Testing random effects in linear mixed-effects models with serially correlated errors

Reza Drikvandi [1,2] and **Sajad Noorian*** [3]

[1] Department of Computing and Mathematics, Manchester Metropolitan University, Manchester, UK
[2] Statistics Section, Department of Mathematics, Imperial College London, London, UK
[3] Department of Statistics, Faculty of Science, University of Qom, Qom, Iran

In linear mixed-effects models, random effects are used to capture the heterogeneity and variability between individuals due to unmeasured covariates or unknown biological differences. Testing for the need of random effects is a non-standard problem because it requires testing on the boundary of parameter space where the asymptotic chi-squared distribution of the classical tests such as likelihood ratio and score tests is incorrect. In the literature several tests have been proposed to overcome this difficulty, however all of these tests rely on the restrictive assumption of i.i.d. measurement errors. The presence of correlated errors, which often happens in practice, makes testing random effects much more difficult. In this paper, we propose a permutation test for random effects in the presence of serially correlated errors. The proposed test not only avoids issues with the boundary of parameter space, but also can be used for testing multiple random effects and any subset of them. Our permutation procedure includes the permutation procedure in Drikvandi et al. (2013) as a special case when errors are i.i.d., though the test statistics are different. We use simulations and a real data analysis to evaluate the performance of the proposed permutation test. We have found that random slopes for linear and quadratic time effects may not be significant when measurement errors are serially correlated.

Key words: Correlated errors; Linear mixed-effects model; Longitudinal data; Permutation test; Random effects; Serial correlation;

1 Introduction

Longitudinal, panel, and clustered data arise in medical, economical, and behavioural studies, when a number of individuals or subjects are followed over time and repeated measurements on each individual are recorded at different time points. Linear mixed-effects models are a routine tool for analysing such data when the response variable is continuous. Linear mixed-effects models incorporate subject-specific random effects into the model to capture the heterogeneity and variability between individuals due to unmeasured covariates or unknown biological differences. Furthermore, they account for the serial correlation among the repeated data within individuals by allowing serially correlated errors.

This paper introduces a permutation test for testing random effects in linear mixed-effects models with serially correlated errors. Unlike the existing tests for random effects, our proposed test does not require the measurement errors to be independent and identically distributed (i.i.d.). This makes the proposed test very useful for practical use, also because the case of i.i.d. errors is a special case of our test when there is no serial correlation (see Section 2).

It is important to test for the need of random effects in linear mixed-effects models to decide which random effects should be included or excluded from the model. While several practical examples on testing random effects are given in (Drikvandi et al., 2012, 2013), there are some theoretical and computational reasons why such a test on random effects is important. For example, if unnecessary random effects

* Correspondence: e-mail: s.noorian@qom.ac.ir, Phone: +98-25-32103041, Fax: +98-25-32103041

are included in the model, the parameter estimates will not be efficient. On the other hand, ignoring an important random effect could substantially affect the estimates of parameters including fixed-effects parameters (see, e.g., Heagerty and Kurland (2001); Drikvandi et al. (2017)). Moreover, adding many random effects will result in a more complicated covariance structure and possibly an overparameterised model.

There is a large literature on testing random effects when measurement errors are assumed to be i.i.d.; see, for example, Stram and Lee (1994); Miller (1977); Verbeke and Molenberghs (2003); Crainiceanu and Ruppert (2004); Fitzmaurice et al. (2007); Saville and Herring (2009); Sinha (2009); Giampaoli and Singer (2009); Lee and Braun (2012); Drikvandi et al. (2012); Drikvandi et al. (2013). It is well understood that the main challenge with testing random effects is that the null hypothesis puts the true values of variance components on the boundary of parameter space, and hence the asymptotic chi-squared distribution of the classical tests such as likelihood ratio, Wald, and score tests is incorrect. The correct asymptotic distribution is a mixture of chi-squared distributions whose weights are generally unknown, except for very special cases such as testing a single random effect (see, e.g., Crainiceanu and Ruppert (2004)).

Testing random effects is much more difficult in the presence of correlated errors. The correct asymptotic distribution of the likelihood ratio or score test statistic for testing random effects in the presence of correlated errors is still unknown. We develop a permutation test for testing random effects in the presence of serially correlated errors which avoids the issues with testing on the boundary of parameter space. Beneficially, our permutation test can be used for testing multiple random effects and any subset of them. Our permutation procedure includes the permutation procedure in Drikvandi et al. (2013) as a special case when errors are i.i.d., though the test statistics are different.

While random effects mainly account for the between-individual variability, they also contribute to the within-individual association. However, with the assumption of i.i.d. errors it is not clear how one can model the serial correlation using random effects. On the other hand, as shown in Chi and el (1989), random effects for time effects (i.e. random slopes) may not be significant when measurement errors are serially correlated. In other words, instead of a model with i.i.d. errors and many random effects, it might be more appropriate to use a model with serially correlated errors and a fewer number of random effects. It is also in line with the parsimony principle. Therefore, a test for random effects in the presence of correlated errors is very helpful to find out which random effects should be present in the model.

Finally, we should point out that Baltagi and Li (1995), Baltagi and Wu (1999), Wooldridge (2002), Baltagi et al. (2010) and Montes-Rojas (2010) have suggested tests for random effects and serial correlation within each spatial unit in a panel data model, where they mainly aimed to test whether the error model is $AR(1)$ or not. The test by Baltagi et al. (2010) is particularly useful to check if the errors are serially correlated. However, the panel data model considered in these papers is a special case of linear mixed-effects models since it contains only one random effect (a random intercept).

2 The linear mixed-effects model with serially correlated errors

Given N individuals, the linear mixed-effects model is expressed as (Laird and Ware (1982))

$$y_{it} = x_{it}^t\beta + z_{it}^t b_i + \varepsilon_{it}, \quad i = 1, \ldots, N, \; t = 1, \ldots, n_i, \tag{1}$$

where y_{it} denotes the response for individual i measured at time t, x_{it} is an $m \times 1$ vector of covariates for individual i, β is an $m \times 1$ vector of regression parameters known as fixed effects, z_{it} is a $q \times 1$ vector of random effects' covariates, b_i is a $q \times 1$ vector of random effects following a normal distribution with mean 0 and covariance matrix D, and ε_{it} is the measurement error at time t. The within-individual measurement errors ε_{it}'s are assumed to be serially correlated with order p (i.e., $AR(p)$), that is, for each time point t,

$$\varepsilon_{it} = \sum_{k=1}^{p} \rho_k \varepsilon_{i(t-k)} + w_{it}, \tag{2}$$

where the ρ_k are unknown coefficients (with $|\rho_k| < 1$), and w_{it}'s are independent error terms, each normally distributed with mean 0 and variance σ^2.

Note that the linear mixed-effects model (1) with serially correlated errors in (2) includes the linear mixed-effects model with i.i.d. errors as a special case when $\rho_k = 0$ for each k.

Let $\theta = (\beta^t, vech^t(D), \sigma^2, \rho_k)^t$ represent all unknown parameters in the linear mixed-effects model (1), where $vech(D)$ denotes the vector of $q(q+1)/2$ unique elements of the symmetric matrix D. The normality assumption on the random effects and errors makes the marginal likelihood function of model (1) available in closed-form, which is as follows (see, e.g., Verbeke and Molenberghs (2009))

$$L(\theta) = \prod_{i=1} (2\pi)^{-n_i/2} |Z_i D Z_i^t + \sigma^2 \Gamma_i|^{-1/2} \times \exp\left\{-\frac{1}{2}(Y_i - X_i\beta)^t (Z_i D Z_i^t + \sigma^2 \Gamma_i)^{-1} (Y_i - X_i\beta)\right\}, \quad (3)$$

where $Y_i = (y_{i1}, \ldots, y_{in_i})^t$, $X_i = [x_{i1}, \ldots, x_{in_i}]^t$, $Z_i = [z_{i1}, \ldots, z_{in_i}]^t$, and $\sigma^2\Gamma_i$ is the residual covariance matrix with Γ_i^{-1} whose (i, j)-th element is given by (Galbraith and Galbraith (1974))

$$\gamma_p^{ij} = \sum_{h=0}^{i-1} \rho_h \rho_{h+j-i} - \sum_{h=p+1-j}^{p+i-j} \rho_h \rho_{h+j-i}, \quad 1 \le i \le j \le p,$$

in which $\rho_0 = -1$.

The maximum likelihood estimates of the model parameters can then be obtained using a standard software package like R or SAS. For this, we use PROC MIXED in SAS which allows to specify multiple random effects and serially correlated errors.

3 Testing all random effects in the presence of serially correlated errors

To test whether or not all the random effects b_i can be excluded from the linear mixed-effects model (1), we need to test $H_0 : D = 0$ versus $H_A : D > 0$, where the inequality $D > 0$ means that D is a positive definite matrix.

The marginal likelihood (3) obtained under the normality assumption on the random effects and errors enables us to utilise the likelihood ratio test statistic, which is defined as

$$\lambda = 2[\log L(\hat{\theta}) - \log L(\hat{\theta}_0)], \quad (4)$$

where $L(\hat{\theta}) = \sup\{L(\theta) : \theta \in \Theta\}$ and $L(\hat{\theta}_0) = \sup\{L(\theta) : \theta \in H_0\}$. Large values of λ lead to the rejection of H_0, indicating that the random effects b_i are needed in the model.

Because the null hypothesis puts the true values of variance components on the boundary of parameter space, the asymptotic chi-squared distribution of the likelihood ratio statistic λ under H_0 is incorrect. To the best of our knowledge, the correct asymptotic distribution of λ (or other test statistics like the Wald and score statistics) is not available for linear mixed-effects models with correlated errors. Note that when the measurement errors are assumed to be i.i.d., the correct asymptotic distribution of the likelihood ratio statistic for testing a single random effect is a mixture of two chi-squared distributions (see, for example, Crainiceanu and Ruppert (2004)). But, in general, for testing multiple random effects the asymptotic distribution is not available even with the assumption of i.i.d. errors.

To avoid the challenges with the boundary of parameter space and to overcome the difficulty with testing multiple random effects in the presence of correlated errors, we propose a permutation procedure to approximate the null distribution of the likelihood ratio statistic λ. For this, we first substitute (2) into

(1) to get

$$y_{it} = x_{it}^t\beta + z_{it}^t b_i + \sum_{k=1}^{p} \rho_k \varepsilon_{i(t-k)} + w_{it},$$

which can be rewritten as

$$y_{it} = x_{it}^t\beta + z_{it}^t b_i + \sum_{k=1}^{p} \rho_k [y_{i(t-k)} - x_{i(t-k)}^t\beta - z_{i(t-k)}^t b_i] + w_{it}. \quad (5)$$

Next, we adjust the observations y_{it}'s to be permutable among individuals for each time point t. To do this, by considering (5) we define

$$y_{it}^* = y_{it} - x_{it}^t\beta - \sum_{k=1}^{p} \rho_k [y_{i(t-k)} - x_{i(t-k)}^t\beta], \quad i = 1, \ldots, N, \quad t = 1, \ldots, n_i, \quad (6)$$

and then, from (5) and (6), we obtain that $y_{it}^* = z_{it}^t b_i - \sum_{k=1}^{p} \rho_k z_{i(t-k)}^t b_i + w_{it}$. Now, under the null hypothesis $H_0 : D = 0$, the random effects b_i will be 0 almost surely, and consequently we get $y_{it}^* = w_{it}$ under H_0. Therefore, under the null hypothesis, y_{it}^*'s are i.i.d. random variables and hence they are exchangeable or permutable.

However, y_{it}^* in (6) depends on the unknown parameters β and ρ_k. We replace β and ρ_k by their maximum likelihood estimates to obtain $\hat{y}_{it}^*$ as an estimate of y_{it}^*. Clearly, $\hat{y}_{it}^*$'s are not i.i.d. variables. Nonetheless, the following theorem shows that, under H_0, the adjusted observations $\hat{y}_{it}^*$'s are permutable among individuals for each time point t. Note that Theorem 1 below for serially correlated errors is an extension of the exchangeability proof for i.i.d. errors in Drikvandi et al. (2013).

Theorem 1: *Under the null hypothesis $H_0 : D = 0$, $\hat{y}_{it}^*$'s are exchangeable among individuals for each time point t.*

Proof: To prove the exchangeability of $\hat{y}_{it}^*$'s for each time point t, we need to show that for each t the joint distribution of $\hat{y}_{1t}^*, \ldots, \hat{y}_{Nt}^*$ is the same for any order of the variables. Let $\hat{\beta}$ and $\hat{\rho}_k$ be the maximum likelihood estimators of β and ρ_k, respectively. Then, under H_0, for each t

$$\begin{aligned} f(\hat{y}_{1t}^*, \ldots, \hat{y}_{Nt}^*) &= \int f(\hat{y}_{1t}^*, \ldots, \hat{y}_{Nt}^* | \hat{\beta} = \beta, \hat{\rho}_k = \rho_k) dF_{\hat{\beta},\hat{\rho}_k}(\beta, \rho_k) \\ &= \int \left[\prod_{i=1}^{N} f(\hat{y}_{it}^* | \hat{\beta} = \beta, \hat{\rho}_k = \rho_k) \right] dF_{\hat{\beta},\hat{\rho}_k}(\beta, \rho_k), \end{aligned} \quad (7)$$

where the second equality is obtained since, under H_0, the random variables $\hat{y}_{1t}^*, \ldots, \hat{y}_{Nt}^*$ are i.i.d. given $\hat{\beta} = \beta$ and $\hat{\rho}_k = \rho_k$. From (7), the exchangeability of $\hat{y}_{it}^*$'s holds for each t.

The exchangeability of $\hat{y}_{1t}^*, \ldots, \hat{y}_{Nt}^*$ for each time point t enables us to conduct a permutation test based on the likelihood ratio test statistic (4). By regarding the adjusted observations $\{\hat{y}_{it}^* : i = 1, \ldots, N, \ t = 1, \ldots, n_i\}$ as the original sample, we set up our permutation test for testing $H : D = 0$ as follows:

1. Calculate the likelihood ratio test statistic (4) for the original sample and denote it by λ_{obs}.
2. For $b = 1, \ldots, B$, repeat the following two steps:
   (i) Obtain a permutation sample under H_0 by randomly permuting the individual indices of $\hat{y}_{it}^*$'s for each t.
   (ii) Calculate the likelihood ratio test statistic (4) for the permutation sample obtained in step (i) and

denote it λ^b.

3. Compute the empirical p-value as the proportion of λ^b exceeding λ_{obs}.

4. Given the significance level α, reject H_0 if α is greater than the empirical p-value.

It should be mentioned that, in our permutation procedure, the number of repeated measurements for each individual is kept the same as in the original sample because for each time point t we permute the individual indices only among those individuals that are measured at time t. In fact, the proposed permutation procedure can be easily applied to both balanced and unbalanced data. Furthermore, none of the covariates are permuted.

4 Testing a subset of random effects in the presence of serially correlated errors

It is often of interest to test for a subset of random effects, for example, to test if a random slope for some covariate (e.g., time) is needed while a random intercept is already present in the model. Finding the correct asymptotic distribution of the likelihood ratio statistic for testing a subset of random effects is more complicated.

In model (1), suppose that $b_i = (b_{1i}^t, b_{2i}^t)^t$ and we wish to test whether the set of random effects b_{2i} can be excluded while the random effects b_{1i} are present in the model. Also, let z_{1it} and z_{2it} be the corresponding random effects's design matrices for b_{1i} and b_{2i}, respectively. Then, similar to (5), we can rewrite the linear mixed-effects model (1) as follows

$$y_{it} = x_{it}^t\beta + z_{1it}^t b_{1i} + z_{2it}^t b_{2i} + \sum_{k=1}^{p} \rho_k [y_{i(t-k)} - x_{i(t-k)}^t\beta - z_{1i(t-k)}^t b_{1i} - z_{2i(t-k)}^t b_{2i}] + w_{it}. \quad (8)$$

Let $D = \begin{pmatrix} D_{11} & D_{12} \\ D'_{12} & D_{22} \end{pmatrix}$ be the covariance matrix of $b_i = (b_{1i}^t, b_{2i}^t)^t$. Then, to test if the random effects b_{2i} can be left out while retaining the random effects b_{1i} in the model is equivalent to testing

$$H_0 : D = \begin{pmatrix} D_{11} & 0 \\ 0 & 0 \end{pmatrix} \text{ versus } H_A : D = \begin{pmatrix} D_{11} & D_{12} \\ D_{12}^t & D_{22} \end{pmatrix}.$$

To develop a permutation test for testing the above null hypothesis, we first define (similar to (6))

$$y^{**}_{it} = y_{it} - x_{it}^t\beta - z_{1it}^t b_{1i} - \sum_{k=1}^{p} \rho_k [y_{i(t-k)} - x_{i(t-k)}^t\beta - z_{1i(t-k)}^t b_{1i}], \quad i = 1, \ldots, N, \quad t = 1, \ldots, n_i. \quad (9)$$

Next, from (8) and (9), we obtain that $y^{**}_{it} = z_{2it}^t b_{2i} - \sum_{k=1}^{p} \rho_k z_{2i(t-k)}^t b_{2i} + w_{it}$, which then becomes $y^{**}_{it} = w_{it}$ under the null hypothesis H_0 above. Therefore, y^{**}_{it} are i.i.d. random variables under the null. However, y^{**}_{it} depends on the unknown parameters β and ρ_k as well as the random effects b_{1i} which are unobservable. We replace β and ρ_k by their maximum likelihood estimates and b_{1i} by their predicted values to obtain $\hat{y}^{**}_{it}$ as an estimate of y^{**}_{it}. Although $\hat{y}^{**}_{it}$'s are not i.i.d. variables, the following theorem shows that, under H_0, they are permutable among individuals for each time point t. Again, Theorem 2 below for serially correlated errors is an extension of the Drikvandi et al. (2013) result of i.i.d. errors.

Theorem 2: *Under the null hypothesis* $H_0 : D = \begin{pmatrix} D_{11} & 0 \\ 0 & 0 \end{pmatrix}$, $\hat{y}^{**}_{it}$*'s are exchangeable among individuals for each time point* t.

Proof: Let $\hat{b}_{1i}$ be the prediction of b_{1i}. Under H_0, for each t

$$f(\hat{y}^{**}_{1t}, \ldots, \hat{y}^{**}_{Nt}) = \int f(\hat{y}^{**}_{1t}, \ldots, \hat{y}^{**}_{Nt} | \hat{\beta} = \beta, \hat{\rho}_k = \rho_k, \hat{b}_{1i} = b_{1i}) dF_{\hat{\beta},\hat{\rho}_k,\hat{b}_{1i}}(\beta, \rho_k, b_{1i})$$
$$= \int \left[\prod_{i=1}^{N} f(\hat{y}^{*}_{it} | \hat{\beta} = \beta, \hat{\rho}_k = \rho_k, \hat{b}_{1i} = b_{1i}) \right] dF_{\hat{\beta},\hat{\rho}_k,\hat{b}_{1i}}(\beta, \rho_k, b_{1i}), \quad (10)$$

where the second equality is obtained since, under H_0, the random variables $\hat{y}^{*}_{1t}, \ldots, \hat{y}^{*}_{Nt}$ are i.i.d. given $\hat{\beta} = \beta$ and $\hat{\rho}_k = \rho_k$ and $\hat{b}_{1i} = b_{1i}$. From (10), the exchangeability of $\hat{y}^{**}_{it}$'s holds for each t.

Now, for testing $H_0 : D = \begin{pmatrix} D_{11} & 0 \\ 0 & 0 \end{pmatrix}$, the proposed permutation algorithm in Section 3 can be used but with $\{\hat{y}^{**}_{it} : i = 1, \ldots, N, t = 1, \ldots, n_i\}$ as the original sample.

5 Simulation study

We conducted a simulation study to investigate the performance and properties of the proposed permutation test in testing all random effects as well as in testing a subset of random effects. In the simulations, we considered the following linear mixed-effects model with random intercepts and random slopes and with serially correlated errors:

$$y_{it} = \beta_0 + \beta_1 x_{1it} + \beta_2 x_{2it} + \beta_3 t + b_{0i} + b_{1i} t + \varepsilon_{it}, \quad i = 1, \ldots, N, \; t = 1, \ldots, n_i, \quad (11)$$

where y_{it} is the outcome (e.g., a disease biomarker) for individual i measured at time t, x_{1it} is a time-invariant binary covariate (e.g., treatment indicator), and x_{2it} is a time-varying continuous covariate (e.g., heart rate). In our simulations, we set $\beta_0 = 10$, $\beta_1 = 2$, $\beta_2 = 3$, $\beta_3 = -2$, and generated x_{1it}'s from a Bernoulli distribution with parameter 0.5, while x_{2it}'s were generated from a normal distribution with mean 80 and variance 5. We generated the random effects (b_{0i}, b_{1i}) from two different distributions: first, a bivariate normal distribution with mean 0 and covariance matrix $D = \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix}$ and, second, a bivariate t-distribution with degree of freedom $df = 3$, mean 0 and scale matrix $(df - 2/df)D$. Note that the bivariate t-distribution was considered to evaluate the performance of our permutation test under misspecification of the random-effects distribution since we fit the model assuming the bivariate normal distribution for (b_{0i}, b_{1i}). Also, the measurement errors ε_{it}'s were assumed to be serially correlated following an autoregressive pattern with order 1, i.e. $\varepsilon_{it} = \rho\varepsilon_{i,t-1} + w_{it}$ where we generated w_{it}'s from a normal distribution with mean 0 and variance $\sigma^2 = 1$.

We first focused on testing whether or not both random effects b_{0i} and b_{1i} can be left out from the model (11). For this, we set the covariance matrix D to $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ (to examine the Type I error), $\begin{pmatrix} 0.05 & 0.02 \\ 0.02 & 0.05 \end{pmatrix}$, $\begin{pmatrix} 0.1 & 0.05 \\ 0.05 & 0.1 \end{pmatrix}$, $\begin{pmatrix} 0.2 & 0.1 \\ 0.1 & 0.2 \end{pmatrix}$, and $\begin{pmatrix} 0.5 & 0.1 \\ 0.1 & 0.5 \end{pmatrix}$ to compute the empirical power of our permutation test at significant level $\alpha = 0.05$. Under each of the two above distributions as well as each of these covariance matrices and for each sample size $N = 10, 30, 50$ and with $n_i = 5$ repeated measurements per individual, we generated 500 data sets from the linear mixed-effects model (11) with the autocorrelation parameter ρ set to 0.3, 0.5 and 0.7, respectively. In each simulation replication, we used $B = 200$ permutation samples to perform the permutation test by determining the rejection rates (power) at the nominal level 0.05.

The simulation results for testing for the need of both random intercepts b_{0i} and random slopes b_{1i} are presented in Table 1 (for the bivariate normal distribution) and Table 2 (for the bivariate t-distribution). The results indicate that the Type I error of the proposed permutation test is stable across the two distributions as well as the three values of ρ and is close to the nominal level 0.05. Also, the power of our test is

high for all the three values of ρ, and the test has a remarkable power when $N \geq 30$. Note that the test

Table 1 The power of our permutation test in testing all random effects at the significance level $\alpha = 0.05$ with $n_i = 5$, $\rho = 0.3, 0.5, 0.7$, and with random effects generated from the bivariate normal distribution.

D	$\rho = 0.3$			$\rho = 0.5$			$\rho = 0.7$		
	$N = 10$	$N = 30$	$N = 50$	$N = 10$	$N = 30$	$N = 50$	$N = 10$	$N = 30$	$N = 50$
$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	0.03	0.04	0.05	0.04	0.05	0.05	0.04	0.05	0.05
$\begin{bmatrix} 0.05 & 0.02 \\ 0.02 & 0.05 \end{bmatrix}$	0.06	0.60	0.84	0.05	0.46	0.81	0.05	0.54	0.83
$\begin{bmatrix} 0.1 & 0.05 \\ 0.05 & 0.1 \end{bmatrix}$	0.13	0.86	1.00	0.12	0.82	0.99	0.21	0.85	0.99
$\begin{bmatrix} 0.2 & 0.1 \\ 0.1 & 0.2 \end{bmatrix}$	0.40	0.99	1.00	0.38	0.97	1.00	0.52	0.99	1.00
$\begin{bmatrix} 0.5 & 0.1 \\ 0.1 & 0.5 \end{bmatrix}$	0.67	1.00	1.00	0.68	1.00	1.00	0.77	1.00	1.00

Table 2 The power of our permutation test in testing all random effects at the significance level $\alpha = 0.05$ with $n_i = 5$, $\rho = 0.3, 0.5, 0.7$, and with random effects generated from the bivariate t-distribution.

D	$\rho = 0.3$			$\rho = 0.5$			$\rho = 0.7$		
	$N = 10$	$N = 30$	$N = 50$	$N = 10$	$N = 30$	$N = 50$	$N = 10$	$N = 30$	$N = 50$
$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	0.03	0.04	0.04	0.03	0.04	0.05	0.04	0.05	0.04
$\begin{bmatrix} 0.05 & 0.02 \\ 0.02 & 0.05 \end{bmatrix}$	0.05	0.50	0.73	0.06	0.42	0.69	0.05	0.48	0.71
$\begin{bmatrix} 0.1 & 0.05 \\ 0.05 & 0.1 \end{bmatrix}$	0.12	0.78	0.93	0.11	0.70	0.87	0.16	0.78	0.93
$\begin{bmatrix} 0.2 & 0.1 \\ 0.1 & 0.2 \end{bmatrix}$	0.25	0.96	1.00	0.23	0.95	1.00	0.38	0.96	1.00
$\begin{bmatrix} 0.5 & 0.1 \\ 0.1 & 0.5 \end{bmatrix}$	0.55	0.98	1.00	0.58	0.99	1.00	0.69	1.00	1.00

shows a lower power for the bivariate t-distribution case compared to the correctly-specified case (bivariate normal), however the power loss is not large and the power is high under this misspecification when N is sufficiently large ($N \geq 30$). The reason is that, in our testing procedure, the assumption of normally distributed random effects is only needed for construction of the likelihood ratio test statistic, while our permutation procedure does not require the normality assumption. Overall, for testing all random effects, the power tends to get closer to 1 as the sample size increases, and it reaches 1 even with the sample size of $N = 50$ suggesting that the test is consistent.

We repeated the above simulation for two other covariance matrices: $\begin{bmatrix} 0.2 & 0 \\ 0 & 0.0005 \end{bmatrix}$ and $\begin{bmatrix} 0.0005 & 0 \\ 0 & 0.2 \end{bmatrix}$, in order to evaluate the power of our test in situations where one of the random effects has a very small variance. The results (not shown here) indicate that the power of the test is low (0.06 for $N = 30$ and 0.10 for $N = 50$) when the random slope has a very small variance, while the power is much higher (0.99 for $N = 30$ and 1.00 for $N = 50$) when the random intercept has a very small variance. This suggests that the test is not powerful enough to detect a significant random intercept when the random slope has a very small variance. So, caution is necessary when applying the proposed test to such a situation.

Next, we examined the behaviour of the proposed test for testing a subset of random effects. For this, we considered testing whether or not the random slopes b_{1i} in model (11) can be left out whilst the random intercepts b_{0i} are present in the model. The null hypothesis of this test is $H_0 : d_{22} = d_{12} = 0, d_{11} >$

Table 3 The power of our permutation test in testing a subset of random effects at the significance level $\alpha = 0.05$ with $n_i = 5$, $\rho = 0.3, 0.5, 0.7$, and with random effects generated from the bivariate normal distribution.

	$\rho = 0.3$			$\rho = 0.5$			$\rho = 0.7$		
D	$N = 10$	$N = 30$	$N = 50$	$N = 10$	$N = 30$	$N = 50$	$N = 10$	$N = 30$	$N = 50$
$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$	0.03	0.04	0.04	0.03	0.05	0.04	0.03	0.04	0.05
$\begin{bmatrix} 1 & 0 \\ 0 & 0.05 \end{bmatrix}$	0.11	0.25	0.31	0.12	0.19	0.30	0.09	0.27	0.35
$\begin{bmatrix} 1 & 0 \\ 0 & 0.1 \end{bmatrix}$	0.22	0.54	0.66	0.20	0.49	0.61	0.21	0.53	0.73
$\begin{bmatrix} 1 & 0 \\ 0 & 0.2 \end{bmatrix}$	0.34	0.88	0.93	0.30	0.86	0.95	0.34	0.94	0.96
$\begin{bmatrix} 1 & 0 \\ 0 & 0.5 \end{bmatrix}$	0.68	0.99	1.00	0.70	1.00	1.00	0.72	1.00	1.00

Table 4 The power of our permutation test in testing a subset of random effects at the significance level $\alpha = 0.05$ with $n_i = 5$, $\rho = 0.3, 0.5, 0.7$, and with random effects generated from the bivariate t-distribution.

	$\rho = 0.3$			$\rho = 0.5$			$\rho = 0.7$		
D	$N = 10$	$N = 30$	$N = 50$	$N = 10$	$N = 30$	$N = 50$	$N = 10$	$N = 30$	$N = 50$
$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$	0.03	0.04	0.04	0.03	0.04	0.05	0.04	0.06	0.05
$\begin{bmatrix} 1 & 0 \\ 0 & 0.05 \end{bmatrix}$	0.11	0.18	0.25	0.09	0.18	0.26	0.10	0.22	0.28
$\begin{bmatrix} 1 & 0 \\ 0 & 0.1 \end{bmatrix}$	0.13	0.39	0.64	0.12	0.43	0.52	0.15	0.43	0.63
$\begin{bmatrix} 1 & 0 \\ 0 & 0.2 \end{bmatrix}$	0.26	0.76	0.85	0.25	0.75	0.86	0.30	0.87	0.88
$\begin{bmatrix} 1 & 0 \\ 0 & 0.5 \end{bmatrix}$	0.55	0.96	1.00	0.59	0.97	1.00	0.67	0.98	1.00

0. However, for simplicity in the simulations, we here assumed that the random effects b_{0i} and b_{1i} are independent, implying that $d_{12} = 0$. Note that this assumption may not be realistic in practice and we would not make such an assumption in our real data analysis in the next section. We fixed $d_{11} = 1$ for the random intercepts b_{0i}, and then varied d_{22} from 0 to 0.5 to examine the power of our test in detecting significant random slopes b_{1i}. The simulation results are reported in Table 3 (for the bivariate normal) and Table 4 (for the bivariate t-distribution). It can be seen that the Type I error rate of our permutation test is close to the nominal 0.05 level across the the two distributions as well as the three values of ρ. Also, the test shows a reasonably high power for all the three values of ρ, and the power increases rapidly when the variance component d_{22} or the sample size N increases. Again, the test shows a lower power for the bivariate t-distribution case compared to the correctly-specified case (bivariate normal), however the power loss is not large when N is sufficiently large ($N \geq 30$).

Further simulations (not reported here) showed that the performance of the test under negative serial correlation values is very similar to the results for the positive ones. Also, by increasing the number of repeated measurements to $n_i = 10$, we observed a generally higher power for our permutation test.

6 Real data example

In this section, we apply the proposed permutation 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 Centre (Zerbe (1979); Zerbe and Murphy (1986)). In the study, standard glucose tolerance tests were administered to three groups of patients: 13 controls, 12 non-hyperinsulinemic obese patients, and 8 hyperinsulinemic obese patients. For each patient, plasma inorganic phosphate measurements were obtained from blood samples taken at 0, 0.5, 1, 1.5, 2, 3, 4, and 5 hours after the glucose challenge. The main objectives were to investigate the changes of plasma level over time and to see whether these changes are treatment-dependent.

From the individual profiles presented in Figure 1, there is a high variability (at baseline and over time) between patients within each group, and further the plasma level exhibits a quadratic response as a function of time. Drikvandi et al. (2013) considered a linear mixed-effect model with linear and quadratic time effects to analyse this dataset. They assumed that the measurement errors are i.i.d., but Chi and el (1989) have shown that there is a significant autocorrelation in the within-individual measurement errors. Therefore, we here consider the following linear mixed-effects model with serially correlated errors:

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

where y_{it} is the plasma level for patient i measured at time t (in hours), the β_l ($l = 1, \ldots, 9$) are fixed-effects parameters, the b_{1i} are random intercepts representing the baseline heterogeneity between patients, the b_{2i} are random slopes representing the heterogeneity between patients over time, and finally the ε_{ij} are measurement errors following an $AR(1)$ process (acording to Chi and el (1989)).

In the above model, there exists no random effect for quadratic time effects because the estimate of its variance component is 0 and consequently it is not included in the model. The maximum likelihood estimate of the covariance matrix of random effects $b_i = (b_{1i}, b_{2i})^t$ in the model, obtained using PROC MIXED, is given by

$$\hat{D}_{ML} = \begin{bmatrix} 0.232 & -0.009 \\ -0.009 & 0.0004 \end{bmatrix}.$$

Note that the existence of autocorrelation in the within-individual measurement errors can also be confirmed from the estimate of the autocorrelation parameter which is 0.51 with a standard error of 0.08 (p-value < 0.0001).

We use 1000 permutation samples in all tests performed in the following. We first test whether all random effects can be removed from the model for the plasma data. For this, our permutation test with a test statistic of 8.52 produces a p-value of 0.10, suggesting that the random effects can be removed from the model. But as shown in the simulations, because the random slope has a very small variance, the test might not have power to detect significant random intercepts. Therefore, we need to test whether or not the random slope b_{2i} can be removed from the model whilst the random intercept b_{1i} is present in the model. For this, the proposed test with a test statistic being equal to 0.43 gives a p-value of 0.53. Hence, the random slope b_{2i} is not significant and should be removed from the model.

The next step is to test whether or not the random intercept b_{1i} is significant. The permutation test with a test statistic of 8.09 produces a p-value of 0.001, confirming that the random intercepts are needed in the model. So, the results of our permutation test suggest that a more appropriate model for the plasma data would be as follows

$$y_{it} = \begin{cases} \beta_1 + \beta_2 t + \beta_3 t^2 + b_{1i} + \varepsilon_{ij}, & \text{if control} \\ \beta_4 + \beta_5 t + \beta_6 t^2 + b_{1i} + \varepsilon_{ij}, & \text{if non-hyperinsulinemic obese} \\ \beta_7 + \beta_8 t + \beta_9 t^2 + b_{1i} + \varepsilon_{ij}, & \text{if hyperinsulinemic obese,} \end{cases} \quad (12)$$

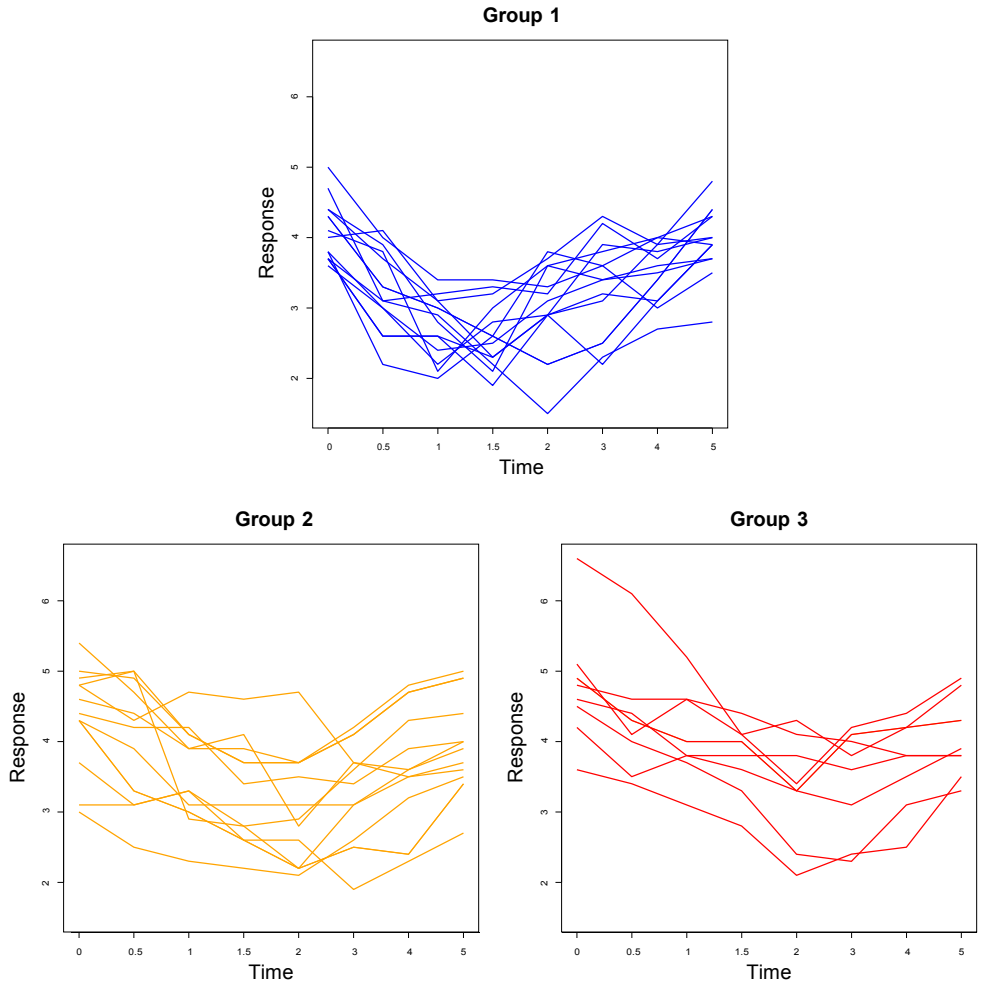

Figure 1 Individual profiles of control and obese patients in the plasma inorganic phosphate experiment.

in which the measurement errors ε_{ij}'s are serially correlated with an $AR(1)$ pattern. The main conclusion here is that the random slopes for linear and quadratic time effects are not needed when measurement errors are serially correlated. This result is also in accordance with the findings of Chi and el (1989), though the validity of the random-effects part is now confirmed using our formal test of random effects in the presence of serially correlated errors. For the final model (12), the maximum likelihood estimates of parameters along with associated standard errors are calculated and reported in Table 5.

7 Discussion

We developed a permutation test for testing random effects in the presence of serially correlated errors, which can also be used when the errors are independent. Our permutation procedure includes the permutation procedure in Drikvandi et al. (2013) as a special case when errors are i.i.d. (i.e., $\rho_k = 0$), though the

Table 5 Plasma data: the maximum likelihood estimates of parameters and associated standard errors obtained from the final model (12) fitted using PROC MIXED in SAS.

Parameter	Estimate	Standard error
Fixed effects:		
β_1	3.840	0.179
β_2	− 0.781	0.119
β_3	0.162	0.022
β_4	4.290	0.186
β_5	− 0.789	0.124
β_6	0.148	0.023
β_7	4.754	0.228
β_8	− 0.908	0.152
β_9	0.154	0.028
Residual variance:		
σ^2	0.251	0.040
Autocorrelation:		
ρ	0.512	0.078
Variance component of b_{1i}:		
d	0.190	0.069
− 2 log-likelihood	355.1	

test statistics are different. The permutation test avoids issues with the boundary of parameter space and can be applied for testing all random effects and any subset of them.

The simulations suggested that the proposed permutation test has Type I error rate close to the nominal level and produces a high power in detecting significant random effects. The power of the test increases rapidly when sample size or variance components increase. Also, in our simulations, the permutation test appears to show a reasonably high power for the different values of the autocorrelation within-individual measurement errors.

We also found that the test has a good power to detect significant random slopes when the random intercept has a very small variance, but it is not powerful enough to detect a significant random intercept when the random slope has a very small variance. So, caution is needed when applying the proposed test to situations where random slopes have very small variances.

Our real data analysis showed that random slopes for linear and quadratic time effects may not be needed when measurement errors are serially correlated. This result is in accordance with the results of Chi and el (1989) and also in line with the parsimony principle.

In our simulations we considered $AR(1)$ errors, however the proposed test can be easily used for autoregressive errors with any order as the permutation procedures in Sections 3 and 4 were developed with $AR(p)$ errors.

Finally, we used the likelihood ratio test statistic which requires some distributional assumption (often normality) on the random effects and errors. However, our permutation procedure works with any other test statistics, especially those obtained from distribution-free methods though power loss in anticipated with distribution-free test statistics. It would be useful to check the normality assumption on random effects before applying the proposed test. Drikvandi et al. (2017) have developed a diagnostic tool for assessing the random-effects distribution which can be applied to mixed models with multiple random effects and correlated errors (see also Drikvandi, 2017).

References

Baltagi, B. H., B. C. Jung, and S. H. Song (2010). Testing for heteroskedasticity and serial correlation in a random effects panel data model. *Journal of Econometrics 154*, 122–124.

Baltagi, B. H. and Q. Li (1995). Testing AR(1) against MA(1) disturbances in an error component model. *Journal of Econometrics 68*(1), 133–151.

Baltagi, B. H. and P. X. Wu (1999). Unequally spaced panel data regressions with AR(1) disturbances. *Econometric Theory 15*(6), 814–823.

Chi, E. M. and G. C. el (1989). Models for longitudinal data with random effects and AR(1) errors. *Journal of the American Statistical Association 84*, 452–459.

Crainiceanu, C. M. and D. Ruppert (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66*(1), 165–185.

Drikvandi, R. (2017). Nonlinear mixed-effects models for pharmacokinetic data analysis: assessment of the random-effects distribution. *Journal of pharmacokinetics and pharmacodynamics 44*(3), 223–232.

Drikvandi, R., A. Khodadadi, and G. Verbeke (2012). Testing variance components in balanced linear growth curve models. *Journal of Applied Statistics 39*(3), 563–572.

Drikvandi, R., G. Verbeke, A. Khodadadi, and V. Partovi Nia (2013). Testing multiple variance components in linear mixed-effects models. *Biostatistics 14*(1), 144–159.

Drikvandi, R., G. Verbeke, and G. Molenberghs (2017). Diagnosing misspecification of the random-effects distribution in mixed models. *Biometrics 73*, 63–71.

Fitzmaurice, G. M., S. R. Lipsitz, and J. G. Ibrahim (2007). A note on permutation tests for variance components in multilevel generalized linear mixed models. *Biometrics 63*(3), 942–946.

Galbraith, R. F. and J. I. Galbraith (1974). On the inverses of some patterned matrices arising in the theory of stationary time series. *Journal of Applied Probability 11*, 63–71.

Giampaoli, V. and J. M. Singer (2009). Likelihood ratio tests for variance components in linear mixed models. *Journal of Statistical Planning and Inference 139*(4), 1435–1448.

Heagerty, P. J. and B. F. Kurland (2001). Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika 88*, 973–985.

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

Lee, O. E. and T. M. Braun (2012). Permutation tests for random effects in linear mixed models. *Biometrics 68*, 486–493.

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.

Montes-Rojas, G. V. (2010). Testing for random effects and serial correlation in spatial autoregressive models. *Journal of Statistical Planning and Inference 140*(4), 1013–1020.

Saville, B. R. and A. H. Herring (2009). Testing random effects in the linear mixed model using approximate bayes factors. *Biometrics 65*(2), 369–376.

Sinha, S. K. (2009). Bootstrap tests for variance components in generalized linear mixed models. *Canadian Journal of Statistics 37*(2), 219–234.

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

Verbeke, G. and G. Molenberghs (2003). The use of score tests for inference on variance components. *Biometrics 59*(2), 254–262.

Verbeke, G. and G. Molenberghs (2009). *Linear mixed models for longitudinal data*. Springer Science & Business Media.

Wooldridge, J. (2002). Econometric analysis of cross section and panel data. econometric analysis of cross section and panel data.

Zerbe, G. O. (1979). Randomization analysis of the completely randomized design extended to growth and response curves. *Journal of the American Statistical Association 74*(365), 215–221.

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