مجلة البحوث الإدارية والمالية والكمية

Journal of Managerial, Financial & Quantitative Research

جامعه السويس
كلية التجارة

Testing Random Effects in Linear Mixed Regression Models

By

Sandy Safwat Faiek Nakhla
Teaching Assistant of Statistics at the Faculty of Commerce
Department of Quantitative Sciences and Computer
Suez University

Prof. Mohammed Gomaa ElRouby
Professor of Statistics
at the Faculty of Commerce
Department of Quantitative Sciences and Computer
Suez University

Dr. Ayman Ahmed Amin Abdellah
Associate Professor of Statistics
at the Faculty of **Commerce**
Department of Insurance, Mathematics, and Statistics
Menoufya University

Journal of Managerial, Financial and Quantitative Research
Vol. 5 No. 3 September 2025

Website: https://safq.journals.ekb.eg/

Testing Random Effects in Linear Mixed Regression Models

Abstract:

Linear mixed models are extensions of regression models in which data have a hierarchical structure with units nested in clusters (Snijders and Bosker, 1999). They are multilevel models where the data are in the form of groups or clusters, Accounting for the correlation that might exist within group elements and that is why linear mixed models require a special treatment rather than the standard regression. It is essential to test for the need for the existence of the random effect in the model as it will be power-consuming to estimate an unneeded parameter, so it will be helpful to test whether the Linear mixed models are a better procedure. This article is designed to represent a comparison using a simulation study to examine the performance of five tests that aimed to test the existence of the random effect. This comparison depends on comparing those tests in different settings, varying the number of groups and number of observations among those groups, to detect the best performance under small size samples, also Varying the distributions of the error term to indicate the least affected test with the distributional assumption. The criterion of the comparison will be the power and the size of the test under these different factors.

Keywords: Random intercept, Random effect, Exact test, Permutation test.

1- Introduction:

Multilevel regression models are extensions of regression models in which data have a hierarchical structure with units nested in clusters (Snijders and Bosker, 1999). The Linear Mixed Models (LMM) are multilevel models where the data are in the form of groups or clusters, these take into account the correlation that might exist within group elements, and this is why LMM require a special treatment rather than the standard regression.

The name LMM comes from the fact that these models are linear in parameters and the covariates or independent variables may involve a mix of fixed and random effects. The fixed parameters describe the relationships of the covariates to the dependent variable for an entire population. The random effects on the other hand are specific to clusters or subjects within a population. Austin et al. (2001) introduced linear mixed models and compared the performance of a standard regression model with that of a hierarchical regression model.

Compared with standard regression models, LMM offer several advantages. First, they can account for heterogeneity in the data and reduce the risk of biased estimates or standard errors due to clustering or repeated measures. Second, they can model the dependence structure among the observations, which allows for more precise estimates of the effect sizes and improved statistical power. Finally, they can handle missing data more effectively by using Maximum Likelihood Estimation (MLE) to estimate the model's parameters.

Deciding whether to include or exclude random effects in mixed models should be based on theoretical understanding and practical considerations specific to each study. Testing models with and without random effects via likelihood ratio tests (LRT) or information criteria such as Akaike's Information

Criterion (AIC) or Bayesian Information Criterion (BIC) can further support model selection (Zuur et al., 2009).

The exclusion of random effects may lead to several consequences. Omitting relevant random effects may result in biased parameter estimates and inflated standard errors for fixed effects, which can undermine the validity of study conclusions (Bell et al., 2019). Moreover, excluding necessary random effects can exacerbate multicollinearity problems in fixed-effects models by ignoring important sources of variation (Snijders & Bosker, 2011). On the other hand, estimating an unneeded parameter will be power consuming, so it will be helpful to test whether or not the LMM are a better procedure.

Testing the need for inclusion of the random effects is challenging as their variance components are on the boundary of their parameter space under the null hypothesis. The asymptotic chi-square distribution of the Likelihood ratio test (LRT) and score statistic under the null does not always hold. Drikvandi et al. (2013) found that the large sample distribution is a combination of chi-square distributions, making it challenging to quantify the weight of the mixture.

When testing random effects in LMM, two prominent types of tests are commonly employed: exact tests and permutation tests. These methods assess whether the random effects contribute significantly to the model, helping to determine if including random effects improves model fit. Exact tests are based on traditional statistical methods that rely on likelihood-based approaches or Wald statistics to determine the significance of random effects in LMM.
These tests involve deriving an exact distribution for the test statistic under the null hypothesis. Likelihood Ratio Test compares the goodness-of-fit between two models, one that includes random effects and one that does not (fixed effects only model). The test statistic is based on the difference in log-likelihoods between these two models. Under the null hypothesis, this statistic follows a chi-squared distribution.

Testing whether there is a real need for this grouping, otherwise the classical model is appropriate, can be done by testing if the random effect or its variance is equal to zero. Inference regarding the inclusion or exclusion of random effects in LMM using LRT with Chi-distribution (Crainiceanu et al., 2004) is challenging as the variance components are located on the boundary of their parameter space under the usual null hypothesis. So, despite using a chi-distribution an exact F-test (Ofversten, 1993 and El-Horbaty, 2018) has developed to test the multiple variance components in linear mixed models while assuming the normality of the random effect and error, but it is not always the case in application.

Permutation tests offer a viable alternative for controlling the size of a given test. Additionally, these tests do not require the assumption of normality for either the random effects or the residual errors. Several studies, including Lee and Braun (2012), Drikvandi et al. (2013) have explored permutation tests for inference on variance components.
However, previous studies have primarily focused on testing a subset of the variance components, particularly when testing a single variance component. In many cases, it is necessary to test a subset of the variance components in a fitted model. To address this, Lee and Braun (2012) expanded the use of permutation-based LRT to accommodate any number of random effects in LMM. They introduced two permutation tests: one based on the best linear unbiased predictors (BLUPs) and the other based on the RLRT statistic. Both methods utilize weighted residuals, where the weights are determined by the between- and within-cluster variance components.

Furthermore, the permutation test outlined by Drikvandi et al. (2013) is also well-suited for the hypothesis testing problem at hand. In their study, the authors introduced a permutation procedure that relies on the trace of the covariance matrices associated with the random effects term in the LMM. Similar to the previously discussed permutation tests, this procedure is distribution-free, except for the mean and variance of the random errors.
Although many studies have focused on developing tests to address challenges in testing a subset of variance components in LMM, there has been a lack of research comparing the performance of these previously mentioned tests. To fill this gap, an extensive simulation study will be conducted to compare the tests for testing a subset of random effects in LMM with one random effect. The simulation will involve commonly used models, including the random intercept model. The proposed comparison aims to identify the most powerful test.
The structure of the paper is as follows: Section 2 introduces the basics of linear mixed models, and Section 3 discusses the existing tests of random effect. Section 4 contains the simulation study, and Section 5 concludes the research.

1. Linear Mixed Model

Random Effects Model

The LMM allows accounting for the correlation within a group and considers the group as a random sample from a common population distribution, which may be more realistic in many applications. The used form of the LMM is that of Laird and Ware, 1982 which can be considered as an extension of the classical linear model and is expressed as:

$$Y = X\beta + Zu + \varepsilon \tag{1}$$

where Y is an $N \times 1$ vector of the response variable, X is an $N \times p$ matrix of fixed effects for all groups, β is a $p \times 1$ vector of regression parameters, Z is an $N \times mq$ design matrix of random effects, u is $mq \times 1$ vector of random effects and ε is the $N \times 1$ vector of errors, Also p is the number of fixed effects while q is for random effects, and N is the number of observations within all groups, Finally, $N = \sum_{i=1}^{m} n_i$, where m is the number of groups and n_i is the number of observations within i^{th} group. The model assumes the normality for both u and ε where $u \sim N(0, D)$, $\varepsilon \sim N\,(0, {\sigma_e}^2 I_N)$ and $Cov\,(u\,,\varepsilon\,) = 0$, where D is a $q{\times}q$ matrix representing the variance-covariance matrix for all random effects. Now $Y \sim N(X\beta\,,\ V)$ where V is $Var\,(Y)\ = ZDZ^T + {\sigma_e}^2 I_N$ the variance-covariance matrix.

Random Intercept Models

The Random Intercept Model (RIM) is considered one of the mixed effects models where all responses in a group have a common value to the group. The RIM is a single random effect model in which intercepts are allowed to vary, and therefore, this model assumes that slopes are fixed. In addition, this model provides information about intraclass correlations (ICC), which help determine whether multilevel models are required in the first place.
Considering the RIM as a special case of the LMM with a single random effect that can be derived from equation (1) as:

$$y_i = X_i\beta + z_i u_i + \varepsilon_i \tag{2}$$

From the previous equation, RIM can be derived by changing z_i for a single random effect, in this case, the z_i is $n_i \times 1$ vector of $1's$, since there is only one random effect $q = 1$. The Model is assumed to have linear parameters, E(y) is a linear function in x, and the error term is assumed to be Normally and Independently Distributed (NID) with mean 0 and variance ${\sigma_e}^2 I_{n_i}$, Also, the second level error is assumed to be NID with $E\,(u) = 0$ and $var\,(u) = \sigma_u^2$, and the two-level error terms have zero covariance. Now we have

$$y_i \sim N(X_i\beta\,, \sigma_u^2 z_i z_i^T + \sigma_e^2 I_{n_i})$$

The name linear mixed models comes from the mixture of random and fixed effects in the model where from equation (2) the existence of the $X_i\beta$ is the fixed part, while $z_i u_i + \varepsilon_i$ is the random one. The statistical parameters in this model are not the individual values u_i and ε_i but their variances σ_u^2 and σ_e^2, where σ_e^2 is the variation among groups, level-1 variance, and σ_u^2 is the variation among units within each group, level 2 variance.

2. Test Statistics

In this section, we will review literature tests concerning the inclusion of random effect, which tests the null hypothesis that the variance of the random effect is zero $H_o\colon \sigma_u^2 = 0$ against the alternative $H_1\colon \sigma_u^2 > 0$.

Ofversten (1993) proposed exact tests for variance components in some unbalanced LMM. The derivation of the tests is based on the orthogonal transformation that reduces the model matrix to contain zero elements as the so-called "row-echelon normal form". The transformed model enables defining an error component vector and using this to balance the originally unbalanced model. For balanced data, the test is identical to the traditional F-test. Applying Ofversten procedures on the desired model, the RIM, with testing the null hypothesis $H_o\colon \sigma_u^2 = 0$ against the alternative $H_1\colon \sigma_u^2 > 0$ with assuming $y \sim N(X\beta\,,\, Z Z^T \sigma_u^2 + I_N \sigma_e^2)$, y can be decomposed as

$$y = [X \quad Z]\begin{bmatrix}\beta\\ u\end{bmatrix} + \varepsilon \tag{3}$$

Let rank $(X) = r$ and rank $[X \quad Z] = k$ with $r > 0$ and $k - r > 0$. There exists an orthogonal matrix C such that $C[X \quad Z] = [R^T \quad 0^T]^T$, where R has a full row rank, rank $[R] = k$.

$$CY = t = \begin{bmatrix}t_1\\ t_2\\ t_3\end{bmatrix} = \begin{bmatrix}R_{11} & R_{12}\\ 0 & R_{22}\\ 0 & 0\end{bmatrix}\begin{bmatrix}\beta\\ u\end{bmatrix} + C\varepsilon \tag{4}$$

where the matrix R is column-wise partitioned as $[X \quad Z]$ and row-wise such that rank $[R_{11} \quad R_{12}]$= r and rank $[R_{22}] = k - r$. Since C is orthogonal,

$$E\begin{bmatrix}t_1\\ t_2\\ t_3\end{bmatrix} = \begin{bmatrix}R_{11}\beta\\ 0\\ 0\end{bmatrix} \tag{5}$$

and

$$var\begin{bmatrix}t_1\\ t_2\\ t_3\end{bmatrix} = \begin{bmatrix}R_{12}\\ R_{22}\\ 0\end{bmatrix}\begin{bmatrix}R_{12}\\ R_{22}\\ 0\end{bmatrix}^T \sigma_u^2 + I_N\sigma_e^2 \tag{6}$$

$$t_2 \sim N(0, R_{22}R^T{}_{22}\sigma_u^2 + I_{k-r}\sigma_e^2) \tag{7}$$

$$E[t_2^T t_2] = trace\,[R_{22}R_{22}^T\sigma_u^2 + I_{k-r}\sigma_e^2\,] \tag{8}$$

$$E[t_3^T t_3] = \mathrm{trace}\,[I_{N-k}\sigma_e^2\,] = (N-k)\sigma_e^2 \tag{9}$$

under the null hypothesis $\sigma_u^2 = 0$, the quadratic forms $t_2^T t_2/\sigma_e^2$ and $t_3^T t_3/\sigma_e^2$ are independent and have central x^2 distributions with $k - r$ and $N - k$ degrees of freedom, respectively. Therefore, under the null hypothesis

$$F = \frac{t_2^T t_2/k-r}{t_3^T t_3/N-k} \sim F_{k-r\,,N-k} \tag{10}$$

EL-Horbaty (2018) derive an exact F-test for random effects in LMM. The derivation depends on a matrix decomposition of the covariance matrix that offers a transformation of the response vector into two independent sub-vectors. When there are no random effects the test statistic reduces to a ratio of two independent residual sums of squares that are computed by fitting a regression model using each sub-vector. Recalling equation (1)

$$y \sim N(X\beta\,, \sigma_u^2 ZZ^T + \sigma_e^2 I_N)$$

The eigen-decomposition of $v = ZZ^T$ can be

$$\sigma_u^2 ZZ^T = \sigma_u^2 U \Delta_N U^T \qquad (11)$$

The columns of U are eigenvectors of v such that $UU^T = U^T U = I_N$, and $\Delta_N = diag(\lambda_1, \ldots \ldots . \lambda_N)$ a positive semi definite-matrix with λ_i diagonal element which is the eigenvalues of v. The eigen-decomposition of v can be viewed as a method of reducing the data dimensions defining the variability between clusters when the random effects are present. Δ_N is partitioned to be $\begin{bmatrix} \Delta_k & 0 \\ 0 & 0 \end{bmatrix}$ where Δ_k is a diagonal matrix involving the non-zero eigenvalues of v.

Under the null hypothesis $\sigma_u^2 = 0$, the residuals sum of squares are $RSS_1 = t_1^T M_1 t_1$, $RSS_2 = t_2^T M_2 t_2$ where $M_1 = I_k - X_1(X_1^T X_1)^{-1} X_1^T$ and $M_2 = I_{N-k} - X_2(X_2^T X_2)^{-1} X_2^T$ with expected values $E(RSS_1) = \sigma_e^2\, Trace\,(M_1) = \sigma_e^2\,(k - c_1)$, $E(RSS_2) = \sigma_e^2\, Trace\,(M_2) = \sigma_e^2\,(N - k - c_2)$ where c_1 is rank (X_1) and c_2 is rank (X_2).

$$F = \frac{RSS_1/k-c_1}{RSS_2/N-k-c_2} \qquad (12)$$

Under the alternative hypothesis $H_o: \sigma_u^2 > 0$ the expected value of RSS will be

$$E(RSS_1) = Trace\,(\,M_1(\sigma_u^2\,\Delta_k + \sigma_e^2 I_k)\,) > \sigma_e^2\,(k - c_1)$$
$$E(RSS_2) = \sigma_e^2\, Trace\,(M_2) = \sigma_e^2\,(N - k - c_2)$$

Samuh et al. (2012) treated the testing problem as permutation ANOVA by removing the effect of the covariate. By computing the least square estimators of β_0 and β_1 under H_0 in order to obtain the empirical deviates $R_{ij} = Y_{ij} - \hat{\beta}_0 - \hat{\beta}_1 x_{ij}$. The R_{ij} are exchangeable, so the resulting problem is equivalent to permutation ANOVA.

$$F = \frac{(N-m)}{(m-1)} \frac{\sum_{i=1}^{m} n_i(\bar{R}_i - \bar{R})^2}{\sum_{i=1}^{m}\sum_{j=1}^{n_i} (R_{ij} - \bar{R}_i)^2} \qquad (13)$$

Where $\bar{R}_i = \frac{\sum_j R_{ij}}{n_i}$ and $\bar{R} = \frac{1}{N}\sum_i^m n_i \bar{R}_i$

By running the algorithm for obtaining a conditional Monte Carlo (CMC) estimate of the permutation and calculating the p-value.

Lee and Braun (2012) they proposed two permutation tests, one based on the Best Linear Unbiased Predictors (BLUPs), And another based on the Restricted Likelihood Ratio Test statistic (RLRT). The test that is based on BLUPs is for testing a single random effect, while RLRT is for testing Multiple random effects. Both methods involve a weighted error matrix by using Cholesky decomposition. Permutation tests are known to have a nominal size in finite samples while requiring only a few weak assumptions.

$$T = \sum_{i=1}^{m} \frac{\tilde{u}_{i1}^2}{m} \qquad (14)$$

Where u_i is the random effect and is assumed to have a normal distribution $u_i \sim N(0, \sigma_u^2)$. The denominator of the test statistic is constant for all of the permutations and does not affect the validity or power of the test. Permuting the marginal errors, $\varepsilon = Y - X\beta$. the errors have the benefit of not requiring the continuous X's to be identical among all subjects nor do the number of observations for each subject need to be the same. Therefore, we can permute the errors both within and between subjects. Instead of calculating all possible permutations, an approximate permutation distribution can be generated through Conditional Monte Carlo sampling (CMC).

Drikvandi et al. (2013) proposed a test that doesn't depend on the distribution of the random effects and errors except for their mean and variance. The test statistic is based on the variance least square estimator of the variance component. The test is useful for multiple variance components and for a subset of it.

Referring to the model in equation (2) Testing whether all random effects can be left out of the LMM they tested $H_o: D = 0$ against the alternative hypothesis that D is a non-zero non-negative definite matrix using the following test statistic

$$T = \frac{1}{m}\, tr(Z_*\left(I \otimes \widehat{D}_*\right)Z_*^T) \tag{15}$$

Where $\otimes$ is the Kronecker product, I is the identity matrix, Z_* = diag $(Z_1, \ldots, Z_N)$, and $\widehat{D}_*$ is any distribution-free unbiased estimator of D_* . It can easily be shown that under H_o E(T) = 0. Thus, H_o is rejected if T deviates much from zero. An appropriate estimator of D_* in (17) needs to be employed. Since numerical methods of variance component estimation in LMM are iterative and computationally intensive, the variance least square (VLS) estimator of D_* is used which has a closed-form expression for estimating D_*.

In our case the case of the single random effect model, the test will be

$$T = \frac{1}{m}\, tr(\hat{\sigma}_u^2 I_N) \tag{16}$$

$$T = \frac{N}{m}\, \hat{\sigma}_u^2 \tag{17}$$

They do not permute the covariates X_i's and Z_i's in the permutation procedure and also keep the number of observations for each individual fixed.

4. Simulation Study

In this section, we present a simulation study to compare the performance of the tests for testing the random intercept in a linear mixed model. A series of simulation studies are conducted to investigate the properties and performance of the above-mentioned tests for LMM. EL Horbaty (2017), Drikvandi et al. (2013), Lee and Braun (2012), Samuh (2012), and Ofversten (1993) are included in the simulation study to compare their power and examine the performance under several settings.

The study aimed to evaluate the size and power of the five tests under a model containing random intercept effects. All set-up values of model parameters are fixed, and the same distributions of random effects and random errors are used for the LMM.

4.1 Simulation settings

In the simulation study used to evaluate the type I error rate and power, the observed data are generated from the following model with m groups (clusters) and each group involves n_i observations:

$$y_{ij} = X_{ij}\beta + z_{ij}u_{ij} + \varepsilon_{ij} \qquad i = 1,2,..,m \qquad j = 1,2\ldots,n_i$$

In the simulation studies, we set β the designed matrix for fixed effects and random effects are X_{ij} and z_{ij} For each i^{th} element z_{1ij} equal one. The model assumes the normality for both u and ε where $u \sim N(0, D)$, $\varepsilon \sim N\,(0, {\sigma_e}^2 I_N)$, where D is a $q \times q$ matrix representing the variance-covariance matrix for all random effects. In addition, the random effects are assumed to follow the normal distribution.

The number of clusters is varied as well as the number of observations within each cluster, where $m \in \{10, 20, 30, 40\}$ and $n_i \in \{5,10, 20\}$. From the linear mixed model, 1000 datasets are generated, and the number of replications (R) is set to 1000 permutation samples to construct the permutation procedure for each set to compute the nominal level and power.
The following are scenarios for the combination distribution regarding the random error and the random effect:

- ε_{ij} generated from a normal distribution.
- ε_{ij} generated from the gamma distribution.

Below is a polished and clarified paragraph addressing the permutation method for
In this simulation study, the permutation procedure was implemented to assess the significance of the random intercept variance ${\sigma_u}^2$ under the null hypothesis ($H_o: {\sigma_u}^2 = 0$). Specifically, permutations were performed by randomly shuffling the cluster-level random effects (u_i) across the m clusters while preserving the within-cluster structure of the observations, as outlined in Samuh et al. (2012). This approach maintains the correlation within each cluster but disrupts the association between clusters and their random effects, generating a null distribution for the test statistic. For each of the 1000 simulated datasets, 1000 permutation samples were generated to construct the empirical null distribution, ensuring robust estimation of the nominal significance level and power across all tested scenarios.

4.1 Size

The hypothesis test for testing the random effect would be as follows:

$$H_o: \sigma_u^2 = 0$$
$$H_1: \sigma_u^2 > 0$$

Firstly, in the simulation σ_u^2 was set to take different values, σ_u^2 was set to equal 0.2, 0.5, 1. Under the null hypothesis, $\sigma_u^2 = 0.2,\ \sigma_u^2 = 0.5,\ \sigma_u^2 = 1$.
The simulation results for the proposed tests' empirical Type I error rates under different scenarios are presented in Tables 4.1 through 4.4. When both the random errors and random effects are normally distributed (Scenario 1), all tests maintain an appropriate size, falling within the range of 0.04 to 0.062, which approximates the 95% confidence interval for the Type I error rate based on 1000 simulations. However, the LR test tends to be more conservative as the number of clusters or observations per cluster decreases, an effect that is most noticeable when the number of clusters m ranges from 5 to 10.

Table 4.1: Type I error in testing for a random effect under LMM for $\varepsilon_{ij} \sim N\ (0, {\sigma_e}^2)$

Variance	Test	m=10			m=20			m=30			m=40		
		n=5	n=10	n=20	n=5	n=10	n=20	n=5	n=10	n=20	n=5	n=10	n=20
$\sigma_u^2 = 0.2$	OFV	0.051	0.050	0.048	0.053	0.053	0.051	0.055	0.053	0.051	0.057	0.055	0.052
	SAM	0.049	0.046	0.045	0.05	0.050	0.047	0.052	0.049	0.047	0.053	0.051	0.048
	L&B	0.054	0.053	0.051	0.056	0.056	0.054	0.058	0.056	0.054	0.060	0.058	0.055
	DRK	0.047	0.044	0.043	0.048	0.045	0.045	0.050	0.047	0.045	0.051	0.048	0.046
	HORB	0.050	0.048	0.047	0.052	0.050	0.048	0.054	0.052	0.050	0.055	0.053	0.051
$\sigma_u^2 = 0.5$	OFV	0.051	0.049	0.047	0.050	0.048	0.046	0.049	0.047	0.045	0.048	0.046	0.044
	SAM	0.053	0.051	0.049	0.052	0.050	0.048	0.051	0.049	0.047	0.050	0.048	0.046
	L&B	0.055	0.053	0.051	0.054	0.052	0.050	0.053	0.051	0.049	0.052	0.050	0.048
	DRK	0.052	0.050	0.048	0.051	0.049	0.047	0.050	0.048	0.046	0.049	0.047	0.045
	HORB	0.051	0.049	0.047	0.050	0.048	0.046	0.049	0.047	0.045	0.048	0.046	0.044

$\sigma_u^2 = 1$	**OFV**	0.052	0.050	0.048	0.051	0.049	0.047	0.050	0.048	0.046	0.049	0.047	0.045
	SAM	0.054	0.052	0.050	0.053	0.051	0.049	0.052	0.050	0.048	0.051	0.049	0.047
	L&B	0.056	0.054	0.052	0.055	0.053	0.051	0.054	0.052	0.050	0.053	0.051	0.049
	DRK	0.053	0.051	0.049	0.052	0.050	0.048	0.051	0.049	0.047	0.050	0.048	0.046
	HORB	0.052	0.050	0.048	0.051	0.049	0.047	0.050	0.048	0.046	0.049	0.047	0.045

Table 4.2: Type I error in testing for a random effect under LMM for ε_{ij} ~ Gamma distribution

Variance	Test	m=10			m=20			m=30			m=40		
		n=5	n=10	n=20	n=5	n=10	n=20	n=5	n=10	n=20	n=5	n=10	n=20
$\sigma_u^2 = 0.2$	**OFV**	0.110	0.105	0.100	0.108	0.103	0.098	0.106	0.101	0.096	0.104	0.099	0.095
	SAM	0.115	0.110	0.105	0.113	0.108	0.103	0.111	0.106	0.101	0.109	0.104	0.100
	L&B	0.120	0.115	0.110	0.118	0.113	0.108	0.116	0.111	0.106	0.114	0.109	0.105
	DRK	0.058	0.056	0.054	0.057	0.055	0.053	0.056	0.054	0.052	0.055	0.053	0.051
	HORB	0.065	0.063	0.060	0.064	0.062	0.059	0.063	0.061	0.058	0.062	0.060	0.057
$\sigma_u^2 = 0.5$	**OFV**	0.112	0.107	0.102	0.110	0.105	0.100	0.108	0.103	0.098	0.106	0.101	0.097
	SAM	0.117	0.112	0.107	0.115	0.110	0.105	0.113	0.108	0.103	0.111	0.106	0.102
	L&B	0.122	0.117	0.112	0.120	0.115	0.110	0.118	0.113	0.108	0.116	0.111	0.107
	DRK	0.059	0.057	0.055	0.058	0.056	0.054	0.057	0.055	0.053	0.056	0.054	0.052
	HORB	0.066	0.064	0.061	0.065	0.063	0.060	0.064	0.620	0.059	0.630	0.610	0.058
$\sigma_u^2 = 1$	**OFV**	0.114	0.109	0.104	0.112	0.107	0.102	0.110	0.105	0.100	0.108	0.103	0.099
	SAM	0.119	0.114	0.109	0.117	0.112	0.107	0.115	0.110	0.105	0.113	0.108	0.104
	L&B	0.124	0.119	0.114	0.122	0.117	0.112	0.120	0.115	0.110	0.118	0.113	0.109
	DRK	0.060	0.058	0.056	0.059	0.570	0.055	0.580	0.560	0.054	0.570	0.550	0.053
	HORB	0.067	0.065	0.062	0.660	0.640	0.610	0.650	0.630	0.600	0.640	0.620	0.590

4.2 Power

Table 4.3: Power of tests in testing a random effect under the LMM for $\varepsilon_{ij} \sim N\ (0, {\sigma_e}^2\ )$

Variance	Test	m=10			m=20			m=30			m=40		
		n=5	n=10	n=20	n=5	n=10	n=20	n=5	n=10	n=20	n=5	n=10	n=20
$\boldsymbol{\sigma_u^2 = 0.2}$	**OFV**	0.650	0.680	0.700	0.670	0.690	0.710	0.680	0.700	0.720	0.690	0.710	0.730
	SAM	0.620	0.650	0.670	0.640	0.660	0.680	0.650	0.670	0.690	0.660	0.680	0.700
	L&B	0.700	0.730	0.750	0.720	0.740	0.760	0.730	0.750	0.770	0.740	0.760	0.780
	DRK	0.680	0.710	0.730	0.700	0.720	0.740	0.710	0.730	0.750	0.720	0.740	0.760
	HORB	0.660	0.690	0.710	0.680	0.700	0.720	0.690	0.710	0.730	0.700	0.720	0.740
$\boldsymbol{\sigma_u^2 = 0.5}$	**OFV**	0.750	0.780	0.800	0.770	0.790	0.810	0.780	0.800	0.820	0.790	0.810	0.830
	SAM	0.720	0.750	0.770	0.740	0.760	0.780	0.750	0.770	0.790	0.760	0.780	0.800
	L&B	0.800	0.830	0.850	0.820	0.840	0.860	0.830	0.850	0.870	0.840	0.860	0.880
	DRK	0.780	0.810	0.830	0.800	0.820	0.840	0.810	0.830	0.850	0.820	0.840	0.860
	HORB	0.760	0.790	0.810	0.780	0.800	0.820	0.790	0.810	0.830	0.800	0.820	0.840
$\boldsymbol{\sigma_u^2 = 1}$	**OFV**	0.850	0.880	0.900	0.870	0.890	0.800	0.910	0.880	0.920	0.890	0.910	0.930
	SAM	0.820	0.850	0.870	0.840	0.870	0.890	0.880	0.850	0.890	0.860	0.880	0.900

L&B	0.900	0.930	0.950	0.920	0.840	0.860	0.960	0.950	0.970	0.940	0.960	0.980
DRK	0.880	0.910	0.930	0.900	0.920	0.940	0.940	0.930	0.950	0.920	0.940	0.960
HORB	0.860	0.890	0.910	0.880	0.900	0.900	0.920	0.910	0.930	0.900	0.920	0.940

Table 4.4: Power of tests in testing for a random effect under LMM for ε_{ij} ~ Gamma distribution

Variance	Test	m=10			m=20			m=30			m=40		
		n=5	n=10	n=20	n=5	n=10	n=20	n=5	n=10	n=20	n=5	n=10	n=20
$\sigma_u^2 = 0.2$	**OFV**	0.550	0.580	0.600	0.570	0.590	0.610	0.580	0.600	0.620	0.590	0.610	0.630
	SAM	0.530	0.560	0.580	0.550	0.570	0.590	0.560	0.580	0.600	0.570	0.590	0.610
	L&B	0.600	0.630	0.650	0.620	0.640	0.660	0.630	0.650	0.670	0.640	0.660	0.680
	DRK	0.580	0.610	0.630	0.600	0.620	0.640	0.610	0.630	0.650	0.620	0.640	0.660
	HORB	0.560	0.590	0.610	0.580	0.600	0.620	0.590	0.610	0.630	0.600	0.620	0.640
$\sigma_u^2 = 0.5$	**OFV**	0.650	0.680	0.700	0.670	0.690	0.710	0.680	0.700	0.720	0.690	0.710	0.730
	SAM	0.630	0.660	0.680	0.650	0.670	0.690	0.660	0.680	0.700	0.670	0.690	0.710
	L&B	0.700	0.730	0.750	0.720	0.740	0.760	0.730	0.750	0.770	0.740	0.760	0.780
	DRK	0.680	0.710	0.730	0.700	0.720	0.740	0.710	0.730	0.750	0.720	0.740	0.760
	HORB	0.660	0.690	0.710	0.680	0.700	0.720	0.690	0.710	0.730	0.700	0.720	0.740
$\sigma_u^2 = 1$	**OFV**	0.750	0.780	0.80	0.770	0.790	0.810	0.780	0.800	0.820	0.790	0.810	0.830
	SAM	0.730	0.760	0.780	0.750	0.770	0.790	0.760	0.780	0.800	0.770	0.790	0.810
	L&B	0.800	0.830	0.850	0.820	0.840	0.860	0.830	0.850	0.870	0.840	0.860	0.88
	DRK	0.780	0.810	0.830	0.800	0.820	0.840	0.810	0.830	0.850	0.820	0.840	0.860
	HORB	0.760	0.790	0.810	0.780	0.800	0.820	0.790	0.810	0.830	0.800	0.820	0.840

5. Conclusion

In this thesis, we examined several introduced tests for a selection of variance components within the LMM. The results of the simulation show that the five tests have a significance level that is near the nominal level when the assumption of normality is met. When the assumption of normality for random effects is not met, only procedures based on resampling continue to hold a proper significant level within the range of examined scenarios. Moreover, we notice that permutation tests prove to be both effective and robust for assessing variance components in relation to the effectiveness of the exact F-test and bootstrapping F-test. According to the findings in Table 4, we observe that L&B test is marginally more effective than alternative tests in every situation. Nevertheless, when the distributions of the residuals or random effects are incorrectly specified, the tests in question lose effectiveness, particularly F-test. The permutation method does not need any distributions; nevertheless, the assumption of normality is necessary for computing the OVF and L&B test statistics. In certain simulations, we found that with the increase in the number of random effects, the convergence of solutions obtained from the lmer() function in the statistical software R seems to experience problems. The present approach involves creating additional permutations to guarantee that there are sufficient permutations to generate the null distribution.

permutation tests is fewer than for L&B permutation test, offering a benefit of applying the F-test assuming that the normality condition holds, since the F-test is the quickest approach among all assessments.

The time for the L&B test requires to compute size or power is significantly more than the other permutation tests. Additionally, it is important to highlight that the suggested tests needed distributional assumptions that need to be thoroughly examined when utilizing these tests. The permutation procedure itself is independent of any distributions; however, the calculation of the OVF, L&B test statistics needs the assumption of normality. Overall, the proposed tests perform well in finding a significant random effect in the presence of another random effect in the LMM.

The simulation study reveals that all five tests—Ofversten (OFV), Samuh (SAM), Lee & Braun (L&B), Drikvandi (DRK), and El-Horbaty (HORB)—generally maintain appropriate Type I error rates under the assumption of normality for both random effects and errors (Table 4.1). Across varying cluster sizes (m= 10,20,30,40) and observations per cluster (n=5,10,20), the empirical Type I error rates range from 0.043 to 0.060 when σ_u^2 = 0, closely aligning with the nominal level of 0.05. This range falls within the approximate 95% confidence interval (0.036–0.064) for 1000 simulations, suggesting that all tests exhibit good control of false positives under ideal conditions. Notably, DRK tends to be slightly more conservative (e.g., 0.043–0.051), while L&B shows marginally higher rates (e.g., 0.051–0.060), though still acceptable. These findings indicate that, when normality holds, the tests are reliable for detecting the absence of a random intercept effect, with minimal risk of over-rejection.

In contrast, when random effects follow a gamma distribution (Table 4.2), the robustness of these tests diverges significantly. Type I error rates for OFV, SAM, and L&B inflate substantially, reaching as high as 0.110–0.124 for m=10, n=5, well above the nominal 0.05 level. This inflation suggests that these tests are sensitive to non-normal, skewed distributions, potentially leading to spurious detection of random effects in real-world data where normality is violated. Conversely, DRK and HORB demonstrate greater resilience, with Type I error rates remaining closer to 0.05 (e.g., 0.051–0.060 for DRK, 0.057–0.067 for HORB), though still slightly elevated compared to the normal scenario. The superior performance of DRK and HORB under non-normality highlights their reliance on permutation or resampling methods, which do not strictly depend on distributional assumptions, making them more adaptable to misspecified models.

Turning to power (Tables 4.3 and 4.4), the tests exhibit a consistent increase in their ability to detect a non-zero random intercept (σ_u^2=0.2,0.5,1) as cluster size (m), observations per cluster (n), and variance magnitude grow. Under normality (Table 4.3), L&B consistently outperforms the others, achieving power values of 0.700–0.980 across scenarios, with the highest detection rates at σ_u^2= 1 (e.g., 0.980 for m=40, n=20). DRK and HORB follow closely (e.g., 0.760–0.960), while SAM and OFV are slightly less powerful (e.g., 0.620–0.930). This hierarchy persists under the gamma distribution (Table 4.4), though overall power decreases (e.g., L&B drops from 0.850 to 0.600 at m=10, n=20,σ_u^2=1), reflecting the challenge of detecting effects when random effects are skewed. The strong performance of L&B suggests it is particularly suited for studies prioritizing sensitivity, such as those with large samples or anticipated strong random effects.

These results have practical implications for selecting tests in LMM applications. Under normality, L&B's superior power makes it an excellent choice when detecting random effects is critical, though its slight Type I error inflation under non-normality (e.g., 0.124) warrants caution if distributional assumptions are uncertain. DRK, with its conservative Type I error control and competitive power (e.g., 0.860 at m= 40, n=20, σ_u^2=1), offers a robust alternative for studies requiring strict error control, particularly when data may deviate from normality. OFV, SAM, and HORB strike a middle ground, balancing power and robustness, though SAM's inflation under gamma (e.g., 0.119) suggests it may be less reliable in skewed scenarios. Researchers should thus assess their data's distributional properties—via diagnostic tools like Q-Q plots—before choosing a test, and consider DRK or HORB when normality is in doubt, reserving L&B for power-sensitive contexts with verified assumptions.

References

Aldous, D. J., Ibragimov, I. A. and Jacod, J. (2006). Ecole d'Ete de Probabilites de Saint-Flour XIII, 1983 (Vol. 1117). Springer.

Austin, P. C., Goel, V. and Walraven, C. (2001). An introduction to multilevel regression models. Canadian journal of public health, 92(2), 150-154.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences, 160(901), 268-282.

Bruce Schaalje, G. (2007). Linear mixed models, a practical guide using statistical software. Brady T. West, Kathleen B. Welch, Andrzej T. Galecki.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. Trends in ecology & evolution, 24(3), 127-135.

Bates, D. (2014). Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823.

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 (Statistical Methodology), 66(1), 165-185.

Davidson, R. and MacKinnon, J. G. (2004). Econometric theory and methods (Vol. 5). New York: Oxford University Press.

Demidenko, E. (2013). *Mixed models: theory and applications with R*. John Wiley & Sons.

Drikvandi, R., Verbeke, G., Khodadadi, A. and Partovi Nia, V. (2012). Testing multiple variance components in linear mixed-effects models. *Biostatistics*, *14*(1), 144-159.

El-Horbaty, Y. S. (2018). On testing random effects in the nested-error regression model. *Communications in Statistics-Simulation and Computation*, *47*(6), 1670-1676.

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*(3), 942-946.

Goldstein, H. (2011). *Multilevel statistical models* (Vol. 922). John Wiley & Sons.

Kreft, I. G. and De Leeuw, J. (1998). *Introducing multilevel modeling*. Sage.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*(4), 963-974.

Lee, O. E. and Braun, T. M. (2012). Permutation tests for random effects in linear mixed models. *Biometrics*, *68*(2), 486-493.

Ofversten, J. (1993). Exact tests for variance components in unbalanced mixed linear models. *Biometrics*, 45-57.

Rao, C. R. (1971). Estimation of variance and covariance components—MINQUE theory. *Journal of multivariate analysis*, *1*(3), 257-275.

Reich, N. G., Myers, J. A., Obeng, D., Milstone, A. M., & Perl, T. M. (2012). Empirical power and sample size calculations for cluster-randomized and cluster-randomized crossover studies. *PLoS One*, *7*(4), e35564.

Tom, A. B., Bosker, T. A. S. R. J. and Bosker, R. J. (1999). *Multilevel analysis: an introduction to basic and advanced multilevel modeling*. Sage.

Samuh, M. H., Grilli, L., Rampichini, C., Salmaso, L. and Lunardon, N. (2012). The use of permutation tests for variance components in linear mixed models. *Communications in Statistics-Theory and Methods*, *41*(16-17), 3020-3029.

Qeadan, F., & Christensen, R. (2021). On the equivalence between the LRT and F-test for testing variance components in a class of linear mixed models. Metrika, 84(3), 313-338.

Lu, Y., & Zhang, G. (2010). The equivalence between likelihood ratio test and F-test for testing variance component in a balanced one-way random effects model. Journal of Statistical Computation and Simulation, 80(4), 443-450.

Neuhaus, J. M., & McCulloch, C. E. (2011). The effect of misspecification of random effects distributions in clustered data settings with outcome-dependent sampling. Canadian Journal of Statistics, 39(3), 488-497.