta_{0w}).$$

We have by Taylor-expanding $\partial L_N^* / \partial \theta_u$ that on $\partial S_{\varepsilon, a}$

$$\sum_{u=s_{a-1}+1}^{s_a} (\theta_u - \theta_{0u}) \frac{\partial L_N^*}{\partial \theta_u}$$

$$= \psi_{a,\theta} + \sum_{v=s_{a-1}+1}^{s_a} \psi_{a,\theta}^{(v)} + \sum_{b<a} \left(\sum_{v=s_{b-1}+1}^{s_b} \psi_{a,\theta}^{(v)} \right) + \sum_{b>a} (\cdots)$$

$$+ \frac{1}{2} \left\{ \sum_{b \le a} \sum_{d \le a} \left(\sum_{v=s_{b-1}+1}^{s_b} \sum_{w=s_{d-1}+1}^{s_d} \psi_{a,\theta,\tilde{\theta}}^{(v,w)} \right) \right.$$

$$\left. + \sum_{b \le a} \sum_{d > a} (\cdots) + \sum_{b > a} \sum_{d \le a} (\cdots) + \sum_{b > a} \sum_{d > a} (\cdots) \right\}$$

$$= I_1 + \cdots + I_4 + \frac{1}{2}(I_5 + \cdots + I_8)$$

$$\left(\tilde{\theta}_u \text{ is between } \underline{\theta}_0 \text{ and } \theta,\ s_{a-1}+1 \le u \le s_a \right).$$

It follows from Lemma 7.5 and the proof of Theorem 4.1 and 4.3 that $\partial L_N^* / \partial \theta_u |_{\underline{\theta}_0} = m_u o_p(1) = m^{(a)} o_p(1)$, $s_{a-1}+1 \le u \le s_a$. So $|I_1| \le m^{(a)} \varepsilon^{c-a+1} o_p(1)$. By the notes we made at the beginning of this proof, it is easy to show

$$|I_3| \le m^{(a)}\varepsilon^{2(c-a+1)+1}O_p(1), \qquad |I_4| \le m^{(a)}\varepsilon^{c-a+2}o_p(1),$$
$$|I_5| \le m^{(a)}\varepsilon^{3(c-a+1)}O_p(1),$$
$$|I_j| \le m^{(a)}\varepsilon^{2(c-a+1)+1}o_p(1), \qquad j = 6,7, \qquad |I_8| \le m^{(a)}\varepsilon^{c-a+3}o_p(1),$$

and (38) implies that $w.p. \to 1$ $I_2 = (\theta - \underline{\theta}_0)' I_{aa}(\theta - \underline{\theta}_0) \ge m^{(a)}\delta p_a \varepsilon^{2(c-a+1)}$. Note that the $O_p(1)$'s and $o_p(1)$'s do not depend on ε. Thus it is easy to conclude that

$$P_{\underline{\theta}_0}\left(\inf_{\theta \in \partial S_{\varepsilon,a}} \left\{ \sum_{u=s_{a-1}+1}^{s_a} (\theta_u - \theta_{0u}) \frac{\partial L_N^*}{\partial \theta_u} \right\} > 0 \right) \to 1, \qquad 1 \le a \le c.$$

The result now follows from Lemma 7.4.

(ii) Suppose the MLE $\hat{\theta}_N^*$ exist w.p. $\to 1$ and are AN. Let the parameters λ and μ_i, $i \in S$, be ordered as $\theta_0, \theta_1, \dots, \theta_s$ and $N-p, m_i$, $i \in S$, ordered correspondingly as $m_0, m_1, \dots, m_s$. Let $I_N(\underline{\theta}_0)$ and $J_N(\underline{\theta}_0)$ be as in Section 3 with $p_i(N) = \sqrt{m_i}$, $0 \le i \le s$.

The consistency of $\hat{\theta}_N^*$ implies, by (i), $p/N \to 0$ and $\tilde{I}_N^*(\underline{\theta}_0) = I_N(\underline{\theta}_0) + o_p(1)$ (see the proof of previous theorems). By Taylor expansion and the note made at the beginning of this proof

$$\begin{aligned}
-Y_{N,u}^* &= -\frac{1}{\sqrt{m_u}} \frac{\partial L_N^*}{\partial \theta_u}\bigg|_{\underline{\theta}_0} \\
&= \sum_{v=0}^{s} \frac{\sqrt{m_v}}{p_v(N)} \left[\tilde{I}_N^*(\underline{\theta}_0)_{uv} + \frac{1}{2} \sum_{w=0}^{s} \frac{1}{\sqrt{m_u m_v}} \frac{\partial^3 L_N^*}{\partial \theta_u\, \partial \theta_v\, \partial \theta_w}\bigg|_{\tilde{\theta}_{N,u}} \left(\hat{\theta}_{Nw}^* - \theta_{0w} \right) \right] \\
&\quad \times p_v(N) \left(\hat{\theta}_{Nv}^* - \theta_{0v} \right),
\end{aligned}$$

where $\tilde{\theta}_{N,u}$ is between $\underline{\theta}_0$ and $\hat{\theta}_N^*$, $0 \le u \le s$. Let $a_N = \max_v \{ p_v(N)^{-1} m_v^{1/2} \}$, $c_{N,v} = p_v(N)^{-1} m_v^{1/2}/a_N$, $\eta_N = (p_v(N)(\hat{\theta}_{Nv}^* - \theta_{0v}))$, then the above equations give $-Y_N^* = a_N X_N$, where $X_N = (I_N(\underline{\theta}_0) + o_p(1))\, \mathrm{diag}(c_{N,v})\eta_N$.

First we assume $I_N(\underline{\theta}_0) \to I(\underline{\theta}_0)$, $J_N(\underline{\theta}_0) \to J(\underline{\theta}_0)$, $M_N(\underline{\theta}_0) \to M(\underline{\theta}_0)$ and $c_{N,v} \to c_v$, $0 \le v \le s$. Then $X_N \to_{\mathscr{L}} N(0, V(\underline{\theta}_0))$, where $V(\underline{\theta}_0) = I(\underline{\theta}_0)\, \mathrm{diag}(c_v)[M(\underline{\theta}_0)' M(\underline{\theta}_0)]^{-1} \mathrm{diag}(c_v) I(\underline{\theta}_0)$; $V(\underline{\theta}_0) \ne 0$ (since that will imply $1 = \max_v c_v = 0$), therefore $V(\underline{\theta}_0)_{uu} \ne 0$ for some u, and $X_{N,u} \to_{\mathscr{L}} N(0, V(\underline{\theta}_0)_{uu})$.

On the other hand, $Y_N^* = Y_N + \delta_N$, where

$$Y_N = \left(\frac{1}{\sqrt{m_u}} \frac{\partial L_N}{\partial \theta_u}\bigg|_{\underline{\theta}_0} \right), \qquad \delta_{N,0} = \frac{p}{\lambda_0 \sqrt{m_0}}, \qquad \delta_{N,u} = \frac{\mathrm{tr}(C_u(\mu_0))}{\sqrt{m_u}}, \qquad u \ne 0.$$

So we have $-Y_{N,u} = a_N X_{N,u} + \delta_{N,u}$. By the proof of Theorem 4.3 and 4.1, $-Y_N \to_{\mathscr{L}} N(0, J(\underline{\theta}_0))$; thus, in particular, $-Y_{N,u} \to_{\mathscr{L}} N(0, J(\underline{\theta}_0)_{uu})$.

Now we can apply Theorem 8.2.3 in Chow and Teicher (1978) to conclude that $a_N \to (J(\underline{\theta}_0)_{uu}/V(\underline{\theta}_0)_{uu})^{1/2}$, so $\{p_v(N)^{-1}m_v^{1/2}\}$, $0 \le v \le s$, are bounded. Thus we can rewrite $-Y_N^* = (I_N(\underline{\theta}_0) + o_p(1))\,\mathrm{diag}(p_v(N)^{-1}m_v^{1/2})\eta_N$ to get $\eta_N = \mathrm{diag}(m_v^{-1/2}p_v(N))\xi_N - b_N$, where $\xi_N = I_N(\underline{\theta}_0)^{-1}(-Y_N + o_p(1))$, $b_N = \mathrm{diag}(m_v^{-1/2}p_v(N))I_N(\underline{\theta}_0)^{-1}\delta_N$.

Since we have $\xi_N \to_{\mathscr{L}} N(0, U(\underline{\theta}_0))$, $\eta_N \to_{\mathscr{L}} N(0, W(\underline{\theta}_0))$, where $U(\underline{\theta}_0) = I(\underline{\theta}_0)^{-1}J(\underline{\theta}_0)I(\underline{\theta}_0)^{-1} > 0$, $W(\underline{\theta}_0) = [M(\underline{\theta}_0)'M(\underline{\theta}_0)]^{-1} > 0$. By considering each component and applying again Theorem 8.2.3 in Chow and Teicher (1978), we get $p_u(N)^{-1}m_u^{1/2} \to (U(\underline{\theta}_0)_{uu}/W(\underline{\theta}_0)_{uu})^{1/2} \in (0, \infty)$ and $b_{N,u} \to 0$, $0 \le u \le s$, which implies $\delta_N \to 0$.

Now we drop the assumption that the limits exist. The result then follows by an a.o.s.; note that $c_{N,v} \le 1$. This completes the proof of the "only if" part (use Lemma 7.5).

The "if" part follows from Theorem 4.4(ii) and Lemma 7.5 (see the proof of Theorem 4.1). □

Acknowledgments. This paper is based on part of the author's Ph.D. dissertation at the University of California at Berkeley. I would sincerely like to thank my advisor, Professor P. J. Bickel for his suggestions and guidance and many helpful discussions. I would also like to thank Professor T. P. Speed for helpful discussions. The author is also grateful to an Associate Editor and two referees for their careful reading of the manuscript and constructive comments.

REFERENCES

ANDERSON, T. W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Ann. Statist.* **1** 135–141.

AZZALINI, A. (1984). Estimation and hypothesis testing for collection of autoregressive time series. *Biometrika* **71** 85–90.

BARNDORFF-NIELSEN, O. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70** 343–365.

BICKEL, P. J. (1993). Estimation in semiparametric model. In Multivariate analysis: Future direction (C. R. Rao, ed.) 55–73. North-Holland, Amsterdam.

BROWN, K. G. (1976). Asymptotic behavior of MINQUE-type estimators of variance components. *Ann. Statist.* **4** 746–754.

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

CHOW, Y. S. and TEICHER, H. (1978). *Probability Theory*. Springer, New York.

COOPER, D. M. and THOMPSON, R. (1977). A note on the estimation of the parameters of autoregressive–moving average process. *Biometrika* **64** 625–628.

CRESSIE, N. (1992). REML estimation in empirical Bayes smoothing of census undercount. *Survey Methodology* **18** 75–94.

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.

DE JONG, P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Theory Related Fields* **75** 261–277.

FOX, R. and TAQQU, M. S. (1985). Noncentral limit theorems for quadratic forms in random variables having long-range dependence. *Ann. Probab.* **13** 428–446.

GLEESON, A. C. and CULLIS, B. R. (1987). Residual maximum likelihood (REML) estimation of a neighbour model for field experiments. *Biometrics* **43** 277–287.

GREEN, P. J. (1985). Linear models for field trials, smoothing and cross-validation. *Biometrika* **72** 527–537.

GUTTORP, P. and LOCKHART, R. A. (1988). On the asymptotic distribution of quadratic forms in uniform order statistics. *Ann. Statist.* **16** 433–449.

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York.

HAMMERSTROM, T. (1978). On the asymptotic optimality of tests and estimates in the presence of increasing numbers of nuisance parameter. Ph.D. dissertation, Univ. California, Berkeley.

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. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika* **61** 383–385.

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

HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York.

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

LAIRD, N. M. and WARE, J. M. (1982). Random effects models for longitudinal data. *Biometrics* **38** 963–974.

LEHMANN, E. H. (1983). *Theory of Point Estimation*. Wiley, New York.

MÄKELÄINEN, T., SCHMIDT, K. and STYAN, G. P. H. (1981). On the existence and uniqueness of the maximum likelihood estimates of a vector-valued parameter in fixed-size samples. *Ann. Statist.* **9** 758–767.

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.

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

PFANZAGL, J. (1993). Incidental versus random nuisance parameters. *Ann. Statist.* **21** 1663–1691.

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

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. *The Statistician* **36** 3–14.

SCHMIDT, W. H. and THRUM, R. (1981). Contributions to asymptotic theory in regression models with linear covariance structure. *Math. Operationsforsch. Statist. Ser. Statist.* **12** 243–269.

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

SPEED, T. P. (1986). Cumulants and partition lattices IV: a.s. convergence of generalized k-statistics. *J. Austral. Math. Soc.* **41** 79–94.

SPEED, T. P. (1991). Comment on "That BLUP is a good thing—The estimation of random effects" by G. K. Robinson. *Statist. Sci.* **6** 42–44.

SZATROWSKI, T. H. and MILLER, J. J. (1980). Explicit maximum likelihood estimates from balanced data in the mixed model of the analysis of variance. *Ann. Statist.* **8** 811–819.

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

VERBYLA, A. P. (1990). A conditional derivation of residual maximum likelihood. *Austral. J. Statist.* **32** 227–230.

WAHBA, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia.

WEISS, L. (1971). Asymptotic properties of maximum likelihood estimators in some nonstandard cases. *J. Amer. Statist. Assoc.* **66** 345–350.

WESTFALL, P. H. (1986). Asymptotic normality of the ANOVA estimates of components of variance in the nonnormal, unbalanced hierarchal mixed model. *Ann. Statist.* **14** 1572–1582.

ZELLNER, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate Student-t error terms. *J. Amer. Statist. Assoc.* **71** 400–405.

DEPARTMENT OF STATISTICS
CASE WESTERN RESERVE UNIVERSITY
10900 EUCLID AVENUE
CLEVELAND, OHIO 44106-7058