Biostatistics (2013), **14**, 1, *pp.* 144–159
doi:10.1093/biostatistics/kxs028
Advance Access publication on August 28, 2012

Testing multiple variance components in linear mixed-effects models

REZA DRIKVANDI*

Department of Statistics, Shahid Beheshti University, 1983963113 G.C., Tehran, Iran
r_drikvandi@sbu.ac.ir

GEERT VERBEKE

I-BioStat, Katholieke Universiteit Leuven, Kapucijnenvoer 35, B-3000 Leuven, Belgium; I-BioStat, Universiteit Hasselt, Agoralaan 1, B-3590 Diepenbeek, Belgium

AHMAD KHODADADI

Department of Statistics, Shahid Beheshti University, 1983963113 G.C., Tehran, Iran

VAHID PARTOVI NIA

Department of Mathematics and Industrial Engineering, Ecole Polytechnique de Montreal, Montreal, Canada H3T 1J4

Summary

Testing zero variance components is one of the most challenging problems in the context of linear mixed-effects (LME) models. The usual asymptotic chi-square distribution of the likelihood ratio and score statistics under this null hypothesis is incorrect because the null is on the boundary of the parameter space. During the last two decades many tests have been proposed to overcome this difficulty, but these tests cannot be easily applied for testing multiple variance components, especially for testing a subset of them. We instead introduce a simple test statistic based on the variance least square estimator of variance components. With this comes a permutation procedure to approximate its finite sample distribution. The proposed test covers testing multiple variance components and any subset of them in LME models. Interestingly, our method does not depend on the distribution of the random effects and errors except for their mean and variance. We show, via simulations, that the proposed test has good operating characteristics with respect to Type I error and power. We conclude with an application of our process using real data from a study of the association of hyperglycemia and relative hyperinsulinemia.

Keywords: Likelihood ratio; Linear mixed-effects model; Permutation test; Variance components; Variance least square estimator.

*To whom correspondence should be addressed.

1. Introduction

The linear mixed-effects (LME) model (Laird and Ware, 1982) is well suited for the analysis of longitudinal, clustered, panel, and other correlated data. Given N distinct individuals, the LME model is expressed as

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i, \quad i = 1, \dots, N, \tag{1.1}$$

where $Y_i = (Y_{i1}, \dots, Y_{in_i})'$ is an $n_i \times 1$ vector of repeated measurements on the ith individual, X_i is an $n_i \times m$ design matrix of explanatory variables, β is an $m \times 1$ vector of population parameters called fixed effects, Z_i is an $n_i \times k$ random effects' design matrix, ε_i is an $n_i \times 1$ random error term with independent components; each of them has zero mean and the within-individual variance σ^2, and b_i is a $k \times 1$ vector of random effects with zero mean and covariance matrix $D_* = \sigma^2 D$, where D is a non-negative definite matrix. All random effects and errors are assumed to be mutually independent.

In many applications of LME models, testing for heteroskedasticity, correlation, and variability among individuals or groups can produce a variety of useful applications; for example, in ecological studies, to investigate whether or not population disease rates are homogeneous in different regions; in genetic epidemiology, to study familial aggregation of a disease; and in clinical trials, to test for institutional variation in the effects of therapy on survival (Lin, 1997). All these instances need equivalent testing to account for random effects in the model. From a statistical perspective, testing for the need of random effects translates into testing the hypothesis that all or some of the variance components of D are zero.

Literature on the subject suggests that the likelihood ratio (LR), score, Wald, and F-tests can be used when working with variance components in LME models; see, for example, Seely and El-Bassiouni (1983), Stram and Lee (1994), Lin (1997), Verbeke and Molenberghs (2003), Demidenko (2004), Molenberghs and Verbeke (2007), and Giampaoli and Singer (2009). These tests are all based on the normality assumption for the random effects and errors which may be violated in practice. On the other hand, the usual asymptotic chi-square distribution of the LR and score statistics under the null is incorrect because $D = 0$ is on the boundary of the parameter space. Instead, the large sample distribution is a mixture of chi-square distributions. Determining the weights of this mixture distribution is difficult especially for testing multiple variance components and a subset of them. For more details, see Miller (1977), Self and Liang (1987), Dean (1992), Stram and Lee (1994), Lin (1997), Gueorguieva (2001), Verbeke and Molenberghs (2003), and Fitzmaurice *and others* (2007).

To bypass the issues with testing hypotheses on the boundary of the parameter space, Crainiceanu and Ruppert (2004) introduced an algorithm to simulate the null finite sample distribution of the LR statistic. Unfortunately, their algorithm can only be used for testing a single random effect in LME models. Fitzmaurice *and others* (2007) proposed another method, namely a permutation procedure, for approximating the finite sample distribution of the LR statistic to test a single variance component. Sinha (2009) recently suggested a bootstrap test instead; however, his bootstrap procedure is difficult to apply to testing multiple variance components or a subset of them. Saville and Herring (2009) developed yet another test based on Bayes factors using a Laplace approximation. Their test falters in the sense that it cannot be easily extended to multiple random effects, and relies on the subjective choice of the prior distribution of parameters. Still others have suggested procedures based on Markov chain Monte Carlo methods (Chen and Dunson, 2003; Kinney and Dunson, 2008), but this approach can be time consuming, especially when the number of random effects is large. In this article, we propose a simple permutation test that does not depend on the distribution of the random effects and errors except for their mean and variance. The advantage with our strategy is its easy application to situations with multiple variance components in LME models. But even more importantly, our method handles testing any subset of variance components.

As a motivating example, we considered the plasma inorganic phosphate data of 13 control and 20 obese patients obtained from a study of the association of hyperglycemia and relative hyperinsulinemia (Zerbe, 1979; Zerbe and Murphy, 1986). The objectives of the study were to investigate the changes of plasma level over time and to see whether these changes are treatment-dependent. In assessing the impact of the treatment on plasma level, we had to consider the heterogeneity among patients with respect to the overall mean and evolutions over time. One might expect patients to have different patterns of plasma levels resulting from biological mechanisms or unmeasured covariates that cause different individual profiles over time. This leads to the task of testing whether to include multiple random effects in the model, such as random coefficients for intercept and random coefficients for slopes over time. Further details of these data and associated analysis using our method are described in Section 4.

The rest of the article is organized as follows. In Section 2, we introduce the test statistic and suggest a permutation procedure to obtain its finite sample distribution for testing all variance components and any subset of them in LME models. In Section 3, we present simulation studies designed to evaluate the behavior of the proposed test in different situations and to compare its efficiency with respect to the LR test and the F-test . We then apply our test to the plasma inorganic phosphate data in Section 4. We conclude with a discussion in Section 5.

2. Our method for testing multiple variance components

2.1 *Testing all variance components*

First, we consider testing whether all random effects can be left out of the LME model (1.1), i.e. we wish to test

$$H_0 : D = 0 \tag{2.1}$$

versus the alternative hypothesis that D is a non-zero non-negative definite matrix.

For this purpose, we propose the test statistic

$$T = \frac{1}{N}\,\mathrm{tr}(Z_*(I \otimes \hat{D}_*)Z'_*), \tag{2.2}$$

where $\otimes$ is the Kronecker product, I is the identity matrix, $Z_* = \mathrm{diag}(Z_1, \ldots, Z_N)$, and $\hat{D}_*$ is any distribution-free unbiased estimator of D_*. One can easily show that under H_0, $E(T) = 0$. Thus, H_0 is rejected, if T deviates much from zero.

An appropriate estimator of D_* in (2.2) needs to be employed. Since numerical methods of variance component estimation in LME models are iterative and computationally intensive, we use the variance least square (VLS) estimator of D_*, which has a closed-form expression for estimating D_*. The idea of this method of estimation comes from least squares on squared residuals suggested by Amemiya (1977). For an excellent treatment of the VLS method, see Demidenko (2004). Let $U_* = \mathrm{vec}(D_*)$, where the vec operator is used to represent matrix D_* as a vector by stacking its vector columns. In other words, U_* denotes the $k^2 \times 1$ vector of all elements of matrix D_*. By first defining that $W = (\sum_{i=1}^{N} X'_i X_i)^{-1}$ and $\hat{e}_i = Y_i - X_i\hat{\beta}_{\mathrm{OLS}}$ where $\hat{\beta}_{\mathrm{OLS}} = W\sum_{i=1}^{N} X'_i Y_i$ is the ordinary least squares estimator of β, an unbiased VLS estimator of D_* for the LME model (1.1) can be explicitly derived from the following equation (see Demidenko, 2004, p. 174):

$$\hat{U}_{*\mathrm{VLS}} = \frac{1}{q}\left(\{qH^{-1} + H^{-1}\mathbf{c}\mathbf{c}'H^{-1}\}\sum_{i=1}^{N}(Z'_i\hat{e}_i \otimes Z'_i\hat{e}_i) - H^{-1}\mathbf{c}\sum_{i=1}^{N}\hat{e}'_i\hat{e}_i\right), \tag{2.3}$$

where $\mathbf{c} = \text{vec}(\sum_{i=1}^{N} \{Z_i'Z_i - Z_i'X_iWX_i'Z_i\})$, $q = \sum_{i=1}^{N} n_i - m - \mathbf{c}'H^{-1}\mathbf{c}$, and

$$H = \sum_{i=1}^{N} (Z_i'Z_i \otimes Z_i'Z_i - Z_i'Z_i \otimes Z_i'X_iWX_i'Z_i - Z_i'X_iWX_i'Z_i \otimes Z_i'Z_i)$$
$$+ \left\{\sum_{i=1}^{N} Z_i'X_iW \otimes Z_i'X_iW\right\}\left\{\sum_{i=1}^{N} X_i'Z_i \otimes X_i'Z_i\right\}.$$

For example, in the case of a balanced random-coefficient model, i.e. $Z_i = X_i = X$, the unbiased VLS estimator of D_* is (see Demidenko, 2004, p. 175)

$$\hat{D}_{*\text{VLS}} = \frac{1}{N-1}(X'X)^{-1}X'\Xi X(X'X)^{-1} - (X'X)^{-1}\hat{\sigma}^2, \tag{2.4}$$

where $\Xi = \sum_{i=1}^{N} (Y_i - X\hat{\beta}_{\text{OLS}})(Y_i - X\hat{\beta}_{\text{OLS}})'$ and

$$\hat{\sigma}^2 = \frac{1}{N(n-m)} \sum_{i=1}^{N} Y_i'(I - X(X'X)^{-1}X')Y_i.$$

Interestingly, by substituting (2.4) into (2.2), the test statistic (2.2) for a balanced random-coefficient model is

$$T = \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - X\hat{\beta}_{\text{OLS}})'P_X(Y_i - X\hat{\beta}_{\text{OLS}}) - m\hat{\sigma}^2, \tag{2.5}$$

where $P_X = X(X'X)^{-1}X'$ is the projection matrix onto the column space of X. For the proof of (2.5), see Appendix A.

Note that an appropriate estimator for β based on the generalized least squares method is

$$\hat{\beta} = \left(\sum_{i=1}^{N} X_i'(\hat{\sigma}^2 I + Z_i\hat{D}_*Z_i')^{-1}X_i\right)^{-1}\left(\sum_{i=1}^{N} X_i'(\hat{\sigma}^2 I + Z_i\hat{D}_*Z_i')^{-1}Y_i\right), \tag{2.6}$$

where $\hat{D}_*$ and $\hat{\sigma}^2$ are suitable estimators of D_* and σ^2, respectively. For $\hat{D}_*$ and $\hat{\sigma}^2$, we use, accordingly, the unbiased VLS estimator derived in (2.3) and the unbiased estimator

$$\hat{\sigma}^2 = \frac{\mathbf{Y}'(I - P_S)\mathbf{Y}}{\text{rank}(I - P_S)},$$

in which $\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix}$, $P_S = S(S'S)^{-}S'$, and $S = [X_*, Z_*]$ with $X_* = \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix}$ and $Z_* = \text{diag}(Z_1, \ldots, Z_N)$. Also, an appropriate predictor for the random effects vector b_i is the empirical best linear unbiased predictor (Robinson, 1991)

$$\hat{b}_i = \hat{D}_*Z_i'(\hat{\sigma}^2 I + Z_i\hat{D}_*Z_i')^{-1}(Y_i - X_i\hat{\beta}). \tag{2.7}$$

The exact or asymptotic distribution under the null hypothesis of the test statistic T is needed to test (2.1) in the LME model (1.1). In general, without any distributional assumption for the random effects and errors, finding the exact distribution of T is difficult. On the other hand, even if the asymptotic distribution of T is derived, the sample size may be inadequate to apply the asymptotic result. To overcome such difficulties, we approximate the finite sample distribution of T using a permutation procedure.

Let $Y_i^* = Y_i - X_i\beta$, $i = 1, \ldots, N$, then we have $Y_i^* = Z_i b_i + \varepsilon_i$. Since under the null $Y_i^* = Z_i b_i + \varepsilon_i$ reduces to $Y_i^* = \varepsilon_i$, random vectors Y_i^*s are i.i.d. under H_0. If we define $Y_i^* = (Y_{i1}^*, \ldots, Y_{in_i}^*)'$, $i = 1, \ldots, N$, then under the null Y_{ij}^*s are i.i.d. random variables for each j. We substitute β with $\hat{\beta}$ from (2.6) to get an estimate for Y_{ij}^*, say $\hat{Y}_{ij}^*$. Although $\hat{Y}_{ij}^*$s are not i.i.d. anymore, under the null they are exchangeable random variables for each j; see Appendix B. Exchangeability under the null allows the use of a permutation procedure. We regard $\{\hat{Y}_{ij}^* : i = 1, \ldots, N;\ j = 1, \ldots, n_i\}$ as the original sample for the permutation procedure, where $\hat{Y}_{ij}^*$ denotes the jth adjusted repeated measurement for the ith individual. The permutation procedure approximates the distribution of T through randomly permuting the individual indices of $\hat{Y}_{ij}^*$s for each fixed j. Under H_0, the individual indices are simply random labels and any permutation of the individual indices is equally likely. Specifically, the individual indices of $\hat{Y}_{ij}^*$s are randomly permuted for each j while the number of repeated measurements for each individual are kept fixed. Using this invariance property under H_0, the proposed test can be set up by the following steps:

(1) Compute the test statistic T for the original sample, denoted by T_{obs}.
(2) Randomly permute the individual indices of $\hat{Y}_{ij}^*$s for each j while holding fixed the number of repeated measurements for each individual, and compute the test statistic T for this permutation sample.
(3) Repeat the process B times, giving B test statistics, say T^b, $b = 1, \ldots, B$.
(4) Compute the empirical p-value being the proportion of permutation samples with T^b greater than or equal to T_{obs}.
(5) Given the significance level α, reject H_0 if α is greater than the empirical p-value.

2.2 *Testing a subset of variance components*

In certain situations, testing if a subset of variance components is zero is necessary, i.e. we are interested in testing whether some of the random effects can be omitted while keeping others in the model. For instance, it may be of interest to test for the need of only random intercept in a model involving both random intercept and random slope. Suppose that the LME model (1.1) is rewritten as

$$Y_i = X_i\beta + Z_i^{(1)} b_i^{(1)} + Z_i^{(2)} b_i^{(2)} + \varepsilon_i, \quad i = 1, \ldots, N,$$

where matrices $Z_i^{(1)}$ and $Z_i^{(2)}$ are, respectively, $n_i \times r$ and $n_i \times (k - r)$, and random vectors $b_i^{(1)}$ and $b_i^{(2)}$ are $r \times 1$ and $(k - r) \times 1$, respectively. Also let

$$\operatorname{cov}\left(\begin{bmatrix} b_i^{(1)} \\ b_i^{(2)} \end{bmatrix}\right) = \sigma^2 \begin{bmatrix} D_{11} & D_{12} \\ D_{12}' & D_{22} \end{bmatrix}.$$

In order to test whether the random effects $b_i^{(2)}$ can be omitted while keeping the random effects $b_i^{(1)}$ in the model, we need to test $H_0 : \left[\begin{smallmatrix} D_{11} & 0 \\ 0 & 0 \end{smallmatrix}\right]$ versus $H_1 : \left[\begin{smallmatrix} D_{11} & D_{12} \\ D_{12}' & D_{22} \end{smallmatrix}\right]$. Similar to (2.2), we use the test statistic

$$T = \frac{1}{N} \operatorname{tr}(Z_*^{(2)} (I \otimes \hat{D}_{*22}) (Z_*^{(2)})'), \tag{2.8}$$

where $Z_*^{(2)} = \operatorname{diag}(Z_1^{(2)}, \ldots, Z_N^{(2)})$ and $\hat{D}_{*22}$ is the appropriate block of the matrix $\hat{D}_*$ derived in (2.3).

The previous permutation procedure described in Section 2.1 with $Y_i^* = Y_i - X_i\beta$ cannot be applied anymore to approximate the distribution of the test statistic (2.8), because Y_i^*s are not i.i.d. random vectors under this null hypothesis. Let $Y_i^* = Y_i - X_i\beta - Z_i^{(1)}b_i^{(1)}$; then new Y_i^*s are i.i.d. random vectors under the null. We substitute β and $b_i^{(1)}$ with their estimates from (2.6) and (2.7), respectively, to get an estimate for Y_i^*, say $\hat{Y}_i^*$. Under H_0, new $\hat{Y}_{ij}^*$s are exchangeable random variables for each j (see Appendix B), hence we apply the previous permutation procedure with this new $\hat{Y}_{ij}^*$ to test H_0.

In the next section, we conduct simulation studies to evaluate the behavior of the proposed test, say T-test, in different situations and to compare its efficiency with respect to the LR test and the F-test.

3. Simulation studies

In this section, we summarize simulation studies conducted with the objective of evaluating the behavior of the proposed test. First, we examine the efficiency of the T-test under different distributions for the random effects and errors. Next, we compare the efficiency of the proposed T-test with respect to the LR test and the F-test. We note that the VLS estimator (2.3) is not necessarily non-negative definite. In the simulations, if this estimator is indefinite, we replace it with $\hat{D}_*^+ = P\Lambda_+P'$, where P and Λ are the matrix of eigenvectors and the diagonal matrix of eigenvalues of $\hat{D}_*$, respectively, and also $\Lambda_+ = \max(0, \Lambda)$. It has been shown that $\hat{D}_*^+$ is the closest matrix to $\hat{D}_*$ among all non-negative definite matrices (see Demidenko, 2004, p. 104).

3.1 *Efficiency of the proposed test*

We first considered the linear trend model with random intercepts and random slopes

$$\begin{aligned} &Y_{ij} = a_{1i} + a_{2i}t_{ij} + \varepsilon_{ij}, \\ &a_{1i} = \beta_1 + b_{1i}, \quad a_{2i} = \beta_2 + b_{2i}, \quad i = 1, \ldots, N, \ j = 1, \ldots, n_i, \end{aligned} \tag{3.1}$$

where t_{ij} is the jth observation time for the ith individual, β_1 and β_2 are fixed effects, and b_{1i} and b_{2i} are the random intercept and the random slope, respectively. For this model, a simulation study was performed to investigate the behavior of the proposed method to test for the need of both random intercept b_{1i} and random slope b_{2i} in the model. In the simulations, we set $\beta_1 = 1$, $\beta_2 = 2$, and $t_{ij} = j$, and assumed that $\varepsilon_{ij} \sim N(0, 1)$. Allowing that $D_* = \left[\begin{smallmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{smallmatrix}\right]$, we assumed two types of distributions for the random effects vector $b_i = (b_{1i}, b_{2i})'$: first, a bivariate normal distribution with zero mean and covariance matrix D_* and, second, a bivariate Student's t distribution with a zero mean, degree of freedom df $= 3$, and scale matrix $(\text{df} - 2/\text{df})D_*$. Under each of these two distributions, we generated 1000 Monte Carlo samples from model (3.1) with different values of D_* for $N = 10$, 15 and $n = 3, 5$. Specifically, we set the covariance matrix D_* equal to $\left[\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right]$ (to estimate the size of the test), $\left[\begin{smallmatrix} 0.05 & 0.02 \\ 0.02 & 0.05 \end{smallmatrix}\right]$, $\left[\begin{smallmatrix} 0.08 & 0.02 \\ 0.02 & 0.08 \end{smallmatrix}\right]$, $\left[\begin{smallmatrix} 0.1 & 0.05 \\ 0.05 & 0.1 \end{smallmatrix}\right]$, and $\left[\begin{smallmatrix} 0.1 & 0.09 \\ 0.09 & 0.1 \end{smallmatrix}\right]$ to investigate the empirical power of the test for a significance level of $\alpha = 0.05$. We also selected $B = 1000$ permutation samples for each setting.

The results, displayed in Tables 1 and 2, indicate that the Type I error of the proposed T-test is stable across the two distributions and is very close to the nominal 0.05 level. Furthermore, the power of the test is high even for these small values of N and n. In additional simulations, not reported here, similar results were obtained when a bivariate log-normal distribution was considered for the random effects vector b_i.

Next, we performed a simulation study for the LME model

$$Y_{ij} = \beta_1 + \beta_2 x_{ij} + b_{1i} + b_{2i}z_{ij1} + b_{3i}z_{ij2} + \varepsilon_{ij}, \quad i = 1, \ldots, N, \ j = 1, \ldots, n_i, \tag{3.2}$$

Table 1. *Rejection rates (expressed as percentages) for the 5% level T-test in the LME model (3.1), with $\varepsilon_{ij} \sim N(0, 1)$ and bivariate normal distribution for $(b_{1i}, b_{2i})'$*

	$N=10$		$N=15$	
Covariance matrix of (b_{1i}, b_{2i})	$n=3$	$n=5$	$n=3$	$n=5$
$D_* = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	5.4	4.5	5.8	5.6
$D_* = \begin{bmatrix} 0.05 & 0.02 \\ 0.02 & 0.05 \end{bmatrix}$	16.8	51.5	23.3	67.5
$D_* = \begin{bmatrix} 0.08 & 0.02 \\ 0.02 & 0.08 \end{bmatrix}$	21.9	65.3	30.5	79.4
$D_* = \begin{bmatrix} 0.1 & 0.05 \\ 0.05 & 0.1 \end{bmatrix}$	35.0	75.1	44.7	89.9
$D_* = \begin{bmatrix} 0.1 & 0.09 \\ 0.09 & 0.1 \end{bmatrix}$	37.5	83.8	51.1	94.4

Table 2. *Rejection rates (expressed as percentages) for the 5% level T-test in the LME model (3.1), with $\varepsilon_{ij} \sim N(0, 1)$ and bivariate Student's t distribution for $(b_{1i}, b_{2i})'$*

	$N=10$		$N=15$	
Covariance matrix of (b_{1i}, b_{2i})	$n=3$	$n=5$	$n=3$	$n=5$
$D_* = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	4.8	4.9	5.0	4.5
$D_* = \begin{bmatrix} 0.05 & 0.02 \\ 0.02 & 0.05 \end{bmatrix}$	17.1	41.4	21.0	55.4
$D_* = \begin{bmatrix} 0.08 & 0.02 \\ 0.02 & 0.08 \end{bmatrix}$	18.4	55.4	25.5	69.0
$D_* = \begin{bmatrix} 0.1 & 0.05 \\ 0.05 & 0.1 \end{bmatrix}$	26.2	63.5	35.2	76.7
$D_* = \begin{bmatrix} 0.1 & 0.09 \\ 0.09 & 0.1 \end{bmatrix}$	30.5	69.9	40.6	84.6

where β_1 and β_2 are fixed effects and b_{1i}, b_{2i}, and b_{3i} are three random effects with zero mean and covariance matrix

$$D_* = \begin{bmatrix} d_{11} & d_{12} & d_{13} \\ d_{12} & d_{22} & d_{23} \\ d_{13} & d_{23} & d_{33} \end{bmatrix}. \tag{3.3}$$

We evaluated the efficiency of the proposed method in testing whether the two random effects b_{2i} and b_{3i} can be omitted while keeping the random effect b_{1i} in the model (3.2). This is equivalent to testing $H_0 : d_{11} > 0$, $d_{22} = d_{33} = d_{12} = d_{13} = d_{23} = 0$ versus the alternative hypothesis that $d_{11} > 0$ and at least one another of the other variance components are non-zero.

In the simulations, we assumed $\beta_1 = 1$, $\beta_2 = 2$, and $\varepsilon_{ij} \sim N(0, 1)$. The covariates x_{ij}s, z_{ij1}s, and z_{ij2}s were all generated from $U(0, 1)$. For simplicity in simulations, we assumed that the random effect b_{1i} is

Table 3. *Rejection rates (expressed as percentages) for the 5% level T-test in the LME model (3.2), with* $d_{11} = 1$, $n = 10$, $\varepsilon_{ij} \sim N(0, 1)$, *and multivariate normal distribution for* $(b_{1i}, b_{2i}, b_{3i})'$

Covariance matrix of (b_{2i}, b_{3i})	$N = 7$	$N = 15$	$N = 25$	$N = 50$
$\begin{bmatrix} d_{22} & d_{23} \\ d_{23} & d_{33} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	8.7	4.0	4.6	6.1
$\begin{bmatrix} d_{22} & d_{23} \\ d_{23} & d_{33} \end{bmatrix} = \begin{bmatrix} 0.2 & 0.1 \\ 0.1 & 0.2 \end{bmatrix}$	7.2	8.1	11.6	24.0
$\begin{bmatrix} d_{22} & d_{23} \\ d_{23} & d_{33} \end{bmatrix} = \begin{bmatrix} 0.5 & 0.1 \\ 0.1 & 0.5 \end{bmatrix}$	11.6	25.1	25.5	44.1
$\begin{bmatrix} d_{22} & d_{23} \\ d_{23} & d_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}$	20.8	21.5	59.4	70.2
$\begin{bmatrix} d_{22} & d_{23} \\ d_{23} & d_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$	16.1	40.5	54.6	86.4

Table 4. *Rejection rates (expressed as percentages) for the 5% level T-test in the LME model (3.2), with* $d_{11} = 1$, $n = 10$, $\varepsilon_{ij} \sim N(0, 1)$, *and multivariate Student's t distribution for* $(b_{1i}, b_{2i}, b_{3i})'$

Covariance matrix of (b_{2i}, b_{3i})	$N = 7$	$N = 15$	$N = 25$	$N = 50$
$\begin{bmatrix} d_{22} & d_{23} \\ d_{23} & d_{33} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	3.0	6.7	5.4	5.8
$\begin{bmatrix} d_{22} & d_{23} \\ d_{23} & d_{33} \end{bmatrix} = \begin{bmatrix} 0.2 & 0.1 \\ 0.1 & 0.2 \end{bmatrix}$	5.2	11.4	20.1	25.6
$\begin{bmatrix} d_{22} & d_{23} \\ d_{23} & d_{33} \end{bmatrix} = \begin{bmatrix} 0.5 & 0.1 \\ 0.1 & 0.5 \end{bmatrix}$	7.0	20.4	30.3	43.6
$\begin{bmatrix} d_{22} & d_{23} \\ d_{23} & d_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}$	17.0	29.6	44.5	65.2
$\begin{bmatrix} d_{22} & d_{23} \\ d_{23} & d_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$	18.0	33.0	54.8	71.2

independent of both the random effect b_{2i} and the random effect b_{3i} (i.e. $d_{12} = d_{13} = 0$). In practice, this assumption may be inappropriate. Similar to the previous model, we considered a multivariate normal distribution and a multivariate Student's t distribution for the vector of random effects $b_i = (b_{1i}, b_{2i}, b_{3i})'$. Under each distribution, we generated 1000 Monte Carlo samples from model (3.2) with different values of D_* for $N = 7, 15, 25, 50$ and $n = 10$. We specifically fixed $d_{11} = 1$ and set the variance components $\begin{bmatrix} d_{22} & d_{23} \\ d_{23} & d_{33} \end{bmatrix}$ equal to $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ (to estimate the size of the test), $\begin{bmatrix} 0.2 & 0.1 \\ 0.1 & 0.2 \end{bmatrix}$, $\begin{bmatrix} 0.5 & 0.1 \\ 0.1 & 0.5 \end{bmatrix}$, $\begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}$, and $\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$ to evaluate the empirical power of the test for a significance level of $\alpha = 0.05$. We also selected $B = 1000$ permutation samples for each setting.

The simulation results for the two distributions are presented in Tables 3 and 4, respectively. For small values of N, the Type I error of the test is not very close to the nominal 0.05 level but it gets closer to the nominal level as N increases. Moreover, the power of the test is reasonably large and increases with N, as expected.

3.2 *Comparison between the proposed test and two existing tests*

We compared the efficiency of the proposed T-test with respect to the LR test and the F-test via a simulation study performed for the balanced mixed one-way ANOVA model

$$Y_{ij} = \beta + b_i + \varepsilon_{ij}, \quad i = 1, \ldots, N, \ j = 1, \ldots, n, \tag{3.4}$$

where β is a general mean and b_i a random effect with $E(b_i) = 0$ and $\text{var}(b_i) = \sigma_b^2$.

The hypothesis (2.1) for this model becomes $H_0 : \sigma_b^2 = 0$ versus $H_1 : \sigma_b^2 > 0$, and the test statistic (2.2) simplifies to (see (2.5)),

$$T = \frac{1}{N-1} \sum_{i=1}^{N} n(\bar{Y}_{i.} - \bar{Y}_{..})^2 - \hat{\sigma}^2,$$

where $\bar{Y}_{i.} = \sum_{j=1}^{n} Y_{ij}/n$, $\bar{Y}_{..} = \sum_{i=1}^{N} \bar{Y}_{i.}/N$, and

$$\hat{\sigma}^2 = \frac{1}{N(n-1)} \sum_{i=1}^{N} \sum_{j=1}^{n} (Y_{ij} - \bar{Y}_{i.})^2.$$

Stram and Lee (1994) showed that the asymptotic distribution of the LR statistic, when $N \to \infty$, is the mixture

$$0.5\chi_0^2 + 0.5\chi_1^2,$$

where χ_0^2 is a point mass at 0 and χ_1^2 denotes a chi-square distribution with one degree of freedom.

For this balanced ANOVA model, under normality, there exists an exact F-test on the basis of the test statistic (see Searle *and others*, 1992)

$$F = \frac{\sum_{i=1}^{N} n(\bar{Y}_{i.} - \bar{Y}_{..})^2/(N-1)}{\sum_{i=1}^{N} \sum_{j=1}^{n} (Y_{ij} - \bar{Y}_{i.})^2/N(n-1)}. \tag{3.5}$$

Under H_0, the test statistic (3.5) has F distribution with degrees of freedom $N-1$ and $N(n-1)$. The proposed test statistic T and the test statistic F are interestingly related as

$$T = \hat{\sigma}^2(F - 1).$$

For simulations, we fixed $\beta = 2$ and assumed that $\varepsilon_{ij} \sim N(0, 1)$, thus allowing us to review three types of random effect distributions, normal, t , and log-normal. For normal random effects, we assumed $b_i \sim N\left(0, \sigma_b^2\right)$, and for non-normal random effects, we assumed that $b_i \sim \{(X - E\,(X))\,/\sqrt{\text{var}\,(X)}\} \times \sigma_b$, where the distribution of random variable X is $t(3)$ and log-normal(0,1), so that $\text{var}(b_i) = \sigma_b^2$. We generated 1000 Monte Carlo samples under model (3.4), for different numbers of individuals $N = 7, 15, 25, 50, 100$ and the number of repeated measurements $n = 5$ for each individual and selected $B = 1000$ permutation samples for each setting. First, we set $\sigma_b^2 = 0$ to examine the sizes of the three tests at the significance level of 0.05. The results, presented in Table 5, indicate that both the T-test and the F-test have Type I error rates much closer to the nominal level than the LR test based on the asymptotic (0.5, 0.5) mixture of chi-square distributions. Although the size of the LR test gets closer to the nominal level of 0.05 as the number of individuals, N, gets larger, it still remains under 0.04 even for $N = 100$. In additional simulations, not reported here, we found that the size of the LR test is approximately 0.05 for $N = 200$.

Since the LR test based on the asymptotic (0.5, 0.5) mixture of chi-square distributions has incorrect Type I error rates in finite samples, comparison between the powers of the T-test and the incorrectly sized LR test may be misleading. One needs to correct the size of the LR test to fairly compare the powers in

Table 5. *Type I error rates (expressed as percentages) of the proposed T-test, the LR test based on the asymptotic $(0.5, 0.5)$ mixture of chi-square distributions, and the F-test for the balanced mixed one-way ANOVA model (3.4), with $n = 5$ and $\varepsilon_{ij} \sim N(0, 1)$*

Distribution of b_i		$N = 7$	$N = 15$	$N = 25$	$N = 50$	$N = 100$
Normal	LR	1.7	1.2	2.3	2.1	3.6
	F	5.5	5.3	5.4	4.9	5.6
	T	6.2	5.7	5.2	4.9	5.5
t	LR	1.9	1.5	1.9	2.8	2.4
	F	4.9	4.6	4.8	5.7	4.3
	T	5.7	4.1	5.4	5.8	4.0
Log-normal	LR	1.8	1.9	1.6	2.1	2.4
	F	5.9	5.0	5.5	5.0	4.4
	T	5.5	5.3	5.4	5.4	4.6

Table 6. *Powers (expressed as percentages) of the proposed T-test, the LR test based on the finite sample approximation using Fitzmaurice and others (2007), and the F-test for the balanced mixed one-way ANOVA model (3.4), with $n = 5$ and $\varepsilon_{ij} \sim N(0, 1)$*

Distribution of b_i		$N = 7$	$N = 15$	$N = 25$	$N = 50$
Normal	LR	17.1	19.8	19.0	22.2
	F	17.8	20.9	19.9	22.0
	T	17.3	20.0	19.3	22.4
t	LR	16.7	17.5	19.2	23.9
	F	15.1	16.0	18.1	22.2
	T	16.1	17.6	19.0	23.4
Log-normal	LR	16.8	23.0	21.8	26.7
	F	15.5	19.2	16.8	21.4
	T	15.0	20.1	17.2	22.7

finite samples. Because there is no exact distribution available for the LR statistic, one can approximate its finite sample distribution using a permutation procedure as in Fitzmaurice *and others* (2007). For testing a single variance component, Fitzmaurice *and others* (2007) have shown that this approximation provides Type I error rates close to the nominal level. Thus, in the simulation study we compared the power of the T-test with the power of the correctly sized LR test based on Fitzmaurice *and others* (2007).

According to Fitzmaurice *and others* (2007) and our experience in the simulation study, we varied σ_b^2 from 0.04 to 0.1, using a smaller value of σ_b^2 in simulation configurations with larger individual numbers, N, to compare the powers at the significance level of 0.05. Specifically, we set σ_b^2 equal to 0.1, 0.07, 0.05, and 0.04 for N equal to 7, 15, 25, and 50, respectively. It is straightforward to compute the LR statistic under normality, but for non-normal random effects, evaluating the LR statistic is computationally demanding in a permutation or bootstrap procedure (see Greven *and others*, 2008); therefore, we used $B = 200$ permutation samples for non-normal random effects as in Fitzmaurice *and others* (2007). The results, displayed in Table 6, suggest that, for the normal random effect, all the three tests perform similarly. But, for non-normal random effects, the proposed T-test appears to be more powerful than the F-test, while the LR test based on Fitzmaurice *and others* (2007) appears to be more powerful than both the T-test and the F-test. We conjecture that this result is due to the LR statistic rather than any differences between the two permutation procedures. Although using the LR statistic in a permutation procedure provides a test that is powerful, this approach requires extensive computation, especially for high-dimensional random effects. For instance, in Table 6, note the simulation for the log-normal random effect with $N = 50$ and $B = 200$ permutation samples. For this simulation, the average computation time for one iteration on a server (Six-Core AMD Opteron Processor 2435, 2.6 GHz) was 0.31 s for the proposed T-test and 5188 s for the LR test based on Fitzmaurice *and others* (2007). Note that in our simulations we used the Monte Carlo integration with 500 samples to compute the likelihood function for the LR test. Furthermore, to evaluate the maximum of the log-likelihood function, we utilized the `nlminb` function in `R`.

Overall, the results of the simulations indicate that the proposed test has the correct Type I error rate, and its efficiency is reasonably well in comparison to the LR test and the F-test. Moreover, our proposed method covers testing multiple variance components and any subset of them due to the permutation procedure we used. The LR test is not easily applied to test the two hypotheses we considered in Section 3.1.

4. APPLICATION: THE PLASMA INORGANIC PHOSPHATE DATA

We apply our test to the plasma inorganic phosphate flux data obtained from a study of the association of hyperglycemia and relative hyperinsulinemia performed in the Pediatric Clinical Research Ward of the University of Colorado Medical Center (Zerbe, 1979; Zerbe and Murphy, 1986). In this study, standard glucose tolerance tests were administered to three groups of subjects: 13 controls, 12 non-hyperinsulinemic obese patients, and 8 hyperinsulinemic obese patients. Plasma inorganic phosphate measurements were obtained from blood samples drawn at 0, 0.5, 1, 1.5, 2, 3, 4, and 5 h after a standard-dose oral glucose challenge. The objectives of the study were to investigate the changes of plasma level over time and to see whether these changes are treatment-dependent. The individual profiles are presented in Figure 1 for each group separately. As discussed in Section 1, in assessing the impact of the treatment on the plasma level, the assessment of the heterogeneity among patients with respect to the overall mean and evolutions over time is particularly important. The profiles show that the plasma level exhibits a quadratic response as a function of hours. Thus, the LME model we favor here is of the form (see Verbeke and Molenberghs, 2000, p. 25)

$$Y_{ij} = \begin{cases} (\beta_1 + b_{1i}) + (\beta_2 + b_{2i})t_{ij} + (\beta_3 + b_{3i})t_{ij}^2 + \varepsilon_{ij}, & \text{if control,} \\ (\beta_4 + b_{1i}) + (\beta_5 + b_{2i})t_{ij} + (\beta_6 + b_{3i})t_{ij}^2 + \varepsilon_{ij}, & \text{if non-hyperinsulinemic obese,} \\ (\beta_7 + b_{1i}) + (\beta_8 + b_{2i})t_{ij} + (\beta_9 + b_{3i})t_{ij}^2 + \varepsilon_{ij}, & \text{if hyperinsulinemic obese,} \end{cases} \tag{4.1}$$

where Y_{ij} is the jth plasma level for the ith subject at time t_{ij} (in hours), $\beta = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9)'$ is the vector of fixed-effects parameters, b_{1i} is a random intercept representing the heterogeneity between subjects with respect to baseline values, b_{2i} and b_{3i} are, respectively, a random slope for the linear time effect and a random slope for the quadratic time effect representing the heterogeneity between subjects with respect to evolutions over time, and ε_{ij} is the random error term.

Employing the covariance matrix D_* in (3.3) for the vector of random effects $b_i = (b_{1i}, b_{2i}, b_{3i})'$, the VLS estimator of D_* is

$$\hat{D}_{*\text{VLS}} = \begin{bmatrix} \hat{d}_{11} & \hat{d}_{12} & \hat{d}_{13} \\ \hat{d}_{12} & \hat{d}_{22} & \hat{d}_{23} \\ \hat{d}_{13} & \hat{d}_{23} & \hat{d}_{33} \end{bmatrix} = \begin{bmatrix} 0.377 & -0.078 & 0.009 \\ -0.078 & 0.079 & -0.011 \\ 0.009 & -0.011 & 0.001 \end{bmatrix}.$$

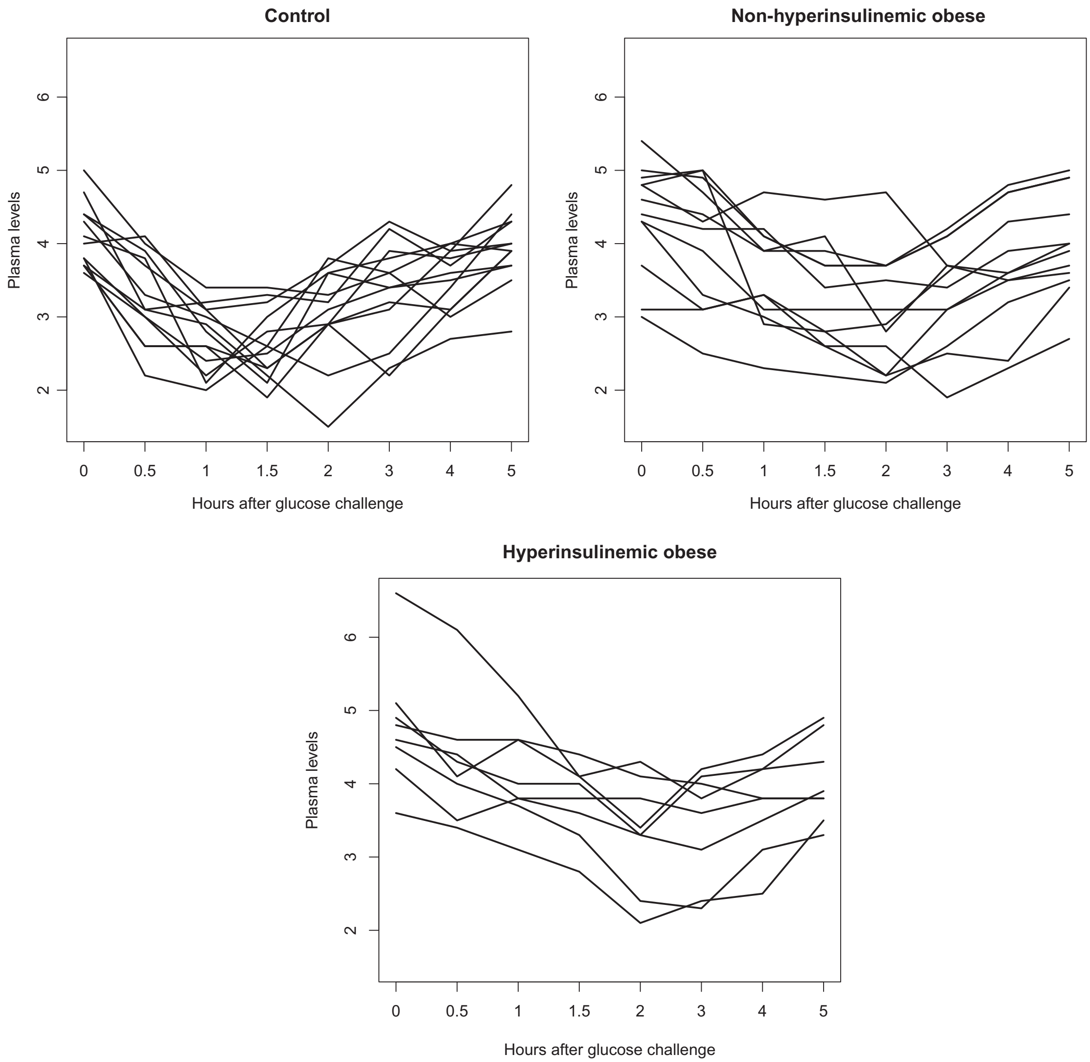

Fig. 1. Individual profiles of control and obese patients in the plasma inorganic phosphate experiment.

In all tests performed in the following, we used 1000 number of permutations. Initially, we determine whether all the three random effects can be left out of the LME model (4.1). The proposed test produces a test statistic of 2.48, giving a p-value of 0.001. The proposed test rejects the null hypothesis at the 5% nominal level, i.e. it allows a random-effects interpretation. Note that both the LR test and the score test cannot be easily used to test this hypothesis.

Next, we examine the baseline heterogeneity between subjects, i.e. we wish to test for the need of only random intercepts in the model, which is equivalent to testing

$$H_0 : \begin{bmatrix} d_{11} & d_{12} & d_{13} \\ d_{12} & d_{22} & d_{23} \\ d_{13} & d_{23} & d_{33} \end{bmatrix} = \begin{bmatrix} d_{11} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

versus the alternative hypothesis that $d_{11} > 0$ and at least one another of the other variance components are non-zero. Our proposed test produces a test statistic of 1.08 with a p-value of 0.035. Thus, the null hypothesis is rejected at the 5% nominal level, i.e. the heterogeneity between subjects with respect to the slopes over time is advocated. The LR and score tests again are not easily applied to test this hypothesis.

In the estimation of model parameters, we attended that the point estimate of d_{33} is tiny ($\hat{d}_{33} = 0.001$). Therefore, we may be interested in testing whether or not the random slope for the quadratic time effect can be omitted from the model, i.e. if the model only requires a random intercept and a random slope for the linear time effect. The corresponding null hypothesis is $H_0: d_{13} = d_{23} = d_{33} = 0$. For testing this hypothesis, our proposed method produces a test statistic of 1.74 with a p-value of 0.649. Assuming normal distribution for both random effects and errors, Stram and Lee (1994) showed that the asymptotic distribution of the LR statistic for testing k versus $k+1$ random effects is a mixture of χ^2_k and χ^2_{k+1}, with equal weights 0.5. Using this result and considering $k = 2$, the p-value of the LR test is a value of 0.22. Hence, both the proposed test and the LR test do not reject this null hypothesis at the 5% nominal level and therefore the random slope for the quadratic time effect can be omitted from the LME model (4.1).

According to the results of our method, a more appropriate model for analyzing the plasma data is

$$Y_{ij} = \begin{cases} (\beta_1 + b_{1i}) + (\beta_2 + b_{2i})t_{ij} + \beta_3 t_{ij}^2 + \varepsilon_{ij}, & \text{if control,} \\ (\beta_4 + b_{1i}) + (\beta_5 + b_{2i})t_{ij} + \beta_6 t_{ij}^2 + \varepsilon_{ij}, & \text{if non-hyperinsulinemic obese,} \\ (\beta_7 + b_{1i}) + (\beta_8 + b_{2i})t_{ij} + \beta_9 t_{ij}^2 + \varepsilon_{ij}, & \text{if hyperinsulinemic obese.} \end{cases}$$

The estimates of variance components of the above model are $\hat{d}_{11} = 0.33$, $\hat{d}_{22} = 0.01$, $\hat{d}_{12} = -0.03$, and $\hat{\sigma}^2 = 0.17$. Also, using (2.6), the estimates of the fixed-effects parameters are $\beta_1 = 3.69(s.e. = 0.13)$, $\beta_2 = -0.72(s.e. = 0.14)$, $\beta_3 = 0.16(s.e. = 0.03)$, $\beta_4 = 4.32(s.e. = 0.12)$, $\beta_5 = -0.86(s.e. = 0.13)$, $\beta_6 = 0.17(s.e. = 0.02)$, $\beta_7 = 4.77(s.e. = 0.14)$, $\beta_8 = -0.94(s.e. = 0.16)$, and $\beta_9 = 0.16(s.e. = 0.03)$, respectively.

5. Discussion

We recommend our approach as a simple way of testing variance components in LME models. Our method is effective in avoiding issues with testing on the boundary of the parameter space, uses a simple test statistic, and alleviates the necessity of any distributional assumptions for the random effects and errors. Our permutation procedure can approximate the finite sample distribution of the test statistic. As an important advantage, our method can further be used to test for multiple variance components and any subset of them. It performs well in a variety of contexts, as illustrated via simulation studies and a real data example. The simulation results suggest that the proposed test has the correct Type I error rate, and its efficiency is reasonably well in comparison to the LR test and the F-test.

Silvapulle and Silvapulle (1995) have shown how a one-sided score test can be defined, both in the scalar as well as in the vector parameter case. They demonstrate that the large sample distribution of the score statistic equals a weighted sum of chi-squared probabilities, but they do not discuss how the weights of this mixture distribution can be calculated in different situations. Although in some particular situations the weights of this mixture distribution are expressible in closed form, Shapiro (1988) shows that for a broad number of cases determining the mixture's weights is a complex and perhaps a numerical task. Verbeke and Molenberghs (2003) used the results of Silvapulle and Silvapulle (1995) and argued that the equivalence of the LR and score tests holds also when testing variance components in LME models. Therefore, it seems that further research is needed to employ score tests for testing multiple variance components or a subset of them.

Finally, in our permutation procedure we utilized the simple test statistic (2.2) rather than the LR statistic because of two reasons: first, computing the likelihood function requires specifying the distribution for both random effects and errors; and second, even after specification of these distributions, evaluating the likelihood is computationally demanding and this is not suitable when performing thousands of permutations with the LR statistic (see Greven *and others*, 2008).

Acknowledgments

Conflict of Interest: None declared.

Funding

Reza Drikvandi would like to thank Professor M. Asgharian and the Department of Mathematics and Statistics, McGill University for their hospitality. His research was partially funded by Prof. Asgharian through his research grant from NSERC, Canada. Vahid Partovi Nia was supported by the NSERC discovery grant 418034-2012.

Appendix A: proof of (2.5)

For $\hat{D}_*$ in (2.4), we have

$$\begin{aligned}
\operatorname{tr}(X\hat{D}_*X') &= \operatorname{tr}\left(\frac{1}{N-1}X(X'X)^{-1}X'EX(X'X)^{-1}X' - X(X'X)^{-1}X'\hat{\sigma}^2\right) \\
&= \operatorname{tr}\left(\frac{1}{N-1}\sum_{i=1}^{N}P_X(Y_i - X\hat{\beta}_{\mathrm{OLS}})(Y_i - X\hat{\beta}_{\mathrm{OLS}})'P_X - P_X\hat{\sigma}^2\right) \\
&= \frac{1}{N-1}\sum_{i=1}^{N}\operatorname{tr}(P_X(Y_i - X\hat{\beta}_{\mathrm{OLS}})(Y_i - X\hat{\beta}_{\mathrm{OLS}})'P_X) - \operatorname{tr}(P_X\hat{\sigma}^2) \\
&= \frac{1}{N-1}\sum_{i=1}^{N}\operatorname{tr}((Y_i - X\hat{\beta}_{\mathrm{OLS}})'P_X(Y_i - X\hat{\beta}_{\mathrm{OLS}})) - m\hat{\sigma}^2 \\
&= \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - X\hat{\beta}_{\mathrm{OLS}})'P_X(Y_i - X\hat{\beta}_{\mathrm{OLS}}) - m\hat{\sigma}^2.
\end{aligned}$$

Hence,

$$\begin{aligned}
T &= \frac{1}{N}\operatorname{tr}(Z_*(I\otimes\hat{D}_*)Z_*') = \frac{1}{N}\sum_{i=1}^{N}\operatorname{tr}(Z_i\hat{D}_*Z_i') = \frac{1}{N}\sum_{i=1}^{N}\operatorname{tr}(X\hat{D}_*X') \\
&= \operatorname{tr}(X\hat{D}_*X') = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - X\hat{\beta}_{\mathrm{OLS}})'P_X(Y_i - X\hat{\beta}_{\mathrm{OLS}}) - m\hat{\sigma}^2,
\end{aligned}$$

and the equation follows.

Appendix B: proof of exchangeability of $\hat{Y}^*_{ij}$s

A finite set of random variables is exchangeable if their joint distribution is the same irrespective of the variables' order. We need to show this property for random variables $\hat{Y}^*_{ij}$s for each j in both testing all variance components and testing a subset of them. First we prove the exchangeability of $\hat{Y}^*_{ij}$s for testing a subset of variance components. Let $b^{(1)} = (b^{(1)}_1, \ldots, b^{(1)}_N)'$, and $\hat{b}^{(1)}$ be the empirical best linear unbiased predictor of $b^{(1)}$ that can be obtained from (2.7). Under H_0, for each fixed j we have

$$\begin{aligned} f(\hat{y}^*_{1j}, \ldots, \hat{y}^*_{Nj}) &= \iint f(\hat{y}^*_{1j}, \ldots, \hat{y}^*_{Nj} | \hat{b}^{(1)}, \hat{\beta})\, \mathrm{d}F(\hat{b}^{(1)}, \hat{\beta}) \\ &= \iint f(\hat{y}^*_{1j}, \ldots, \hat{y}^*_{Nj} | \hat{b}^{(1)}, \hat{\beta})\, \mathrm{d}F(\hat{b}^{(1)} | \hat{\beta})\, \mathrm{d}F(\hat{\beta}) \\ &= \iint \left\{ \prod_{i=1}^{N} f(\hat{y}^*_{ij} | \hat{b}^{(1)}, \hat{\beta}) \right\} \mathrm{d}F(\hat{b}^{(1)} | \hat{\beta})\, \mathrm{d}F(\hat{\beta}), \end{aligned}$$

where the last equality is derived using the fact that under H_0 random variables $\hat{Y}^*_{ij}$s, for each j, are i.i.d. given $\hat{b}^{(1)}$ and $\hat{\beta}$. Hence, $\hat{Y}^*_{ij}$s are exchangeable random variables for each j. The proof of exchangeability for testing all variance components is similar, except that $\mathrm{d}F(\hat{b}^{(1)}|\hat{\beta})$ vanishes, and $f(\hat{y}^*_{ij}|\hat{b}^{(1)}, \hat{\beta})$ changes to $f(\hat{y}^*_{ij}|\hat{\beta})$. Note that we do not permute the covariates X_is and Z_is in the permutation procedure and also keep the number of observations for each individual fixed.

References

Amemiya, T. (1977). A note on a heteroscedastic model. *Journal of Econometrics* **6**, 365–370.

Chen, Z. and Dunson, D. B. (2003). Random effects selection in linear mixed models. *Biometrics* **59**, 762–769.

Crainiceanu, C. M. and Ruppert, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society, Series B* **66**, 165–185.

Dean, C. (1992). Testing for overdispersion in Poisson and binomial regression models. *Journal of the American Statistical Association* **87**, 451–457.

Demidenko, E. (2004). *Mixed Models: Theory and Applications*. New York: Wiley.

Fitzmaurice, G. M., Lipsitz, S. R. and Ibrahim, J. G. (2007). A note on permutation tests for variance components in multilevel generalized linear mixed models. *Biometrics* **63**, 942–946.

Giampaoli, V. and Singer, J. M. (2009). Likelihood ratio tests for variance components in linear mixed models. *Journal of Statistical Planning and Inference* **139**, 1435–1448.

Greven, S., Crainiceanu, C. M., Kuchenhoff, H. and Peters, A. (2008). Restricted likelihood ratio testing for zero variance components in linear mixed models. *Journal of Computational and Graphical Statistics* **17**, 870–891.

Gueorguieva, R. (2001). A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family. *Statistical Modelling* **1**, 177–193.

Kinney, S. K. and Dunson, D. B. (2008). Fixed and random effects selection in linear and logistic models. *Biometrics* **63**, 690–698.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963–974.

Lin, X. (1997). Variance component testing in generalised linear models with random effects. *Biometrika* **84**, 309–326.

Miller, J. J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *The Annals of Statistics* **5**, 746–762.

Molenberghs, G. and Verbeke, G. (2007). Likelihood ratio, score, and Wald tests in a constrained parameter space. *The American Statistician* **61**, 22–27.

Robinson, G. K. (1991). That BLUP is a good thing: estimation of random effects. *Statistical Science* **6**, 15–51.

Saville, B. R. and Herring, A. H. (2009). Testing random effects in the linear mixed model using approximate Bayes factors. *Biometrics* **65**, 369–376.

Searle, S. R, Casella, G. and McCulloch, C. M. (1992). *Variance Components*. New York: Wiley.

Seely, J. F. and El-Bassiouni, Y. (1983). Applying Wald's variance component test. *The Annals of Statistics* **11**, 197–201.

Self, S. G. and Liang, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82**, 605–610.

Shapiro, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis. *International Statistical Review* **56**, 49–62.

Silvapulle, M. J. and Silvapulle, P. (1995). A score test against one-sided alternatives. *Journal of the American Statistical Association* **90**, 342–349.

Sinha, S. K. (2009). Bootstrap tests for variance components in generalized linear mixed models. *The Canadian Journal of Statistics* **37**, 219–234.

Stram, D. O. and Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics* **50**, 1171–1177.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Berlin: Springer.

Verbeke, G. and Molenberghs, G. (2003). The use of score tests for inference on variance components. *Biometrics* **59**, 254–262.

Zerbe, G. O. (1979). Randomization analysis of the completely randomized design extended to growth and response curves. *Journal of the American Statistical Association* **74**, 215–221.

Zerbe, G. O. and Murphy, J. R. (1986). On multiple comparisons in the randomization analysis of growth and response curves. *Biometrics* **42**, 795–804.

[*Received January 16, 2012; revised July 17, 2012; accepted for publication July 18, 2012*]