Statistica Sinica **8**(1998), 861-885

ASYMPTOTIC PROPERTIES OF THE EMPIRICAL BLUP AND BLUE IN MIXED LINEAR MODELS

Jiming Jiang

Case Western Reserve University

Abstract: We show in a general mixed model the best linear unbiased estimators (BLUE) of fixed effects, with unknown variance components substituted by the REML estimates, are jointly asymptotically normal with the REML estimates. We also prove that given sufficient information the empirical distributions of the best linear unbiased predictors (BLUP) of random effects, again with REML-estimated variance components, converge to the true distributions of the corresponding random effects. As a consequence, we obtain a consistent estimate of the asymptotic variance-covariance matrix of the REML estimates. The results require neither that the data is normally distributed nor that the model is hierarchical (nested).

Key words and phrases: Asymptotic normality, empirical BLUP and BLUE, empirical distributions, mixed models, REML.

1. Introduction

This paper explores asymptotic properties of the empirical best linear unbiased predictors (BLUP) and best linear unbiased estimators (BLUE), i.e., BLUP and BLUE with estimated variance components, in a general mixed linear model. The BLUE and BLUP are methods of estimating fixed and random effects in a mixed model and have important applications in animal breeding, survey sampling, and many other fields (e.g., Henderson (1963, 1973), Harris et al. (1989), Ghosh and Rao (1994), Liski and Nummi (1995)). Robinson (1991) gives a very wide-ranging account of the BLUP and BLUE with examples and applications.

A general mixed linear model can be expressed as

$$y = X\beta + Z_1\alpha_1 + \cdots + Z_s\alpha_s + \epsilon, \tag{1.1}$$

where y is a $N \times 1$ vector of observations, β is a $p \times 1$ vector of fixed effects, α_i is a $m_i \times 1$ vector of random effects, $1 \le i \le s$, ϵ is a $N \times 1$ vector of errors; X is a $N \times p$ design matrix of full rank p, Z_i is a $N \times m_i$ design matrix, $1 \le i \le s$. The components of α_i are i.i.d. $\sim F_i$ with mean 0 and variance $\sigma_i^2, 1 \le i \le s$, the components of ϵ are i.i.d. $\sim F_0$ with mean 0 and variance $\sigma_0^2 > 0$, and $\alpha_1, \ldots, \alpha_s, \epsilon$ are mutually independent.

Under model (1.1), the BLUE for β and BLUP for α_i are given by (e.g., Robinson (1991), Speed (1991))

$$\tilde{\beta} = (X'V_\mu^{-1}X)^{-1}X'V_\mu^{-1}y, \tag{1.2}$$

and

$$\tilde{\alpha_i} = \mu_i Z_i' V_\mu^{-1}(y - X\tilde{\beta}), i = 1, \ldots, s, \tag{1.3}$$

where $V_\mu = I_N + \mu_1 Z_1 Z_1' + \cdots + \mu_s Z_s Z_s', \mu_i = \sigma_i^2/\sigma_0^2, 1 \le i \le s$. Since the above expressions depend on the dispersion parameters or variance components $\mu_i, 1 \le i \le s$, they can not be calculated unless the μ_i's are known. A situation of more practical interest is when the variance components are unknown, in which case one has to substitute the variance components in (1.2) and (1.3) by their estimates, say $\hat{\mu}$. The resulting expressions are usually referred to as the empirical BLUE (EBLUE) and empirical BLUP (EBLUP) (Harville (1991)), given respectively by

$$\hat{\beta} = (X'V_{\hat{\mu}}^{-1}X)^{-1}X'V_{\hat{\mu}}^{-1}y, \tag{1.4}$$

$$\hat{\alpha_i} = \hat{\mu}_i Z_i' V_{\hat{\mu}}^{-1}(y - X\hat{\beta}), \qquad i = 1, \ldots, s. \tag{1.5}$$

Note both the EBLUE and EBLUP are no longer linear functions of the observations y.

There has been much concern about the performance of the EBLUE and EBLUP (e.g., Robinson (1991), Ghosh and Rao (1994) and discussions following the two articles). Asymptotically, it is generally believed that good behavior in estimation of the variance components will be transmitted to the EBLUE and EBLUP. For example, if the variance components estimates are asymptotically normal, one would expect $\hat{\beta}$ to be likewise. There is more to say about the asymptotic properties of the EBLUP. One conjecture is that the empirical distribution (e.d.) of the EBLUPs (i.e., components of $\hat{\alpha_i}$) converges to the true distribution of the corresponding random effects. Such a question arises when there is a need to estimate not just the variance but the entire distribution of some random effects. An important problem in estimating the variance components is to obtain a consistent estimate of the asymptotic variance-covariance matrix (AVCM) and hence the MSE of the estimates. Since (in a non-normal situation) the AVCM may contain unknown parameters such as the third and forth moments of the random effects as well as the variance components, having consistent variance component estimates is not enough to provide a good approximation to the AVCM. One promising idea is to use moments of the e.d. of the EBLUPs to estimate those of the corresponding random effects. Another important problem in mixed model analysis is model diagnostics. In practice, methods are needed for checking the basic model assumptions such as normality, independence, and linearity. Unlike standard regression diagnostics, however, mixed model diagnostics techniques are not well-developed (see Ghosh and Rao (1994), §7.1 for a summary of literature). Since the EBLUPs are thought, in some way, to resemble the random effects, it is natural to consider the use of the EBLUPs to check, for example, whether the random effects are distributed as they are assumed (e.g., Lange and Ryan (1989), Calvin and Sedransk (1991)). Asymptotic results for the e.d. of the EBLUPs will certainly provide a theoretical basis for such methods. Note that we will not assume the random effects or errors are normally or even symmetrically distributed. Despite the high expectation, these are fundamental issues that have to be addressed rather seriously.

In (1.4) and (1.5) we did not specify what the estimate $\hat{\mu}$ was. There are certainly many choices. Two of the most frequently used estimates for the μ_i's and $\lambda = \sigma_0^2$ are the maximum likelihood estimates (MLE) and the restricted maximum likelihood (REML) estimates. As some authors have pointed out (e.g., Harville (1977), Thompson (1980), Fellner (1986), and Speed (1991)), REML and BLUP are intimately connected. In fact, REML equations can be derived by simply equating observed with expected sum of squares of BLUPs. On the other hand, the MLE for λ, μ and β can be obtained by finding the MLE for λ and μ first and then calculating $\hat{\beta}$ by (1.4), using the MLE for μ as $\hat{\mu}$. Miller (1977) considered a special class of the mixed model (1.1) — models having a standard ANOVA structure. Under normality assumption, he proved that asymptotically there exists a sequence of roots to the maximum likelihood (ML) equations for λ, μ and β which are (jointly) asymptotically normal. Recently, Richardson and Welsh (1994) considered asymptotic normality of $\hat{\beta}$, where $\hat{\mu}$ was chosen to be the REML estimate. A main feature of the later work is that the authors did not assume that the data is normally distributed. However, their study was restricted to hierarchical (nested) models. The literature on asymptotic properties of the EBLUP is very limited. Assuming normality, Lange and Ryan (1989) considered asymptotic normality of the e.d. of the (standardized) empirical Bayes estimators (EBE) of the random effects, which in the normal situation are equivalent to the EBLUP (e.g., Harville (1991)). One of the key assumptions they made was that the e.d. of the EBE with unknown parameters replaced by their true values rather than estimates is jointly asymptotically normal with estimates of the unknown parameters, which may not be easy to check.

In this work, we shall consider both the joint asymptotic normality of $\hat{\lambda}, \hat{\mu}$ and $\hat{\beta}$ and the convergence of the e.d. of the EBLUPs for a general mixed model (1.1), where $\hat{\lambda}$ and $\hat{\mu}$ are chosen to be the REML estimates. Since normality is not assumed, we have to make it clear what is meant by the REML estimates. The REML estimates for λ and $\mu_i, 1 \leq i \leq s$ are defined as solutions of the following

REML equations with the requirement that they belong to the parameter space $\Theta = \{\theta : \lambda > 0, \mu_i \geq 0, 1 \leq i \leq s\}$, where $\theta = (\lambda, \mu_1, \ldots, \mu_s)'$. The REML equations under normality are equivalent to

$$z'V(A,\mu)^{-1}z = \lambda(N-p), \tag{1.6}$$

$$z'V(A,\mu)^{-1}A'Z_iZ_i'AV(A,\mu)^{-1}z = \lambda \text{tr}(Z_i'AV(A,\mu)^{-1}A'Z_i), \quad 1 \leq i \leq s, \tag{1.7}$$

where $z = A'y$, A is a $N \times (N-p)$ matrix such that

$$\text{rank}(A) = N - p, \quad A'X = 0, \tag{1.8}$$

and

$$V(A,\mu) = A'A + \sum_{i=1}^{s} \mu_i A'Z_iZ_i'A. \tag{1.9}$$

(e.g., Searle, Casella and McCulloch (1992)). Note that our definition of the REML estimates is the same as that of Richardson and Welsh (1994) except that they did not require $\theta \in \Theta$.

We shall state our main results in §3 and interpret the conditions. In §4 we use examples to illustrate the application of the main theorems. The proofs of the theorems are given in §5, and a concluding remark is made in §6. Some notation will be used throughout, which we summarize in §2. Of course, the best way of using the notation is to skip §2 for the moment, and come back to it when needed.

2. Notation

In addition to the notation already introduced, we define the following.

For an integer n, let I_n and 1_n be the n-dimentional identity matrix and vector of 1's, respectively. Let $A, B, A_1, \ldots, A_s$ be matrices. Define $\|A\| = \lambda_{\max}^{1/2}(A'A), \|A\|_R = \text{tr}^{1/2}(A'A); \text{Cor}(A_1, \ldots, A_s) = (\text{cor}(A_i, A_j))$ if $A_1, \ldots, A_s \neq 0$, and 0 otherwise, where $\text{cor}(A_i, A_j) = \text{tr}(A_i'A_j)/\|A_i\|_R\|A_j\|_R$; $\text{diag}(A_1, \ldots, A_s)$ or $\text{diag}(A_i)$ = the block diagonal matrix with $A_1, \ldots, A_s$ on its diagonal; A_{ll} = the l'th diagonal element of A.

Let X_j be the j's column of $X, 1 \leq j \leq p$; $m = m_1 + \cdots + m_s$, $\mu = (\mu_1 \cdots \mu_s)'$; $b(\mu) = (I_N \sqrt{\mu_1}Z_1 \cdots \sqrt{\mu_s}Z_s)'$, $V(\mu) = AV(A,\mu)^{-1}A'$, $V_0(\mu) = b(\mu)V(\mu)b(\mu)'$, $V_i(\mu) = b(\mu)V(\mu)Z_iZ_i'V(\mu)b(\mu)'$; $U_0 = I_N/\lambda_0$, $V_0 = I_{N-p}/\lambda_0$, $U_i = V_{\mu_0}^{-1/2}Z_iZ_i'\,V_{\mu_0}^{-1/2}$, $V_i = V(A,\mu_0)^{-1/2}A'Z_iZ_i'AV(A,\mu_0)^{-1/2}, 1 \leq i \leq s$. Let $B(\mu) = (X'V_\mu^{-1}X)^{-1}X'V_\mu^{-1}$.

Let $p_i(N), 0 \leq i \leq s$ be sequences of positive numbers. For $0 \leq i, j \leq s$, denote $I_{ij}^{(N)} = \text{tr}(V_iV_j)/p_i(N)p_j(N), K_{ij}^{(N)} = \sum_{l=1}^{N+m}(EW_{Nl}^4 - 3)V_i(\mu_0)_{ll}V_j(\mu_0)_{ll}/\lambda_0^{1_{(i=0)}+1_{(j=0)}}$

$p_i(N)p_j(N)$, where

$$W_{Nl} = \begin{cases} \epsilon_l/\sqrt{\lambda_0}, & 1 \le l \le N, \\ \alpha_{il-N-\sum_{k<i} m_k}/\sqrt{\lambda_0 \mu_{0i}}, & N+\sum_{k<i} m_k+1 \le l \le N+\sum_{k\le i} m_k, 1 \le i \le s. \end{cases}$$

Let $I_N(\theta_0) = (I_{ij}^{(N)}), K_N(\theta_0) = (K_{ij}^{(N)}), J_N(\theta_0) = 2I_N(\theta_0) + K_N(\theta_0), M_N = M_N(\theta_0) = J_N^{-1/2}(\theta_0)I_N(\theta_0)$; Finally, the abbreviation w.p. $\to 1$ stands for "with probability tending to one".

3. Main Results

We shall not assume that the random effects and errors are normally distributed. However, it is required that

$$E\epsilon_1^4 < \infty, E\alpha_{i1}^4 < \infty, \ 1 \le i \le s. \tag{3.1}$$

Since we shall consider the joint asymptotic normality of estimates of both the fixed effects and the variance components, for which a central limit theorem for quadratic forms of random variables is needed, condition (3.1) is necessary. The following definitions are given and explained in Jiang (1996).

Definition 3.1. Model (1.1) is said to be asymptotically-identifiable and infinitely-informative under the invariant class (AI^4) if

$$\liminf \lambda_{\min}(\mathrm{Cor}(V_0, V_1, \ldots, V_s)) > 0 \text{ and } \lim \|V_i\|_R = \infty, 0 \le i \le s.$$

Definition 3.2. Model (1.1) is said to be non-degenerate (ND) if

$$\lambda_{\min}(\mathrm{Var}\,(\binom{\epsilon_1^2}{\epsilon_1})) \wedge \min_{1\le i\le s} \lambda_{\min}(\mathrm{Var}\,(\binom{\alpha_{i1}^2}{\alpha_{i1}})) > 0.$$

Note that the definition of non-degeneracy here is more restrictive than that in Jiang (1996), which is necessary when considering the joint asymptotic normality of both $\hat{\theta}$ and $\hat{\beta}$.

3.1. Asymptotic property of the EBLUE

Let $p_0(N) = \sqrt{N-p}$, $p_i(N)$ be any sequence $\sim \|V_i\|_R$, $1 \le i \le s$. Define (see §2) $p(N) = \mathrm{diag}(p_0(N)\cdots p_s(N))$, $P_N = M_N p(N)$, $Q_N = (X'V_{\theta_0}^{-1}X)^{1/2}$, and $S_N = \begin{pmatrix} I_{s+1} & R_N \\ R_N' & I_p \end{pmatrix}$, where $R_N = J_N^{1/2} T_N C_N$,

$$T_N = (V_i(\mu_0)_{ll} E W_{Nl}^3 / \lambda_0^{1_{(i=0)}} p_i(N))_{0\le i\le s, 1\le l\le N+m},$$

$$C_N = b(\mu_0)V_{\mu_0}^{-1/2}X(X'V_{\mu_0}^{-1}X)^{-1/2} = (C_{N,1}\cdots C_{N,N+m})' = (c_{lj})_{1\le l\le N+m, 1\le j\le p}.$$

Theorem 3.1. *Suppose model* (1.1) *is AI^4 and ND, and* (3.1) *is satisfied. Furthermore, suppose*

$$\max_{1\le l\le N+m} |C_{N,l}| \longrightarrow 0. \tag{3.2}$$

Then there exist w.p. $\to 1$ REML estimates $\hat\lambda_N, \hat\mu_{Ni}, 1 \le i \le s$ such that with $\hat\theta_N = (\hat\lambda_N, (\hat\mu_{Ni})'_{1\le i\le s})'$ and $\hat\beta_N = B(\hat\mu_N)y$ (see §2),

$$S_N \begin{pmatrix} P_N & 0 \\ 0 & Q_N \end{pmatrix} \begin{pmatrix} \hat\theta_N - \theta_0 \\ \hat\beta_N - \beta_0 \end{pmatrix} \stackrel{\mathcal{L}}{\longrightarrow} N(0, I_{s+p+1}). \tag{3.3}$$

Remark 1. Condition (3.2) corresponds to negligibility. It is easy to show by examples that this condition can not be dropped. In particular, when $s = 0$, i.e., when there is no random effect, model (1.1) reduces to a regression model and condition (3.2) to condition (4.3) of Lai and Wei (1982).

Remark 2. The AI^4 condition is required for estimating the variance components. Some may wonder why there is no similar assumption for the estimation of the fixed effects, as there were in Millar (1977) and Richardson and Welsh (1994). Here is the explanation:

(1) The REML estimates are defined as solution of the REML equations, which do not have an analytic solution in general. Asymptotic-identifiability ensures the existence (w.p.$\to$ 1) of a sequence of roots to the REML equations, which are consistent if infinite-informativity is further satisfied. AI^4, plus non-degeneracy, also implies negligibility (i.e., (26) in Jiang (1996)), which leads to asymptotic normality. Note that AI^4 implies $m_i \to \infty$, $1 \le i \le s$.

(2) Unlike $\hat\theta_N, \hat\beta_N$ has an explicit expression (1.4) after the variance components being estimated. Furthermore, (3.3) is a very general form of asymptotic normality, which does not necessarily imply consistency unless one specifies the order of Q_N (see Corollary 3.1 below). Therefore all one needs here is, in addition to AI^4 and ND, the negligibility (3.2).

Remark 3. The sequence $\hat\theta_N$ in Theorem 3.1 can be identified under a strengthening of AI^4 (Jiang (1997)).

We say a sequence of matrices $\{B_N\}$ is bounded from above if $\limsup \|B_N\| < \infty$, and bounded from below if $\limsup \|B_N^{-1}\| < \infty$.

Corollary 3.1. *Under the assumption of Theorem* 3.1*, if there are sequences $l_j(N), 1 \le j \le p$ such that*

$$\begin{aligned} 0 &< \liminf \lambda_{\min}(l(N)^{-1}(X'V_{\mu_0}^{-1}X)l(N)^{-1}) \\ &\le \limsup \lambda_{\max}(l(N)^{-1}(X'V_{\mu_0}^{-1}X)l(N)^{-1}) < \infty, \end{aligned} \tag{3.4}$$

where $l(N) = \mathrm{diag}(l_1(N) \cdots l_p(N))$, *then*

$$S_N \begin{pmatrix} M_N & 0 \\ 0 & H_N \end{pmatrix} \begin{pmatrix} p(N)(\hat{\theta}_N - \theta_0) \\ l(N)(\hat{\beta}_N - \beta_0) \end{pmatrix} \xrightarrow{\mathcal{L}} N(0, I_{s+p+1}), \tag{3.5}$$

where $H_N = Q_N l(N)^{-1}$, *and the normalizing matrix* $S_N \begin{pmatrix} M_N & 0 \\ 0 & H_N \end{pmatrix}$ *is bounded from above and below.*

Thus, under the conditions of Corollary 3.1, $\hat{\beta}_N$ is consistent if and only if $l_j(N) \to \infty$, $1 \le j \le p$.

By (3.5), the asymptotic covariance matrix of $\begin{pmatrix} \hat{\theta}_N - \theta_0 \\ \hat{\beta}_N - \beta_0 \end{pmatrix}$ is given by

$$\begin{aligned} V^{(N)} &= \begin{pmatrix} P(N)^{-1} M_N^{-1} M_N^{'-1} P(N)^{-1} & P(N)^{-1} M_N^{-1} R_N H_N^{'-1} l(N)^{-1} \\ l(N)^{-1} H_N^{-1} R_N' M_N^{'-1} P(N)^{-1} & l(N)^{-1} H_N^{-1} H_N^{'-1} l(N)^{-1} \end{pmatrix} \\ &= \begin{pmatrix} \tilde{I}_N^{-1} \tilde{J}_N \tilde{I}_N^{-1} & \tilde{I}_N^{-1} \tilde{C}_N Q_N^{-1} \\ Q_N^{-1} \tilde{C}_N' \tilde{I}_N^{-1} & (X' V_{\theta_0}^{-1} X)^{-1} \end{pmatrix}, \end{aligned} \tag{3.6}$$

where $\tilde{I}_N = (\mathrm{tr}(V_i V_j))_{0 \le i,j \le s}$, $\tilde{J}_N = 2\tilde{I}_N + \tilde{K}_N$ with

$$\tilde{K}_N = \Big(\sum_{l=1}^{N+m} (EW_{NL}^4 - 3) V_i(\mu_0)_{ll} V_j(\mu_0)_{ll} / \lambda_0^{1_{(i=0)} + 1_{(j=0)}} \Big)_{0 \le i,j \le s},$$

and $\tilde{C}_N = (\sum_{l=1}^{N+m} EW_{Nl}^3 V_i(\mu_0)_{ll} c_{lj} / \lambda_0^{1_{(i=0)}})_{0 \le i \le s, 1 \le j \le p}$. Note that $V^{(N)}$ does not depend on the normalizing sequences $p(N)$ and $l(N)$. In particular, if $E\epsilon_1^3 = 0, E\alpha_{i1}^3 = 0, 1 \le i \le s$, which are true when the errors and random effects are symmetrically distributed, then

$$V^{(N)} = \begin{pmatrix} (\tilde{I}_N)^{-1} \tilde{J}_N (\tilde{I}_N)^{-1} & 0 \\ 0 & (X' V_{\theta_0}^{-1} X)^{-1} \end{pmatrix}, \tag{3.7}$$

i.e. the $\hat{\theta}_N$ and $\hat{\beta}_N$ are *asymptotically independent*. If furthermore $\kappa_0^{(4)} = E\epsilon_1^4/\sigma_{00}^4 - 3 = 0, \kappa_i^{(4)} = E\alpha_{i1}^4/\sigma_{0i}^4 - 3 = 0, 1 \le i \le s$, which hold under normality, then $(\tilde{I}_N)^{-1} \tilde{J}_N (\tilde{I}_N)^{-1} = 2(\tilde{I}_N)^{-1}$.

3.2. Asymptotic property of the EBLUP

Before stating any asymptotic result, let us speculate about what to expect. One conjecture is that the e.d.'s of the EBLUPs converge to the true distributions of their corresponding random effects, given general conditions that ensure the consistency of the REML estimates. This is, however, not true.

Example 3.1. Consider a simple random effects model $y_i = \alpha_i + \epsilon_i, i = 1, \ldots, m$; and $y_i = \epsilon_i, i = m+1, \ldots, 2m$, where the random effects $\alpha_1, \ldots, \alpha_m$ are i.i.d. $\sim N(0, \sigma_\alpha^2)$, and the errors $\epsilon_1, \ldots, \epsilon_{2m}$ i.i.d. $\sim N(0, \sigma_\epsilon^2)$. The EBLUPs for the α's are

$$\hat{\alpha}_i = \frac{\hat{\mu}}{1+\hat{\mu}} y_i = \frac{\hat{\mu}}{1+\hat{\mu}} (\alpha_i + \epsilon_i), i = 1, \ldots, m.$$

Therefore

$$\frac{1}{m} \sum_{i=1}^{m} 1_{(\hat{\alpha}_i \le x)} \xrightarrow{P} \Phi\Big(\frac{\sqrt{1+\mu}}{\sqrt{\lambda}\mu} x\Big)$$

for all x. But $\Phi(\frac{\sqrt{1+\mu}}{\sqrt{\lambda}\mu} x) \neq P(\alpha_1 \le x) = \Phi(\frac{1}{\sqrt{\lambda\mu}} x)$ no matter what λ and μ.

Example 3.2. Consider the following random effects model $y_{ij} = \beta + \alpha_i + \epsilon_{ij}, i = 1, \ldots, m, j = 1, \ldots, n (m, n \ge 2)$, where β is an unknown mean, α's and ϵ's are random effects and errors, respectively. The EBLUPs are

$$\hat{\alpha}_i = \frac{\hat{\mu} n}{1 + \hat{\mu} n} (\alpha_i - \bar{\alpha}_. + \bar{\epsilon}_{i.} - \bar{\epsilon}_{..}), \quad i = 1, \ldots, m.$$

It follows that $\frac{1}{m} \sum_{i=1}^{m} 1_{(\hat{\alpha}_i \le x)} \xrightarrow{P} P(\alpha_1 \le x)$for all x if and only if $m,\ n \to \infty$.

Example 3.3. This time we consider a two-way balanced random effects model $y_{ijk} = \alpha_i + \gamma_{ij} + \epsilon_{ijk}, i = 1, \ldots, m, j = 1, \ldots, n, k = 1, \ldots, l$. By direct calculation we obtain

$$\hat{\alpha}_i = \frac{\hat{\mu}_1 nl}{1 + \hat{\mu}_2 l + \hat{\mu}_1 nl} (\alpha_i + \bar{\gamma}_{i.} + \bar{\epsilon}_{i..}),\ i = 1, \ldots, m,$$
$$\hat{\gamma}_{ij} = \frac{\hat{\mu}_2 l}{1 + \hat{\mu}_2 l} \Big(\gamma_{ij} + \frac{1 + \hat{\mu}_2 l}{1 + \hat{\mu}_2 l + \hat{\mu}_1 nl} \alpha_i - \frac{\hat{\mu}_1 nl}{1 + \hat{\mu}_2 l + \hat{\mu}_1 nl} \bar{\gamma}_{i.} + \bar{\epsilon}_{ij.} - \bar{\epsilon}_{i..}\Big),$$
$$i = 1, \ldots, m, j = 1, \ldots, n.$$

For the e.d. of the $\hat{\alpha}$'s and $\hat{\gamma}$'s to converge to the true distributions of the α's and γ's, one needs m, n, l all $\to \infty$.

All of the above examples satisfy the AI^4 condition, therefore the REML estimates for the variance components are consistent (Jiang (1996)). However, the e.d. of the EBLUPs will either not converge to the true distribution of the corresponding random effects, or converge only under further restriction on the way that the size of the design is growing. It is not hard to see why in Example 3.1 and 3.2 (when n does not $\to \infty$) the e.d. of the $\hat{\alpha}$'s does not converge to the distribution of the α's: if the number of repetition of each random effect does not increase, how can we distinguish the distribution of random effects from that of the errors. An analogous explanation may be given for Example 3.3 but we leave this for later discussion.

The following theorem gives sufficient conditions for the convergence of the e.d.'s of the EBLUPs to the true distributions of their corresponding random effects. For $1 \leq i \leq s$, let $V_{i-}(A,\mu) = A'A + \sum_{j \neq i} \mu_j A' Z_j Z_j' A, V_{i-}(\mu) = AV_{i-}^{-1}(A,\mu)A', H_i(\mu) = Z_i' V_{i-}(\mu) Z_i$; and Θ^o be the interior of Θ.

Theorem 3.2. *Suppose model* (1.1) *is* AI^4 *and* $\theta_0 \in \Theta^o$, *and* (3.1) *is satisfied. Furthermore, suppose*

$$\frac{1}{m_i \wedge m_j} \mathrm{tr}((Z_i' V(\mu_0) Z_j)(Z_i' V(\mu_0) Z_j)') \longrightarrow 0, 1 \leq i \neq j \leq s, \tag{3.8}$$

$$\frac{1}{m_i} \mathrm{tr}((I_{mi} + \mu_{0i} H_i(\mu_0))^{-1}) \longrightarrow 0,\ 1 \leq i \leq s. \tag{3.9}$$

Then there exists w.p.$\to$ 1 REML estimates $\hat{\lambda}_N, \hat{\mu}_{Ni}, 1 \leq i \leq s$ such that $\{(\|V_0\|_R(\hat{\lambda}_N - \lambda_0), (\|V_i\|_R(\hat{\mu}_{Ni} - \mu_{0i}))'_{1\leq i \leq s})'\}$ is bounded in probability and for $1 \leq i \leq s$

$$\frac{1}{m_i} \sum_{k=1}^{m_i} 1_{(\hat{\alpha}_{ik} \leq x)} \xrightarrow{P} F_i(x) = P(\alpha_{i1} \leq x),\ \ x \in C_{F_i}, \tag{3.10}$$

$$\frac{1}{m_i} \sum_{k=1}^{m_i} \hat{\alpha}_{ik}^q \xrightarrow{P} E\alpha_{i1}^q, \qquad q = 1,2,3,4, \tag{3.11}$$

where $\hat{\alpha}_i = (\hat{\alpha}_{ik})_{1 \leq k \leq m_i}, 1 \leq i \leq s$ are the BLUP's with the unknown variance components substituted by the above REML estimates, and $C_F = \{x \in R : F$ is continuous at $x\}$. If moreover, F_i is continuous for some i $(1 \leq i \leq s)$, then the convergence in (3.10) is uniform in $x \in R$ for the same i.

Remark 1. As in Remark 3 following Theorem 3.1, the sequences $\hat{\lambda}_N, \hat{\mu}_{Ni}, 1 \leq i \leq s$ may be identified.

Remark 2. Taking $q = 2$ in (3.11), the theorem provides (another set of) consistent estimates of the variance components $\sigma_i^2 = E\alpha_{i1}^2, 1 \leq i \leq s$.

To see what conditions (3.8) and (3.9) actually mean, let us consider an important special case of model (1.1) — the balanced case.

Example 3.4. (balanced mixed models) A balanced r-factor mixed model of the analysis of variance can be written as (1.1) where each design matrix is a Kronecker product. Namely, $X = 1_{n_1}^{d_1} \otimes \cdots \otimes 1_{n_{r+1}}^{d_{r+1}}, Z_i = 1_{n_1}^{i_1} \otimes \cdots \otimes 1_{n_{r+1}}^{i_{r+1}}$, where $d = (d_1 \cdots d_{r+1}) \in S_{r+1} = \{0,1\}^{r+1}, i = (i_1 \cdots i_{r+1}) \in S \subset S_{r+1}; 1_n^0 = I_n, 1_n^1 = 1_n$ (e.g., Jiang (1996)). Among our previous examples, Example 3.2 and 3.3 are balanced. The following can be shown by Lemma 7.3 in Jiang (1996) and an argument similar to its proof:

(i) (3.8) is true in the balanced case provided for $1 \le q \le r+1, n_q \to \infty$ if $i_q = 0$ for some $i \in S$ (i.e., $n_q \to \infty$ if factor q appears in the indexes of some random effects).

(ii) (3.9) is true provided the model is unconfounded and $m_i/N \longrightarrow 0, \forall i \in S$, which means the number of appearances of each random effect (i.e., each component of any vector $\alpha_i, i \in S$) must go to ∞.

Note. (ii) in the above agrees with the result of some recent work by Verbeke and Lesaffre (1996), where they found that the EBLUP is not asymptotically accurate if the number of repeated measurements corresponding to the same random effect is not growing.

Applying (i) and (ii) to Example 3.3, we see (3.8) and (3.9) are true provided $m, n, l \to \infty$.

From Theorem 3.2 we obtain consistent estimates of the kurtosis. namely, $\hat{\kappa}_i^{(3)} = (\frac{1}{m_i}\sum_{k=1}^{m_i}\hat{\alpha}_{ik}^3)/(\hat{\lambda}\hat{\mu}_i)^{3/2}$, $\hat{\kappa}_i^{(4)} = (\frac{1}{m_i}\sum_{k=1}^{m_i}\hat{\alpha}_{ik}^4)/(\hat{\lambda}\hat{\mu}_i)^2 - 3, 1 \le i \le s$, where $\hat{\lambda}$ and $\hat{\mu}_i'$s are the REML estimates. Another set of consistent estimates may be obtained by replacing $\hat{\lambda}\hat{\mu}_i$ by $\frac{1}{m_i}\sum_{k=1}^{m_i}\hat{\alpha}_{ik}^2, 1 \le i \le s$.

However, in order to get an approximation of the asymptotic covariance matrix (3.6), one also needs to estimate $\kappa_0^{(r)}, r = 3, 4$ consistently. This can be achieved by defining the following "EBLUP" for the errors $\epsilon_i, 1 \le i \le N$,

$$\hat{\epsilon} = y - X\hat{\beta} - Z_1\hat{\alpha}_1 - \cdots - Z_s\hat{\alpha}_s = V(\hat{\mu})y, \tag{3.12}$$

which generalizes the "residuals" in linear regression. A similar result to Theorem 3.2 can be proved for $\hat{\epsilon}$.

Lemma 3.1. *If in addition to the conditions of Theorem* 3.2 *we have*

$$m_i/N \longrightarrow 0, 1 \le i \le s, \tag{3.13}$$

then

$$\frac{1}{N}\sum_{k=1}^{N} 1_{(\hat{\epsilon}_k \le x)} \xrightarrow{P} F_0(x) = P(\epsilon_1 \le x), \ \ x \in C_{F_0}, \tag{3.14}$$

$$\frac{1}{N}\sum_{k=1}^{N}\hat{\epsilon}_k^q \xrightarrow{P} E\epsilon_1^q, \ \ q = 1, 2, 3, 4, \tag{3.15}$$

where $\hat{\epsilon} = (\hat{\epsilon}_k)_{1 \le k \le N}$ *is given by* (3.12) *with the unknown variance components substituted by the REML estimates in Theorem* 3.2. *If moreover,* F_0 *is continuous, then the convergence in* (3.14) *is uniform in* $x \in R$.

Thus, $\hat{\kappa_0}^{(3)} = (\frac{1}{N}\sum_{k=1}^{N}\hat{\epsilon}_k^3)/\hat{\lambda}^{3/2}, \hat{\kappa_0}^{(4)} = (\frac{1}{N}\sum_{k=1}^{N}\hat{\epsilon}_k^4)/\hat{\lambda}^2 - 3$ are consistent estimates of $\kappa_0^{(r)}, r = 3, 4$. As before, one may replace the REML estimates $\hat{\lambda}$ by $\frac{1}{N}\sum_{k=1}^{N}\hat{\epsilon}_k^2$ and still maintain consistency.

4. Examples

The first two examples are used to illustrate the asymptotics for the EBLUE.

Example 4.1. Consider, once again, Example 3.2. Since the model is balanced and unconfounded, the AI^4 condition is satisfied provided $m \to \infty$ (Jiang (1996)). By direct calculation, it can be shown that

$$\max_{1\le l\le N+m} |C_{N,l}| \le (\frac{n^{-1} \vee \mu_{0n}}{m(1+\mu_{0n})})^{1/2} \longrightarrow 0 \text{ as } m \to \infty,$$

and $X'V_{\theta_0}^{-1}X = [\lambda_0(1+\mu_{0n})]^{-1}mn$. Thus, by Corollary 3.1 the REML estimates for θ and EBLUE for β are jointly asymptotically normal. Further calculation shows, by (3.5), that

$$\begin{pmatrix} \sqrt{mn-1}(\hat\lambda-\lambda_0) \\ \sqrt{m}(\hat\mu-\mu_0) \\ \sqrt{m}(\hat\beta-\beta_0) \end{pmatrix} \xrightarrow{\mathcal{L}} N(0, \begin{pmatrix} \lambda_0^2(2+\kappa_{0\epsilon}^{(4)}) & 0 & 0 \\ 0 & \mu_0^2(2+\kappa_{0\alpha}^{(4)}) & \lambda_0^{1/2}\mu_0^{3/2}\kappa_{0\alpha}^{(3)} \\ 0 & \lambda_0^{1/2}\mu_0^{3/2}\kappa_{0\alpha}^{(3)} & \lambda_0\mu_0 \end{pmatrix}), \tag{4.1}$$

where $\kappa_{0\epsilon}^{(4)} = E\epsilon_{11}^4/\sigma_{0\epsilon}^4 - 3$, $\kappa_{0\alpha}^{(4)} = E\alpha_1^4/\sigma_{0\alpha}^4 - 3$, and $\kappa_{0\alpha}^{(3)} = E\alpha_1^3/\sigma_{0\alpha}^3$.

One may compare Example 4.1 with the same example discussed by Miller (1977) under normality (therefore $\kappa_{0\epsilon}^{(r)} = \kappa_{0\alpha}^{(r)} = 0, r = 3,4$), where the variance components were estimated by the MLE. A generalization of Example 4.1 was also considered by Richardson and Welsh (1994) (hereafter RW), although they did not work out the details. Note that one condition is missing in both Theorem 1 and 2 of RW: There should be an extra assumption in condition B of RW which ensures the nonsingularity of F there. For example, in Example 4.1, a sufficient condition for $|F| \neq 0$ is that Var $(\epsilon_{11}^2) > 0$ and Var $(\alpha_1^2) > 0$. In our case, since we are considering the joint asymptotic normality of $\hat\theta$ and $\hat\beta$, the corresponding condition is Var $((\begin{smallmatrix}\epsilon_{11}^2 \\ \epsilon_{11}\end{smallmatrix})) > 0$ and Var $((\begin{smallmatrix}\alpha_1^2 \\ \alpha_1\end{smallmatrix})) > 0$, i.e., ND.

Example 4.2. In this example, we show that condition $A(ii), B(i)$ and $B(iii)$ of RW for hierarchical models imply (3.2). In fact, under the hierarchical structure of RW (and keeping the notation there)

$$C_N = \frac{\sigma_{0c}}{\sqrt{n}} \begin{pmatrix} I_n \\ \sigma_{01}Z_1' \\ \vdots \\ \sigma_{0c-1}Z_{c-1}' \end{pmatrix} \begin{pmatrix} V_{01}^{-1}X_1 \\ \vdots \\ V_{0g}^{-1}X_g \end{pmatrix} (n^{-1}X'V_0^{-1}X)^{-1/2}.$$

Let $Z(i,k)' = (Z_{i1k}' \cdots Z_{igk}')$, where Z_{ijk} is a m_j-vector corresponding to $V_j, 1 \le$

$j \le g$, then

$$|Z(i,k)'V_0^{-1}X(n^{-1}X'V_0^{-1}X)^{-1/2}| = |(n^{-1}X'V_0^{-1}X)^{-1/2}\sum_{j=1}^{g}X_j'V_{0j}^{-1}Z_{ijk}|$$

$$\le \|(n^{-1}X'V_0^{-1}X)^{-1/2}\|\sum_{j=1}^{g}|X_jV_{oj}^{-1}Z_{ijk}|$$

$$\le \sigma_{0c}^{-2}\|\cdots\|(n^{-1}\sum_{j=1}^{g}\|X_j\|_R^{2+\delta})^{\frac{1}{2+\delta}}n^{\frac{1}{2+\delta}}\sum_{j=1}^{g}|Z_{ijk}|^2$$

$$\le \sigma_{0c}^{-2}Q\|(n^{-1}X'V_0^{-1}X)^{-1/2}\|(n^{-1}\sum_{j=1}^{g}\|X_j\|_R^{2+\delta})^{\frac{1}{2+\delta}}n^{\frac{1}{2+\delta}},\ 1\le k\le p_i, 1\le i\le c-1,$$

by A(ii) of RW and the fact that $|Z_{ijk}|^2$ is 0 or a positive integer. Also

$$\|V_{0j}^{-1}X_j(n^{-1}X'V_0^{-1}X)^{-1/2}\|_R$$
$$\le \sigma_{0c}^{-1}\|(n^{-1}X'V_0^{-1}X)^{-1/2}\|\Big(n^{-1}\sum_{j=1}^{g}\|X_j\|_R^{2+\delta}\Big)^{\frac{1}{2+\delta}}n^{\frac{1}{2+\delta}},\ 1\le j\le g.$$

Thus $\max_{1\le l\le N+m}|C_{N,l}| \le \sigma_{0c}^{-1}(Q\vee 1)\lambda_{\max}((n^{-1}X'V_0^{-1}X)^{-1})\ (n^{-1}\sum_{j=1}^{g}\|X_j\|_R^{2+\delta})^{\frac{1}{2+\delta}}n^{-\frac{\delta}{2(2+\delta)}} \longrightarrow 0$ by conditions B(i) and B(ii) of RW.

Our third example is used to demonstrate the asymptotics for the EBLUP.

Example 4.3. Consider an unbalanced mixed model $y_{ijk} = \beta_i + \alpha_{ij} + \epsilon_{ijk}$, $i = 1,\ldots,p, j = 1,\ldots,m_i, k = 1,\ldots,n_i$, where the β's, α's and ϵ's are fixed, random effects and errors, respectively. Suppose the model is AI^4 and $\theta_0 \in \Theta^o$. Write the model as $y = X\beta+Z\alpha+\epsilon$, where $X = \text{diag}(1_{m_i}\otimes 1_{n_i}), Z = \text{diag}(I_{m_i}\otimes 1_{n_i})$. The EBLUPs can be derived as

$$\hat{\alpha}_{ij} = \frac{\hat{\mu}n_i}{1+\hat{\mu}n_i}(\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot}), \tag{4.2}$$

$i = 1,\ldots,p, j = 1,\ldots,m_i$, where $\bar{y}_{ij\cdot} = \sum_{k=1}^{n_i}y_{ijk}/n_i, \bar{y}_{i\cdot\cdot} = \sum_{j=1}^{m_i}\sum_{k=1}^{n_i}y_{ijk}/m_in_i$.

On the other hand, it is easy to show that $m/N = \sum_{i=1}^{p}m_i/\sum_{i=1}^{p}m_in_i \to 0$, and

$$\frac{1}{m}\text{tr}((I_m+\mu_0H(\mu_0))^{-1}) = \sum_{i=1}^{p}\frac{m_i+\mu_0n_i}{1+\mu_0n_i}/\sum_{i=1}^{p}m_i \longrightarrow 0 \tag{4.3}$$

provided $n_i \longrightarrow \infty, 1\le i\le p$ and $m = \sum_{i=1}^{p}m_i \longrightarrow \infty$. Thus, by Theorem 3.2 and Lemma 3.1, the e.d.'s of the $\hat{\alpha}'$s and $\hat{\epsilon}'$s converge to the true distributions of the α's and ϵ's, respectively, where

$$\hat{\epsilon}_{ijk} = y_{ijk} - \bar{y}_{i\cdot\cdot} - \hat{\alpha}_{ij} = y_{ijk} - \frac{\hat{\mu}n_i}{1+\hat{\mu}n_i}\bar{y}_{ij\cdot} - \frac{1}{1+\hat{\mu}n_i}\bar{y}_{i\cdot\cdot}, \tag{4.4}$$

$i = 1, \ldots, p, j = 1, \ldots, m_i, k = 1, \ldots, n_i$. The estimated kurtoses

$$\hat{\kappa}_\alpha^{(3)} = \Big(\frac{1}{m}\sum_{i=1}^{p}\sum_{j=1}^{m_i}\hat{\alpha}_{ij}^3\Big)/(\hat{\lambda}\hat{\mu})^{3/2},\ \hat{\kappa}_\alpha^{(4)} = \Big(\frac{1}{m}\sum_{i=1}^{p}\sum_{j=1}^{m_i}\hat{\alpha}_{ij}^4\Big)/(\hat{\lambda}\hat{\mu})^{2} - 3,$$

$$\hat{\kappa}_\epsilon^{(3)} = \Big(\frac{1}{N}\sum_{i=1}^{p}\sum_{j=1}^{m_i}\sum_{k=1}^{n_i}\hat{\epsilon}_{ijk}^3\Big)/\hat{\lambda}^{3/2},\ \hat{\kappa}_\epsilon^{(4)} = \Big(\frac{1}{N}\sum_{i=1}^{p}\sum_{j=1}^{m_i}\sum_{k=1}^{n_i}\hat{\epsilon}_{ijk}^4\Big)/\hat{\lambda}^{2} - 3$$

are consistent, where $\hat{\lambda}$ and $\hat{\mu}$ are the REML estimates of $\lambda = \sigma_\epsilon^2$ and $\mu = \sigma_\alpha^2/\sigma_\epsilon^2$. Therefore by (3.6) an approximation to the asymptotic covariance matrix of the REML estimates and the EBLUE is given by:

$$\mathrm{Var}\,(\begin{pmatrix}\hat{\lambda}\\ \hat{\mu}\\ \hat{\beta}\end{pmatrix}) \sim \begin{pmatrix}\hat{I}^{-1}\hat{J}\hat{I}^{-1} & \hat{I}^{-1}\hat{F}\\ \hat{F}'\hat{I}^{-1} & \hat{G}\end{pmatrix}, \tag{4.5}$$

where $\hat{I} = (\hat{I}_{ij})_{0\le i,j\le 1}$, $\hat{J} = (\hat{J}_{ij})_{0\le i,j\le 1}$, $\hat{F} = (\hat{F}_{ij})_{0\le i\le 1, 1\le j\le p}$, $\hat{G} = \mathrm{diag}(\hat{\lambda}(1 + \hat{\mu}n_i)/m_i n_i)$;

$$\hat{I}_{00} = (N-p)/\hat{\lambda}^2, \hat{I}_{01} = \hat{I}_{10} = \sum_{t=1}^{p}(m_t - 1)(\frac{\hat{\mu}n_t}{1+\hat{\mu}n_t})/\hat{\lambda}\hat{\mu}, \hat{I}_{11}$$
$$= \sum_{t=1}^{p}(m_t - 1)(\frac{\hat{\mu}n_t}{1+\hat{\mu}n_t})^2/\hat{\mu}^2;$$
$$\hat{J}_{00} = 2\hat{I}_{00} + \hat{\lambda}^{-2}\hat{\kappa}_\epsilon^{(4)}\sum_{t=1}^{p}\frac{m_t n_t}{(1+\hat{\mu}n_t)^2}(1 + \hat{\mu}(n_t - 1) - \frac{1}{m_t n_t})^2$$
$$+\hat{\lambda}^{-2}\hat{\kappa}_\alpha^{(4)}\sum_{t=1}^{p} m_t(1 - \frac{1}{m_t})^2(\frac{\hat{\mu}n_t}{1+\hat{\mu}n_t})^2,$$
$$\hat{J}_{01} = \hat{J}_{10} = 2\hat{I}_{01} + \hat{\lambda}^{-1}\hat{\kappa}_\epsilon^{(4)}\sum_{t=1}^{p}\frac{m_t n_t}{(1+\hat{\mu}n_t)^3}(1 + \hat{\mu}(n_t - 1) - \frac{1}{m_t n_t})(1 - \frac{1}{m_t})$$
$$+\hat{\lambda}^{-1}\hat{\mu}^{-2}\hat{\kappa}_\alpha^{(4)}\sum_{t=1}^{p} m_t(1 - \frac{1}{m_t})^2(\frac{\hat{\mu}n_t}{1+\hat{\mu}n_t})^3,$$
$$\hat{J}_{11} = 2\hat{I}_{11} + \hat{\kappa}_\epsilon^{(4)}\sum_{t=1}^{p}\frac{m_t n_t}{(1+\hat{\mu}n_t)^4}(1 - \frac{1}{m_t})^2 + \hat{\mu}^{-4}\hat{\kappa}_\alpha^{(4)}\sum_{t=1}^{p}(1 - \frac{1}{m_t})^2(\frac{\hat{\mu}n_t}{1+\hat{\mu}n_t})^4;$$
$$\hat{F}_{0j} = \hat{\lambda}^{-1/2}\hat{\kappa}_\epsilon^{(3)}\frac{1 + \hat{\mu}(n_j - 1) - m_j^{-1}n_j^{-1}}{(1+\hat{\mu}n_j)^{1/2}} + (\hat{\lambda}^{-1}\hat{\mu})^{1/2}\hat{\kappa}_\alpha^{(3)}\frac{\hat{\mu}n_j(1 - m_j^{-1})}{(1+\hat{\mu}n_j)^{1/2}},$$
$$\hat{F}_{1j} = \hat{\lambda}^{1/2}\hat{\kappa}_\epsilon^{(3)}\frac{1 - m_j^{-1}}{(1+\hat{\mu}n_j)^{3/2}} + (\hat{\lambda}\hat{\mu})^{1/2}\hat{\kappa}_\alpha^{(3)}\frac{n_j^2(1 - m_j^{-1})}{(1+\hat{\mu}n_j)^{3/2}}.$$

Note that (4.5) is derived without normality or any symmetry assumption for the distributions of the α's and ϵ's (otherwise it would be much simpler).

5. Proofs

The proof of Theorem 3.1 is based on a generalization of Theorem 5.1 in Jiang (1996), i.e., a central limit theorem for combination of quadratic and linear forms of random variables.

For each n, let $X_{n1},\ldots,X_{nk_n}$ be independent with mean 0, $A_n=(a_{nij})_{1\le i,j\le k_n}$ be a symmetric matrix, and $b_n=(b_{ni})_{1\le i\le k_n}$ be a vector. Let $\mathcal{X}_n=(X_{n1},\ldots,X_{nk_n})'$, $\mathcal{A}_n=\{1\le i\le k_n : a_{nii}\ne 0\}$.

Theorem B. *Suppose*

$$\inf_n \min_{1\le i\le k_n} \lambda_{\min}\Big(\mathrm{Var}\,(\binom{X_{ni}^2}{X_{ni}})\Big) > 0, \tag{5.1}$$

$$\sup_n\big(\max_{i\in\mathcal{A}_n} EX_{ni}^4 1_{(|X_{ni}|>x)}\big)\vee\big(\max_{1\le i\le k_n} EX_{ni}^2 1_{(|X_{ni}|>x)}\big)\underset{x\to\infty}{\longrightarrow} 0. \tag{5.2}$$

Then

$$\frac{\mathcal{X}_n'A_n\mathcal{X}_n+b_n'\mathcal{X}_n-E\mathcal{X}_n'A_n\mathcal{X}_n}{[\mathrm{Var}\,(\mathcal{X}_n'A_n\mathcal{X}_n+b_n'\mathcal{X}_n)]^{1/2}}\xrightarrow{\mathcal{L}} N(0,1) \tag{5.3}$$

provided

$$\frac{\lambda_{\max}(A_n^2)\vee\max_{1\le i\le k_n} b_{ni}^2}{\mathrm{tr}(A_n^2)+\sum_{i=1}^{k_n} b_{ni}^2}\longrightarrow 0. \tag{5.4}$$

Theorem B is a special case of a more general theorem. Let $A_n^o=A_n-\mathrm{diag}(a_{nii},1\le i\le k_n)$. Given numbers $\{L_{ni},1\le i\le k_n,n\ge 1\}$, define $\gamma_{ni}^{(1)}=EX_{ni}^4 1_{(|X_{ni}|\le L_{ni})},\gamma_{ni}^{(2)}=E(X_{ni}^2-1)^4 1_{(|X_{ni}|\le L_{ni})};\delta_{ni}^{(1)}=EX_{ni}^2 1_{(|X_{ni}|>L_{ni})},\delta_{ni}^{(2)}=E(X_{ni}^2-1)^2 1_{(|X_{ni}|>L_{ni})};\gamma_{nij}=\gamma_{ni}^{(1)}\gamma_{nj}^{(1)}$ if $i\ne j$, and $\gamma_{ni}^{(2)}$ if $i=j$; $\delta_{nij}=\frac12(\delta_{ni}^{(1)}+\delta_{nj}^{(1)})$ if $i\ne j$, $\delta_{ni}^{(2)}$ if $i=j\in\mathcal{A}_n$, and 0 otherwise.

Theorem A. *Suppose* $EX_{ni}^2=1, 1\le i\le k_n$ *and*

$$\frac{1}{\sigma_n^2}\Big\{\sum_{i,j=1}^{k_n} a_{nij}^2\delta_{nij}+\sum_{i=1}^{k_n} b_{ni}^2\delta_{ni}^{(1)}\Big\}\longrightarrow 0, \tag{5.5}$$

$$\frac{1}{\sigma_n^4}\Big\{\sum_{i,j=1}^{k_n} a_{nij}^4\gamma_{nij}+\sum_{i=1}^{k_n}[(\sum_{j\ne i} a_{nij}^2)^2+b_{ni}^4]\gamma_{ni}^{(1)}\Big\}\longrightarrow 0, \tag{5.6}$$

where $\sigma_n^2=\mathrm{Var}\,(\mathcal{X}_n'A_n\mathcal{X}_n+b_n'\mathcal{X}_n)$. *Then*

$$\frac{\mathcal{X}_n'A_n\mathcal{X}_n+b_n'\mathcal{X}_n-E\mathcal{X}_n'A_n\mathcal{X}_n}{[\mathrm{Var}\,(\mathcal{X}_n'A_n\mathcal{X}_n+b_n'\mathcal{X}_n)]^{1/2}}\xrightarrow{\mathcal{L}} N(0,1) \tag{5.7}$$

provided

$$\frac{\lambda\text{max}((A_n^o)^2)}{\sigma_n^2} \longrightarrow 0. \tag{5.8}$$

Proof of Theorem A. As in the proof of Theorem 5.2 in Jiang (1996) (hereafter PT5.2)

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

where

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

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

$U_{ni}, V_{ni}, u_{ni}, v_{ni}$ are as in PT5.2. (5.5) and (5.6) imply that

$$E\Big(\frac{1}{\sigma_n}\sum_{i=1}^{k_n} b_{ni}v_{ni}\Big)^2 \le \frac{1}{\sigma_n^2}\sum_{i=1}^{k_n} b_{ni}^2\delta_{ni}^{(1)} \longrightarrow 0,$$

$$E\Big(\frac{1}{\sigma_n^4}\max_{1\le i\le k_n} b_{ni}^4 u_{ni}^4\Big) \le \frac{16}{\sigma_n^4}\sum_{i=1}^{k_n} b_{ni}^4\gamma_{ni}^{(1)} \longrightarrow 0.$$

Thus by PT5.2 $\sum_{i=1}^{k_n}\eta_{ni} \stackrel{L^2}{\longrightarrow} 0$, and $\max_{1\le i\le k_n}|\xi_{ni}|$ is bounded in L^2 and $\to 0$ in probability. Now

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

where

$$U_1 = \sigma_n^{-2}\sum_{i=1}^{k_n}[(a_{nii}U_{ni} + b_{ni}u_{ni})^2 - E(a_{nii}U_{ni} + b_{ni}u_{ni})^2]$$

$$U_2 = 4\sigma_n^{-2}\sum_{i=1}^{k_n}(\sum_{j<i} a_{nij}u_{nj})[a_{nii}(U_{ni}u_{ni} - E(U_{ni}u_{ni})) + b_{ni}(u_{ni}^2 - Eu_{ni}^2)],$$

$$V_1 = \sigma_n^{-2}\sum_{i=1}^{k_n} E(a_{nii}U_{ni} + b_{ni}u_{ni})^2,$$

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

and U_3 and V_3 are as in PT5.2. Note that we now have a new expression for σ_n^2:

$$\sigma_n^2 = \sum_{j=1}^{k_n} \text{Var}\,(a_{nii}X_{ni}^2 + b_{ni}X_{ni}) + 2\sum_{i\neq j} a_{nij}^2. \tag{5.9}$$

By (5.6) it is easy to show, as in PT5.2, that $U_i \xrightarrow{L^2} 0, i = 1, 2, 3$. And by (5.5) and (5.9)

$$V_1 = \sigma_n^{-2} \sum_{i=1}^{k_n} \text{Var}\,(a_{nii}X_{ni}^2 + b_{ni}X_{ni}) + o(1).$$

Finally, by Lemma 5.1 in Jiang (1996), (5.8), (5.9), and similar argument as in PT5.2 one can show

$$EV_2^2 \le (32/\sqrt{2})(\lambda_{\max}((A_n^o)^2)/\sigma_n^2)^{1/2}(1 + o(1)) \longrightarrow 0.$$

The result now follows as in PT5.2, observing the same identity for V_3:

$$V_3 = 2\sigma_n^{-2} \sum_{i\neq j} a_{nij}^2 + o_p(1).$$

Proof of Theorem B. This is very similar to that of Theorem 5.1 in Jiang (1996). Note (5.1) implies there exist $\delta, M > 0$ such that

$$\delta(a_{nii}^2 + b_{ni}^2) \le \text{Var}\,(a_{nii}X_{ni}^2 + b_{ni}X_{ni}) \le M(a_{nii}^2 + b_{ni}^2),$$

$1 \le i \le k_n, n \ge 1$.

In the following lemma, let $\hat\beta = B(\hat\mu)y$, where $\{\hat\mu\}$ is a sequence of estimates such that

$$\|U_i\|_R|\hat\mu_i - \mu_{0i}| = O_p(1),\ 1 \le i \le s. \tag{5.10}$$

Lemma 5.1. *Suppose* $\|U_i\|_R \to \infty$, $1 \le i \le s$. *Then*

$$(X'V_{\theta_0}^{-1}X)^{1/2}(\hat\beta - \beta_0) = (X'V_{\theta_0}^{-1}X)^{-1/2}X'V_{\theta_0}^{-1}(y - X\beta_0) + o_p(1).$$

Proof. Let $H(\mu) = X'V_\mu^{-1}X$. Then

$$H(\mu_0)^{1/2}(\hat\beta - \beta_0) = H(\mu_0)^{-1/2}X'V_{\mu_0}^{-1}(y - X\beta_0) + I_1 + I_2, \tag{5.11}$$

where

$$I_1 = H(\mu_0)^{1/2}H(\hat\mu)^{-1}X'(V_{\hat\mu}^{-1} - V_{\mu_0}^{-1})(y - X\beta_0)$$
$$I_2 = H(\mu_0)^{1/2}[H(\hat\mu)^{-1} - H(\mu_0)^{-1}]X'V_{\mu_0}^{-1}(y - X\beta_0).$$

By identity

$$V_{\hat\mu}^{-1} = V_{\mu_0}^{-1} + \sum_{i=1}^{s}(\mu_{0i} - \hat\mu_i)V_{\hat\mu}^{-1}Z_iZ_i'V_{\mu_0}^{-1}, \tag{5.12}$$

it follows that

$$I_1 = \sum_{i=1}^{s}(\mu_{0i} - \hat\mu_i)H(\mu_0)^{1/2}H(\hat\mu)^{-1}X'V_{\hat\mu}^{-1}Z_iZ_i'V_{\mu_0}^{-1}(y - X\beta_0). \tag{5.13}$$

By a similar identity

$$H(\hat\mu)^{-1} = H(\mu_0)^{-1} + \sum_{j=1}^{s}(\hat\mu_j - \mu_{0j})H(\hat\mu)^{-1}X'V_{\hat\mu}^{-1}Z_jZ_j'V_{\mu_0}^{-1}XH(\mu_0)^{-1}, \tag{5.14}$$

we have

$$\begin{aligned}
&H(\mu_0)^{1/2}H(\hat\mu)^{-1}X'V_{\hat\mu}^{-1}Z_iZ_i'V_{\mu_0}^{-1}(y-X\beta_0) = H(\mu_0)^{-1/2}X'V_{\mu_0}^{-1}Z_iZ_i'V_{\mu_0}^{-1}(y-X\beta_0)\\
&+\sum_{j=1}^{s}(\hat\mu_j-\mu_{0j})H(\mu_0)^{1/2}H(\hat\mu)^{-1}X'V_{\hat\mu}^{-1}Z_jZ_j'V_{\mu_0}^{-1}XH(\mu_0)^{-1}X'V_{\mu_0}^{-1}Z_iZ_i'V_{\mu_0}^{-1}(y-X\beta_0)\\
&+\sum_{j=1}^{s}(\mu_{0j} - \hat\mu_j)H(\mu_0)^{1/2}H(\hat\mu)^{-1}X'V_{\hat\mu}^{-1}Z_jZ_j'V_{\mu_0}^{-1}Z_iZ_i'V_{\mu_0}^{-1}(y - X\beta_0).
\end{aligned} \tag{5.15}$$

$$\begin{aligned}
&|H(\mu_0)^{1/2}H(\hat\mu)^{-1}X'V_{\hat\mu}^{-1}Z_jZ_j'V_{\mu_0}^{-1}Z_iZ_i'V_{\mu_0}^{-1}(y - X\beta_0)|\\
&\le \|H(\mu_0)^{1/2}H(\hat\mu)^{-1}X'V_{\hat\mu}^{-1/2}\|\ |V_{\hat\mu}^{-1/2}Z_jZ_j'V_{\mu_0}^{-1}Z_iZ_i'V_{\mu_0}^{-1}(y - X\beta_0)|\\
&= \lambda_0^{1/2}\mu_{0i}^{-1}\|U_j\|_R O_p(1),
\end{aligned} \tag{5.16}$$

since $w.p. \to 1$, $\|H(\mu_0)^{1/2}H(\hat\mu)^{-1}X'V_{\hat\mu}^{-1/2}\|^2 \le 2$,

$$\begin{aligned}
|V_{\hat\mu}^{-1/2}Z_jZ_j'V_{\mu_0}^{-1}Z_iZ_i'V_{\mu_0}^{-1}(y-X\beta_0)|^2 &\le 2|V_{\mu_0}^{-1/2}Z_jZ_j'V_{\mu_0}^{-1}Z_iZ_i'V_{\mu_0}^{-1}(y-X\beta_0)|^2, \text{and}\\
E|V_{\mu_0}^{-1/2}Z_jZ_j'V_{\mu_0}^{-1}Z_iZ_i'V_{\mu_0}^{-1}(y-X\beta_0)|^2 &\le \lambda_0\lambda_{\max}^2(Z_i'V_{\mu_0}^{-1}Z_i)\mathrm{tr}((Z_j'V_{\mu_0}^{-1}Z_j)^2)\\
&\le \lambda_0\mu_{0i}^{-2}\|U_j\|_R^2.
\end{aligned} \tag{5.17}$$

Therefore the 3rd term in (5.15) is $O_p(1)$. Similarly one can show the 2nd term in (5.15) is $O_p(1)$. Thus $I_1 = o_p(1)$ by (5.13), (5.15) and the fact that as in (5.17)

$$E|H(\mu_0)^{-1/2}X'V_{\mu_0}^{-1}Z_iZ_i'V_{\mu_0}^{-1}(y - X\beta_0)|^2 \le \lambda_0\mu_{0i}^{-2}p.$$

Again by (5.14) and (5.12) and a similar argument

$$I_2 = \sum_{i=1}^{s}(\hat\mu_i - \mu_{0i})H(\mu_0)^{1/2}H(\hat\mu)^{-1}X'V_{\hat\mu}^{-1}Z_iZ_i'V_{\mu_0}^{-1}XH(\mu_0)^{-1}X'V_{\mu_0}^{-1}(y - X\beta_0), \tag{5.18}$$

$$
\begin{aligned}
&H(\mu_0)^{1/2}H(\hat{\mu})^{-1}X'V_{\hat{\mu}}^{-1}Z_iZ_i'V_{\mu_0}^{-1}XH(\mu_0)^{-1}X'V_{\mu_0}^{-1}(y-X\beta_0)\\
&= H(\mu_0)^{1/2}H(\hat{\mu})^{-1}X'V_{\mu_0}^{-1}Z_iZ_i' \cdots\\
&\quad + \sum_{j=1}^{s}(\mu_{0j}-\hat{\mu}_j)H(\mu_0)^{1/2}H(\hat{\mu})^{-1}X'V_{\hat{\mu}}^{-1}Z_jZ_j'V_{\mu_0}^{-1}Z_iZ_i' \cdots\\
&= H(\mu_0)^{-1/2}X'V_{\mu_0}^{-1}Z_iZ_i' \cdots\\
&\quad + \sum_{j=1}^{s}(\hat{\mu}_j-\mu_{0j})H(\mu_0)^{1/2}H(\hat{\mu})^{-1}X'V_{\hat{\mu}}^{-1}Z_jZ_j'V_{\mu_0}^{-1}XH(\mu_0)^{-1}X'V_{\mu_0}^{-1}Z_iZ_i' \cdots\\
&\quad + \sum_{j=1}^{s}(\mu_{0j}-\hat{\mu}_j)H(\mu_0)^{1/2}H(\hat{\mu})^{-1}X'V_{\hat{\mu}}^{-1}Z_jZ_j'V_{\mu_0}^{-1}Z_iZ_i'\cdots, \qquad (5.19)
\end{aligned}
$$

$$
\begin{aligned}
&|H(\mu_0)^{1/2}H(\hat{\mu})^{-1}X'V_{\hat{\mu}}^{-1}Z_jZ_j'V_{\mu_0}^{-1}Z_iZ_i'\cdots|\\
&\le \|H(\mu_0)^{1/2}H(\hat{\mu})^{-1}X'V_{\hat{\mu}}^{-1/2}\||V_{\hat{\mu}}^{-1/2}Z_jZ_j'V_{\mu_0}^{-1}Z_iZ_i'V_{\mu_0}^{-1}XH(\mu_0)^{-1}X'V_{\mu_0}^{-1}(y-X\beta_0)|\\
&\le \sqrt{2}\cdot\sqrt{2}\lambda_0^{1/2}\lambda_{\max}(Z_i'V_{\mu_0}^{-1}Z_i)\|U_j\|_R O_p(1)\\
&\le 2\lambda_0^{1/2}\mu_{0i}^{-1}\|U_j\|_R O_p(1), \qquad (5.20)
\end{aligned}
$$

$$
\begin{aligned}
&|H(\mu_0)^{1/2}H(\hat{\mu})^{-1}X'V_{\hat{\mu}}^{-1}Z_jZ_j'V_{\mu_0}^{-1}XH(\mu_0)^{-1}X'V_{\mu_0}^{-1}Z_iZ_i'\cdots|\\
&\le \sqrt{2}|V_{\hat{\mu}}^{-1/2}Z_jZ_j'V_{\mu_0}^{-1}XH(\mu_0)^{-1}X'V_{\mu_0}^{-1}Z_iZ_i'V_{\mu_0}^{-1}XH(\mu_0)^{-1}X'V_{\mu_0}^{-1}(y-X\beta_0)|\\
&\le 2\lambda_0^{1/2}\mu_{0i}^{-1}\|U_j\|_R O_p(1), \qquad (5.21)
\end{aligned}
$$

and finally

$$
\begin{aligned}
&E|H(\mu_0)^{-1/2}X'V_{\mu_0}^{-1}Z_iZ_i'V_{\mu_0}^{-1}X'H(\mu_0)^{-1}X'V_{\mu_0}^{-1}(y-X\beta_0)|^2\\
&\le \lambda_0\lambda_{\max}(V_{\mu_0}^{-1/2}XH(\mu_0)^{-1}X'V_{\mu_0}^{-1/2})\\
&\quad \mathrm{tr}(V_{\mu_0}^{-1/2}XH(\mu_0)^{-1}X'V_{\mu_0}^{-1/2}(V_{\mu_0}^{-1/2}Z_iZ_i'V_{\mu_0}^{-1/2})^2V_{\mu_0}^{-1/2}XH(\mu_0)^{-1}X'V_{\mu_0}^{-1/2})\\
&\le \lambda_0\mu_{0i}^{-2}p. \qquad (5.22)
\end{aligned}
$$

So $I_2 = o_p(1)$ by combing (5.18) - (5.22).

Proof of Theorem 3.1. By Theorem 4.3 of Jiang (1996), there exist REML estimates $\hat{\lambda}_N$, $\hat{\mu}_{Ni}$, $1 \le i \le s$ such that $\|V_i\|_R(\hat{\mu}_{Ni}-\mu_{0i}) = O_p(1)$, $1 \le i \le s$. We now show that the same is true with V_i replaced by U_i. By a well-known identity (e.g., Searle, Casella and McCulloch (1992), page 451)

$$A(A'V_{\mu_0}A)^{-1}A' = V_{\mu_0}^{-1} - V_{\mu_0}^{-1}X(X'V_{\mu_0}^{-1}X)^{-1}X'V_{\mu_0}^{-1},$$

we have $Z_i'AV(A,\mu_0)^{-1}A'Z_i \le Z_i'V_{\mu_0}^{-1}Z_i$. Therefore $\|U_i\|_R^2 = \mathrm{tr}((Z_i'V_{\mu_0}^{-1}Z_i)^2) \ge \mathrm{tr}((Z_i'AV(A,\mu_0)^{-1}A'Z_i)^2) = \|V_i\|_R^2$. Note that $A, B \ge 0$, $A \le B \Rightarrow \mathrm{tr}(A^2) \le$

$\text{tr}(B^2)$ but does not $\Rightarrow A^2 \le B^2$ (e.g., Chan and Kwong (1985)). On the other hand, let $A_i = Z_i' V_{\mu_0}^{-1} X (X' V_{\mu_0}^{-1} X)^{-1} X' V_{\mu_0}^{-1} Z_i$, $B_i = (X' V_{\mu_0}^{-1} X)^{-1/2} X' V_{\mu_0}^{-1} Z_i Z_i' V_{\mu_0}^{-1} X (X' V_{\mu_0}^{-1} X)^{-1/2}$. Since $Z_i Z_i' \le \mu_{0i}^{-1} V_{\mu_0}$, we have $B_i \le \mu_{0i}^{-1} I_p$. Thus by the above fact $\text{tr}(B_i^2) \le \mu_{0i}^{-2} p$, hence $\|U_i\|_R = \|Z_i' V_{\mu_0}^{-1} Z_i\|_R = \|Z_i' A V(A,\mu_0)^{-1} A' Z_i + A_i\| \le \|Z_i' A V(A,\mu_0)^{-1} A' Z_i\|_R + \|A_i\|_R = \|V_i\|_R + \|B_i\|_R \le \|V_i\|_R + \mu_{0i}^{-1}\sqrt{p}$. Since p is fixed, and $\min_{1\le i\le s} \|V_i\|_R \to \infty$, we have $\|U_i\|_R \sim \|V_i\|_R$, $1 \le i \le s$, and hence $\min_{1\le i\le s} \|U_i\|_R \to \infty$.

For any $u = (u_i)_{0\le i\le s} \in R^{s+1}, v \in R^p$, let $\xi_{N,u} = J_N^{-1/2} u, d_{N,v} = C_N v$. By Lemma 5.1, the proof (outline) of Lemma 7.2 in Jiang (1996) (bottom of page 273, i.e. $-A_N(\theta_0) = I_N(\theta_0) p(N)(\hat\theta_N - \theta_0) + o_p(1)$), and the proof of Theorem 4.3 (ii) in Jiang (1996) (top of page 276, note that a and b_N there correspond to u and $\xi_{N,u}$ here), we have

$$\begin{aligned} &u' P_N(\hat\theta_N - \theta_0) + v' Q_N(\hat\beta_N - \beta_0) \\ &= \mathcal{W}_N' B_{N,u} \mathcal{W}_N + d_{N,v}' \mathcal{W}_N - E\mathcal{W}_N B_{N,u} \mathcal{W}_N + u' o_p(1) + v' o_p(1), \end{aligned} \tag{5.23}$$

where $B_{N,u} = B_N(\xi_{N,u}), B_N(\xi) = \sum_{i=0}^s \xi_i V_i(\mu_0)/q_i(N)$ with $q_i(N) = \lambda_0^{1_{(i=0)}} p_i(N)$, $0 \le i \le s$. Write $I_N = I_N(\theta_0), J_N = J_N(\theta_0)$. It is easy to see that

$$\text{tr}(B_{N,u}^2) = u' J_N^{-1/2} I_N J_N^{-1/2} u, \tag{5.24}$$

$$|d_{N,v}|^2 = |v|^2. \tag{5.25}$$

Let $B_{N,u} = (b_{kl}), d_{N,v} = (d_l)$, then

$$\begin{aligned} &\text{Var}\,(\mathcal{W}_N' B_{N,u} \mathcal{W}_N + d_{n,v}' \mathcal{W}_N) \\ &= 2\sum_{k\ne l} b_{kl}^2 + \sum_{l=1}^{N+m} (b_{ll} d_l) \text{Var}\,((W_{Nl}^2 W_{Nl})') \binom{b_{ll}}{d_l} \\ &\ge (2 \wedge \delta)(\text{tr}(B_{N,u}^2) + |d_{N,v}|^2), \end{aligned} \tag{5.26}$$

where $\delta = \inf_{N,l} \text{Var}\,((W_{Nl}^2 W_{Nl})') > 0$.

On the other hand,

$$\begin{aligned} &\text{Var}\,(\mathcal{W}_N' B_{n,u} \mathcal{W}_N + d_{N,v}' \mathcal{W}_N) \\ &= 2\text{tr}(B_{N,u}^2) + \sum_{l=1}^{N+m} b_{ll}^2 (EW_{Nl}^4 - 3) + 2\sum_{l=1}^{N+m} b_{ll} d_l EW_{Nl}^3 + |d_{N,v}|^2 \\ &= |u|^2 + 2u' R_N v + |v|^2 = (u'v') S_N \binom{u}{v}. \end{aligned} \tag{5.27}$$

Combing (5.26), (5.27) and (5.24), (5.25),

$$(u'v') S_N \binom{u}{v} \ge (2 \wedge \delta)(u' J_N^{-1/2} I_N J_N^{-1/2} u + |v|^2) > 0 \tag{5.28}$$

provided $|u|^2 + |v|^2 \neq 0$. Thus S_N is positive definite.

For any $a \in R^{s+1}, b \in R^p$ such that $|a|^2 + |b|^2 \neq 0$, let $(a'_N b'_N) = (a'b')S_N^{-1/2}$, $B_N = B_{N,a_N}, d_N = d_{N,b_N}$. Then by (5.23)

$$\begin{aligned}&(a'b')S_N^{-1/2}\begin{pmatrix} P_N & 0 \\ 0 & Q_N \end{pmatrix}\begin{pmatrix} \hat{\theta}_N - \theta_0 \\ \hat{\beta}_N - \beta_0 \end{pmatrix}\\ &= a'_N P_N(\hat{\theta}_N - \theta_0) + b'_N Q_N(\hat{\beta}_N - \beta_0)\\ &= \mathcal{W}'_N B_N \mathcal{W}_N + d'_N \mathcal{W}_N - E\mathcal{W}'_N B_N \mathcal{W}_N + o_p(1).\end{aligned}$$

Since $|a_N|^2 + |b_N|^2 \leq \lambda_{\min}^{-1}(S_N)(|a|^2 + |b|^2)$, by (5.28) it is easy to show that

$$\lambda_{\min}(S_N) \geq (2 \wedge \delta)\Big[((2 \vee M)^{-1}\frac{\lambda_{\min}(I_N)}{\lambda_{\max}(I_N)}) \wedge 1\Big], \tag{5.29}$$

where $M = \sup_N \max_{1\leq l\leq N+m} \mathrm{Var}\,(W_{Nl}^2) < \infty$. Note for any $\xi \in R^{s+1}$

$$\xi' J_N \xi = \mathrm{Var}\,(\mathcal{W}'_N B_N \mathcal{W}_N) \leq (2 \vee M)\mathrm{tr}(B_N^2(\xi)) = (2 \vee M)\xi' I_N \xi. \tag{5.30}$$

Again, by (5.30)

$$\mathrm{tr}(B_N^2) = \mathrm{tr}(B_N^2(\xi_{N,a_N})) \geq \frac{1}{2 \vee M}\xi'_{N,a_N} J_N \xi_{N,a_N} = \frac{|a_N|^2}{2 \vee M},$$

$$\begin{aligned}\lambda_{\max}(B_N^2) &= \|B_N(\xi_{N,a_N})\|^2\\ &\leq (\lambda_0^{-2}p_0^{-2}(N) + \sum_{i=1}^{s}\mu_{0i}^{-2}p_i^{-2}(N))|\xi_{N,a_N}|^2 \leq \delta_N |a_N|^2,\end{aligned}$$

where $\delta_N = \lambda_{\min}^{-1}(J_N)(\lambda_0^{-2}p_0^{-2}(N) + \sum_{i=1}^{s}\mu_{0i}^{-2}p_i^{-2}(N))$.

Also, by (5.25), $|d_N|^2 = |b_N|^2$, and by Cauchy-Schwarz $\max_{1\leq l\leq N+m} d_{Nl}^2 \leq \rho_N|b_N|^2$, where $\rho_N = \max_{1\leq l\leq N+m}|C_{N,l}|^2$. Thus

$$\frac{\lambda_{\max}(B_N^2) \vee \max_{1\leq l\leq N+m} d_{Nl}^2}{\mathrm{tr}(B_N^2) + |d_N|^2} \leq (2 \vee M)(\delta_N \vee \rho_N) \longrightarrow 0$$

(note $|a_N|^2 + |b_N|^2 \neq 0$).

The result now easily follows from Theorem B, using the fact that $\mathrm{Var}\,(\mathcal{W}'_N B_N \mathcal{W}_N + d'_N \mathcal{W}_N) = |a|^2 + |b|^2$.

The proof of Theorem 3.2 requires the following lemmas whose proofs are mostly straightforward (e.g., Billingsley (1986), §20).

Lemma 5.2. *Let $\xi_{Nk} = \eta_k + \zeta_{Nk}, k = 1, \ldots, m_N$, where $\eta_1, \eta_2, \ldots,$ are i.i.d. $\sim F$ and $m_N \to \infty$. If there is $p > 1$ such that $E|\eta_1|^p < \infty$ and $\frac{1}{m_N}\sum_{k=1}^{m_N}|\zeta_{Nk}|^p \xrightarrow{P} 0$, then*

(i) $\frac{1}{m_N}\sum_{k=1}^{m_N} 1_{(\xi_{Nk}\le x)} \xrightarrow{P} F(x), \forall x \in C_F$;

(ii) $\frac{1}{m_N}\sum_{k=1}^{m_N} \xi_{Nk}^q \xrightarrow{P} E\eta_1^q, \forall$ *positive integer* $q \le p$.

Corollary 5.1. *If in Lemma* 5.2 F *is continuous, then the convergence in* (i) *is uniform in* $x \in R$.

Lemma 5.3. *Let* $\xi = (\xi_1, \ldots, \xi_n)'$, *where* $\xi_1, \ldots, \xi_n$ *are i.i.d. with* $E\xi_1 = 0$ *and* $B = (B_1', \ldots, B_m')'$ *is* $m \times n$.

(I) *If* $E\xi_1^2 < \infty$, *then* $E|B\xi|^2 = \sum_{k=1}^m E(B_k\xi)^2 = E\xi_1^2 \mathrm{tr}(BB')$;

(II) *If* $E\xi_1^4 < \infty$, *then* $\sum_{k=1}^m E(B_k\xi)^4 \le 3E\xi_1^4 \lambda_{\max}(BB')\,\mathrm{tr}(BB')$.

Proof of Theorem 3.2. First we note that (3.8) is equivalent to

$$\frac{1}{m_i}\mathrm{tr}((Z_i'V(\mu_0)Z_j)(Z_i'V(\mu_0)Z_j)') \longrightarrow 0,\ \forall i \ne j.$$

Theorem 4.3 of Jiang (1996) guarantees the existence of such REML estimates. For simplicity, write $\mu = \mu_0, \hat\mu = \hat\mu_N, H_i = H_i(\mu_0)$. We have

$$\begin{aligned}\hat\alpha_i &= \alpha_i - (I_{m_i} + \mu_i H_i)^{-1}\alpha_i + (\hat\mu_i - \mu_i)Z_i'V(\mu)Z_i\alpha_i + \hat\mu_i(Z_i'V(\mu)\epsilon + \sum_{j\ne i} Z_i'V(\mu)Z_j\alpha_j)\\ &\quad + \hat\mu_i \sum_{j=1}^s (\mu_j - \hat\mu_j) Z_i'AV(A,\hat\mu)^{-1}A'Z_jZ_j'AV(A,\mu)^{-1}z\\ &= \alpha_i + \sum_{l=1}^4 \xi^{(l)}. \end{aligned} \tag{5.31}$$

By a similar argument as in the proof of Lemma 5.1 it is easy to show that $|\xi^{(4)}| = O_p(1)$, so that

$$\sum_{k=1}^{m_i}(\xi_k^{(4)})^4 \le (\sum_{k=1}^{m_i}(\xi_k^{(4)})^2)^2 = |\xi^{(4)}|^4 = O_p(1),$$

which implies

$$\frac{1}{m_i}\sum_{k=1}^{m_i}(\xi_k^{(4)})^4 \xrightarrow{P} 0. \tag{5.32}$$

Denoting the k'th row of a matrix B by B_k, we have by Lemma 5.3

$$\begin{aligned}\sum_{k=1}^{m_i} E((Z_i'V(\mu))_k\epsilon)^4 &\le 3E\epsilon_1^4\lambda_{\max}(Z_i'(V(\mu))^2Z_i)\mathrm{tr}(Z_i'(V(\mu))^2Z_i)\\ &\le 3\mu_i^{-1}E\epsilon_1^4\mathrm{tr}(Z_i'(V(\mu))^2Z_i). \end{aligned} \tag{5.33}$$

Using a well-known identity (e.g., Rao (1965), Page 33) for the first step in the following, we have

$$\begin{aligned} Z_i'V(\mu)Z_i &= \mu_i^{-1}(I_{m_i} - (I_{m_i} + \mu_i H_i)^{-1}) \\ &= Z_i'(V(\mu))^2 Z_i + \mu_i(Z_i'V(\mu)Z_i)^2 + \sum_{j\neq i}\mu_j Z_i'V(\mu)Z_jZ_j'V(\mu)Z_i. \quad (5.34) \end{aligned}$$

It follows from (5.33), (5.34), (3.8) and (3.9) that

$$\frac{1}{m_i}\sum_{k=1}^{m_i}((Z_i'V(\mu))_k\epsilon)^4 \xrightarrow{L^1} 0. \quad (5.35)$$

Again, by Lemma 5.3

$$\begin{aligned} \sum_{k=1}^{m_i} E((Z_i'V(\mu)Z_j)_k\alpha_j)^4 &\leq 3E\alpha_{j1}^4\lambda_{\max}((Z_i'V(\mu)Z_j)(Z_i'V(\mu)Z_j)') \\ &\quad \cdot \mathrm{tr}((Z_i'V(\mu)Z_j)(Z_i'V(\mu)Z_j)') \\ &\leq 3(\mu_i\mu_j)^{-1}E\alpha_{j1}^4\mathrm{tr}((Z_i'V(\mu)Z_j)(Z_i'V(\mu)Z_j)'). \end{aligned}$$

Thus by (3.8) and (5.35)

$$\begin{aligned} \frac{1}{m_i}\sum_{k=1}^{m_i}(\xi_k^{(3)})^4 &\leq s^3\hat{\mu}_i^4[\frac{1}{m_i}\sum_{k=1}^{m_i}((Z_i'V(\mu))_k\epsilon)^4 \\ &\quad + \sum_{j\neq i}\frac{1}{m_i}\sum_{k=1}^{m_i}((Z_i'V(\mu)Z_j)_k\alpha_j)^4] \xrightarrow{P} 0. \quad (5.36) \end{aligned}$$

Similarly,

$$\begin{aligned} \sum_{k=1}^{m_i} E((Z_i'V(\mu)Z_i)_k\alpha_i)^4 &\leq 3E\alpha_{i1}^4\lambda_{\max}^2(Z_i'V(\mu)Z_i)\mathrm{tr}((Z_i'V(\mu)Z_i)^2) \\ &\leq 3\mu_i^{-4}E\alpha_{i1}^4 m_i, \end{aligned}$$

which implies

$$\frac{1}{m_i}\sum_{k=1}^{m_i}(\xi_k^{(2)})^4 = 3(\hat{\mu}_i - \mu_i)^4\mu_i^{-4}E\alpha_{i1}^4 O_p(1) \xrightarrow{P} 0. \quad (5.37)$$

Finally,

$$\begin{aligned} \sum_{k=1}^{m_i} E(((I_{m_i} + \mu_i H_i)^{-1})_k\alpha_i)^4 &\leq 3E\alpha_{i1}^4\lambda_{\max}^2((I_{m_i} + \mu_i H_i)^{-1})\mathrm{tr}((I_{m_i} + \mu_i H_i)^{-2}) \\ &\leq 3E\alpha_{i1}^4\mathrm{tr}((I_{m_i} + \mu_i H_i)^{-1}), \end{aligned}$$

so that by (3.9)

$$\frac{1}{m_i}\sum_{k=1}^{m_i}(\xi_k^{(1)})^4 \xrightarrow{L^1} 0. \tag{5.38}$$

The results now follow from (5.31), (5.32), (5.36)—(5.38), Lemma 5.2, and Corollary 5.1.

Proof of Lemma 3.1. Similar to (5.31) we have

$$\begin{aligned}\hat{\epsilon} &= \epsilon - P_X\epsilon - P_{X^\perp}Z(Z'P_{X^\perp}Z + M^{-1})^{-1}Z'P_{X^\perp}\epsilon \\ &\quad + \sum_{j=1}^{s} V(\mu)Z_j\alpha_j + \sum_{j=1}^{s}(\mu_j - \hat{\mu_j})AV(A,\hat{\mu})^{-1}A'Z_jZ_j'AV(A,\mu)^{-1}z \\ &= \epsilon + \sum_{l=1}^{4}\eta^{(l)},\end{aligned}$$

where $P_X = X(X'X)^{-1}X', P_{X^\perp} = I_N - P_X, Z = (Z_1,\ldots,Z_s), M = \text{diag}(\mu_i I_{m_i})$. As in the proof of Theorem 3.2,

$$\begin{aligned}\frac{1}{N}\sum_{k=1}^{N}(\eta_k^{(4)})^4 &\le \frac{1}{N}|\eta^{(4)}|^4 = \frac{1}{N}O_p(1) \xrightarrow{P} 0, \\ \frac{1}{N}\sum_{k=1}^{N}(\eta^{(3)})^4 &\le s^3(\sum_{j=1}^{s}\frac{1}{N}\sum_{k=1}^{N}((V(\mu)Z_j)_k\alpha_j)^4) \xrightarrow{L^1} 0.\end{aligned}$$

And by Lemma 5.3

$$\begin{aligned}\sum_{k=1}^{N}E(\eta_k^{(2)})^4 &\le 3E\epsilon_1^4\lambda_{\max}^2((Z'P_{X^\perp}Z)^{1/2}(Z'P_{X^\perp}Z+M^{-1})^{-1}(Z'P_{X^\perp}Z)^{1/2}) \\ &\qquad \text{tr}(((Z'P_{X^\perp}Z)^{1/2}(Z'P_{X^\perp}Z+M^{-1})^{-1}(Z'P_{X^\perp}Z)^{1/2})^2) \\ &\le 3E\epsilon_1^4(m_1+\cdots+m_s),\end{aligned}$$

$\sum_{k=1}^{N}E(\eta_k^{(1)})^4 \le 3E\epsilon_1^4\lambda_{\max}(P_X)\text{tr}(P_X) \le 3E\epsilon_1^4 p$, which imply $\frac{1}{N}\sum_{k=1}^{N}(\eta_k^{(l)})^4 \xrightarrow{L^1} 0, l = 1,2$ by (3.13). The conclusions therefore hold, by Lemma 5.2 and Corollary 5.1.

6. Concluding Remark

From our proof in §5 we see that under certain conditions the EBLUPs may be expressed as the corresponding random effects plus something that is asymptotically negligible, i.e., $\hat{\alpha}_i = \alpha_i + o_p(1),\ i = 1,\ldots,s$. This is important because although the EBLUPs are not independent, they may be somehow regarded, asymptotically, as i.i.d.. Such an observation may support the idea of using the EBLUPs in mixed model diagnostics. A topic of further research seems to be the asymptotic normality of the e.d. of the EBLUPs. Some discussion of the topic under normality can be found in Lange and Ryan (1989).

Acknowledgements

The auther would like to thank Professor P. J. Bickel for helpful discussions.

References

Billingsley, P. (1986). *Probability and Measure.* John Wiley.

Calvin, J. A. and Sedransk, J. (1991). Bayesian and frequentist predictive inference for the patterns of care studies. *J. Amer. Statist. Assoc.* **86**, 36-48.

Chan, N. N. and Kwong, M. K. (1985). Hermitian matrix inequalities and a conjecture. *Amer. Math. Monthly* **92**, 533-541.

Fellner, W. H. (1986). Robust estimation of variance components. *Technometrics* **28**, 51-60.

Ghosh, M. and Rao, J. N. K. (1994). Small Area estimation: An appraisal. *Statist. Sci.* **9**, 55-93.

Harris, D. L., Lofgren, D. L., Stewart, T. S. and Schinckel, A. P. (1989). Adapting best linear unbiased prediction (BLUP) for timely genetic evaluation: I. Progeny traits in a single contemporary group for each sex. *J. Anim. Sci.* **67**, 3209-3222.

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

Harville, D. A. (1991). Comment on Robinson: The estimation of random effects. *Statist. Sci.* **6**, 35-39.

Henderson, C. R. (1963). Selection index and expected genetic advance. In *Statist. Genet. and Plant Breed* (Edited by W. D. Hanson and H. F. Robinson), 141-163. NAS-NRC Publ. 982.

Henderson, C. R. (1973). Sire evaluation and genetic trends. *Proc. Anim. Breed. Genet. Symp.* in honor of Dr. Jay L. Lush., 10-41, Am. Soc. Anim. Sci., Champaign, IL.

Jiang, J. (1996). REML estimation: Asymptotic behavior and related topics. *Ann. Statist.* **24**, 255-286.

Jiang, J. (1997). Wald consistency and the method of sieves in REML estimation. *Ann. Statist.* **25**, 1781-1803.

Lai, T. L. and Wei, C. Z. (1982). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Ann. Statist.* **10**, 154-166.

Lange, N. and Ryan, L. (1989). Assessing normality in random effects models. *Ann. Statist.* **17**, 624-642.

Liski, E. P. and Nummi, T. (1995). Prediction of tree stems to improve efficiency in automatized harvesting of forests. *Scand. J. Statist.* **22**, 255-269.

Lofgren, D. L., Harris, D. L., Stewart, T. S. and Schinckel, A. P. (1989). Adapting best linear unbiased prediction (BLUP) for timely genetic evaluation: II. Progeny traits in multiple contemporary groups within a herd. *J. Anim. Sci.* **67**, 3223-3242.

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

Rao, C. R. (1965). *Linear Statistical Inference and Its Applications.* Wiley, New York.

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, G. K. (1991). That BLUP is a good thing: The estimation of random effects. *Statist. Sci.* **6**, 15-51.

Searle, S. R., Casella, G. and McCulloch, C. E. (1992). *Variance Components.* John Wiley.

Speed, T. (1991). Comment on Robinson: The estimation of random effects. *Statist. Sci.* **6**, 42-44.

Thompson, R. (1980). Maximum likelihood estimation of variance components. *Math. Operforsch, Statist. Ser. Statist.* **11**, 545-561.

Verbeke, G. and Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *J. Amer. Statist. Assoc.* **91**, 217-221.

Department of Statistics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, Ohio 44106-7054, U.S.A.

E-mail: jiang@eureka.cwru.edu

(Received October 1996; accepted August 1997)