The Annals of Statistics
2001, Vol. 29, No. 4, 1137–1164

GOODNESS-OF-FIT TESTS FOR MIXED MODEL DIAGNOSTICS[1]

BY JIMING JIANG

Case Western Reserve University

A simple goodness of fit test is proposed for checking distributional assumptions involved in a mixed linear model. An estimated critical value of the test statistic is derived, and is shown to be asymptotically correct under mild conditions. As a special case, the test may be applied to linear regression models to formally check distribution of the errors. Finite sample performance of the proposed test is examined and compared with that of a previously proposed test by simulations.

1. Introduction. Model-based statistical inference relies on careful checking of the assumed models, including the distributional assumptions involved, in order to find suitable models that fit the data well, and for the methods used to be valid. The aim of this paper is to develop a method of goodness-of-fit tests for checking distributional assumptions involved in mixed linear models. A mixed linear model is defined as

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

where $y = (y_i)_{1\le i\le N}$ is a vector of observations; X is a $N \times p$ matrix of known covariates such that $\text{rank}(X) = p$; $\beta = (\beta_j)_{1\le j\le p}$ is a vector of unknown regression coefficients (the fixed effects); $Z_1, \ldots, Z_s$ are known design matrices; $\alpha_1, \ldots, \alpha_s$ are vectors of unobservable random variables (the random effects); and ε is a vector of errors. Suppose that for $1 \le r \le s$, $\alpha_r = (\alpha_{rk})_{1\le k\le m_r}$, where the α_{rk}'s are i.i.d. with mean 0, variance σ_r^2 which is unknown, and continuous distribution $F_r = F_r(\cdot\,|\sigma_r)$; and $\varepsilon = (\varepsilon_i)_{1\le i\le N}$, where the ε_i's are i.i.d. with mean 0, variance τ^2 which is unknown, and continuous distribution $G = G(\cdot\,|\tau)$; and $\alpha_1, \ldots, \alpha_s, \varepsilon$ are independent. We are interested in testing the following hypothesis:

$$H_0 : F_r(\cdot|\sigma_r) = F_{0r}(\cdot|\sigma_r),\ 1 \le r \le s \quad \text{and} \quad G(\cdot|\tau) = G_0(\cdot|\tau); \tag{1.2}$$

that is, the distributions of the random effects and errors, up to a set of unknown variance components $\sigma_1^2, \ldots, \sigma_s^2, \tau^2$, are as assumed.

Such distributional assumptions are vital in many applications of mixed linear models, and this is true even in large sample situations. For example, in many cases the prediction of a mixed effect is of main interest. Consider, for example, a nested error regression model, a special case of mixed linear models, which is useful in small area estimation [e.g., Battese, Harter and

Received June 1999; revised February 2001.

[1]Supported in part by NSA Grant MDA904-98-1-0038 and NSF Grant SES-99-78101.

AMS 2000 *subject classifications*. Primary 62J20.

Key words and phrases. asymptotic distribution, cell frequencies, χ^2 tests, eigenvalues, mixed models, REML.

Fuller (1988), Prasad and Rao (1990), Ghosh and Rao (1994), Arora, Lahiri and Mukherjee (1997)]: $y_{ij} = x'_{ij}\beta + \alpha_i + \varepsilon_{ij}$, $i = 1, \ldots, m$, $j = 1, \ldots, n_i$, where x_{ij} is a known vector of covariates, β is an unknown vector of regression coefficients, α_i is a random effect associated with the ith small area, and ε_{ij} is an error. A mixed effect may be in the form $\eta = x'\beta + \alpha_i$, where x is known. If the sample size is large (i.e., m is large), one may consistently estimate β and even obtain an asymptotic confidence interval for it, and this does not rely on distributional assumptions such as normality. However, large sample results may not help, for example, in obtaining a prediction interval for η, because the effective sample size for estimating α_i is n_i, the sample size for the ith small area, which is often very small. Therefore, unless one knows the form of the distribution of α_i (e.g., *normal*), an accurate prediction interval for η cannot be obtained no matter how large m is (provided that n_i is small). To see another example, consider the estimation of the mean squared error (MSE) of the empirical best linear unbiased predictor (EBLUP). Prasad and Rao (1990) gives approximation formulas for MSE of EBLUP in the context of small area estimation, which are correct to the order $o(m^{-1})$. Although their results are asymptotic, assuming that m is large, a normality distributional assumption remains critical for the validity of their approximations.

In their comprehensive review on small area estimation, Ghosh and Rao (1994) wrote the following: "... However, the literature on diagnostics for mixed linear models involving random effects is not extensive, unlike standard regression diagnostics. Only recently have some useful diagnostic tools been proposed...." Most of the methods proposed for diagnosing mixed linear models, including checking the distributional assumptions such as normality, involve the use of EBLUP or empirical Bayes estimators [e.g., Dempster and Ryan (1985), Lange and Ryan (1989), Calvin and Sedransk (1991)]. However, formal tests are not given [on a rigorous basis, e.g., Verbeke and Lesaffre (1996), last paragraph of Section 4]. Note that the EBLUP method may be regarded as an extension of the residual method in linear regression with i.i.d. errors. While the latter works well in regression diagnostics, the former has been less successful in mixed model diagnostics [e.g., Verbeke and Lesaffre (1996)]. Jiang (1998) provides a theoretical interpretation of why there is such a difference. He noticed that the very reason that the residual method works in checking the error distribution is that the empirical distribution of the residuals converges in probability to the true distribution of the errors. However, similar results hold for the EBLUP only under some strong conditions [see Theorem 3.2 and Lemma 3.1 of Jiang (1998)]. For the most part, these conditions require that the number of observations corresponding to each random effect goes to ∞ (e.g., in the earlier nested error regression model, $n_i \to \infty$). While this may be true in some cases, it certainly does not hold in many typical situations where mixed linear models are used. Recently, Jiang, Lahiri and Wu (1998) developed asymptotic theory of Pearson's χ^2-test with estimated cell frequencies, and applied the method to mixed model diagnostics. They consider a special case of the mixed model (1.1) in which $s = 1$, and give a formal test for checking distributional assumptions about the random

effects and errors. However, their procedure requires splitting the data into two parts, one used for estimation and the other for testing, and hence reduces the power of the test.

The present paper is devoted to developing a method that applies to a general mixed linear model, and does not involve a less pleasant step such as splitting the data. The basic idea may be traced back to Pearson's χ^2-test, one of the celebrated results in classical statistics. Let $X_1, \ldots, X_N$ be i.i.d. observations from an unknown distribution F. Let $E_1, \ldots, E_M$ be a partition of R, the real line. Pearson's χ^2-test for the hypothesis: H_0: $F = F_0$ is based on Pearson's χ^2-statistic:

$$\hat{\chi}_P^2 = \sum_{k=1}^{M} \frac{(N_k - \mathrm{E}N_k)^2}{\mathrm{E}N_k}, \tag{1.3}$$

where $N_k = \sum_{i=1}^{N} 1_{(X_i \in E_k)} = \#\{1 \le i \le N : X_i \in E_k\}$ is the observed count for cell E_k, and $\mathrm{E}N_k = NP(X_1 \in E_k)$ is the expected count under the null hypothesis. Here it is assumed that the cell probabilities $P(X_1 \in E_k)$, $1 \le k \le M$ are completely specified under the null. In such a case, the asymptotic null distribution of $\hat{\chi}_P^2$ is χ^2 with $M-1$ degrees of freedom. In many cases, however, the cell probabilities are not completely specified under the null but involve unknown parameters. In such a case, it is customary to replace the unknown parameters by their estimators. The result is Pearson's χ^2-test with estimated cell probabilities. However, this is the case where the asymptotic theory gets more complicated. In a simple problem of assessing the goodness of fit to a *Poisson* or *multinomial* distribution, it is known that when the unknown parameters are estimated by the maximum likelihood estimators (MLE), the asymptotic null-distribution of $\hat{\chi}_P^2$ is χ_{M-q-1}^2, where q is the number of (independent) parameters being estimated. This is the famous "subtract one degree of freedom for each parameter estimated" rule taught in many elementary statistics books. However, an easy mistake could be made here when one tries to generalize this rule to other cases, because it is not even true in testing fit to a *normal* distribution with unknown mean and variance. Chernoff and Lehmann (1954) showed that when MLE based on the original observations (not the cell frequencies) are used, the asymptotic null-distribution of $\hat{\chi}_P^2$ is not χ^2 but instead a "weighted χ^2" in which the weights are eigenvalues of some nonnegative definite matrix. All of the above results, and much more, can be found in a nice historical review by Moore (1978) on χ^2-tests.

Our test statistic is similar to Pearson's χ^2-statistic with estimated cell probabilities. Let n be a suitable normalizing constant that increases with N. Define

$$\hat{\chi}^2 = \frac{1}{n} \sum_{k=1}^{M} (N_k - \mathrm{E}_{\hat{\theta}} N_k)^2, \tag{1.4}$$

where $N_k = \sum_{i=1}^{N} 1_{(y_i \in E_k)} = \#\{1 \le i \le N : y_i \in E_k\}$, and $\hat{\theta}$ is an estimator of the vector of parameters involved in model (1.1), which we shall specify later. Despite the similarity, there are several major differences. First and

most importantly, the observed count N_k is *not* a sum of independent random variables. In Pearson's χ^2-test, one deals with i.i.d. observations, so that N_k is a sum of i.i.d. random variables, and hence the asymptotic result follows from the classic central limit theorem (CLT). In a mixed linear model, however, the observations are correlated. Therefore, the classic CLT cannot handle the asymptotics. Second, unlike (1.3), the normalizing constant in (1.4) is the same for all the squares in the sum. The choice of the normalizing constants, $\mathrm{E}N_k = NP(X_1 \in E_k)$, in (1.3) is such that the asymptotic distribution is χ^2. However, as noted earlier, even in the i.i.d. case, the asymptotic distribution of (1.3) is not necessarily χ^2, if the parameters are to be estimated; in fact, it may never be χ^2 no matter what normalizing constants might be used to replace $NP(X_1 \in E_k)$, $1 \le k \le M$. Thus, for simplicity, we choose a unified n. Note that, because of the dependence among the observations, n may not be increasing at the same rate as N. Third, in a mixed linear model, the number of fixed effects may be allowed to increase with N [e.g., Jiang (1996); see, also, Portnoy (1984) for the case of linear regression]. As a consequence, the dimension of θ may increase with N. This shows, from another angle, that one can no longer expect an asymptotic distribution like χ^2_{M-q-1}, where q is the number of (independent) parameters being estimated.

There have been extensive studies on the estimation of θ in the literature [e.g., Searle, Casella and McCulloch (1992)]. Among various procedures, the main difference seems to be in the estimation of the variance components. In this paper, we choose to use the restricted maximum likelihood (REML) estimators of the variance components, which have been well-accepted in this field. (However, see the last paragraph in Section 5.) Let $\phi = \tau^2$, $\gamma_r = \sigma_r^2/\tau^2$, $1 \le r \le s$. Define $V_\gamma = I + \sum_{r=1}^s \gamma_r Z_r Z_r^t$ with $\gamma = (\gamma_r)_{1\le r\le s}$. Following Jiang (1996), the REML estimators of ϕ, $\gamma_1, \ldots, \gamma_s$ are defined as solution to the Gaussian REML equations (i.e., REML equations derived under normality), even if the data is not normal. The REML equations are

$$y^t A(A^t V_\gamma A)^{-1} A^t y = \phi(N-p), \tag{1.5}$$

$$\begin{aligned} &y^t A(A^t V_\gamma A)^{-1} A^t Z_r Z_r^t A(A^t V_\gamma A)^{-1} A^t y \\ &\quad = \phi \mathrm{tr}(Z_r^t A(A^t V_\gamma A)^{-1} A^t Z_r), \qquad 1 \le r \le s, \end{aligned} \tag{1.6}$$

where A is any $N \times (N-p)$ matrix such that $\mathrm{rank}(A) = N-p$ and $A^t X = 0$. Once the variance components have been estimated, the fixed effects are then estimated by the empirical best linear unbiased estimator:

$$\hat{\beta} = (X^t V_{\hat\gamma}^{-1} X)^{-1} X^t V_{\hat\gamma}^{-1} y, \tag{1.7}$$

where $\hat\gamma$ is the REML estimator of γ. It is known that the REML estimators have nice asymptotic properties even if the data is not normal, and $p \to \infty$ as $N \to \infty$ [Jiang (1996, 1997)]. A key condition for consistency and asymptotic normality of the REML estimators is the AI[4] condition given below [see Jiang (1996)].

The main result of this paper is based on the asymptotic distribution of the statistic (1.4). Before giving the proof for a general result, we first take a look, in the next two sections, at two special cases: the case of no random effects, that is, linear regression; and the case of single random effect factor. In the first special case, the problem of testing goodness of fit for the error distribution has been studied by Pierce and Kopecky (1979); also see Gray and Pierce (1985), using a quite different approach. Let $F_0(\cdot)$ be the distribution of the standardized errors under the null hypothesis, and $\hat{u}_i = F_0(\hat{e}_i)$, where $\hat{e}_i$ is the Studentized residual, $\hat{e}_i = (y_i - x_i^t\tilde{\beta})/\tilde{\sigma}$ with $(\tilde{\beta}, \tilde{\sigma})$ being the MLE. Let $\hat{F}_N(\cdot)$ be the empirical distribution of the $\hat{u}_i$'s. Pierce and Kopecky showed that, under regularity conditions $\sqrt{N}(\hat{F}_N(t) - t)$ converges to a Gaussian process as $N \to \infty$. Thus, the result may be used for checking the distribution of the errors. However, it is not clear whether such a result still holds if $p \to \infty$, because, as is well-known, in such a case the MLE may not be consistent [Neyman and Scott (1948)]. Nevertheless, it would be interesting to compare Pierce and Kopecky's procedure with ours. The second special case has been considered by Jiang, Lahiri and Wu (2001), as discussed earlier. In this case we use a special property of mixed models with single random effect factor, that is, the observations may be divided into independent groups. Such a technique, however, may not apply to mixed models with multiple random effect factors, and this general case is considered in Section 4. For simplicity, in this paper we mainly consider the case where p is fixed. However, the techniques used here do allow extension of the results to the case where p is not necessarily fixed or bounded. See the discussion in Section 5 and Jiang (2001b) for details. Section 5 also contains discussion and further remarks on a number of issues regarding the application of the results in this paper. A simulated example is considered in Section 6. In addition to those that have been introduced, a number of notations will be used throughout, which we summarize below.

NOTATION. Let $x \vee y$ represent $\max(x, y)$ for real numbers x, y. Let b be a vector and B a matrix. Then, b_l represents the lth component of b and $b_{i,j}$, or simply $B_{i,j}$, = the (i, j) element of B. Define $|b| = (b^tb)^{1/2}$, $\|b\| = \max_l |b_l|$; $|B| = (\mathrm{tr}(B^tB))^{1/2}$, $\|B\| = (\lambda_{\max}(B^tB))^{1/2}$, where $\lambda_{\max}$ represents the largest eigenvalue (and similarly $\lambda_{\min}$ the smallest one). Let $B_1, \ldots, B_k$ be matrices. Define $\mathrm{Cor}(B_1, \ldots, B_k) = (\mathrm{cor}(B_i, B_j))$ if $B_1, \ldots, B_k \neq 0$ and 0 otherwise, where $\mathrm{cor}(B_i, B_j) = \mathrm{tr}(B_i^tB_j)/|B_i||B_j|$; and $\mathrm{diag}(B_1, \ldots, B_k) =$ the block diagonal matrix with $B_1, \ldots, B_k$ on its diagonal.

Let $\psi = (\phi, \gamma^t)^t$, $\theta = (\beta^t, \psi^t)^t$. Let θ_0 be the true θ. For notation simplicity, we shall write, in short, $\theta = \theta_0$, when there is no confusion.

Let $b(\gamma) = (I_{(N)}, \sqrt{\gamma_1}Z_1, \ldots, \sqrt{\gamma_s}Z_s)^t$, where $I_{(k)}$ represents the k-dimensional identity matrix, and $b(\psi) = \phi^{1/2}b(\gamma)$; $V(\gamma) = A(A^tV_\gamma A)^{-1}A^t$, where A is as in (1.5). Let $V_0(\gamma) = b(\gamma)V(\gamma)b(\gamma)^t$, $V_0 = I_{(N-p)}/\phi$ and $p_0(N) = \sqrt{N-p}$; $V_r(\gamma) = b(\gamma)V(\gamma)Z_rZ_r^tV(\gamma)b(\gamma)^t$, $V_r = (A^tV_\gamma A)^{-1/2}A^tZ_rZ_r^tA(A^tV_\gamma A)^{-1/2}$ and $p_r(N) = |V_r|$, $1 \leq r \leq s$.

Let $p(N) = \operatorname{diag}(p_0(N), \ldots, p_s(N))$; $I_N = (\operatorname{tr}(V_u V_v)/p_u(N)p_v \times (N))_{0\le u,v\le s}$; $K_N = (K_{u,v})_{0\le u,v\le s}$, where

$$K_{u,v} = \sum_{l=1}^{N+m_1+\cdots+m_s} (\mathrm{E}\omega_{N,l}^4 - 3)\frac{V_u(\gamma)_{l,l}V_v(\gamma)_{l,l}}{\phi^{1_{(u=0)}+1_{(v=0)}}p_u(N)p_v(N)},$$

in which the random variables $\omega_{N,l}$ are defined as follows:

$$\omega_{N,l} = \begin{cases} \varepsilon_l/\tau, & 1 \le l \le N, \\ \alpha_{r\,l-N-\sum_{u<r}m_u}/\sigma_r, & N + \sum_{u<r} m_u + 1 \le l \le N + \sum_{u\le r} m_u, \end{cases}$$

$1 \le r \le s$, that is, the standardized errors and random effects. Let $J_N = 2I_N + K_N$ and $P_N = J_N^{-1/2} I_N p(N)$; $V_\psi = \phi V_\gamma$, $Q_N = (X^t V_\psi^{-1} X)^{1/2}$, and $C_N = b(\psi)V_\psi^{-1}XQ_N^{-1}$, whose lth row is denoted by $C_{N,l}^t$.

The following definition is needed [see Jiang (1996)]. Model (1.1) is said to be asymptotically-identifiable and infinitely-informative under the invariant class, or AI^4, if

$$\liminf \lambda_{\min}(\operatorname{Cor}(V_0, V_1, \ldots, V_s)) > 0 \quad \text{and} \quad \lim |V_r| = \infty, \qquad 0 \le r \le s.$$

Finally, if ξ is a random variable, $\xi \sim F$ means that ξ is distributed as F. If a_N, b_N are sequences of positive numbers, $a_N \sim b_N$ means that $\liminf(a_N/b_N) > 0$ and $\limsup(a_N/b_N) < \infty$. We use the following abbreviations: w.p. $\to 1$ for "with probability tending to one," and w.r.t. for "with respect to."

2. No random effect factor (linear regression). We first consider the case in which there is no random effect, that is, the linear regression model:

$$y_i = x_i^t\beta + \varepsilon_i, \tag{2.1}$$

$i = 1, \ldots, N$, where $x_i = (x_{ij})_{1\le j\le p}$ is a vector of known covariates, β is an unknown vector of regression coefficients, and $\varepsilon_1, \ldots, \varepsilon_N$ are i.i.d. random errors with mean 0, variance τ^2, and continuous distribution $G = G(\cdot\,|\tau)$. Let $X = (x_i^t)_{1\le i\le N}$; and $\theta = (\beta^t, \tau^2)^t$. We are interested in testing the following hypothesis:

$$H_0 : G(\cdot|\tau) = G_0(\cdot|\tau). \tag{2.2}$$

Let $E_1 = (-\infty, c_1]$, $E_2 = (c_1, c_2]$, $\ldots$, $E_{M-1} = (c_{M-2}, c_{M-1}]$, $E_M = (c_{M-1}, \infty)$, where $c_1 < c_2 < \cdots < c_{M-1}$. For convenience, let $c_0 = -\infty$, $c_M = \infty$. Let N_k be as in (1.4). A well-known estimator of θ is $\hat\theta = (\hat\beta\ \hat\tau^2)^t$, where $\hat\beta$ is the least-squares estimator and $\hat\tau^2$ is the unbiased estimator based on residual sum of squares:

$$\hat\beta = (X^tX)^{-1}X^ty,\ \hat\tau^2 = \frac{|\hat\varepsilon|^2}{N-p}, \tag{2.3}$$

where $\hat{\varepsilon} = y - X\hat{\beta}$. Note that $\hat{\tau}^2$ is, in fact, the REML estimator of the variance component τ^2 obtained by maximizing the Gaussian likelihood based on $z = A^t y$, where A is as in (1.5).

Let $\xi_N = (\xi_{N,k})_{1\le k\le M}$, where $\xi_{N,k} = N_k - \mathrm{E}_{\hat{\theta}} N_k$. Define

$$h_{N,i} = (1_{(y_i\in E_k)} - p_{ik}(\theta))_{1\le k\le M} - \left(\sum_i \frac{\partial}{\partial\beta^t} p_i(\theta)\right)(X^tX)^{-1}x_i\varepsilon_i$$

$$-\frac{1 - x_i^t(X^tX)^{-1}x_i}{N-p}\left(\sum_i \frac{\partial}{\partial\tau^2} p_i(\theta)\right)\varepsilon_i^2,$$

where $p_i(\theta) = (p_{ik}(\theta))_{1\le k\le M}$, $p_{ik}(\theta) = P_\theta(y_i \in E_k)$, and $\Sigma_N = \Sigma_N(\theta) = N^{-1}\sum_{i=1}^N \mathrm{Var}(h_{N,i})$. Note that

$$\frac{\partial}{\partial\beta_j} p_{ik}(\theta) = x_{ij}\{G'(c_{k-1} - x_i^t\beta|\tau) - G'(c_k - x_i^t\beta|\tau)\}, \qquad 1\le j\le p, \tag{2.4}$$

$$\frac{\partial}{\partial\tau^2} p_{ik}(\theta) = \frac{\partial}{\partial\tau^2} G(c_k - x_i^t\beta|\tau) - \frac{\partial}{\partial\tau^2} G(c_{k-1} - x_i^t\beta|\tau), \tag{2.5}$$

where $G'(x|\tau)$ means $(\partial/\partial x)G(x|\tau)$. In particular, if G has density $\tau^{-1}g(\cdot/\tau)$, then

$$\frac{\partial}{\partial\beta_j} p_{ik}(\theta) = \frac{x_{ij}}{\tau}\left[g\left(\frac{c_{k-1} - x_i^t\beta}{\tau}\right) - g\left(\frac{c_k - x_i^t\beta}{\tau}\right)\right], \qquad 1\le j\le p,$$

and it is easy to show that

$$\frac{\partial}{\partial\tau^2} p_{ik}(\theta) = (2\tau^3)^{-1}\left[(c_{k-1} - x_i^t\beta)g\left(\frac{c_{k-1} - x_i^t\beta}{\tau}\right) - (c_k - x_i^t\beta)g\left(\frac{c_k - x_i^t\beta}{\tau}\right)\right],$$

where the corresponding terms are understood as 0 if $k = 1$ or M. Let $P_i(\theta)$ be the $M\times(p+1)$ matrix whose (k, j) element is given by the right side of (2.4), and $(k, p+1)$ element that of (2.5). We now derive an expression for Σ_N. It is easy to show that

$$\Sigma_N = \sum_{r=1}^3 \Sigma_{N,rr}(\theta) + \sum_{1\le r<s\le 3}\{\Sigma_{N,rs}(\theta) + \Sigma_{N,rs}(\theta)^t\}, \tag{2.6}$$

where the Σ matrices on the right side of (2.6) are defined as follows:

$$\Sigma_{N,11}(\theta) = \mathrm{diag}\left(\frac{1}{N}\sum_{i=1}^N p_i(\theta)\right) - \frac{1}{N}\sum_{i=1}^N p_i(\theta)p_i(\theta)^t$$

(note that the kth component of $p_i(\theta)$ is $G(c_k - x_i^t\beta|\tau) - G(c_{k-1} - x_i^t\beta|\tau)$);

$$\Sigma_{N,22}(\theta) = \frac{\tau^2}{N}\left\{\sum_{i=1}^N P_{i,1}(\theta)\right\}(X^tX)^{-1}\left\{\sum_{i=1}^N P_{i,1}(\theta)^t\right\}$$

with $P_{i,1}(\theta)$ being the matrix consisting of the first p columns of $P_i(\theta)$;

$$\Sigma_{N,33}(\theta) = \frac{\int x^4 dG(x|\tau) - \tau^4}{N} \left[\sum_{i=1}^N \left\{ \frac{1 - x_i^t (X^t X)^{-1} x_i}{N-p} \right\}^2 \right] \times \left\{ \sum_{i=1}^N P_{i,2}(\theta) \right\} \left\{ \sum_{i=1}^N P_{i,2}(\theta)^t \right\}$$

with $P_{i,2}(\theta)$ being the $(p+1)$th column of $P_i(\theta)$;

$$\Sigma_{N,12}(\theta) = -\frac{1}{N} \left\{ \sum_{i=1}^N \left(\int_{c_{k-1}-x_i^t\beta}^{c_k - x_i^t\beta} x dG(x|\tau) \right)_{1 \le k \le M} x_i^t \right\} (X^t X)^{-1} \left\{ \sum_{i=1}^N P_{i,1}(\theta)^t \right\};$$

$$\Sigma_{N,13}(\theta) = -\frac{1}{N} \left\{ \sum_{i=1}^N \frac{1 - x_i^t (X^t X)^{-1} x_i}{N-p} \times \left(\int_{c_{k-1}-x_i^t\beta}^{c_k - x_i^t\beta} (x^2 - \tau^2) dG(x|\tau) \right)_{1 \le k \le M} \right\} \left\{ \sum_{i=1}^N P_{i,2}(\theta)^t \right\};$$

$$\Sigma_{N,23}(\theta) = \frac{\int x^3 dG(x|\tau)}{N} \left[\sum_{i=1}^N \frac{1 - x_i^t (X^t X)^{-1} x_i}{N-p} \times \left\{ \sum_{i=1}^N P_{i,1}(\theta) \right\} (X^t X)^{-1} x_i \right] \left\{ \sum_{i=1}^N P_{i,2}(\theta)^t \right\}.$$

Note that, although Σ_N has a "closed-form" expression, given θ it is often more convenient to compute Σ_N by a Monte-Carlo method, using the original definition: $\Sigma_N = N^{-1} \sum_{i=1}^N \mathrm{Var}(h_{N,i})$ (see the eighth paragraph in Section 5).

Let $T_N = T_N(\theta)$ be an orthogonal matrix such that $T_N^t \Sigma_N T_N = D_N = \mathrm{diag}(\lambda_{N,1}, \ldots, \lambda_{N,M})$, where $\lambda_{N,1} \ge \cdots \ge \lambda_{N,M}$ are the eigenvalues of Σ_N. In order to obtain critical values for $\hat{\chi}^2$, we need to replace the $\lambda_{N,k}$'s by their estimators. Let $\hat{\lambda}_1 \ge \cdots \ge \hat{\lambda}_M$ be the eigenvalues of $\hat{\Sigma}_N = \Sigma_N(\hat{\theta})$. Note that these eigenvalues are random and dependent on N, even if $\Sigma_N(\theta)$ does not depend on N. For any $\lambda_1 \ge \cdots \ge \lambda_M$ and $0 < \alpha < 1$, let $c_\alpha(\lambda_1, \ldots, \lambda_M)$ be the α-critical value of the random variable $\xi = \sum_{k=1}^M \lambda_k Z_k^2$, where $Z_1, \ldots, Z_M$ are independent $N(0,1)$ random variables, that is, $P(\xi > c_\alpha(\lambda_1, \ldots, \lambda_M)) = \alpha$. Let $\hat{c}_\alpha = c_\alpha(\hat{\lambda}_1, \ldots, \hat{\lambda}_M)$. We now give conditions under which $\hat{c}_\alpha$ is the asymptotic α-critical value of $\hat{\chi}^2$.

ASSUMPTION 2.1. $\mathrm{E}\varepsilon_i^4 < \infty$, $\mathrm{E}\varepsilon_i^2 > 0$.

ASSUMPTION 2.2. *$G(x|\tau)$ is twice continuously differentiable w.r.t. both x and τ, and $\int x^\lambda dG(x|\tau)$ is continuously differentiable w.r.t. τ, $\lambda = 3, 4$.*

ASSUMPTION 2.3. *x_{ij}, $1 \le i \le N$, $1 \le j \le p$ and $N/\lambda_{\min}(X^tX)$ are bounded.*

THEOREM 2.1. *Under Assumptions 2.1 – 2.3, $P(\hat{\chi}^2 > \hat{c}_\alpha) \longrightarrow \alpha$ as $N \to \infty$.*

PROOF. First assume that, in addition, the following limit exists:

$$\Sigma_N \longrightarrow \Sigma. \tag{2.7}$$

We show that, in this case,

$$N^{-1/2} T_N^t \xi_N \xrightarrow{d} N(0, D), \tag{2.8}$$

where $D = \text{diag}(\lambda_1, \ldots, \lambda_M)$ with $\lambda_1 \ge \cdots \ge \lambda_M$ being the eigenvalues of Σ. It follows that

$$\hat{\chi}^2 = \frac{1}{N} \sum_{k=1}^{M} (N_k - \mathrm{E}_{\hat{\theta}} N_k)^2 \xrightarrow{d} \sum_{k=1}^{M} \lambda_k Z_k^2, \tag{2.9}$$

where $Z_1, \ldots, Z_M$ are independent $N(0, 1)$ random variables.

For any $\lambda \in R^M$, let $\lambda_{(N)} = T_N \lambda$. By (4.5) (with n replaced by N) and the Taylor expansion, we have

$$\begin{aligned} \lambda^t (N^{-1/2} T_N^t \xi_N) &= \sum_{i=1}^{N} \eta_{N,i} - N^{-1/2} \lambda_{(N)}^t \left(\sum_i \frac{\partial}{\partial \theta^t} p_i(\theta) \right) (\hat{\theta} - \theta) + o_P(1) \\ &= \sum_{i=1}^{N} \eta_{N,i} - d_N^t \omega_N - (\omega_N^t B_N \omega_N - \mathrm{E} \omega_N^t B_N \omega_N) + o_P(1) \\ &= \sum_{i=1}^{N} Y_{N,i} + o_P(1), \end{aligned} \tag{2.10}$$

where $\eta_{N,i} = N^{-1/2} \sum_{k=1}^{M} \lambda_{(N),k} (1_{(y_i \in E_k)} - p_{ik}(\theta))$, $\omega_N = \varepsilon/\tau$ with $\varepsilon = (\varepsilon_i)_{1 \le i \le N}$,

$$d_N = \frac{\tau}{\sqrt{N}} X (X^tX)^{-1} \left(\sum_i \frac{\partial}{\partial \beta^t} p_i(\theta) \right)^t \lambda_{(N)},$$

$$B_N = \frac{\tau^2}{(N-p)\sqrt{N}} \left(\sum_i \frac{\partial}{\partial \tau^2} p_i(\theta) \right)^t \lambda_{(N)} P_{X^\perp},$$

$$P_{X^\perp} = I - P_X = I - X(X^tX)^{-1} X^t$$

and

$$Y_{N,i} = \eta_{N,i} - d_{N,i} \omega_{N,i} - b_{N,i,i} (\omega_{N,i}^2 - 1) - 2 \left(\sum_{i' < i} b_{N,i,i'} \omega_{N,i'} \right) \omega_{N,i}.$$

Let $\mathscr{F}_{N,i} = \sigma(\varepsilon_{i'}, i' \le i)$. Then, $\{Y_{N,i}, \mathscr{F}_{N,i}, 1 \le i \le N\}$ is an array of martingale differences. The rest of the proof for (2.8) is similar to the corresponding part in the proof of Theorem 4.1 (but easier).

We now show that (2.7) and the conditions imply

$$\max_{1\le k\le M} |\hat{\lambda}_k - \lambda_k| \xrightarrow{P} 0. \tag{2.11}$$

First we note that the conditions are sufficient for the consistency of $\hat{\theta}$ [this can be seen from (2.10), or proved directly]. By Weyl's eigenvalue perturbation theorem [e.g., Bhatia (1997), page 63], we have

$$\max_{1\le k\le M} |\hat{\lambda}_k - \lambda_{N,k}| \le \|\Sigma_N(\hat{\theta}) - \Sigma_N(\theta)\|. \tag{2.12}$$

On the other hand, it can be shown that all the partial derivatives of $\sigma_{N,kl}(\theta)$ w.r.t. components of θ are bounded in a neighborhood of θ [see Jiang (2001) for details], where $\sigma_{N,kl}(\theta)$ is the (k,l) element of $\Sigma_N(\theta)$. Thus, w.p.$\to 1$, we have $\sup_N |\sigma_{N,kl}(\hat{\theta}) - \sigma_{N,kl}(\theta)| \le c|\hat{\theta} - \theta|$, where c is a constant. It follows that the right side of (2.12) converges to 0 in probability. Again, by Weyl's theorem, we have

$$\max_{1\le k\le M} |\lambda_{N,k} - \lambda_k| \le \|\Sigma_N(\theta) - \Sigma\|, \tag{2.13}$$

and the right side of (2.13) converges to 0 by (2.7). This proves (2.11).

Next, we show that $c_\alpha(\lambda_1, \ldots, \lambda_M)$ is continuous in $\lambda_1, \ldots, \lambda_M$. Let the latter be fixed. Suppose that $|\tilde{\lambda}_k - \lambda_k| < \delta$, $1 \le k \le M$. Write $\xi = \sum_{k=1}^M \lambda_k Z_k^2$, $\tilde{\xi} = \sum_{k=1}^M \tilde{\lambda}_k Z_k^2$, $c_\alpha = c_\alpha(\lambda_1, \ldots, \lambda_M)$ and $\tilde{c}_\alpha = c_\alpha(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_M)$. Then, since the distribution of ξ is nondegenerate, i.e., $P(a < \xi \le b) > 0$ for any $0 \le a < b$, we have $P(\xi > c_\alpha + \varepsilon) < \alpha$ for any $\varepsilon > 0$. If $\tilde{c}_\alpha \ge c_\alpha + 2\varepsilon$, we have $\alpha = P(\tilde{\xi} > \tilde{c}_\alpha) \le P(\xi > c_\alpha + \varepsilon) + P(\tilde{\xi} - \xi > \varepsilon) \le P(\xi > c_\alpha + \varepsilon) + M\delta/\varepsilon$, by Chebychev's inequality. Thus, by making δ sufficiently small we would have $\alpha < \alpha$, a contradiction. Therefore, with such a δ we must have $\tilde{c}_\alpha - c_\alpha < 2\varepsilon$. Similarly, we have $\tilde{c}_\alpha - c_\alpha > -2\varepsilon$, if δ is sufficiently small. This proves continuity of c_α.

For any $\varepsilon > 0$, let $\delta > 0$ be such that $\max_{1\le k\le M} |\tilde{\lambda}_k - \lambda_k| < \delta$ implies $|\tilde{c}_\alpha - c_\alpha| < \varepsilon$. Then, we have $P(\hat{\chi}^2 > \hat{c}_\alpha) = P(\hat{\chi}^2 > \hat{c}_\alpha, \max_{1\le k\le M} |\hat{\lambda}_k - \lambda_k| < \delta) + P(\hat{\chi}^2 > \hat{c}_\alpha, \max_{1\le k\le M} |\hat{\lambda}_k - \lambda_k| \ge \delta) \le P(\hat{\chi}^2 > c_\alpha - \varepsilon) + P(\max_{1\le k\le M} |\hat{\lambda}_k - \lambda_k| \ge \delta)$. It follows, by (2.9) and (2.11), that $\limsup P(\hat{\chi}^2 > \hat{c}_\alpha) \le P(\xi > c_\alpha - \varepsilon)$. Similarly, we have $\liminf P(\hat{\chi}^2 > \hat{c}_\alpha) \ge P(\xi > c_\alpha + \varepsilon)$. Thus, by the arbitrariness of $\varepsilon > 0$, we have $P(\hat{\chi}^2 > \hat{c}_\alpha) \to P(\xi > c_\alpha) = \alpha$.

Finally, we drop the additional assumption (2.7). We then use the following argument of subsequences. First, it can be shown that all elements of Σ_N are bounded [see Jiang (2001b) for details]. Therefore, for any subsequence of $\{N\}$ there is a further subsequence, for which Σ_N converges to a limit Σ, that is, (2.7) holds. It then follows that, for the further subsequence, $P(\hat{\chi}^2 > \hat{c}_\alpha) \to \alpha$. Because the limit does not depend on the subsequence, the proof is complete. □

We now consider a very special case in which the proposed test may be related to a classic one.

EXAMPLE 2.1. Suppose that $y_i = \mu + \varepsilon_i$, $i = 1, \ldots, N$, where the ε_i's are i.i.d. $\sim G$ with density $\tau^{-1} g(\cdot/\tau)$. Then, the testing of (2.2) reduces to a classic

problem, that is, goodness-of-fit test based on i.i.d. observations. Suppose that $g'(\cdot)$ is bounded on any compact subset of R, and Assumption 2.1 is satisfied. Then, it is obvious that Assumptions 2.2 and 2.3 are satisfied. Therefore, the conclusion of Theorem 2.1 holds. Note that, here, $\Sigma_N = \Sigma = \mathrm{Var}(h_1)$, where $h_1 = \zeta_1 - \Delta_u(\varepsilon_1/\tau) - (1/2)\Delta_v(\varepsilon_1/\tau)^2$; $\zeta_1 = (1_{(y_1 \in E_k)} - p_k)_{1 \le k \le M}$ with $p_k = p_{1k}(\theta)$; $\Delta_u = (u_{k-1} - u_k)_{1 \le k \le M}$, $\Delta_v = (v_{k-1} - v_k)_{1 \le k \le M}$ with $u_k = g((c_k - \mu)/\tau)$, $v_k = \{(c_k - \mu)/\tau\}u_k$, $1 \le k \le M - 1$ and $u_k = v_k = 0$, $k = 0$ or M.

To see how the test is related to Pearson's χ^2-test, write $\mathrm{Var}(h_1) = \mathrm{Var}(\zeta_1) + \Delta$, where Δ represents the difference between the covariance matrices of h_1 and ζ_1. Note that Δ is the result of estimation of θ. If the parameters were known, in which case there is no need to estimate θ, then, it is easy to show that (2.8) holds, where $D = T^t \Sigma T$ with $\Sigma = \mathrm{Var}(\zeta_1)$, that is, $\Delta = 0$. Note that, since Σ_N does not depend on N, neither does T_N, the orthogonal matrix, that is, $T_N = T$. Furthermore, it is easy to show that $\mathrm{Var}(\zeta_1) = P - \underline{p}\underline{p}^t$, where $P = \mathrm{diag}(p_k, 1 \le k \le M)$, $\underline{p} = (p_k)_{1 \le k \le M}$. It follows that $Q = P^{-1/2}\mathrm{Var}(\zeta_1)P^{-1/2} = I - (\underline{p}^{1/2})(\underline{p}^{1/2})^t$, where $\underline{p}^{1/2} = (p_k^{1/2})_{1 \le k \le M}$. It is easy to see that Q is idempotent with $\mathrm{rank}(Q) = M - 1$. Thus, by (2.8), we have $N^{-1/2}P^{-1/2}\xi_N = P^{-1/2}T(N^{-1/2}T^t\xi_N) \xrightarrow{d} N(0, Q)$. Therefore,

$$\sum_{k=1}^{M} \frac{(N_k - Np_k)^2}{Np_k} = N^{-1}|P^{-1/2}\xi_N|^2 \xrightarrow{d} \chi^2_{M-1}. \tag{2.14}$$

The left side of (2.14) is exactly Pearson's χ^2-statistic (1.3) when θ is known, hence the cell probabilities are completely specified under H_0. When θ is unknown and therefore has to be estimated, the asymptotic distribution of $\hat{\chi}^2_P$ is more complicated and, in particular, may not be χ^2 (see our brief review in Section 1).

3. Single random effect factor. We now consider the mixed linear model (1.1) with one random effect factor (i.e., $s = 1$). Suppose that the model can be expressed as

$$y_{ij} = x_{ij}^t\beta + \alpha_i + \varepsilon_{ij}, \tag{3.1}$$

$i = 1, \ldots, m$, $j = 1, \ldots, n_i$ ($n_i \ge 1$), where $x_{ij} = (x_{ijk})_{1 \le k \le p}$ is a known vector of covariates, β is an unknown vector of regression coefficients; α_i is a random effect; and ε_{ij} is a random error. For a model that allows some extension, see Jiang (2001b). Suppose that the α_i's are i.i.d. with mean 0, variance σ^2 and continuous distribution $F = F(\cdot\,|\sigma)$; the ε_{ij}'s are i.i.d. with mean 0, variance τ^2 and continuous distribution $G = G(\cdot\,|\tau)$; and the α's and ε's are independent. We are interested in testing the following hypothesis:

$$H_0 : F(\cdot|\sigma) = F_0(\cdot|\sigma) \text{ and } G(\cdot|\tau) = G_0(\cdot|\tau). \tag{3.2}$$

Let E_k, $1 \le k \le M$ be defined as in Section 2. Consider

$$N_k = \sum_{i,j} 1_{(y_{ij} \in E_k)} = \#\{(i, j) : y_{ij} \in E_k\}, \qquad 1 \le k \le M.$$

Define ξ_N as in Section 2. Let

$$p_1(N) = (\mathrm{tr}[(Z^tV(\gamma)Z)^2])^{1/2}, \qquad \rho_N = \mathrm{tr}[Z^tV(\gamma)Z]/[p_0(N)p_1(N)].$$

Note that $p_1(N)$ defined here is $\sim$ to the $p_1(N)$ defined at the end of Section 1 (in fact, the ratio of the two $\to$ 1). Let $P_{ij}(\theta)$ be the $M \times (p+2)$ matrix whose (k, r) element is

$$\frac{\partial}{\partial\theta_r}\int\{G(c_k - x_{ij}^t\beta - u|\tau) - G(c_{k-1} - x_{ij}^t\beta - u|\tau)\}dF(u|\sigma),$$

where $\theta_r = \beta_r$ if $1 \le r \le p$, $\theta_{p+1} = \phi$, and $\theta_{p+2} = \gamma$. Let $P_{ij}[r](\theta)$ be the rth column of $P_{ij}(\theta)$ and $P_{ij}[1, p](\theta)$ the matrix consist of the first p columns of $P_{ij}(\theta)$. Define the following matrices:

$$\Phi_N = \frac{1}{1-\rho_N^2}\begin{pmatrix}\phi^2 & -\phi\rho_N\\ -\phi\rho_N & 1\end{pmatrix}\begin{pmatrix}\dfrac{1}{p_0(N)}\displaystyle\sum_{i,j}P_{ij}[p+1](\theta)^t\\ \dfrac{1}{p_1(N)}\displaystyle\sum_{i,j}P_{ij}[p+2](\theta)^t\end{pmatrix} = \begin{pmatrix}\Phi_{N,0}^t\\ \Phi_{N,1}^t\end{pmatrix},$$

$$\Psi_N = b(\psi)V_\psi^{-1}XQ_N^{-2}\sum_{i,j}P_{ij}[1, p](\theta)^t = (\Psi_{N,l}^t)_{1\le l\le N+m}.$$

Let $S_i = \{l : \sum_{i'<i} n_{i'} + 1 \le l \le \sum_{i'\le i} n_{i'}\} \cup \{N+i\}$. Write $\omega_N(i) = (\omega_{N,l})_{l\in S_i}$, $V_r(i, i') = (V_r(\gamma)_{l,l'})_{l\in S_i, l'\in S_{i'}}$, $r = 0, 1$ and $\Psi_N(i) = (\Psi_{N,l}^t)_{l\in S_i}$. Note that

$$\Psi_{N,l}^t = b_l^tV_\psi^{-1}X(X^tV_\psi^{-1}X)^{-1}\sum_{ij}P_{ij}[1, p](\theta)^t,$$

where b_l^t is the lth row of $b(\psi)$. Let

$$h_{N,i} = \left(\sum_{j=1}^{n_i}[1_{(y_{ij}\in E_k)} - p_{ijk}(\theta)]\right)_{1\le k\le M} - \Psi_N^t(i)\omega_N(i)$$

$$- \sum_{r=0}^{1}\frac{\omega_N^t(i)V_r(i, i)\omega_N(i)}{\phi^{1-r}p_r(N)}\Phi_{N,r}.$$

Let $H_N = \sum_{i=1}^m \mathrm{Var}(h_{N,i})$; $R_N = (r_{N,r,r'})_{0\le r,r'\le 1}$, where

$$r_{N,r,r'} = \frac{\sum_{i=1}^m \mathrm{tr}(V_r(i, i)V_{r'}(i, i))}{\phi^{2-r-r'}p_r(N)p_{r'}(N)};$$

and $\Sigma_N = \Sigma_N(\theta) = n^{-1}[H_N + 2\Phi_N^t(I_N - R_N)\Phi_N]$, where $n = n(N)$ is a sequence of normalizing constants. It is possible, as in Section 2, to obtain a "closed-form" expression for Σ_N [see Jiang (2001b) for details]. Again, it is often more convenient to compute Σ_N by a Monte-Carlo method.

Define T_N and D_N as in Section 2. In the following, we assume that the n_i' are bounded. Under such an assumption, the conditions under which our asymptotic result holds is rather simple and intuitive. In particular, one may let $n = m$ in the definition of Σ_N. Note that, in most applications of

mixed linear models there is limited information about individual random effects, which means that the n_i's are small. For example, in small area estimation [e.g., Ghosh and Rao (1994)], the n_i's correspond to sample sizes for the small areas, which are small. Therefore, such an assumption (i.e., that the n_i's are bounded) is not impractical. Nevertheless, this assumption may be weakened, although the corresponding conditions are less neat [see Jiang (2001b) for details, or, similarly, Theorem 4.1 in the sequel]. Define $G_d(x|\tau) = \int_{-\infty}^{x} v^d dG(v|\tau)$, $d = 0, 1, 2$ [so that $G_0(x|\tau) = G(x|\tau)$]. The following are analogues of Assumptions 2.1–2.3 in Section 2.

ASSUMPTION 3.1. $\mathrm{E}\alpha_i^4 < \infty$, $\mathrm{E}\alpha_i^2 > 0$; $\mathrm{E}\varepsilon_{ij}^4 < \infty$, $\mathrm{E}\varepsilon_{ij}^2 > 0$.

ASSUMPTION 3.2. *$\int G(c_k - x_{ij}^t\beta - u|\tau)dF(u|\sigma)$ is twice differentiable w.r.t. β, σ, and τ; $\int G(c_k - x_{ij}^t\beta - u|\tau)G(c_{k'} - x_{ij'}^t\beta - u|\tau)dF(u|\sigma)$, $\int u^c G_d(c_k - x_{ij}^t\beta - u|\tau)dF(u|\sigma)$, where $c, d \in \{0, 1, 2\}$ such that $c + d = 1$ or 2, are differentiable w.r.t. β, σ and τ; and $\int u^\lambda dF(u|\sigma)$, $\int v^\lambda dG(v|\tau)$, $\lambda = 3, 4$ are differentiable w.r.t. σ and τ, respectively. The partial derivatives are bounded in a neighborhood of θ.*

ASSUMPTION 3.3. *x_{ijk}, n_i, $1 \le i \le m$, $1 \le j \le n_i$, $1 \le k \le p$ and $N/\lambda_{\min}(X^t V_\psi^{-1} X)$ are bounded.*

In addition, we assume the following.

ASSUMPTION 3.4. *Model* (3.1) *is* AI^4, *and* $\max_l |C_{N,l}| = o(1)$.

The AI^4 condition implies that there exists w.p. $\to 1$ REML estimator $\hat{\psi}$ which is consistent and asymptotically normal [Jiang (1996)]. Note that, by Jiang (1997), such a REML estimator can be identified. Define $\hat{c}_\alpha$ $(0 < \alpha < 1)$ the same way as in Section 2, where $\hat{\psi}$ is the REML estimator and $\hat{\beta}$ given by (1.7).

THEOREM 3.1. *Under Assumptions* $3.1 - 3.4$, $P(\hat{\chi}^2 > \hat{c}_\alpha) \longrightarrow \alpha$ *as* $N \to \infty$.

PROOF. The procedure is similar to proof of Theorem 2.1. First assume the existence of $\lim \Sigma_N$ and show (2.8) (with N replaced by m). We defer the details to the corresponding part in proof of Theorem 4.1. The first term in (4.5) can be expressed as $\sum_{i=1}^m \eta_{N,i}$, where $\eta_{N,i} = n^{-1/2} \sum_{k=1}^M \lambda_{(N),k} \sum_{j=1}^{n_i} [1_{(y_{ij} \in E_k)} - p_{ijk}(\theta)]$, and the second term in (4.5) $= d_N^t \omega_N + \omega_N^t B_N \omega_N - \mathrm{E}(\omega_N^t B_N \omega_N) + o_P(1)$. Write $d_N(i) = (d_{N,l})_{l \in S_i}$, and $B_N(i, i') = (b_{N,l,l'})_{l \in S_i, l' \in S_{i'}}$. Then, $d_N^t \omega_N = \sum_{i=1}^m d_N^t(i)\omega_N(i)$ and $\omega_N^t B_N \omega_N = \sum_{i,i'=1}^m \omega_N^t(i') B_N(i', i)\omega_N(i)$. Thus,

$$\lambda^t(n^{-1/2} T_N^t \xi_N) = \sum_{i=1}^m Y_{N,i} + o_P(1), \tag{3.3}$$

where

$$Y_{N,l} = \eta_{N,i} - d_N^t(i)\omega_N(i) - [\omega_N^t(i)B_N(i,i)\omega_N(i) - \mathrm{tr}(B_N(i,i))] - 2\left[\sum_{i'<i}\omega_N^t(i')B_N(i',i)\right]\omega_N(i).$$

Let $\mathscr{F}_{N,i} = \sigma(\omega_N(i'), i' \le i)$. Then $\{Y_{N,i}, \mathscr{F}_{N,i}, 1 \le i \le m\}$ is an array of martingale differences. The following facts will be used to complete the proof, where c represents a constant, and $u_N(i) = (u_{N,l})_{l\in S_i}$ with $u_{N,l}$ defined as in the proof of Theorem 4.1:

(i) $\mathrm{E}(d_N^t(i)u_N(i))^4 \le c\|d_N\|^2|d_N(i)|^2$;
(ii) $\mathrm{E}(\zeta_i - \mathrm{E}\zeta_i)^4 \le c[\|B_N\|^2 + \|B_N\|^2 t_N^4]|B_N(i,i)|^2$, where $\zeta_i = u_N^t(i)B_N(i,i)u_N(i)$;
(iii) $\mathrm{E}|d_N^t(i)u_N(i)B_N(i',i)u_N(i) - \mathrm{E}\cdots|^2 \le c\|d_N\|^2|B_N(i',i)|^2$;
(iv) $\mathrm{E}|B_N(i'i)u_N(i)(\zeta_i - \mathrm{E}\zeta_i) - \mathrm{E}\cdots|^2 \le c[\|B_N\|^2 + \|B_N\|^2 t_N^2]|B_N(i',i)|^2$.

We also use Lemma 3.1 below which is an extension of Lemma 5.1 of Jiang (1996).

The rest of the proof is exactly the same as the corresponding part of Theorem 2.1 once we establish the boundedness of elements of Σ_N and their partial derivatives. However, it can be shown that, under Assumptions 3.1–3.4, the elements of Σ_N are bounded, and their partial derivatives w.r.t. components of θ are bounded in a neighborhood of θ. See Jiang (2001b) for details. □

LEMMA 3.1. *Let $B = (B_{ij}1_{(i>j)})$ be a block lower triangular matrix. Then*

$$|B^tB| \le \sqrt{2}\|B^t + B\||B|.$$

We now consider a simple example.

EXAMPLE 3.1. Consider a balanced one-way random effects model: $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, $i = 1,\ldots,m$, $j = 1,\ldots,l$, where the α_i's are i.i.d. $\sim F$ with density $\sigma^{-1}f(\cdot/\sigma)$, the ε_{ij}'s are i.i.d. $\sim G$ with density $\tau^{-1}g(\cdot/\tau)$, and the α's and ε's are independent. Suppose that Assumption 3.1 is satisfied, and that the second derivatives of $p_{11k}(\theta) = P_\theta(y_{11} \in E_k)$ w.r.t. θ are bounded in a neighborhood near θ. In addition we assume that $1 \le k < M$ and $2 \le l \le L$ for some M and L. Then, Assumptions 3.2 and 3.3 are satisfied. Assumption 3.4 is satisfied, because the model is balanced, unconfounded, and with positive variance components [see Jiang (1996)]. Therefore, the conclusion of Theorem 3.1 holds.

4. Multiple random effect factors. Finally, we consider the mixed linear model (1.1) with s ($s \ge 2$) random effect factors. Suppose that there is a multiple index i such that the components of (1.1) can be expressed as

$$y_i = x_i^t\beta + \sum_{r=1}^{s} z_{r,i,i_{(r)}}\alpha_{r,i_{(r)}} + \varepsilon_i, \tag{4.1}$$

where $i = (i_1, \ldots, i_q, i_{q+1})$, $1 \le i_r \le m_r$, $1 \le r \le q$ and $1 \le i_{q+1} \le n_{i_1,\ldots,i_q}$. Here we allow $n_{i_1,\ldots,i_q}$ to be 0, which means no observation for the cell. Thus, the total number of observations $N = \sum_{i_1=1}^{m_1} \cdots \sum_{i_q=1}^{m_q} n_{i_1,\ldots,i_q}$. Furthermore, suppose

$$\sum_{r=1}^{s} z_{r,i,i_{(r)}} \alpha_{r,i_{(r)}} = \sum_{r=1}^{a} z_{r,i,i_r} \alpha_{r,i_r} + \sum_{r=a+1}^{s} z_{r,i,i_{(r)}} \alpha_{r,i_{(r)}}, \tag{4.2}$$

that is, $i_{(r)} = i_r$, $1 \le r \le a$. We assume that $i_{(r)} = (i_{r,1}, \ldots, i_{r,d_r})$ with $d_r \ge 2$, and $\{i_{r,l}, 1 \le l \le d_r\} \subset \{i_r, 1 \le r \le b\}$, $a+1 \le r \le s$, where $a \le b \le q$, and $b > 1$. Intuitively, α_r, $1 \le r \le a$ correspond to random main effects, while α_r, $a+1 \le r \le s$ correspond to (random) interactions and nested random effects. For example, in the model $y_{ijk} = \mu + t_i + u_j + v_{ij} + w_{ik} + e_{ijk}$, t_i and u_j are main effects, v_{ij} is an interaction, w_{ik} is a nested effect, and e_{ijk} is an error. Finally, suppose $\alpha_r = (\alpha_{r,i_{(r)}})$ ($i_{(r)}$ ranges according to the above), where the $\alpha_{r,i_{(r)}}$'s are i.i.d. with mean 0, variance σ_r^2, and continuous distribution $F_r = F_r(\cdot\,|\sigma_r)$, $1 \le r \le s$; $\varepsilon = (\varepsilon_i)$ (i ranges according to the above), where the ε_i's are i.i.d. with mean 0, variance τ^2, and continuous distribution $G = G(\cdot\,|\tau)$; and $\alpha_1, \ldots, \alpha_s, \varepsilon$ are independent. Although not every mixed linear model (1.1) can be expressed as (4.1), the latter does cover most mixed linear models that are used in practice. We are interested in testing the following hypothesis:

$$H_0 : F_r(\cdot|\sigma_r) = F_{0r}(\cdot|\sigma_r), \qquad 1 \le r \le s \quad \text{and} \qquad G(\cdot|\tau) = G_0(\cdot|\tau). \tag{4.3}$$

Let E_k, $1 \le k \le M$ be defined as in Section 2. Consider

$$N_k = \sum_i 1_{(y_i \in E_k)} = \#\{i : y_i \in E_k\},\ 1 \le k \le M.$$

Define ξ_N as in Section 2. Similarly, let $P_i(\theta)$ be the $M \times (p+s+1)$ matrix whose (k, r) element is

$$\frac{\partial}{\partial \theta_r}\{P_\theta(y_i \in E_k)\} = \frac{\partial}{\partial \theta_r} \int \cdots \int \{G(c_k - \eta_i(u)|\tau) - G(c_{k-1} - \eta_i(u)|\tau)\} \times dF_1(u_1|\sigma_1) \cdots dF_s(u_s|\sigma_s)$$

($1 \le k \le M, 1 \le r \le p+s+1$), where $\theta_j = \beta_j$, $1 \le j \le p$, $\theta_{p+1} = \phi$ and $\theta_{p+1+j} = \gamma_j$, $1 \le j \le s$; $u = (u_1, \ldots, u_s)$ and $\eta_i(u) = x_i^t \beta + \sum_{r=1}^{s} z_{r,i,i_{(r)}} u_r$. Let

$$\Phi_N = I_N^{-1} p(N)^{-1} \sum_i P_i[p+1, p+s+1](\theta)^t,$$

$$\Psi_N = b(\psi) V_\psi^{-1} X Q_N^{-2} \sum_i P_i[1, p](\theta)^t,$$

where $P_i[c, d](\theta)$ represents the matrix consist of the $c, \ldots, d$ columns of $P_i(\theta)$. Furthermore, let $P'_{i,j}[c, d](\theta)$ denote the matrix $(\partial/\partial\theta_j) P_i[c, d](\theta)$, $1 \le j \le p+s+1$. Note that the (k, r) element of $P'_{i,j}[1, p](\theta)$ is

$$\frac{\partial^2}{\partial \theta_j \partial \theta_r} \int \cdots \int \{G(c_k - \eta_i(u)|\tau) - G(c_{k-1} - \eta_i(u)|\tau)\} \times dF_1(u_1|\sigma_1) \cdots dF_s(u_s|\sigma_s),$$

where $1 \le r \le p$; and the (k, r) element of $P'_{i,j}[p+1, p+s+1](\theta)$ has the same expression but $p+1 \le r \le p+s+1$. Let the rows of Φ_N be numbered as $0, \ldots, s$, and denote the rth row of Φ_N by $\Phi^t_{N,r}$. Also, let $\Psi^t_{N,l}$ be the lth row of Ψ_N, $1 \le l \le L$, where $L = N + \sum_{r=1}^s m_r$. Write $\eta_i = y_i - \varepsilon_i$, $d_k(\eta_i) = P(y_i \in E_k|\alpha) = G(c_k - \eta_i|\tau) - G(c_{k-1} - \eta_i|\tau)$ and $\zeta_{k,i} = \mathrm{E}(d_k(\eta_i)|\alpha^{(1)}) - \mathrm{E}d_k(\eta_i)$, where $\alpha^{(1)} = (\alpha_r)_{1 \le r \le a}$. Let $S_r(l) = \{i : i_r = l\}$, and $S^{(l)} = S_r(l - \sum_{r' < r} m_{r'})$, if $\sum_{r' < r} m_{r'} + 1 \le l \le \sum_{r' \le r} m_{r'}$, $1 \le r \le a$ ($m_0 = N$). Let $K = N + \sum_{r=1}^a m_r$ and

$$h_{N,l} = \left(\sum_{i \in S^{(l)}} \mathrm{E}(\zeta_{k,i}|\omega_{N,l}) \right)_{1 \le k \le M} 1_{(N+1 \le l \le K)} - \omega_{N,l} \Psi_{N,l} - \omega^2_{N,l} \sum_{r=0}^s \frac{V_r(\gamma)_{l,l}}{\phi^{1_{(r=0)}} p_r(N)} \Phi_{N,r}.$$

Let $H_N = \sum_{l=1}^L \mathrm{Var}(h_{N,l})$; $R_N = (r_{N,r,r'})_{0 \le r,r' \le s}$, where

$$r_{N,r,r'} = \frac{\sum_{l=1}^L V_r(\gamma)_{l,l} V_{r'}(\gamma)_{l,l}}{\phi^{1_{(r=0)}+1_{(r'=0)}} p_r(N) p_{r'}(N)}.$$

Define Σ_N as in Section 3 (right below the definition of R_N therein). It can be shown that, with R_N defined here, $I_N - R_N$ is nonnegative definite (although this may not be true for the R_N defined in Section 3). Again, it is possible to obtain a "closed-form" expression for Σ_N [see Jiang (2001b) for details]. On the other hand, it is often more convenient to compute Σ_N by a Monte-Carlo method.

Define T_N and D_N as in Section 2. Also, Let $\mathscr{I}_1 = \{(i, i') : i_{(r)} = i'_{(r)} \text{ for some } a+1 \le r \le s\}$; $\mathscr{I}_2 = \{(i, i') : i_r = i'_r, i_{r'} = i'_{r'} \text{ for some } 1 \le r, r' \le a, r \ne r'\}$; and $H_{i,k,r}(u_r) = \int \cdots \int G(c_k - \eta_i(u)|\tau) \prod_{t \ne r} dF_t(u_t|\sigma_t)$. The following assumptions are similar to Assumptions 3.1–3.4, but somewhat more general.

ASSUMPTION 4.1. $\mathrm{E}\alpha^4_{r,i_{(r)}} < \infty$, $\mathrm{E}\alpha^2_{r,i_{(r)}} > 0$, $1 \le r \le s$ *and* $\mathrm{E}\varepsilon_i^4 < \infty$, $\mathrm{E}\varepsilon_i^2 > 0$.

ASSUMPTION 4.2. *The integral* $\int \cdots \int G(c_k - \eta_i(u)|\tau) \prod_{t=1}^s dF_t(u_t|\sigma_t)$ *is twice differentiable w.r.t. components of* θ, *the integrals* $\int H_{i,k,r}(u_r) H_{i',k',r}(u_r) d F_r(u_r|\sigma_r)$, $\int \cdots \int u_r^c G(c_k - \eta_i(u)|\tau) \prod_{t=1}^s dF_t(u_t|\sigma_t)$, $c = 1, 2$ *are differentiable w.r.t. components of* θ, *where* $i, i' \in S^{(l)}$ *for some* $N+1 \le l \le K$; *and* $\int u_r^\lambda dF_r(u_r|\sigma_r)$, $1 \le r \le s$ *and* $\int v^\lambda dG(v|\tau)$, $\lambda = 3, 4$ *are differentiable w.r.t.* σ_r, $1 \le r \le s$ *and* τ, *respectively. The partial derivatives are bounded in a neighborhood of* θ.

ASSUMPTION 4.3. *The following are bounded:* $n^{-1} \sum_{r=1}^a \sum_{l=1}^{m_r} |S_r(l)|^2$,

$$\begin{aligned} &n^{-1/2} \left\| l(N)^{-1} \sum_i P_i[1, p](\theta)^t \right\|, \\ &n^{-1/2} \left\| p(N)^{-1} \sum_i P_i[p+1, p+s+1](\theta)^t \right\|, \end{aligned} \tag{4.4}$$

where $l(N) = \text{diag}(l_j(N), 1 \le j \le p)$ *satisfies* $\|l(N)^{-1}Q_N\| \vee \|Q_N^{-1}l(N)\| = O(1)$*, and, with* P_i *replaced by* $P'_{i,j}$ $(1 \le j \le p+s+1)$*, (4.4) are bounded in a neighborhood of* θ*. Furthermore, the following are* $o(1)$*:* $n^{-1}(N + |\mathscr{I}_1| + |\mathscr{I}_2|)$*,* $n^{-1/2}\max_{1\le r\le a}\max_{1\le l\le m_r}|S_r(l)|$*, and* $n^{-1/2}N[(\min_r p_r(N))^{-2} + (\min_j l_j(N))^{-2}]$.

ASSUMPTION 4.4. *Model* (4.1) *is* AI^4*, and* $\max_l |C_{N,l}| = o(1)$.

Assumption 4.3 specifies orders of quantities that may increase with N, including that of $n = n(N)$. To see that these are reasonable assumptions, consider the following special case of model (4.1): Suppose that m_r, $b < r \le q$ and $n_{i_1,\ldots,i_q}$ are bounded. Note that only m_r, $1 \le r \le b$ are associated with the numbers of levels of the random effects. Furthermore, suppose that $m_r \to \infty$, $1 \le r \le b$ at the same rate, that is, $m_r \sim m_{r'}$, $1 \le r \ne r' \le b$. Let m be their common rate. Then, it is easy to show that $N \sim m^b$, $|\mathscr{I}_1| = O(m^{2b-2})$, $|\mathscr{I}_2| \sim m^{2b-2}$, and, typically, $|S_r(l)| \sim N/m \sim m^{b-1}$. Furthermore, according to Jiang (1996), under regularity conditions, $p_0(N) \sim m^{b/2}$, and $p_r(N) \sim \sqrt{m_r}$, so $p_r(N) \sim m^{1/2}$, if $1 \le r \le a$, and $p_r(N) \sim m^{d_r/2}$, if $a < r \le s$. Also, under regularity conditions, $l(N) = O(\sqrt{N}) = O(m^{b/2})$. Thus, if one chooses n such that $n \sim m^{2b-1}$, these assumptions are either obvious or satisfied under regularity conditions, such as AI^4. The latter implies that there exists w.p. $\to 1$ REML estimator $\hat{\psi}$ which is consistent and asymptotically normal [Jiang (1996)]. Such a REML estimator can be identified, according to Jiang (1997). Define $\hat{c}_\alpha$ $(0 < \alpha < 1)$ the same way as in Section 2, where $\hat{\psi}$ is the REML estimator and $\hat{\beta}$ given by (1.7).

THEOREM 4.1. *Under Assumptions* 4.1 – 4.4, $P(\hat{\chi}^2 > \hat{c}_\alpha) \longrightarrow \alpha$ *as* $N \to \infty$.

PROOF. Again, the procedure is similar to proof of Theorem 2.1. First, assume that the limit (2.7) exists, and show (2.8) holds with N replaced by n. In the following, c represents a positive constant that may have different values at different places. For any $\lambda \in R^M$, write $\lambda_{(N)} = T_N\lambda$. Then,

$$\begin{aligned}\lambda^t(n^{-1/2}T_N^t\xi_N) &= n^{-1/2}\lambda_{(N)}^t\xi_N \\ &= n^{-1/2}\sum_{k=1}^M \lambda_{(N),k}(N_k - \mathrm{E}_\theta N_k) \\ &\quad - n^{-1/2}\sum_{k=1}^M \lambda_{(N),k}(\mathrm{E}_{\hat{\theta}}N_k - \mathrm{E}_\theta N_k).\end{aligned} \tag{4.5}$$

To make the rest of the proof easier to follow, we break it into several steps, each presented in the form of a lemma, all under the conditions of Theorem 4.1.

LEMMA 4.1.

$$n^{-1/2}\sum_{k=1}^{M}\lambda_{(N),k}(N_k-\mathrm{E}_\theta N_k)=\sum_{r=1}^{a}\sum_{l=1}^{m_r}\eta_{N,r,l}+o_P(1), \tag{4.6}$$

where $\eta_{N,r,l}=\sum_{k=1}^{M}\lambda_{(N),k}n^{-1/2}\sum_{i\in S_r(l)}\mathrm{E}(\zeta_{k,i}|\alpha_{r,l})$. □

PROOF. If we define $\delta_{k,i,0}=1_{(y_i\in E_k)}-P(y_i\in E_k|\alpha)$; $\delta_{k,i,1}=d_k(\eta_i)-\mathrm{E}(d_k(\eta_i)|\alpha^{(1)})$; and $\delta_{k,i,2}=\zeta_{k,i}-\sum_{r=1}^{a}\mathrm{E}(\zeta_{k,i}|\alpha_r)$, then,

$$N_k-\mathrm{E}_\theta N_k=\sum_{j=0}^{2}\sum_i\delta_{k,i,j}+\sum_{r=1}^{a}\sum_i\mathrm{E}(\zeta_{k,i}|\alpha_{r,i_r}). \tag{4.7}$$

Note that $\mathrm{E}(\zeta_{k,i}|\alpha_r)=\mathrm{E}(\zeta_{k,i}|\alpha_{r,i_r})$. Given α, the $\delta_{k,i,0}$'s are independent. Thus,

$$\begin{aligned}\mathrm{E}\Big(\sum_i\delta_{k,i,0}\Big)^2&=\mathrm{E}(\mathrm{E}[(\cdots)^2|\alpha])\\&=\sum_i\mathrm{E}(\mathrm{var}(1_{(y_i\in E_k)}|\alpha))\le N/4.\end{aligned} \tag{4.8}$$

If $i_{(r)}\neq i'_{(r)}$, $a+1\le r\le s$, then, given $\alpha^{(1)}$, $\delta_{k,i,1}$ and $\delta_{k,i',1}$ are independent, hence $\mathrm{E}\delta_{k,i,1}\delta_{k,i',1}=\mathrm{E}(\mathrm{E}(\delta_{k,i,1}|\alpha^{(1)})\mathrm{E}(\delta_{k,i',1}|\alpha^{(1)}))=0$. It follows that

$$\mathrm{E}\Big(\sum_i\delta_{k,i,1}\Big)^2=\sum_{(i,i')\in\mathscr{I}_1}\mathrm{E}(\delta_{k,i,1}\delta_{k,i',1})\le|\mathscr{I}_1|. \tag{4.9}$$

If $i_r\neq i'_r$, $1\le r\le a$, $\delta_{k,i,2}$ and $\delta_{k,i',2}$ are independent, hence $\mathrm{E}\delta_{k,i,2}\delta_{k,i',2}=0$. If $i_r=i'_r$ but $i_{r'}\neq i'_{r'}$, $1\le r'\le a$, $r'\neq r$, then, given α_r, $\delta_{k,i,2}$ and $\delta_{k,i',2}$ are independent, hence $\mathrm{E}\delta_{k,i,2}\delta_{k,i',2}=\mathrm{E}(\mathrm{E}(\delta_{k,i,2}|\alpha_r)\mathrm{E}(\delta_{k,i',2}|\alpha_r))=0$. It follows that

$$\mathrm{E}\Big(\sum_i\delta_{k,i,2}\Big)^2=\sum_{(i,i')\in\mathscr{I}_2}\mathrm{E}(\delta_{k,i,2}\delta_{k,i',2})\le c|\mathscr{I}_2|. \tag{4.10}$$

(4.6) then follows by combining (4.7)–(4.10) and Assumption 4.3. □

Note that the sequence $\{\eta_{N,r,l},1\le r\le a,1\le l\le m_r\}$ can be expressed according to the order of $S^{(l)}$ as $\{\eta_{N,l},N+1\le l\le K\}$, where $\eta_{N,l}$ is a function of $\omega_{N,l}$. Let $\eta_{N,l}=0$ if $l\le N$ or $l>K$.

LEMMA 4.2.

$$\begin{aligned}\lambda^t(n^{-1/2}T_N^t\xi_N)&=\sum_{l=N+1}^{K}\eta_{N,l}-d_N^t\omega_N-\omega_N^tB_N\omega_N\\&\quad+\mathrm{E}(\omega_N^tB_N\omega_N)+o_P(1)\\&=\sum_{l=1}^{L}Y_{N,l}+o_P(1),\end{aligned} \tag{4.11}$$

where $B_N = B_N(n^{-1/2}\Phi_N T_N \lambda)$, $B_N(\mu) = \sum_{r=0}^{s} \mu_r V_r(\gamma)/\phi^{1_{(r=0)}} p_r(N)$; $d_N = n^{-1/2}\Psi_N T_N \lambda$; *and* $Y_{N,l} = \eta_{N,l} - d_{N,l}\omega_{N,l} - b_{N,l,l}(\omega_{N,l}^2 - 1) - 2(\sum_{l'<l} b_{N,l,l'}\omega_{N,l'})\omega_{N,l}$.

PROOF. By the Taylor expansion and the conditions, it is easy to show that

$$\begin{aligned} &n^{-1/2}\sum_{k=1}^{M} \lambda_{(N),k}(\mathrm{E}_{\hat\theta} N_k - \mathrm{E}_\theta N_k) \\ &\quad = n^{-1/2}\lambda_{(N)}^t \left(\sum_i \frac{\partial}{\partial \theta^t} p_i(\theta)\right)(\hat\theta - \theta) + o_P(1). \end{aligned} \tag{4.12}$$

Let $u_N^t = n^{-1/2}\lambda_{(N)}^t(\sum_i (\partial/\partial\psi^t) p_i(\theta)) P_N^{-1}$, $v_N^t = n^{-1/2}\lambda_{(N)}^t(\sum_i (\partial/\partial\beta^t) p_i(\theta)) Q_N^{-1}$. Then, by (5.23) of Jiang (1998), we have the asymptotic expansion [note that, under the conditions, both $|u_N|$ and $|v_N|$ are $O(1)$]:

$$\begin{aligned} &n^{-1/2}\lambda_{(N)}^t \left(\sum_i \frac{\partial}{\partial \theta^t} p_i(\theta)\right)(\hat\theta - \theta) \\ &\quad = u_N^t P_N(\hat\psi - \psi) + v_N^t Q_N(\hat\beta - \beta) \\ &\quad = d_N^t \omega_N + \omega_N^t B_N \omega_N - \mathrm{E}(\omega_N^t B_N \omega_N) + o_P(1). \end{aligned} \tag{4.13}$$

(4.11) then follows by (4.5), (4.6), (4.12) and (4.13). □

It is easy to show that

$$\begin{aligned} \sigma_N^2 &= \mathrm{E}\left(\sum_{l=1}^{L} Y_{N,l}\right)^2 \\ &= \sum_{l=1}^{L} \mathrm{var}(\eta_{N,l} - d_{N,l}\omega_{N,l} - b_{N,l,l}\omega_{N,l}^2) + 2\sum_{l\neq l'} b_{N,l,l'}^2 \\ &= \lambda^t T_N^t \Sigma_N T_N \lambda \longrightarrow \lambda^t D \lambda. \end{aligned} \tag{4.14}$$

So, if $\lambda^t D\lambda = 0$, then by (4.11), (4.14),

$$\lambda^t (n^{-1/2} T_N^t \xi_N) \xrightarrow{P} 0 = (\lambda^t D\lambda)^{1/2}\xi, \tag{4.15}$$

where $\xi \sim N(0,1)$. We now deal with the case $\lambda^t D\lambda > 0$.

LEMMA 4.3. *Suppose that* $\lambda^t D\lambda > 0$. *Then,*

$$\sum_{l=1}^{L} Y_{N,l} \xrightarrow{d} N(0, \lambda^t D\lambda). \tag{4.16}$$

PROOF. Let $t_N = (\|B_N\| \vee \|d_N\|)^{-\delta}$, where $0 < \delta < 1/2$. Then, $Y_{N,l} = Y_{N,l}^{(1)} + Y_{N,l}^{(2)}$, where

$$Y_{N,l}^{(1)} = \eta_{N,l} - b_{N,l,l}U_{N,l} - \left(d_{N,l} + 2\sum_{l'<l} b_{N,l,l'}u_{N,l'}\right)u_{N,l},$$

$$Y_{N,l}^{(2)} = -b_{N,l,l}V_{N,l} - 2\Bigg(\sum_{l'<l} b_{N,l,l'}v_{N,l'}\Bigg)u_{N,l} - \Bigg(d_{N,l} + 2\sum_{l'<l} b_{N,l,l'}\omega_{N,l'}\Bigg)v_{N,l},$$

$$U_{N,l} = (\omega_{N,l}^2 - 1)1_{(|\omega_{N,l}|\le t_N)} - \mathrm{E}(\omega_{N,l}^2 - 1)1_{(|\omega_{N,l}|\le t_N)},$$

$V_{N,l} = \omega_{N,l}^2 - 1 - U_{N,l}$; $u_{N,l} = \omega_{N,l}1_{(|\omega_{N,l}|\le t_N)} - \mathrm{E}\omega_{N,l}1_{(|\omega_{N,l}|\le t_N)}$ and $v_{N,l} = \omega_{N,l} - u_{N,l}$. Given the conditions, it is easy to show, as in the proofs of Theorem 5.2 of Jiang (1996) and Theorem A of Jiang (1998) (hereafter, PTT), that $\sum_{l=1}^{L} Y_{N,l}^{(2)} \xrightarrow{L^2} 0$. Note that the conditions imply that $\|d_N\| \to 0$ and $\|B_N\| \to 0$. On the other hand, for a suitable array of σ-fields, $\{Y_{N,l}^{(1)}, 1 \le l \le L\}$ is an array of martingale differences. By Theorem 3.2 of Hall and Heyde [(1980); see the remarks therein following the theorem], to show that $\sum_{l=1}^{L} Y_{N,l}^{(1)} \xrightarrow{d} N(0, \lambda^t D\lambda)$ we need to verify three things: (i) $\max_l |Y_{N,l}^{(1)}| \xrightarrow{P} 0$; (ii) $\sum_l [Y_{N,l}^{(1)}]^2 \xrightarrow{P} \lambda^t D\lambda$; and (iii) $\mathrm{E}(\max_l [Y_{N,l}^{(1)}]^2)$ is bounded. By PTT, it is easy to show (i) and (iii), so it remains to show (ii). Let $Z_{N,l} = \eta_{N,l} - Y_{N,l}^{(1)}$. Then,

$$\sum_{l=1}^{L} [Y_{N,l}^{(1)}]^2 = \sum_{l=1}^{L} \eta_{N,l}^2 - 2\sum_{l=1}^{L} \eta_{N,l}Z_{N,l} + \sum_{l=1}^{L} Z_{N,l}^2. \tag{4.17}$$

By PTT, it can be shown that

$$\sum_{l=1}^{L} Z_{N,l}^2 = \sum_{l=1}^{L} \mathrm{var}(d_{N,l}\omega_{N,l} + b_{N,l,l}\omega_{N,l}^2) + 2\sum_{l\ne l'} b_{N,l,l'}^2 + o_P(1). \tag{4.18}$$

Also, we have $\eta_{N,l}Z_{N,l} = \sum_{j=1}^{3} t_{N,l}^{(j)}$, where $t_{N,l}^{(1)} = \eta_{N,l}(d_{N,l}u_{N,l} + b_{N,l,l}U_{N,l})$, $t_{N,l}^{(2)} = 2(\cdots)(\eta_{N,l}u_{N,l} - \mathrm{E}\eta_{N,l}u_{N,l})$ and $t_{N,l}^{(3)} = 2(\cdots)\mathrm{E}\eta_{N,l}u_{N,l}$. Here $\cdots$ represents $\sum_{l'<l} b_{N,l,l'}u_{N,l'}$. We have

$$\begin{aligned}
&\mathrm{E}\Bigg(\sum_l [t_{N,l}^{(1)} - \mathrm{E}t_{N,l}^{(1)}]\Bigg)^2 \\
&\quad = \sum_l \mathrm{var}(t_{N,l}^{(1)}) \\
&\quad \le \Bigg(\sum_l \mathrm{E}\eta_{N,l}^4\Bigg)^{1/2} \Bigg\{\sum_l \mathrm{E}(d_{N,l}u_{N,l} + b_{N,l,l}U_{N,l})^4\Bigg\}^{1/2} \\
&\quad \le c\Bigg(n^{-2}\sum_{r=1}^{a}\sum_{l=1}^{m_r} |S_r(l)|^4\Bigg)^{1/2} \Big\{\|d_N\|^2 + (\|B_N\| \vee \|d_N\|)^{2(1-2\delta)}\Big\}^{1/2} \\
&\quad = o(1),
\end{aligned}$$

$$\mathrm{E}\Bigg(\sum_l t_{N,l}^{(2)}\Bigg)^2 = 4\sum_l \mathrm{E}(\cdots)^2 \mathrm{var}(\eta_{N,l}u_{N,l})$$

$$\le c\left\{\sum_l\left(\sum_{l'<l} b_{N,l,l'}^2\right)^2\right\}^{1/2}\left(\sum_l \mathrm{E}\eta_{N,l}^4\right)^{1/2}$$
$$\le c\|B_N\|\left(\sum_{l'<l} b_{N,l,l'}^2\right)^{1/2}\left(n^{-2}\sum_{r=1}^{a}\sum_{l=1}^{m_r}|S_r(l)|^4\right)^{1/2}$$
$$= o(1)$$

and, by the proof of Theorem 5.2 of Jiang [(1996), top of page 272], letting $B = (b_{N,l,l'}1_{(l>l')})$, $v = (\mathrm{E}\eta_{N,l}u_{N,l})$,

$$\mathrm{E}\left(\sum_l t_{N,l}^{(3)}\right)^2 \le c\|B^t + B\|\cdot|B|\cdot|v|^2$$
$$\le c\|B_N\|\left(\sum_{l'<l} b_{N,l,l'}^2\right)^{1/2}\left(n^{-1}\sum_{r=1}^{a}\sum_{l=1}^{m_r}|S_r(l)|^2\right)$$
$$= o(1).$$

Therefore, we have

$$\sum_{l=1}^{L}\eta_{N,l}Z_{N,l} = \sum_{l=1}^{L}\mathrm{E}t_{N,l}^{(1)} + o_P(1)$$
$$= \sum_{l=1}^{L}\mathrm{cov}(\eta_{N,l}, d_{N,l}\omega_{N,l} + b_{N,l,l}\omega_{N,l}^2) + o_P(1). \tag{4.19}$$

Finally, it is easy to show that $\sum_l(\eta_{N,l}^2 - \mathrm{E}\eta_{N,l}^2) \xrightarrow{L^2} 0$, thus,

$$\sum_{l=1}^{L}\eta_{N,l}^2 = \sum_{l=1}^{L}\mathrm{var}(\eta_{N,l}) + o_P(1). \tag{4.20}$$

(ii) then follows by combining (4.17)–(4.20). Therefore, (4.16) follows.

By (4.11), (4.16) and recalling (4.15) for the case $\lambda^t D\lambda = 0$, we have, for any λ,

$$\lambda^t(n^{-1/2}T_N^t\xi_N) \xrightarrow{d} (\lambda^t D\lambda)^{1/2}\xi, \tag{4.21}$$

where $\xi \sim N(0,1)$. It follows by the arbitrariness of λ that $n^{-1/2}T_N^t\xi_N \xrightarrow{d} N(0, D)$.

The rest of the proof is exactly the same as the corresponding part of Theorem 2.1 once we establish the boundedness of elements of Σ_N and their partial derivatives. The verifications of the latter are very similar to those in the previous proofs [see Jiang (2001b) for details]. □

Again, we consider a simple example.

EXAMPLE 4.1. Consider a two-way random effects model with crossed classification: $y_{ij} = \mu + u_i + v_j + e_{ij}$, $i = 1,\ldots,m_1$, $j = 1,\ldots,m_2$, where the u_i's

are i.i.d. $\sim F_1$ with density $\sigma_1^{-1} f_1(\cdot/\sigma_1)$; the v_j's are i.i.d. $\sim F_2$ with density $\sigma_2^{-1} f_2(\cdot/\sigma_2)$; the e_{ij}'s are i.i.d. $\sim G$ with density $\tau^{-1} g(\cdot/\tau)$; and u, v and e are independent. Suppose that the fourth moments of the random effects and errors are finite, and σ_1, σ_2 and τ are positive; and the second derivatives of $p_{11k}(\theta)$ w.r.t. θ are bounded in a neighborhood near θ. Furthermore, suppose, for simplicity, that $m_1, m_2 \to \infty$ such that $m_1 \sim m_2$. Then, it is easy to show that Assumptions 4.1 – 4.3 are satisfied with $n = (m_1 m_2)^{3/2}$. Note that, by Jiang (1996), the model is AI[4], because it is balanced and unconfounded with positive variance components, hence it is easy to show that Assumption 4.4 is satisfied. Furthermore, it can be shown that $n^{-1}\Phi_N^t(I_N - R_N)\Phi_N = o(1)$, hence $\Sigma_N \sim n^{-1} H_N$.

5. Remarks and discussion. In the previous three sections we have assumed that p, the dimension of β, is fixed. It follows, by an argument of subsequences (see the last paragraph in the proof of Theorem 2.1) that the same results hold if p is bounded but not necessarily fixed. Furthermore, it is possible to extend the results to the case in which p may increase with the sample size N. However, it is necessary that p increase at a slower rate than N because, otherwise, the asymptotic behavior of our goodness-of-fit test may change. A similar feature has been noticed by Mammen (1996), who studied the asymptotic behavior of empirical process of residuals in linear models with i.i.d. errors [i.e., (1.1) without the α's] when the dimension of the covariates [i.e., p in (1.1)] goes to infinity with the sample size, say, n. It was shown that, in such a case, the asymptotic expansion of the empirical process of the residuals has an extra term which is of the order $p/\sqrt{n}$ [Mammen (1996), Theorem 1]. Thus, if $p/\sqrt{n}$ does not go to 0, the asymptotic behavior of the process is changed and, as Mammen noted, using a goodness-of-fit test developed under fixed p may lead to wrong rejection with high probability of the true error distribution. However, if $p/\sqrt{n} \to 0$, the asymptotic expansion of the process remains the same, therefore the goodness-of-fit test should still apply.

The techniques we use allow us to extend the previous results to the case $p \to \infty$. First, the AI[4] condition does not require that p is fixed or bounded [Jiang (1996)]. Second, the asymptotic results in this paper remain valid under similar restrictions to Mammen (1996) on the rate at which $p \to \infty$. For example, Theorem 2.1 remains valid if $p/\sqrt{N} \to 0$. See Jiang (2001b) for more details.

The normalizing constant n in (1.4) is chosen such that it equals to N for the case of linear regression, it is m for the case of single random effect factor, and it satisfies Assumptions 4.2 – 4.4 for the case of multiple random effect factors. It is clear that the choice of n is not unique. For example, if n satisfies these conditions, so does $2n$, with Σ replaced by $(1/2)\Sigma$. Although this provides flexibility, it also creates uncertainty about what n to use in practice, especially in the case of multiple random effect factors. Therefore, in the following we propose a choice of n that in many cases either uniquely determines n or at least narrows the choices. Note that there are a number of integers that contribute to the total sample size N, for example, m, l in

Example 3.1, and m_1, m_2 in Example 4.1. Usually, n is a function of these integers. We require that n depends on these integers in a way that is as simple as possible. In particular, no unnecessary constant is allowed in the expression of n. We call such a choice of n a *natural choice*. A natural choice of n can be found by examining the leading term in the expression of $H_N + \Delta_N$, where $\Delta_N = 2\Phi_N^t(I_N - R_N)\Phi_N$ (see the definition of Σ_N in Section 3). For example, in Example 3.1, it can be shown that $H_N + \Delta_N = ml^2[\mathrm{Var}(h_1) + o(1)]$. Thus, a natural choice for n is ml^2 (not, e.g., $2ml^2$ or $ml^2 + 1$). Similarly, in Example 4.1, a natural choice for n is $(m_1 m_2)^{3/2}$.

One question that is often associated with χ^2 goodness-of-fit tests is how to choose the number of intervals (or cells) M and the cutoff points c_k, $1 \le k \le M - 1$. Studies on this problem can be traced back to Mann and Wald (1942). One criterion for choosing the optimal M and cutoff points is to maximize the power of the test at certain alternative [e.g., Borovkov (1977), Rayner (1982), Kallenberg, Oosterhoff and Schriever (1985)]. Strictly speaking, when the underlying distribution involves unknown parameters, the theory of such a χ^2 test with the optimal M and cutoff points is more complicated than it might appear. To see this, suppose that one chooses M and c_k's by maximizing the power of the test at a certain alternative, and the alternative distribution involves θ, a vector of unknown parameters. It is easy to see that the optimal choice may depend on θ, that is, $M = M(\theta)$ and $c_k = c_k(\theta)$, $1 \le k \le M(\theta) - 1$. Since θ is unknown, it is customary to replace it by $\hat{\theta}$, an estimator, as in the χ^2 test, so that $M = M(\hat{\theta})$ and $c_k = c_k(\hat{\theta})$, $1 \le k \le M(\hat{\theta}) - 1$. But by doing so, the asymptotic distribution of the test may change, and it is not clear if it is still a weighted χ^2.

This is exactly the situation that we are dealing with. A theoretical investigation of this problem will be left for future study, but here we would like to give some thoughts on how to choose the number of cells for our goodness-of-fit test in practice. Note that our goal is to provide a goodness-of-fit test for checking distributional assumptions about the random effects and errors, and, as in many applications of mixed linear models, no specific alternative is given. In other words, we are mainly interested in *omnibus* rather than *directional* tests [e.g., Rayner and Best (1989), Section 1.1]. Therefore, the first thing we want to make sure is that the test has the right size. In fact, in many cases, a test with the right size also have greater power, in general, than one with smaller-than-expected size. (For example, at 10% level, a test with size 0.10 is often more powerful than one with size .07.) Of course, the choice of cutoff points may also make a difference. One option is "equal probability intervals," but this will again make the c_k's dependent on θ. In addition, as pointed out by Lehmann [(1999), Section 5.7], the equiprobable intervals are mainly for convenience. They are not optimal even for the classic Pearson's χ^2-test (1.3). Thus, to avoid complexity, we shall consider "equal length intervals." More specifically, let $[A, B]$ be an interval that is expected to cover the y_i's (considered as random variables) with high probability. Then, given M, the cutoff points are $c_k = A + (k/M)(B - A)$, $1 \le k \le M - 1$. We now consider the choice of M. First, to ensure good power M should not be too small, because,

roughly speaking, the less cells one has the more difficult it is to distinguish two distributions. On the other hand, if there are too many cells, the size of the test may become a problem. This is because the asymptotic distribution of the test is based on a M-dimensional central limit theorem (CLT). It has been found that, when an M increases with n, the sample size involved in the CLT, the latter may break down. Senatov (1980) studied this problem, and showed that, to ensure the CLT, it is necessary that $M/n^{1/5} \to 0$ [this can be derived from Theorem 1 on page 747 of Senatov (1980), where the author used k instead of M]. Although Senatov's result is for the i.i.d. case, a similar problem is expected to occur in the current situation. Therefore, we propose $n^{1/5}$ as an upper bound for M, where n is the natural choice of n that appears in our M-dimensional CLT (i.e., Theorem 3.1 and 4.1; see the previous remark for the natural choice), and $n = N$ in Theorem 2.1. In conclusion, the best strategy is to choose M as large as possible but not to exceed this upper bound, that is, $M = [n^{1/5}]$, where $[x]$ means the largest integer $\le x$. In the next section we shall consider a simulated example, in which the thoughts given here will be tested.

Sometimes a "nonstandard" problem may be converted into a standard one by making transformations. We illustrate this by an example. Suppose that $y_{ij} = x_{ij}^t\beta + \alpha_i + \varepsilon_{ij}$, $i = 1, \ldots, m$, $j = 1, \ldots, d$, where x_{ij} is a known vector of covariates, β is an unknown vector of regression coefficients; the random effects $\alpha_1, \ldots, \alpha_m$ may not be i.i.d., but the ε_{ij}'s are. Suppose that one is interested in checking the distribution of the errors. This problem is not standard in the sense that the random effects are not i.i.d. (and hence, in particular, do not have a common distribution), and that one is only interested in checking the distribution of the errors. Let $y_i = (y_{ij})_{1\le j\le d}$, $X_i = (x_{ij}^t)_{1\le j\le d}$ and $\varepsilon_i = (\varepsilon_{ij})_{1\le j\le d}$. Then, the model can be written as $y_i = X_i\beta + 1_d\alpha_i + \varepsilon_i$, where 1_d is the d-dimensional vector of 1's. Let $a = (a_j)_{1\le j\le d}$ be a nonzero vector such that $a_\cdot = \sum_{j=1}^d a_j = 0$. Then, we have $z_i = w_i^t\beta + \eta_i$, where $z_i = a^t y_i$, $w_i = X_i^t a$ and $\eta_i = a^t\varepsilon_i$. Thus, the problem reduces to checking the distribution of the η_i's, and the result of Section 2 may be applied. However, some components of β may not be identifiable using the reduced model. Also, by using the reduced model, the information about the random effects α is lost.

Why cell frequencies? The test statistic is in the form

$$n^{-1}\sum_{k=1}^{M}\left(\sum_i [\psi_k(y_i) - \mathrm{E}_{\hat\theta}\psi_k(y_i)]\right)^2,$$

where $\psi_k(y_i) = 1_{(y_i\in E_k)}$. This raises the question: why does $\psi_k(\cdot)$ have to be chosen this way? Although the choice of $\psi_k(\cdot)$ may not be unique, the point is that the expectation $\mathrm{E}_\theta\psi_k(y_i)$ has to be "sensitive" to the distributions of the random effects and errors. For example, $\psi_k(y_i) = y_i$ is not a good choice, because $\mathrm{E}_\theta\psi_k(y_i) = x_i^t\beta$, which has nothing to do with the distributions of the random effects and errors. Therefore, the choice of $\psi_k(\cdot)$ cannot be arbitrary either.

To obtain the critical values of the test one needs to compute the eigenvalues of $\hat{\Sigma} = \Sigma(\hat{\theta})$. In practice, it is often straightforward to evaluate $\hat{\Sigma}$ by Monte-Carlo method. For example, in Section 2,

$$\begin{aligned}\hat{\Sigma} &= N^{-1}\sum_{i=1}^{N}\widehat{\mathrm{Var}(h_{N,i})}\\ &\approx N^{-1}\sum_{i=1}^{N}L^{-1}\sum_{l=1}^{L}(h_{N,i}^{(l)}-\overline{h_{N,i}})(h_{N,i}^{(l)}-\overline{h_{N,i}})^t\\ &= L^{-1}\sum_{l=1}^{L}N^{-1}\sum_{i=1}^{N}(h_{N,i}^{(l)}-\overline{h_{N,i}})(h_{N,i}^{(l)}-\overline{h_{N,i}})^t,\end{aligned}$$

where $h_{N,i}^{(l)}$ is $h_{N,i}$ with θ replaced by $\hat{\theta}$, ε_i by e_l, and y_i by $x_i^t\beta + e_l$, $\overline{h_{N,i}} = L^{-1}\sum_{l=1}^{L}h_{N,i}^{(l)}$ and $e_1,\ldots,e_L$ are i.i.d. random variables generated from $F_0(\cdot\,|\hat{\sigma})$.

Suppose that one applies our goodness-of-fit test and it rejects the null hypothesis. What should one do then? The first thing one can do is to be cautious when making inference about the model. As mentioned in Section 1, distributional assumptions are important to the analysis of mixed linear models. On the other hand, not all parts of the inference about a mixed linear model depend equally heavily on the distributional assumptions. For example, assuming that p is bounded, then consistency of the REML or ML estimators of the fixed effects and variance components does not rely on a normality assumption. Although asymptotic normality of these estimators also does not depend on normality, the asymptotic covariance matrix may be different when normality does not hold [Jiang (1996, 1998)]. Finally, even in a large sample situation, prediction intervals for a mixed effect depend heavily on the distributional assumptions (see the second paragraph of Section 1). Thus, even if the goodness-of-fit test rejects the null hypothesis, one should still be confident about the point estimators, if the sample size is large. One may still use results such as confidence intervals and p-values for the fixed parameters, provided that the asymptotic covariance matrix is modified by estimating the kurtoses of the random effects and errors [see Jiang (2001a)]. As for prediction of the mixed effects, one may still use point predictors such as EBLUP, because its justification does not require normality. However, one should not use a prediction interval for a mixed effect, if it is based on a rejected distributional assumption; and one should not use a Prasad-Rao approximation formula for MSE of EBLUP (see Section 1), if the normality assumption is rejected. Furthermore, one may work on finding distributions that fit the data better. For example, sometimes there is indication that the distribution of the random effects may be heavy-tail. Then, a t-distribution might fit better than a normal one. Of course, whatever new distributions one proposes also need to go through the same goodness-of-fit test.

Finally, although in this paper we consider REML estimators of the parameters, the same asymptotic results hold for MLE as well, provided that p is bounded. This is because, when p is bounded, the ML and REML estimators are asymptotically equivalent [see Jiang (1996)]. In particular, the REML estimators and MLE have the same asymptotic covariance matrix, and the same asymptotic expansion (4.13) holds when the REML estimator of θ is replaced by the MLE.

6. Simulations. In this section, we consider simulations associated with the one-way random effects model of Example 3.1. It is also a simple case of the nested error regression model mentioned in Section 1, which has been used in small area estimation.

The sample size configuration considered in all simulations is $m = 400$ and $l = 4$. We intentionally make l small to mimic the small area situation. The true parameters are chosen as $\mu = \sigma = \tau = 1.0$. Two testing problems are considered. The first is to test a normal hypothesis, that is, H_0: $F = G = \textit{normal}$; the second is to test a nonnormal hypothesis, namely, H_0: $F = G = \textit{logistic}$. In each case, the size of the test based on the simulations is reported. In addition, in each case, the simulated power of the test at two alternatives are reported. The alternatives for the normal case are H_1: $F = t_3$, $G = \textit{normal}$; H_2: $F = G = t_3$; and those for the nonnormal case are H_1: $F = NM(-1.8, 2.2, .45)$, $G = \textit{logistic}$; H_2: $F = NM(-2, 3, .4)$, $G = \textit{logistic}$, where $NM(\mu_1, \mu_2, p)$ represents a mixture of $N(\mu_1, 1)$ and $N(\mu_2, 1)$ with probability $1 - p$ and p, respectively. As can be seen that the alternatives get away from the null as the index increases, and, in particular, in the nonnormal case, the alternatives are asymmetric and the asymmetry increases as p gets away from 1/2. All distributions are suitably scaled so that they have variances specified above. The nominal level of the test is chosen as $\alpha = .10$. The parameters are estimated by REML method.

According to our discussion in Section 5, the best M, the number of cells for the test, is chosen as $[n^{1/5}]$, where $n = ml^2$, that is, $M = 5$. This M is compared with its nearest neighbors, that is, $M = 4$ and $M = 6$. The interval $[A, B]$ is taken as $[-4, 6]$. The asymptotic covariance matrix Σ is computed by the Monte-Carlo method with $L = 10,000$ (see the eighth paragraph of Section 5; note that here $h_{N,i} = h_1, \forall i$). The critical values of the test are obtained by the method described in Section 2. Finally, our goodness-of-fit test is compared with the test of Jiang, Lahiri and Wu (1998) obtained through data-splitting. All the results reported are based on 1000 simulations. In Table 1 and Table 2, GOFT refers to our goodness-of-fit test, while JLW to the test of Jiang, Lahiri and Wu (2001).

Both tests appear to have approximately the right size for all cases of M. The "best choice" of M for GOFT ($= 5$) seems to work well–it has the greatest power in all cases. Our goodness-of-fit test appears to be more powerful than the test of Jiang, Lahiri and Wu (1998), which, of course, is also expected. What might be a little surprising is that the power changes so dramatically even though there is a small change in M. This leads to our concluding remark.

TABLE 1
Size and power for the normal case

	M=4		M=5		M=6	
Hypothesis	**GOFT**	**JLW**	**GOFT**	**JLW**	**GOFT**	**JLW**
H_0	.092	.078	.104	.090	.088	.088
H_1	.309	.230	.895	.627	.767	.637
H_2	.405	.249	1.000	.743	.997	.726

TABLE 2
Size and power for the nonnormal case

	M=4		M=5		M=6	
Hypothesis	**GOFT**	**JLW**	**GOFT**	**JLW**	**GOFT**	**JLW**
H_0	.087	.089	.098	.091	.093	.097
H_1	.224	.081	.993	.833	.473	.406
H_2	.735	.279	.997	.908	.894	.684

In addition to the rules we give in Section 5 about choosing M, in practice, to avoid choosing an M with particularly low power, it is always a good idea to run a small experiment, just as what we do in this simulation, to gain some experience. The estimated parameters, whose consistency does not rely on the distributional assumption being tested [Jiang (1996, 1998)], may be used as the true parameters in such an experiment. Of course, this is only necessary when the null hypothesis is not rejected, and one is concerned with the power of the test against certain alternative which one has in mind.

Acknowledgments. The author is grateful to an Associate Editor and two referees whose constructive comments lead to substantial improvement of the manuscript.

REFERENCES

ARORA, V., LAHIRI, P. and MUKHERJEE, K. (1997). Empirical Bayes estimation of finite population means from complex surveys. *J. Amer. Statist. Assoc.* **92** 1555–1562.

BATTESE, G. E., HARTER, R. M. and FULLER, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc.* **83** 28–36.

BHATIA, R. (1997). *Matrix Analysis*. Springer, New York.

BOROVKOV, A. A. (1977). On the power of the χ^2 test with increasing number of class intervals. *Theory Probab. Appl.* **22** 366–370.

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.

CHERNOFF, H. and LEHMANN, E. L. (1954). The use of maximum-likelihood estimates in χ^2 tests for goodness of fit. *Ann. Math. Statist.* **25** 579–586.

DEMPSTER, A. P. and RYAN, L. M. (1985). Weighted normal plots. *J. Amer. Ststist. Assoc.* **80** 845–850.

GHOSH, M. and RAO, J. N. K. (1994). Small area estimation: An appraisal. *Statist. Sci.* **9** 55–93.

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

GRAY, R. J. and PIERCE, D. A. (1985). Goodness-of-fit tests for censored survival data. *Ann. Statist.* **13** 552–563.

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.

JIANG, J. (1998). Asymptotic properties of the empirical BLUP and BLUE in mixed linear models. *Statist. Sinica* **8** 861–885.

JIANG, J. (2001a). Empirical method of moments and its applications. *J. Statist. Plann. Inference.* To appear.

JIANG, J. (2001b). Goodness-of-fit tests for mixed model diagnostics: details and extensions. Technical report, Dept. Statistics, Case Western Reserve Univ.

JIANG, J., LAHIRI, P. and WU, C. (2001). A generalization of the Pearson's χ^2 goodness-of-fit test with estimated cell frequencies. *Sankhyā Ser. A.* To appear.

KALLENBERG, W. C., ODSTERHOFF, J. and SCHRIEVER, B. F. (1985). The number of classes in chi-squared goodness-of-fit tests. *J. Amer. Statist. Assoc.* **80** 959–968.

LANGE, N. and RYAN, L. (1989). Assessing normality in random effects models. *Ann. Statist.* **17** 624–642.

LEHMANN, E. L. (1999). *Elements of Large-Sample Theory*. Springer, New York.

MAMMEN, E. (1996). Empirical process of residuals for high-dimensional linear models. *Ann. Statist.* **24** 307–335.

MANN, H. B. and WALD, A. (1942). On the choice of the number of intervals in the application of the chi-squared test. *Ann. Math. Statist.* **13** 306–317.

MOORE, D. S. (1978). Chi-square tests. In *Studies in Statistics* (R. V. Hogg, ed.) 66–106. Math. Assoc. America, Washington, DC.

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

PIERCE, D. A. and KOPECKY, K. J. (1979). Testing goodness of fit for the distribution of errors in regression models. *Biometrika* **66** 1–5.

PORTNOY, S. (1984). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. I. *Ann. Statist.* **12** 1298–1309.

PRASAD, N. G. N. and RAO, J. N. K. (1990). The estimation of the mean squared error of small-area estimators. *J. Amer. Statist. Assoc.* **85** 163–171.

RAYNER, J. C. W. (1982). The choice of class probabilities and number of classes for the simple χ^2 goodness of fit test. *Sankhyā Ser. B* **44** 28–37.

RAYNER, J. C. W. and BEST, D. J. (1989). *Smooth Tests of Goodness of Fit*. Oxford Univ. Press.

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

SENATOV, V. V. (1980). Uniform estimates of the rate of convergence in the multi-dimensional central limit theorem. *Theory Probab. Appl.* **25** 745–759.

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
UNIVERSITY OF CALIFORNIA
360 KERR HALL
ONE SHIELDS AVENUE
DAVIS, CALIFORNIA 95616