On the performance of the exact F-test in linear mixed models involving multiple variance components

Yahia S. El-Horbaty[1]

Department of Mathematics, Insurance, and Applied Statistics, Helwan University

Abstract

Testing zero variance components in linear mixed models can be performed using an exact F-test. The test statistic is derived using a decomposition that is valid regardless of the unknown covariance structure of the random effects. The decomposition splits the log-likelihood function into two functions that represent the between and within cluster variability in a traditional analysis of variance problem. Results from numerical simulation studies reveal that the F-test has a competing power compared to distribution-free permutation tests. An application to a real data set is provided.

Keywords random effects, exact test, model selection.

1. Introduction

Linear mixed models (LMM) represent a popular class of models that contain both fixed and random effects and are practically used in various longitudinal studies. The outcome variable usually takes on correlated observations within the same individual or cluster due to the presence of random effects in the model. Checking whether the random effects are needed is often translated to a hypothesis testing problem of zero variance components. However, the difficulty in testing for zero variance components lies in the fact that the null value of the variance components lies on the boundary of the parameter space. Thus, the chi-square approximation to the classical score, and likelihood ratio tests is not appropriate.

One way to classify the relevant tests that exist in the literature is as tests for models with a single variance component and models including multiple variance components in the same model. Tests for a single variance component use mixtures of chi-square distribution (Self and Liang, 1987; Stram and Lee, 1994), an exact simulation-based distribution (Crainiceanu and Ruppert, 2004), an exact F-test (El-Horbaty, 2017) or using parametric bootstrap tests (Datta et al., 2011). Tests for multiple variance components use distribution-free permutation tests based on the variance components (Drikvandi et al., 2013), restricted likelihood ratio tests via permutation tests (Lee and Braun, 2012), restricted likelihood ratio tests using parametric bootstrap (Greven et al., 2008). Focus in this paper is devoted to models involving multiple variance components.

[1] yahia_mohamed@commerce.helwan.edu.eg

Both the restricted likelihood ratio test via permutation tests or parametric bootstrap suffer from two basic difficulties. The first difficulty is the extensive running time of their procedures due to the need to evaluate higher number of integrals as the number of random effects increase. The second difficulty can be summarized as the possibility to obtain unstable results in terms of the estimates of the profile or quasi-likelihood when more and more variance components are set at the boundary of their parameter space.

The interesting feature in considering an F-test is that it avoids the estimation of the variance components meanwhile can detect the departure from the null hypothesis of zero variance components. Thus, beside the exactness of the distribution of the test statistic under the null hypothesis, the test does not suffer from the computational complexities that are present under some likelihood-based tests. The objective of this article is to derive the exact F-test and assess its capability of capturing the presence of the random effects when all the variance components are tested altogether.

As will be shown in the sequel, the log-likelihood function of LMM can be decomposed into two functions that represent the between-cluster variability and within-cluster variability in the traditional analysis of variance problem in some way. This decomposition is provided in Section 3 and is utilized in testing zero variance components. Under the null hypothesis, each of the two decomposed functions is individually evaluated at its maximum likelihood estimates and is independent of the other function. This can be used to formulate the construction of an exact test.

Both El-Horbaty (2017) and Datta et al. (2011) provide a special representation of the test statistic for models with single variance components. The test statistic represented in Section 3 can be shown as a natural extension of F-test that had been proposed under the balanced one-way mixed ANOVA model (Searle et al, 1992). Importantly, a simulation study is needed to compare the performance of the F-test. We focus on the comparison with a recently introduced permutation test in Drikvandi et al. (2013) as it requires only minimal assumptions about the estimation of the variance components. Indeed, the test depends on a distribution-free estimation of the covariance structure under the alternative hypothesis. The test, for example, avoids the problem that stems from the failure to evaluate the log-likelihood function under the alternative hypothesis specially when the random effects are correlated as is the case with likelihood-based tests.

El-Horbaty (2017) conducted simulation comparisons including the permutation test in Lee and Braun (2012) under models with single variance component. The F-test in those comparisons has shown a competing performance. Thus, we focus in this article on models involving more than one variance components to assess the F-test.

The rest of this paper is organized as follows. The LMM is presented in Section 2 and Section 3 introduces the general decomposition of the likelihood function in a way that serves the testing problem. Section 4 provides an explicit formula of the test statistic under models involving a single variance component. The results of the simulation study are given in Section 5. An application to a real data set is provided in Section 6 and the article is concluded in Section 7.

2. Linear Mixed Models

Consider a quite general class of LMM. At the cluster level, the model is given by

[1] $$Y_i = X_i\beta + Z_i u_i + e_i$$

where $i = 1, \dots, m$ clusters and, for the i^{th} cluster, Y_i is a vector of responses of order $n_i \times 1$, $n = \sum_{i=1}^{m} n_i$, X_i is a design matrix for the fixed effects of order $n_i \times p$, Z_i is a matrix for the random effects of order $n_i \times q$, β is a vector of fixed effects of order $p \times 1$, u_i is a vector of random cluster effects of order $q \times 1$, and e_i is a vector of residual errors of order $n_i \times 1$. Further, let G be a $q \times q$ positive definite matrix such that $u_i \sim N(0, G)$ and let u_i be independent from e_i where $e_i \sim N(0, \sigma^2 I_{n_i})$ and I_{n_i} is the $n_i \times n_i$ identity matrix. The covariance matrix of Y_i becomes

[2] $$Var(Y_i) = Z_i G Z_i^T + \sigma^2 I_{n_i} = \sigma^2 V_i$$

where $V_i = \sigma^{-2} Z_i G Z_i^T + I_{n_i}$. Model [1] can be compactly represented for all clusters as

[3] $$Y = X\beta + Zu + e$$

where $Y = [Y_1^T, \dots, Y_m^T]^T$, $X = [X_1^T, \dots, X_m^T]^T$, $Z = diag(Z_1, \dots, Z_m)$, $u = [u_1^T, \dots, u_m^T]^T$, and $e = [e_1^T, \dots, e_m^T]^T$. It is further assumed that $u \sim N(0, G^*)$, G^* is a block diagonal of m matrices, each equal to G, and $e \sim N(0, \sigma^2 I)$. The covariance matrix of Y is

[4] $$Var(Y) = ZG^*Z^T + \sigma^2 I = \sigma^2 V$$

where $V = \sigma^{-2} ZG^*Z^T + I$. Based on the normality assumption of both the random effects and the residual errors, the loglikelihood function underlying model [3] is represented next.

3. Likelihood Decomposition for Testing the Random Effects

3.1 *Likelihood Decomposition*

Let $f(Y|\beta, \sigma^2, G)$ be the marginal density of Y, then

$$-2\log f(Y|\beta, \sigma^2, G) = n\log(2\pi\sigma^2) + \sum_{i=1}^{m} \log|V_i| + \sum_{i=1}^{m} (Y_i - X_i\beta)^T V_i^{-1} (Y_i - X_i\beta)/\sigma^2$$

Since $V_i^{-1} = I_{n_i} - Z_i(\sigma^2 I_{n_i} + GZ_i^T Z_i)^{-1} GZ_i^T$, the above expression of the likelihood function can be decomposed into

[5] $$-2\log f(Y|\beta, \sigma^2, G) = g_1(Y|\beta, \sigma^2) + g_2(Y|\beta, \sigma^2, G)$$

where

$$g_1(Y|\beta, \sigma^2) = (N-m)\log(2\pi\sigma^2) + \textstyle\sum_{i=1}^m B_i^T B_i / \sigma^2,$$

$$g_2(Y|\beta, \sigma^2) = m\log(2\pi\sigma^2) + \textstyle\sum_{i=1}^m \log|V_i| + \sum_{i=1}^m C_i^T V_i^{-1} C_i / \sigma^2,$$

$$B_i = K_i(Y_i - X_i\beta),\ C_i = P_i(Y_i - X_i\beta),\ K_i = I_{n_i} - P_i\ ,\ P_i = Z_i(Z_i^T Z_i)^{-1} Z_i^T.$$

This decomposition utilizes the fact that $K_i V_i^{-1} = K_i$ and thus, $g_1(Y|\beta, \sigma^2)$ doesn't depend on G. The result holds because $B_i^T V_i^{-1} C_i = 0$. A hypothesis test for the need for all random effects in [3] can be developed by utilizing this result as shown below.

3.2 *Testing Random Effects*

The following hypothesis testing problem about the variance components is of concern. Let

[6] $$H_0 : \ G = 0 \ \text{against} \ H_1 : \ G > 0$$

The development of an exact F-test is motivated by the decomposition in [5]. Under the null hypothesis [6], note that from [4] and [5] we have $V_i^{-1} = I_{n_i}$ and $P_i V_i^{-1} = P_i$. Further, the following three results hold.

First, $\sum_{i=1}^m \widehat{B}_i^T \widehat{B}_i / \sigma^2 = \sum_{i=1}^m (Y_i - X_i\widehat{\beta}_w)^T K_i (Y_i - X_i\widehat{\beta}_w) / \sigma^2$ follows chi-square distribution with $trace(M_w)$ degrees of freedom where $\widehat{\beta}_w = \sum_{i=1}^m (X_i^T K_i X_i)^- X_i^T K_i Y_i$, $M_w = K - KX(X^T KX)^- X^T K$ and $K = diag(K_1, \dots, K_m)$.

Second, $\sum_{i=1}^m \widehat{C}_i^T V_i^{-1} \widehat{C}_i / \sigma^2 = \sum_{i=1}^m (Y_i - X_i\widehat{\beta}_B)^T P_i (Y_i - X_i\widehat{\beta}_B) / \sigma^2$ follows chi-square distribution with $trace(M_B)$ degrees of freedom where $\widehat{\beta}_B = \sum_{i=1}^m (X_i^T P_i X_i)^- X_i^T P_i Y_i$, $M_B = P - PX(X^T PX)^- X^T P$ and $P = diag(P_1, \dots, P_m)$. Note that $\widehat{\beta}_w$ and $\widehat{\beta}_B$ are the maximum likelihood estimators under the null hypothesis of no random effects.

Third, $F_w = \sum_{i=1}^m \widehat{B}_i^T \widehat{B}_i / \sigma^2$ and $F_B = \sum_{i=1}^m \widehat{C}_i^T V_i^{-1} \widehat{C}_i / \sigma^2$ are independent quadratic forms because M_B and M_W are orthogonal. Thus, the test statistic

[7] $$F = \frac{F_B/trace(M_B)}{F_W/trace(M_W)} = \frac{Y^T M_B Y/trace(M_B)}{Y^T M_W Y/trace(M_W)}$$

has F distribution where the degrees of freedom are $trace(M_B)$ and $trace(M_w)$ respectively. It is easy to show that the $E_0(F_B) \le E_1(F_B)$ where $E_0(.)$ and $E_1(.)$ denote the expected value under the null and the alternative hypothesis respectively. See for example the analytical evaluation of $E_0(F_B)$ and $E_1(F_B)$ under the random intercept model in El-Horbaty (2017). Observing that $E(F_W)$ remains unchanged under both the null and the alternative hypothesis, the test statistic in [7] provides a useful test for the need for random effects in LMM. An illustration of the decomposition and the derivation of test statistic in explicit formulas is given under the famous NER model as shown next.

4. Formulations under the Nested Error Regression Model

The NER model is a popular example of LMM that appears in many fields of biomedical, demographic, and small area estimation applications. The standard analysis of variance method for evaluating the need for clustering, which is equivalent to testing zero variance components, is obvious when applied under this model. Closely related tests that already exist in the literature of NER models are addressed in the sequel.

The NER model is given by

[8] $$Y_i = X_i\beta + j_{n_i}u_i + e_i$$

where $j_{n_i} = (1, \ldots, 1)^T$ and $u_i \sim N(0, \sigma_u^2)$. The covariance matrix of Y_i is

[9] $$Var(Y_i) = \sigma_u^2 j_{n_i} j_{n_i}^T + \sigma^2 I_{n_i} = \sigma^2 V_i(\psi)$$

where $V_i(\psi) = \psi j_{n_i} j_{n_i}^T + I_{n_i}$ and $\psi = \sigma_u^2/\sigma^2$. Let $Z = diag(j_{n_1}, \ldots, j_{n_m})$, the NER model reduces to [3] with covariance matrix $Var(Y) = \sigma_u^2 ZZ^T + \sigma^2 I = \sigma^2 V(\psi)$ where $V(\psi) = diag[V_1(\psi), \ldots, V_m(\psi)]$.

Let $f(Y|\beta, \sigma^2, \sigma_u^2)$ be the marginal density of Y, it can be shown that

$$-2\log f(Y|\beta, \sigma^2, \sigma_u^2) = g_1(Y|\beta, \sigma^2) + g_2(\bar{Y}|\beta, \sigma^2, \sigma_u^2)$$

where for $\gamma_i = \gamma_i(\psi) = 1/(1 + n_i\psi)$ and $\bar{Y} = (\bar{y}_1, \ldots, \bar{y}_m)^T$ and $\bar{y}_i$ represents the arithmetic mean of the i^{th} cluster. Then, the two functions

$$g_1(Y|\beta, \sigma^2) = (n-m)\log(2\pi\sigma^2) + \sum_{i=1}^{m}\sum_{j=1}^{n_i}\{(y_{ij} - \bar{y}_i) - (x_{ij} - \bar{x}_i)^T\beta\}^2/\sigma^2, \text{ and}$$

$$g_2(\overline{Y}|\beta,\sigma^2,\sigma_u^2) = m\log(2\pi\sigma^2) - \sum_{i=1}^{m}\log\gamma_i(\psi) + \sum_{i=1}^{m} n_i\,\gamma_i(\psi)\,(\bar{y}_i - \overline{x}_i{}^T\beta)^2/\sigma^2,$$

correspond to the within and between parts of analysis of variance, respectively. This result has been utilized under the context of small area estimation in Kubokawa and Erdembat (2010). Under the null hypothesis [6], which now reduces to $H_0 : \sigma_u^2 = 0$, note that the above two functions are, respectively, proportional to

$$\sigma^2 F_1 = \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left\{(y_{ij} - \bar{y}_i) - (x_{ij} - \overline{x}_i)^T\beta\right\}^2 = \sum_{i=1}^{m}(Y_i - X_i\beta)^T K_i (Y_i - X_i\beta)$$

$$\sigma^2 F_2 = \sum_{i=1}^{m} n_i\,(\bar{y}_i - \overline{x}_i{}^T\beta)^2 = \sum_{i=1}^{m}(Y_i - X_i\beta)^T P_i (Y_i - X_i\beta)$$

where $\boldsymbol{P}_i = \boldsymbol{j}_{n_i}(\boldsymbol{j}_{n_i}^T\boldsymbol{j}_{n_i})^{-1}\boldsymbol{j}_{n_i}^T$ and $\boldsymbol{K}_i = \boldsymbol{I}_{n_i} - \boldsymbol{P}_i$. Based on the above representation, then the above two quadratic forms are independent and each follows a chi-square distribution. Thus, the construction of the F-test is analogously given by defining the statistic

$$[10] \qquad F_{NER} = \frac{F_2/(m-(p+m-rank(Q)))}{F_1/(n-rank(Q))}$$

where the augmented design matrix $\boldsymbol{Q}$ is such that $\boldsymbol{Q} = [X\,Z]$. Thus, under the null hypothesis, F_{NER} has an F distribution with $m - (p + m - rank(Q))$ and $n - rank(Q)$ degrees of freedom.

The decomposition in Section 3 and the consequent construction of test statistic highlights the fact that testing zero variance components can be shown as an analysis of variance problem in a simple fashion. Thus, this result extends the work in Searle et al. (1992) under the balanced a one-way mixed ANOVA model. El-Horbaty (2017) emphasizes that the test statistic is derived for testing random effects and cannot be applied to test for absent fixed effects. However, the model transformation used therein is valid under the restriction that the covariance matrix of the random effects is known up to a constant. The constant is σ_u^2 in this case.

Datta et al. (2011) consider only the numerator of F_{NER} in [10] as another test statistic for testing the same hypothesis. The test statistic is given by $F_{Datta} = F_2(m - (p + m - rank(Q)))$. Since $\boldsymbol{\beta}$ and σ^2 are unknown, their estimates are given by $\widehat{\boldsymbol{\beta}}_{OLS} = (X^TX)^{-1}X^TY$ and $\hat{\sigma}_{OLS}^2 = Y^TM_XY/tr(M_X)$ where $M_X = I - X(X^TX)^{-1}X^T$. Since the asymptotic distribution of F_{Datta} is unknown, this problem is overcome by approximating the distribution using the method of parametric bootstrap.

A numerical investigation of the comparative performance of the F-test to recent resampling-based tests doesn't exist except under the NER model context in El-Horbaty (2017) and Drikvandi et al. (2013). We bridge this gap in the next section by conducting extensive simulation studies under LMM involving multiple random effects.

5. Simulation Study

In order to distinguish the numerical study conducted in this section from other studies that involved the use of the F-test in under models involving only a single variance component, we take two points into consideration. First, we run the test that is presented in Section 3 to models involving multiple variance components. Specifically, the test is applied to a linear trend model involving two random effects as well as to a LMM involving three random effects. Second, the test is compared to a recently proposed permutation test (Drikvandi et al., 2013) to show how the F-test is more reliable in detecting the presence of the random effects. Note that both tests are reliable even if the distribution of the random effects is misspecified. The latter fact is emphasized in the runs as shown in the next paragraphs.

5.1 Models with two random effects

First consider the linear trend model with random intercepts and random slopes

[11]
$$Y_{ij} = a_{1i} + a_{2i}t_{ij} + \varepsilon_{ij}$$
$$a_{1i} = \beta_1 + u_{1i}, \quad a_{2i} = \beta_2 + u_{2i}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n_i$$

where t_{ij} is the j^{th} observation time for the i^{th} individual, β_1 and β_2 are fixed effects, and u_{1i} and u_{2i} are the random intercept and the random slope, respectively. For this model, the objective is to test for the need of both the random intercept u_{1i} and random slope u_{2i} in the model.

In the simulations, let $\beta_1 = 1$, $\beta_2 = 2$, $t_{ij} = j$, and assume that $\varepsilon_{ij} \sim N(0,1)$. Let $G = \begin{bmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{bmatrix}$, and consider two types of distributions for the random effects vector $u_i = (u_{1i}, u_{2i})^T$. First, a bivariate normal distribution with zero mean and covariance matrix G and, second, a bivariate Student's t-distribution with a zero mean, degree of freedom $df = 3$, and scale matrix $(df - 2/df)G$. Under each of these two distributions, 1000 Monte Carlo samples are generated from model [11] with different values of G for $m = 10, 15$ groups and $n = 3, 5$ observations per group. Specifically, we set the covariance matrix G_1 equal to $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ to compute the size of the tests, $G_2 = \begin{bmatrix} 0.05 & 0.02 \\ 0.02 & 0.05 \end{bmatrix}$, $G_3 = \begin{bmatrix} 0.08 & 0.02 \\ 0.02 & 0.08 \end{bmatrix}$, $G_4 = \begin{bmatrix} 0.1 & 0.05 \\ 0.05 & 0.1 \end{bmatrix}$ and $G_5 = \begin{bmatrix} 0.1 & 0.09 \\ 0.09 & 0.1 \end{bmatrix}$ to investigate the empirical power of the test at a nominal level α =0.05. $B = 1000$ is chosen as the number of permutation samples for the permutation test in Drikvandi et al. (2013) denoted by T.

Table (1): Size and power of the T-test and the F-test in model [11] with $\varepsilon_{ij} \sim N(0,1)$ and bivariate normal distribution for $(b_{1i}, b_{2i})^T$

Covariance matrix of $(b_{1i}, b_{2i})^T$	$m = 10$				$m = 15$			
	$n = 3$		$n = 5$		$n = 3$		$n = 5$	
	T	F	T	F	T	F	T	F
G_1	5.6	5.7	4.0	5.8	5.4	5.8	5.6	5.8
G_2	13.0	16.6	47.4	73.3	18.0	73.3	58.6	73.3
G_3	17.0	22.9	59.6	87.8	24.6	87.8	71.6	87:8
G_4	21.8	33.5	67.6	95.1	32.6	95.1	77.8	95.1
G_5	27.2	38.8	67.4	96.5	38.4	96.5	76.6	96.5

- T represent the permutation test.
- F represent the exact F-test.

Table (2): Size and power of the T-test and the F-test in model [11] with $\varepsilon_{ij} \sim N(0,1)$ and bivariate Student's t distribution for $(b_{1i}, b_{2i})^T$

Covariance matrix of $(b_{1i}, b_{2i})^T$	$m = 10$				$m = 15$			
	$n = 3$		$n = 5$		$n = 3$		$n = 5$	
	T	F	T	F	T	F	T	F
G_1	4.6	5.2	6.0	5.4	4.6	6.4	3.4	5.0
G_2	8.2	15.4	39.8	57.7	15.6	20.4	46.2	73.9
G_3	10.2	20.7	50.0	72.0	21.0	27.5	60.4	87.7
G_4	15.2	26.7	56.0	81.5	25.8	36.9	68.4	92.8
G_5	17.2	30.7	58.2	84.7	27.6	43.2	68.2	96.0

- T represent the permutation test.
- F represent the exact F-test.

The results from Table 1 and 2 indicate that the F-test is superior to the T-test under all settings. Although the difference in the power between the two tests could be quite small when the group size and the number of groups are both small, for example m = 10 and n = 3, the F-test shows larger power as the G departs from its null value. This is observed by comparing the power of the two tests as G moves from G_2 up to G_5. As both m and n increase the gap between the powers increases to about 30% difference between the two tests. Whether the random effects follow multivariate normal distribution or a multivariate t-distribution does not change the above conclusion.

5.2 Models with three random effects

Next, we perform a simulation study for the LMM given by

[12] $$Y_{ij} = \beta_1 + \beta_2 x_{ij} + b_{1i} + b_{2i} z_{ij1} + b_{3i} z_{ij2} + \varepsilon_{ij}, \quad i = 1, \dots, m, \quad j = 1, \dots, n_i$$

$$G = \begin{bmatrix} g_{11} & g_{12} & g_{13} \\ g_{12} & g_{22} & g_{23} \\ g_{13} & g_{23} & g_{33} \end{bmatrix}$$

The performance of the F-test and the permutation test is assessed when all the random effects u_{1i}, u_{2i}, and u_{3i} are tested under model [12]. In the simulations, let $\beta_1 = 1$, $\beta_2 = 2$, and assume that $\varepsilon_{ij} \sim N(0,1)$. The covariates x_{ij}, z_{ij1}, and z_{ij2} are all generated from $U(0,1)$.

Considered a multivariate normal distribution and a multivariate Student's t distribution for the vector of random effects $\boldsymbol{u}_i = (u_{1i}, u_{2i}, u_{3i})^T$. Under each distribution, we generated 1000 Monte Carlo samples from model [12] with different values of G for $m = 7, 15$ and $n = 5,10$. We set the covariance matrix G_1 equal to $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ to calculate the size of the tests, $G_2 = \begin{bmatrix} 0.2 & 0.1 & 0.1 \\ 0.1 & 0.2 & 0.1 \\ 0.1 & 0.1 & 0.2 \end{bmatrix}$, $G_3 = \begin{bmatrix} 0.5 & 0.1 & 0.1 \\ 0.1 & 0.5 & 0.1 \\ 0.1 & 0.1 & 0.5 \end{bmatrix}$, and $G_4 = \begin{bmatrix} 1 & 0.2 & 0.2 \\ 0.2 & 1 & 0.2 \\ 0.2 & 0.2 & 1 \end{bmatrix}$ to evaluate the empirical power of the tests at a nominal level of $\alpha = 0.05$.

Table (3): Size and power of the T-test and the F-test in model [12] with $\varepsilon_{ij} \sim N(0,1)$ and multivariate normal distribution for $(b_{1i}, b_{2i}, b_{3i})^T$

Covariance matrix of	$m = 7$				$m = 15$			
	$n = 5$		$n = 10$		$n = 5$		$n = 10$	
$(b_{1i}, b_{2i}, b_{3i})^T$	T	F	T	F	T	F	T	F
G_1	4.6	4.8	5.6	4.9	3.4	5.1	5.0	4.9
G_2	21.0	25.9	34.2	74.7	27.8	58.2	49.2	96.7
G_3	30.6	47.2	55.2	92.1	39.0	86.2	68.8	99.9
G_4	43.4	73.4	78.8	98.5	45.8	99	80.4	100

- T represent the permutation test.
- F represent the exact F-test.

The results from Table 3 and 4 confirm the previous conclusion made under the linear trend model. Under a the LMM with three random effects in [12], the F-test is superior to the T-test under all settings. The only case where the two tests show a quite small difference in power is when the number of groups, group size, and the matrix $G > \mathbf{0}$ are set at their smallest values. As both m and n increase the gap between the powers can increase to about 46% between the two tests as is the case when m = 15 and n = 5.

Table (4): Size and power of the T-test and the F-test in model [12] with $\varepsilon_{ij} \sim N(0,1)$ and multivariate Student's t distribution for $(b_{1i}, b_{2i}, b_{3i})^T$

Covariance matrix of	$m = 7$				$m = 15$			
	$n = 5$		$n = 10$		$n = 5$		$n = 10$	
$(b_{1i}, b_{2i}, b_{3i})^T$	T	F	T	F	T	F	T	F
G_1	3.8	4.9	5.0	4.9	4.8	4.8	5.0	5.1
G_2	12.8	23	30.4	57.3	20.0	43.9	44.8	85.4
G_3	19.8	36.8	47.2	77.1	31.8	67.3	61.0	96.9
G_4	29.2	58.6	66.0	91.9	41.6	90	73.6	99.8

- T represent the permutation test.
- F represent the exact F-test.

6. Application

The F-test is applied to a data on the plasma inorganic phosphate flux. The data set is 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 and Murphy, 1986). In that 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 hours after a standard-dose oral glucose challenge. The objective of the study is 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.

The assessment of the heterogeneity among patients with respect to the overall mean and evolutions over time is particularly important for the sake of assessing the impact of the treatment on the plasma level. The profiles show that the plasma level exhibits a quadratic response as a function of hours. Thus, the LMM suggested below is similar to that in Verbeke and Molenberghs, (2000, p. 25).

Each individual profile (subject) shown in Figure 1 is modelled by a quadratic function over time, where time is expressed as hours. Thus, the proposed regression model in the first stage is

[13] $$Y_{ij} = \beta_{1i} + \beta_{2i} t_{ij} + \beta_{3i} t_{ij}^2 + \varepsilon_{ij}$$

where Y_{ij} is the j^{th} plasma level for the i^{th} subject (patient) at time t_{ij} (in hours). In the second stage, the subject-specific intercept and a linear as well as a quadratic time effects are related to the glucose tolerance groups of subjects (controls, non-hyperinsulinemic patients and hyperinsulinemic). Then the model in the second stage becomes

$$\begin{aligned} \beta_{1i} &= \beta_1 C_i + \beta_4 N_i + \beta_7 H_i + b_{1i}\,, \\ \beta_{2i} &= \beta_2 C_i + \beta_5 N_i + \beta_8 H_i + b_{2i}\,, \\ \beta_{3i} &= \beta_3 C_i + \beta_6 N_i + \beta_9 H_i + b_{3i} \end{aligned} \qquad [14]$$

where C_i, N_i, and H_i are indicator functions such that C_i refers to i^{th} subject in control group, N_i represents the i^{th} subject in non-hyperinsulinemic group and H_i is the i^{th} subject in hyperinsulinemic group.

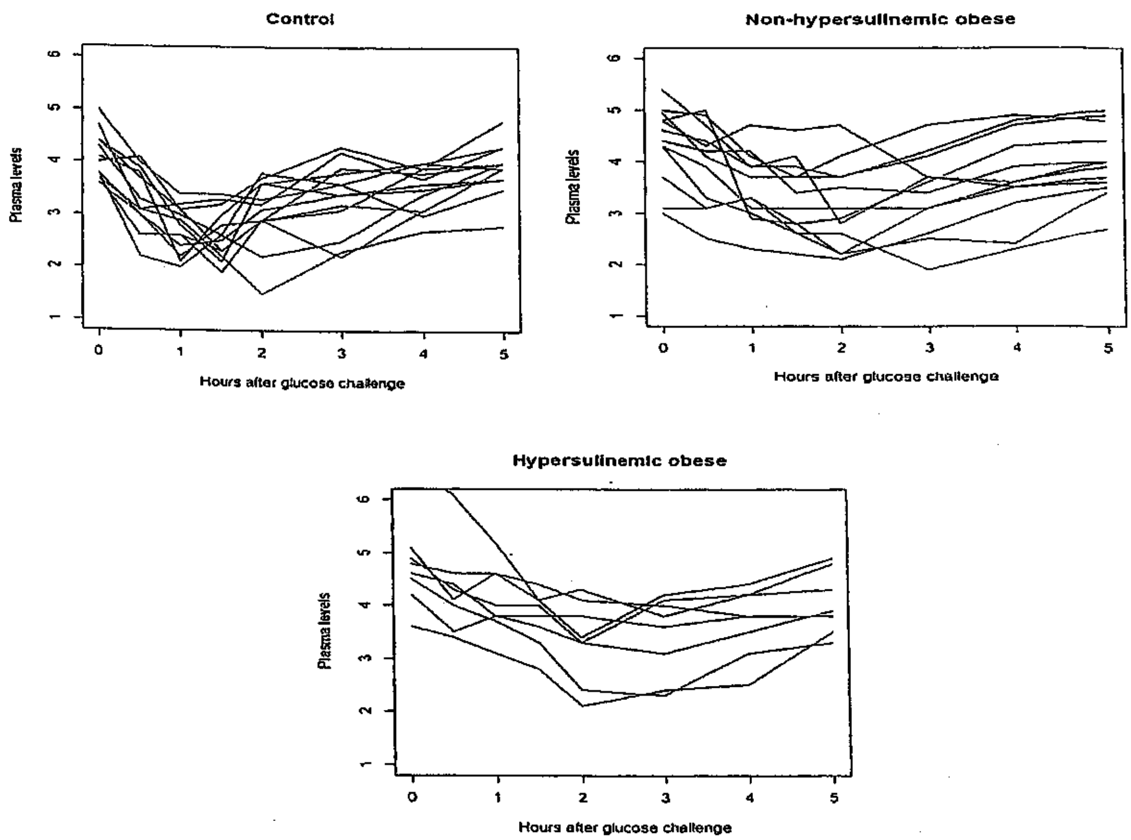

Figure-1: Individual profiles of control and obese patients in the plasma inorganic phosphate experiment

Combining the models [13] and [14] for a two-stage analysis of the plasma inorganic phosphate data yields

$$[15]\quad Y_{ij} = (\beta_1 C_i + \beta_4 N_i + \beta_7 H_i + b_{1i}) + (\beta_2 C_i + \beta_5 N_i + \beta_8 H_i + b_{2i})t_{ij} + (\beta_3 C_i + \beta_6 N_i + \beta_9 H_i + b_{3i})t_{ij}^2 + \varepsilon_{ij}$$

which can be reformulated as

$$Y_{ij} = \begin{cases} (\beta_1 + b_{1i}) + (\beta_2 + b_{2i})t_{ij} + (\beta_3 + b_{3i})t_{ij}^2 + \varepsilon_{ij} & \text{if Control,} \\ (\beta_4 + b_{1i}) + (\beta_5 + b_{2i})t_{ij} + (\beta_6 + b_{3i})t_{ij}^2 + \varepsilon_{ij} & \text{if Non} - \text{Hyperinsulinemic,} \\ (\beta_7 + b_{1i}) + (\beta_8 + b_{2i})t_{ij} + (\beta_9 + b_{3i})t_{ij}^2 + \varepsilon_{ij} & \text{if Hyperinsulinemic} \end{cases}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9)^T$ 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 variance least squares estimator of matrix $\boldsymbol{G}$ defined for the vector of random effects $\boldsymbol{u}_i = (u_{1i}, u_{2i}, u_{3i})^T$

$$\widehat{G} = \begin{bmatrix} \hat{g}_{11} & \hat{g}_{12} & \hat{g}_{13} \\ \hat{g}_{12} & \hat{g}_{22} & \hat{g}_{23} \\ \hat{g}_{13} & \hat{g}_{23} & \hat{g}_{33} \end{bmatrix} = \begin{bmatrix} 0.364 & -0.074 & 0.008 \\ -0.074 & 0.08 & -0.011 \\ 0.008 & -0.011 & 0.001 \end{bmatrix}$$

is essential to apply the T-test. Nevertheless, this is not required under the F-test as it employs the test statistic only under the null hypothesis of no random effects. The independence of estimating $\boldsymbol{G}$ in calculating the F statistic is explicitly shown in [7], which does not depend on $\boldsymbol{G}$. The estimate of σ^2 is $\hat{\sigma} = 0.17$. Also, the estimates of the fixed-effects parameters are $\hat{\beta}_1 = 3.69$, $\hat{\beta}_2 = -0.72$, $\hat{\beta}_3 = 0.16$, $\hat{\beta}_4 = 4.28$, $\hat{\beta}_5 = -0.819$, $\hat{\beta}_6 = 0.158$, $\hat{\beta}_7 = 4.78$, $\hat{\beta}_8 = -0.95$ and $\hat{\beta}_9 = 0.161$ respectively. The previous estimates are based on the presence of the variance components.

Using 1000 as the number of permutations, the T-test yields a test statistic equal to 2.496 with a p-value of 0.001. Thus, the test suggests that the variance components estimates in $\widehat{G}$ are nontrivial. The F-test produces a test statistic value equal to 5.84, with the degrees of freedom are 90 and 165, with a p-value less than 0.0001. Hence, both tests conclude that the variance components deviate from zero and that the random individual effects are needed in the proposed model. The conclusion is made at any nominal level of bigger than or equal to 0.1%.

6. Conclusion

This article provides a new derivation to the exact F-test that is applicable under unbalanced LMM. The relation between the derived statistic and existing methods for the same testing problem in the literature is clarified. The main focus is given to the situation where all the random effects are tested under the null hypothesis to have zero variances. Simulation studies show that the F-test is superior to a distribution-free permutation test by significant power difference. This result shows the reliability of the F-test even if the random effects distribution is not normal.

However, this exact test suffers from two basic problems that might be met in practice. One problem is the need to assume the normality of the residual error terms under the null hypothesis. The other problem is its incapability for testing only a subset of variance components for a wide range of LMM. Since it is common to test all random effects at once, the second problem does not deprive the test from its popularity and usefulness. The users of LMM are to be advised to use the exact F-test if it possesses a reasonably good power relative to the other tests that are based on computed-intensive methods.

References

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

Datta, G. S., Hall, P. and Mandal, A. (2011). Model selection by testing for the presence of small-area effects, and application to area-level data. *Journal of American Statistical Association*, 106: 362-374.

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

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

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

Kubokawa, T. and Erdembat, N. (2010). On testing linear hypothesis in a nested error regression model. *Communications in Statistics – Theory and Methods*, 39: 1552-1562

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

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

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

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*. Springer.

Zerbe, G. O. and Murphy, J. R. (1986). On Multiple Comparisons in the Randomization Analysis of Growth and Response Curves. *International Biometric Society*, 42: 795-804.

عن آداء اختبار F في النماذج الخطية المختلطة التي تضم العديد من مكونات التباين

د. يحيى السيد محمد الهربيطي

قسم الرياضيات والتأمين والإحصاء التطبيقي – جامعة حلوان

مستخلص

اختبار F يمثل واحداً من الطرق المستخدمة في اختبار صفرية مكونات التباين الإضافية في النماذج الخطية المختلطة. في هذا البحث يتم اشتقاق احصاءة اختبار من خلال تقسيم دالة الإمكان بطريقة لا تتأثر بطبيعة هيكل مصفوفة التباين-التغاير للمؤثرات العشوائية داخل النموذج. تقسيم دالة الإمكان إلى دوال جزئية ينتج عنه دالتين تعبر احداهما تباين المشاهدات داخل المجموعات والأخرى تعبر عن التباين فيما بين المجموعات، كما هو الحال في مشكلة تحليل التباين. وتفصح نتائج دراسات المحاكاة التي تم تنفيذها عن اكتساب اختبار F قوة اختبار أكبر لرفض صفرية مكونات التباين، وذلك إذا ما قورن باختبارات التقليب التي لا تعتمد على التوزيع الاحتمالي لمكونات خطأ تلك النماذج. تم تطبيق الطريقة المقترحة على بيانات واقعية.

الكلمات المفتاحية: المؤثرات العشوائية – اختبار دقيق – اختيار النموذج

On the performance of the exact F-test in linear mixed models involving multiple variance components

Yahia S. El-Horbaty[1]

Department of Mathematics, Insurance, and Applied Statistics, Helwan University

Abstract

Testing zero variance components in linear mixed models can be performed using an exact F-test. The test statistic is derived using a decomposition that is valid regardless of the unknown covariance structure of the random effects. The decomposition splits the log-likelihood function into two functions that represent the between and within cluster variability in a traditional analysis of variance problem. Results from numerical simulation studies reveal that the F-test has a competing power compared to distribution-free permutation tests. An application to a real data set is provided.

Keywords random effects, exact test, model selection.

[1] yahia_mohamed@commerce.helwan.edu.eg