



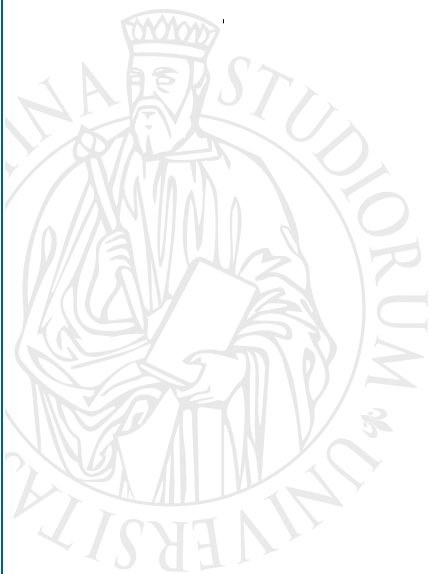
UNIVERSITÀ
DEGLI STUDI
FIRENZE

DISIA

DIPARTIMENTO DI STATISTICA,
INFORMATICA, APPLICAZIONI
"GIUSEPPE PARENTI"

**An *F*-type multiple testing approach
for assessing randomness
of linear mixed models**

Marco Barnabani



**DISIA WORKING PAPER
2019/09**

© Copyright is held by the author(s).

An F –type multiple testing approach for assessing randomness of linear mixed models

Marco Barnabani
Department of Statistics, Informatics, Applications
V.le Morgagni, 59
50134 Florence, Italy.
e-mail: marco.barnabani@unifi.it

Abstract

In linear mixed models the assessing of the significance of all or a subset of the random effects is often of primary interest. Many techniques have been proposed for this purpose but none of them is completely satisfactory. One of the oldest methods for testing randomness is the F –test but it is often overlooked in modern applications due to poor statistical power and non-applicability in some important situations. In this work a two-step procedure is developed for generalizing an F –test and improving its statistical power. In the first step, by comparing two covariance matrices of a least squares statistics, we obtain a "repeatable" F –type test. In the second step, by changing the projected matrix which defines the least squares statistic we apply the test repeatedly to the same data in order to have a set of correlated statistics analyzed within a multiple testing approach. The resulting test is sufficiently general, easy to compute, with an exact distribution under the null and alternative hypothesis and, perhaps more importantly, with a strong increase of statistical power with respect to the F –test.

keywords: Linear Mixed Models; Hypothesis testing; Comparison of matrices; F -distribution; Beta binomial distribution.

1 Introduction

In longitudinal studies with subjects measured repeatedly across time there has been increasingly more attention on linear mixed-effects models (Laird and Ware, 1982) because they can incorporate within-cluster and between-cluster variations. Linear mixed effect models (*LME* models) can be viewed as an extension of linear regression models (*LR* models) where one or more subject-specific latent variables are included to account for within-subject dependency. Typically, an additional random effect is included for each regression coefficient which is expected to vary among subjects and it becomes important to assess the randomness of all or a subset of parameters. A linear mixed model can be regarded as a two-stage model (Laird, 2004) where in the first stage it may be viewed as a set of standard regression models with the matrix of covariates and the random effects design matrix "merged" in a unique matrix and the parameter vector which includes both fixed and random parameters or the sum of both (Rocha and Singer, 2017). In the second stage a specification of the mean and the variance of the random effects are assumed.

When faced with this representation, we can ask whether the "enlarged" parameters vector is fixed, random or has both fixed and random elements. In order to address the issue of which model is more suitable, one might use standard model selection measures based on information criteria such as the widely used Akaike Information Criteria (*AIC*; Akaike (1973)), the Bayesian Information Criteria (*BIC*; Schwarz (1978)) or the conditional Akaike Information Criterion (*cAIC*, Vaida and Blanchard (2005)). We refer to the paper of Muller et al. (2013) for a review of these approaches and other methods such as shrinkage methods like the LASSO (Tibshirani, 1996), Fence methods (Jiang et al., 2008) and Bayesian methods. The validity of all the methods proposed depends on the underlying assumptions. The review paper of Muller et al. (2013) gives an overview of the limits and most important findings of above-mentioned approaches, extracting information from some published simulation results. As is known, one of the major drawbacks of these approaches is that they fail to give any measure of the degree

of uncertainty of the model chosen. The value they produce does not mean anything by itself.

Alternatively, because model selection is closely related to hypothesis testing, the choice between an *LR* model and an *LME* model and the evaluation of its uncertainty could be conducted by assessing the significance of all or a subset of the random effects. This normally involves the use of hypothesis tests to detect whether one or more variance components are equal to zero. Extensive research has been conducted into testing the significance of random effects in linear mixed models. Arguably, the main challenge has been how to deal with the fact that under the null hypothesis, the variance lies on the boundary of the parameter space, meaning that the likelihood ratio as well as the score and Wald tests are not asymptotically chi-squared distributed. Consequently, in large samples they lead to a power lower than that of the standard case and in finite samples they tend to produce conservative tests.

Over the past two decades, these difficulties in conservatism and the somewhat strict model assumptions, along with improvements in statistical power, over the past two decades, have spurred the development of a number of testing procedures which predominantly rely on simulations to determine the null distribution; see for instance Fitzmaurice et al. (2007), Sinha (2009), Samuh et al. (2012) and Drikvandi et al. (2013) among many others.

One of the oldest methods for testing random effects in linear mixed models is the F -test proposed, originally by Wald (1947), for testing all random effects, and subsequently extended by Seely and El-Bassiouni (1983) for testing subsets of random effects. Several authors observe that the F -test has some interesting advantages with respect to other approaches (Hui et al., 2019), nevertheless it is often overlooked in applications to linear mixed models mainly because empirical evidence shows that in some situations this test can have poor power (Scheipl et al., 2008) partly because it is not sufficiently general or "flexible" for being applied in modern applications.

With our aim of generalizing and improving the statistical power of the F -test, we propose a test statistic that can be set up with the following steps:

1. Compute a "repeatable" F -type test as follows

- 1.1. Define a least squares statistic.
- 1.2. Compute the covariance matrices under the null and the alternative hypothesis of the above statistic.
- 1.3. Define a test computing the trace of the product of the two covariance matrices.
2. Repeat steps 1.1 – 1.3, changing the projected matrix of the least squares statistics to obtain a set of different tests.
3. Analyze this set simultaneously in a multiple testing approach.

Hypothesis testing approaches based on the equality of two positive definite matrices have a distinguished history in multivariate statistics, see for example, Roy (1953), Pillai (1955), Pillai and Jayachandran (1968) and Nagao (1973). The multiple testing procedures refer to any instance that involve the simultaneous testing of several hypotheses (Hunt et al., 2009).

Some of the main advantages of this approach (which are in part the same as those in an F -test) include: (i) its generality, being applicable to a general formulation of linear mixed models with or without knowledge of the design matrix, (ii) its exactness having a known distribution under the null and the alternative hypothesis with every sample size, (iii) its ease of computation, it does not require any estimate of the covariance matrix of random components. (iv) its statistical power, our evidence shows a greater power than the F -test.

The paper is organized as follows. Section 2 introduces some notations and defines the two stage linear mixed model. Section 3 motivates the F -type test statistic as a comparison between two positive semidefinite matrices. Section 4 deals with the F -type test and is divided into three subsections. Subsection 4.1 re-examines the F -test as the product of two covariance matrices, subsection 4.2 develops the F -type test statistic and subsection 4.3 examines the exact distribution function of the F -type test under the null and the alternative hypothesis. Section 5 outlines the method proposed for increasing the power of the test. This section is divided into three subsections. Subsection 5.1 describes the base scenarios for all simulations, subsection 5.2 discusses the statistical power of the F -type test, and the multiple testing approach used to improve the power of this test is analyzed in subsection 5.3. Section 6 contains

the conclusions.

2 Two-Stage Random Effects Model: Definitions and Notations

The linear mixed model for longitudinal data can be described as follows: $y_i = X_i^* \beta + Z_i v_i + u_i$, $i = 1, \dots, n$ where y_i is a $t_i \times 1$ vector of repeated measurements, X_i^* is a $t_i \times l$ matrix of explanatory variables, linked to the unknown $l \times 1$ fixed effect β , Z_i are the observed $t_i \times q$ covariates linked to the unknown $q \times 1$ random effects $v_i \sim N(0, \Omega_q)$, Ω_q is a $q \times q$ positive semidefinite matrix, $\Omega_q \succeq 0$, $u_i \sim N(0, \sigma^2 I_{t_i})$. The u_{ij} 's are *iid* so can be thought of as measurement error. We assume that u_i and v_i are independent.

Following Rocha and Singer (2017) we re-express the linear mixed model as a two-stage random coefficients model Laird (2004),

$$y_i = X_i \beta_i + u_i, \quad i = 1, \dots, n \quad (1)$$

where X_i is a matrix with k columns obtained from the elements of X_i^* and Z_i ; the columns of X_i are those common to X_i^* and Z_i plus those that are unique either to X_i^* or Z_i . The matrix Z_i is a subset of X_i , $Z_i = X_i R'$, where R is a $q \times k$ matrix containing ones and zeros. The elements of β_i are given by $\beta_j + v_{ij}$ if column j is common to X_i^* and Z_i , by β_j if column j is unique to X_i^* or by v_{ij} if column j is unique to Z_i . We can therefore write $\beta_i = \beta^* + v_i^*$, where null elements may be added to the original β and v_i vectors so that they have the same dimension.

Regarding (1) as a two stage model, it follows that $y_i | v_i \sim N(X_i \beta_i; \sigma^2 I_{t_i})$ is the first stage model and can be considered as a set of separate regression models for each unit. So in the first stage we may be able to obtain estimates of β_i and σ^2 using just the data from the i -th subject, i.e., $b_i = (X_i' X_i)^{-1} X_i' y_i$ and $s^2 = \frac{1}{df} \sum_{i=1}^n (t_i - k) s_i^2$, with $(t_i - k) s_i^2 = y_i' (I_{t_i} - X_i (X_i' X_i)^{-1} X_i') y_i$ and $df = N_t - nk = \sum_{i=1}^n (t_i - k)$. The estimated parameters, b_i 's, are independent and normally

distributed with mean β_i and variance-covariance matrix $\sigma^2(X_i'X_i)^{-1}$.

The β_i 's are random variables; to specify population parameters, at Stage 2 we assume that $\beta_i \sim N(\beta^*, \Omega_k)$, where $\Omega_k = R'\Omega_q R$ consists of Ω_q augmented with null rows and/or columns corresponding to the null elements in the random vectors v_i^* so that the marginal distribution of b_i is $N(\beta^*; \sigma^2(X_i'X_i)^{-1} + \Omega_k)$. We refer to the model described by the two-stage as linear mixed model, $H_1 : \Omega_k \succeq 0$, and to the model with $H_0 : \Omega_k = 0$ as Linear Regression Model.

Before closing this section, we introduce some additional definitions. Let $\bar{b} = \frac{1}{n} \sum_{i=1}^n b_i$ be the sample average of the individual least squares estimators. By hypothesis \bar{b} is normally distributed with mean β and variance $var(\bar{b}) = \frac{\sigma^2}{n} \bar{V} + \frac{1}{n} \Omega_k$ where $\bar{V} = n^{-1} \sum_{i=1}^n (X_i'X_i)^{-1}$. Simple algebra allows to show that $(b_i - \bar{b}) \sim N(0, \sigma^2 V_{ii} + \frac{n-1}{n} \Omega_k)$, $V_{ii} = \frac{1}{n} \bar{V} + \frac{n-2}{n} (X_i'X_i)^{-1}$ and $E(b_i - \bar{b})(b_j - \bar{b})' = \sigma^2 V_{ij} + h_{ij} \Omega_k$ with $V_{ij} = \frac{1}{n} \bar{V} - \frac{1}{n} (X_i'X_i)^{-1} - \frac{1}{n} (X_j'X_j)^{-1}$ and $h_{ij} = \frac{n-1}{n}$ if $i = j$, $h_{ij} = -\frac{1}{n}$ if $i \neq j$. V_{ii} and V_{ij} are $k \times k$ matrices. Let denote with V the $nk \times nk$ matrix whose (i, j) -th block is V_{ij} . V is positive semidefinite and symmetric with rank $(n-1)k$. Let $V^+ = X_D'(I - P_X)X_D$ be the Moore-Penrose pseudoinverse of V with block matrices V^{ij} , $X_D = diag(X_1, \dots, X_n)$.

3 The motivation of the test statistic

Denote with b a statistic linear in y and such that $E(b) = 0$. Let $A = Var(b|H_0)$ and $Var(b|H_1)$ be the covariance matrices of b when $\Omega_k = 0$ and $\Omega_k \succeq 0$ respectively. Let us suppose that b is defined so that $Var(b|H_1)$ can be written as a sum of two matrices, $A + B(\Omega_k)$ where $B(\Omega_k)$ denotes a covariance matrix depending on Ω_k which is zero if and only if $\Omega_k = 0$. Define the following parameter,

$$\theta = \frac{1}{r} tr [Var(b|H_0)]^+ Var(b|H_1) = \frac{1}{r} tr [A^+ (A + B)] \quad (2)$$

where $r = \text{rank}[\text{Var}(b|H_0)]$ and $B = B(\Omega_k)$ for notational simplicity. The parameter θ can be interpreted as a measure of the relative change of the covariance matrix of b with respect to the (pseudo)inverse covariance matrix of b under H_0 . In the scalar case θ reduces to $1 + B/A$ which highlights the interpretation of the above measure.

We can show the equality: $\frac{1}{r} \text{tr} [A^+ (A + B)]^+ = \frac{1}{r} \text{tr} A (A + B)^+$. The quantity $\frac{1}{r} \text{tr} A (A + B)^+$ can be interpreted as a measure of the share of A in $(A + B)^+$ given Ω_k . This expression has been proposed and analyzed by Theil (1963) in the estimation of regression coefficients with incomplete prior information.

In this work we construct a test statistic that can be viewed as an estimator of θ and is such that its expected value is proportional to θ . A test of this type is developed by defining an "appropriate" statistic b the outer product matrix of which, $S_b = b b'$, is such that $E(S_b|H_0) = \text{Var}(b|H_0)$ is known unless a scalar σ^2 and $E(S_b|H_1)$ can be written as the sum of $E(S_b|H_0)$ plus an unknown covariance matrix capturing randomness of parameters, $E(S_b|H_1) = E(S_b|H_0) + B(\Omega_k)$. By (2) we define

$$T = \frac{1}{r} \text{tr} [E(S_b|H_0)]^+ S_b \quad (3)$$

with the expected value equal to

$$E(T) = 1 + \frac{1}{r} \text{tr} [E(S_b|H_0)]^+ B(\Omega_k) \equiv \theta \quad (4)$$

When $\Omega_k = 0$ the parameter θ is equal to 1, $E(T|H_0) = 1$ and T moves around 1. If $\Omega_k \succeq 0$, θ is greater than 1, $E(T|H_1) > 1$ and T deviates from 1. Because the minimum eigenvalue of Ω_k is greater than or equal to zero and $\text{tr} E(S_b|H_0)^+ > 0$, $\frac{1}{r} \text{tr} [E(S_b|H_0)]^+ B(\Omega_k) \geq 0$. The greater this quantity, the farther θ is from one and the greater the deviation of T from 1 (everything else being equal). Larger values correspond to less "null-like" alternatives. As we shall see, the expression $\frac{1}{r} \text{tr} [E(S_b|H_0)]^+ B(\Omega_k)$ plays the same role as a "non-centrality parameter" of an F distribution.

Given the close relationship between θ and semidefiniteness of Ω_k , the set of hypotheses can also be written as

$$H_0 : \theta \leq 1 \ (\Omega_k = 0) \quad \text{against} \quad H_1 : \theta > 1 \ (\Omega_k \succeq 0) \quad (5)$$

The feasibility and the performance of T depend on an appropriate statistic b that we compute by two consecutive OLS regressions: OLS regression of a projection of a linear transformation of X_D to get \widehat{X}_D followed by OLS of y on \widehat{X}_D . Different projection matrices and different linear transformations of X_D in the first stage produce test-statistics with a different performance.

4 The F –type test statistic

This section describes a test motivated by a comparison of two covariance matrices of a least squares statistic and can be viewed as a generalization of the F –test. This section is structured as follows: Subsection 4.1 re-examines the F –test motivated by a comparison of two covariance matrices, subsection 4.2 describes the F –type test statistic and subsection 4.3 (with Appendix A) studies the exact distribution of the test under the null and the alternative hypothesis.

4.1 The F_W -test Statistic

Among the many expression proposed in the literature, in this paper we refer to the F –test given by Demidenko et al. (2012) and Demidenko (2013) which is used for comparative purposes.

$$F_W = \frac{y'(P_W - P_X)y}{y'(I - P_W)y} \frac{(N_t - \text{rank}(P_W))}{(n-1)q} = \frac{1}{(n-1)q} \frac{y'(P_W - P_X)y}{\widehat{s}^2} \quad (6)$$

where $y = [y'_1, y'_2, \dots, y'_n]'$, $P_W = W(W'W)^+W$, $P_X = X(X'X)^+X$, $W = [X | Z_D]$ is a block matrix with $X = [X'_1, X'_2, \dots, X'_n]'$ and $Z_D = \text{diag}(Z_1, \dots, Z_n)$.

The F_W -test given by (6) can also be formulated by comparing the covariance matrices under H_0 and H_1 of a statistic b_F computed by regressing y on $\widehat{X}_D = (P_W - P_X)X_D$. The least squares statistic

$$b_F = (X_D'(P_W - P_X)X_D)^+ X_D'(P_W - P_X)y$$

has covariance matrices under H_0 and H_1 given by $E(b_F b_F' | H_0) = \sigma^2 [X_D'(P_W - P_X)X_D]^+ = \sigma^2 V_F^+$ and $E(b_F b_F' | H_1) = \sigma^2 V_F^+ + V_F^+ V^+ (I_n \otimes \Omega_k) V^+ V_F^+$ respectively. Then, the expression (6) can also be obtained as

$$F_W = \frac{1}{q(n-1)} \text{tr} E(b_F b_F' | H_0) + \frac{b_F b_F'}{\widehat{s}^2} \quad \text{with} \quad E(F_W) = \frac{N_t - \text{rank}(P_W)}{[N_t - \text{rank}(P_W)] - 2} \theta_F \quad (7)$$

where $q = \text{rank}(Z_i)$ and $\theta_F = 1 + \frac{1}{(n-1)q} \text{tr} V^+ V_F^+ V^+ \left(I_n \otimes \frac{\Omega_k}{\sigma^2} \right)$.

When $\Omega_k = 0$, $\theta_F = 1$ and F_W takes values around the expected value of an F distribution with $(n-1)q$ and $N_t - \text{rank}(P_W)$ degrees of freedom. If $\Omega_k \succeq 0$, $\theta_F > 1$ and F_W deviates from one. The greater F_W the stronger the evidence against an LR model.

The F_W statistic tests randomness in "relative" terms in the sense that the alternative hypothesis is determined by the ratio between "randomness" and σ^2 . That is, it depends on the factor $\left(I_n \otimes \frac{\Omega_k}{\sigma^2} \right)$.

Motivated by a comparison of two covariance matrices of a least squares statistic, the next subsection describes an F -type test that can be seen as a generalization of F_W .

4.2 The F -type test: T

Let define $Q_D = I_n \otimes Q_{p \times k}$ where $Q_{p \times k}$, $p < k$, is a semi-orthogonal matrix such that $Q Q' = I_p$. Let compute a statistic following a two-stage least squares approach. In the first stage we project the block diagonal matrix $X_D Q_D'$ on the kernel of X getting $\widehat{X}_D = (I - P_X) X_D Q_D'$. In the

second stage we regress y on \widehat{X}_D obtaining the following statistic,

$$b = (Q_D V^+ Q_D')^+ Q_D X_D' (I - P_X) y \quad (8)$$

According to the assumptions of the random model (section 2), b is normally distributed with

$$E(b) = 0, \quad \text{Var}(b|H_0) = \sigma^2 (Q_D V^+ Q_D')^+ = \sigma^2 V_Q^+ \quad (9)$$

$$\text{Var}(b|H_1) = \sigma^2 V_Q^+ + C (I_n \otimes \Omega_k) C', \quad \text{with } C = V_Q^+ Q_D V^+ \quad (10)$$

In applications Q is a zero-one matrix appropriately defined for extracting p columns from the matrix X_i . Observe that if $Q = I_k$, $\widehat{X}_D = (I - P_X) X_D$ and $b = V X_D' (I - P_X) y$ is the least squares statistic obtained by stacking $b_i - \bar{b}$, $i = 1, \dots, n$ one under the other where b_i is the vector of OLS estimator computed on the i -th unit and $\bar{b} = \frac{1}{n} \sum_{i=1}^n b_i$ (section 2). If the matrix of covariates Z_i is specified, then Q can be defined so that $X_i Q' = Z_i$, that is $Q_D' = (X_D' X_D)^{-1} X_D' Z_D$. In this case the statistic b allows to obtain a test statistic very close to F_W . In the absence of additional information on Z_i , the matrix Q can be defined as a unit vector, $Q' = (X_i' X_i)^{-1} X_i' x_j$ where x_j is the j -th column of X_i . As we shall see this is the instrument used to increase the power of the test statistic.

Then, given b we compute the covariances matrices under H_0 , under H_1 and the test statistic

$$T = \frac{1}{(n-1)p} \text{tr} V_Q \frac{b b'}{s^2} = \frac{1}{(n-1)p} b' \left(\frac{V_Q}{s^2} \right) b \quad (11)$$

where the sample variance s^2 is defined in Section 2. The parameter θ is given by

$$\theta = 1 + \frac{1}{(n-1)p} \text{tr} V^+ L V^+ \left(I_n \otimes \frac{\Omega_k}{\sigma^2} \right) \quad (12)$$

where $L = Q_D' V_Q^+ Q_D$ and $p = \text{tr} V_Q V_Q^+ = \text{rank}(V_Q V_Q^+) = \text{rank}(Q)$. The trace of the matrix $V^+ L V^+ \left(I_n \otimes \frac{\Omega_k}{\sigma^2} \right)$ can be written as the trace of the product of two matrices $S \frac{\Omega_k}{\sigma^2}$, $S = \sum_{i=1}^n L^i$

where L^{ii} is the (i, i) -th block matrix of V^+LV^+ . Dividing S by $(n-1)$, we get an "average" matrix \bar{S} .

Let $\eta_i \geq 0, i = 1, \dots, k$ be the eigenvalues of the product $\bar{S}\Omega$. The η_i can be interpreted as eigenvalues "adjusted" in magnitude so that a comparison with σ^2 makes sense. We define $\frac{1}{p}tr\bar{S}\Omega_k = \frac{1}{p}\sum_{i=1}^k \eta_i = M(\eta_i)$ as an "average" measure of randomness "adjusted" by the covariance matrix V . Observe that this "average" becomes a "true" arithmetic if $rank(\Omega_k) = rank(V_QV_Q^+)$. As we shall see, a reduction of the difference between these two ranks determines an improvement of the power of the test.

In the light of these considerations, the parameter θ (formula (12)) can also be written as

$$\theta = 1 + \frac{M(\eta_i)}{\sigma^2}, \quad \text{with} \quad E(T) = \frac{df}{df-2} \theta, \quad df = N_t - nk \quad (13)$$

When $\Omega_k = 0$, $M(\eta_i)$ is equal to zero, $\theta = 1$, $E(T) = \frac{df}{df-2}$ and T takes values around the expected value of an $F_{(n-1)p, N_t - nk}$ (see next section). If $\Omega_k \succeq 0$ then $\frac{M(\eta_i)}{\sigma^2} > 0$ and θ is greater than 1. T deviates from 1 and the farther $\frac{M(\eta_i)}{\sigma^2}$ from zero, the greater T , everything else being equal. The greater T the stronger the evidence against an LR model. The parameter θ plays the same role as the non centrality parameter of an F -distribution. As we shall see, if θ increases, the shape of the distribution of T shifts to the right and a larger percentage of the curve moves to the right of the critical value by improving the statistical power. An expression similar to (13) for the non centrality parameter of a non-central F distribution can be traced in the book of Searle (1971) (p. 51).

4.3 Probability density function of T

In Appendix A we show that for any matrix Q , the sample statistic T has the same distribution as the random variable

$$W = \frac{T^*}{\frac{df}{df}(s^2/\sigma^2)} = \frac{df T^*}{df(s^2/\sigma^2)} \quad \text{where} \quad T^* = \text{tr} V_Q \frac{bb'}{p(n-1)\sigma^2} \quad (14)$$

If $\Omega = 0$, $T \sim F((n-1)p, N_t - nk)$, if $\Omega \succeq 0$, we define the probability density function of T using the series representation of Moschopoulos (1985) expressed here in terms of a generalized F -distribution (GF -distribution). This representation results particularly useful for deriving the distribution function and for computing quantiles after switching the order of summation and integration. Appendix A shows that the probability density function of W can be expressed as

$$f_T(w) = \sum_{k=0}^{\infty} p_k GF\left(\rho + k, \frac{df}{2}, \frac{\beta_1}{2}\right) \quad (15)$$

where the weights, p_k and the other notations are described in In Appendix A.

The distribution function of the random variable T , $F_T(w) = P(T \leq w)$, is readily available from (15) by term-by-term integration, i.e.

$$F_T(w) = \sum_{k=0}^{\infty} p_k \int_0^w GF\left(\rho + k, \frac{df}{2}, \frac{\beta_1}{2}\right) \quad (16)$$

The interchange of the integration and summation above is justified from the uniform convergence

Quantiles of the test statistic T are easily obtained by finding a root, w , of $F_T(w) = \alpha$, $0 \leq \alpha \leq 1$ is the probability on the left tail. The computation is done with the function "uniroot.all" of the package "rootSolve" of R software (Soetaert and Herman, 2009).

In most statistical software there is a function that computes the generalized F -distribution. In this paper computations are made with R (R Core Team, 2014) where a library (GB2) (or

flexsurv) allows us to compute density, distribution function, quantile function and random generation for the GF -distribution.

5 Improving statistical power

The construction of the F -type test, T , is based on the following steps: (i) Define a matrix Q and compute the least squares statistic b ; (ii) compute the covariance matrices of b under the null and the alternative hypothesis; (iii) derive T as the trace of the product of the two covariance matrices of b . For any matrix Q , T is an exact test, very flexible with a statistical power at least as large as the power of the F -test given by (7). The next subsections discuss a method to improve the power. More precisely, subsection 5.1 describes the "base" scenario for all simulations (unless otherwise specified), subsection 5.2 analyzes the statistical power of T and discusses how to improve it, and finally, subsection 5.3 defines a test working in a multiple testing approach.

5.1 "Base" Scenario for simulations

To allow the maximum of generality and arbitrariness, we define the following scenario for simulations of T , unless otherwise specified.

- (i) We set the number of parameters $k = 6$ and the number of units $n = 8$. The number of observations per units, t_i , $i = 1, \dots, n$, are drawn randomly from a uniform distribution, $U(k + 1, 3k)$. The vector of "fixed" regression coefficients, β , is generated randomly from a $N(10, 2)$.
- (ii) For each units, the columns of X_i are drawn from an $N(\text{mean}, \text{sqrt})$ where the *mean* is random from a uniform distribution, $U(10, 20)$ and *sqrt* is random from $U(2, 10)$. All the elements in the first column are 1. Given X_i $i = 1, \dots, n$ we construct the variance covariance matrix, V , the pseudoinverse, V^+ and the block matrices V^{ij} , $\forall ij$.

- (iii) The two-stage random effect model is specified as follows: first we choose the column rank of Z_i , q , by sampling a number between zero and k , second, each column of Z_i is the square of the random variable generated by the uniform distribution on the interval $[0, 1]$. The q columns of X_i are replaced by Z_i so that $Z_i \subseteq X_i$.
- (iv) The matrix Ω_k is defined starting from a positive definite matrix, Ψ , computed as follows. First, we randomly generate eigenvalues from a uniform distribution with a prefixed mean. Following, the columns of a randomly generated orthogonal matrix are used as eigenvectors. Ψ is then constructed by diagonalization (Qiu and Joe., 2015). The matrix Ω_k is obtained from Ψ by selecting the q columns and rows concerning random components and zero elsewhere. Ω_k so defined has rank q but a rank less than q is allowed. This approach enables us to simulate fixing a prior mean of eigenvalues of Ω_k .
- (v) Given Ω_k and V^{ii} we construct the eigenvalues of Ω_k in the metric S^{-1} , after which the arithmetic mean $M(\eta_i)$ is computed.
- (vi) Let $\tau = M(\eta_i)/\sigma^2$ be the ratio of "randomness" on σ^2 . Then, σ^2 is computed indirectly, fixing in advance $\tau = 0.1, 0.5, 1, 1.5, \dots$. Given $M(\eta_i)$, a small value of τ implies a large σ^2 and an *LME* model is dominated by an *LR* model. The larger the variance σ^2 , the lower the power of the test, and the greater the probability of failing to reject the null hypothesis everything else being equal. It is vice-versa when τ is large.

5.2 Discussion of statistical power

Due to the fact that many factors influence the power of test statistic T , an exhaustive analysis is too complex. In this section we limit our analysis to those factors that modify the parameter θ . In particular we discuss σ^2 , Ω_k and the number of units n by using the "base" scenario for simulations unless otherwise specified. The results compare F_W with T_j $j = 1, \dots, k$.

1. Given Ω_k and σ^2 , for any matrix Q , the larger the number of units, the greater the power of the test, everything else being equal. Note that a larger n reduces the variance of T . As

the sample size increases, the sampling distribution of T under the null and the alternative hypotheses is concentrated around θ . Fig.:1(a) shows the size-power tradeoff curves of the test statistic T for different n with $Q = I_k$, $\text{rank}(\Omega_k) = 4$, $\tau = \frac{M(\eta_i)}{\sigma^2} = 0.35$. This means that on average, the randomness of the model is 35% of σ^2 . The plot shows the consistency and unbiasedness of the test.

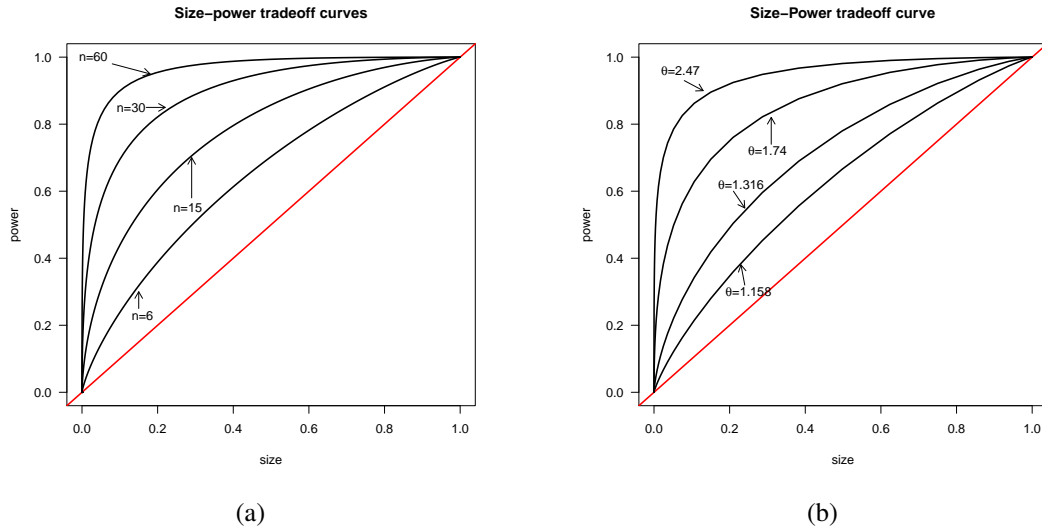


Figure 1: Size-power tradeoff curves of T : (a) different n , $\tau = 0.35$ and $\text{rank}(\Omega) = 4$; (b) different alternative hypotheses, θ , with $n = 7$ and $\text{rank}(\Omega) = 4$; .

2. Given Ω_k , the larger the variance σ^2 , the closer θ is to one, and the lower the power of the test. Conversely, the smaller the variance, the farther θ is from one, and the greater the power of the test. Fig.:1(b) shows the size-power tradeoff curves of the test T for different values of θ and $n = 7$. The reciprocal of the parameter θ , $\theta^{-1} = \sigma^2 / (\sigma^2 + M(\eta_i))$, may be viewed as a measure of the share of σ^2 in "total" variability. It ranges between zero and one. When data come from an LR model $\theta^{-1} = 1$, when at least one eigenvalue is greater than zero, $\theta^{-1} < 1$. The closer θ^{-1} is to zero, the stronger the evidence against an LR model. In applications θ^{-1} may have a more immediate interpretation than θ . Fig.:1(b) shows a share that moves from $\theta^{-1} = 1/1.158 = 0.86$ to $\theta^{-1} = 1/2.47 = 0.4$.
3. Given σ^2 , if $\Omega_k \succeq 0$, $\frac{M(\eta_i)}{\sigma^2} > 0$, θ is greater than 1 and T deviates from 1. The farther $M(\eta_i)$ is from zero, the greater T is and the stronger the evidence against an LR model.

We recall that the magnitude of $M(\eta_i) = \frac{1}{p} \sum_{i=1}^k \eta_i$ depends both on the rank of Ω_k (on how many eigenvalues are zero) and on the rank of the projector $V_Q V_Q^+$ given by the rank of the matrix Q (the denominator, p). The quantity $M(\eta_i)$ "appropriately" captures the randomness of parameters when it is "true" arithmetic mean, that is, when $\text{rank}(\Omega_k) = \text{rank}(V_Q V_Q^+)$. If the rank of Ω_k is less than the rank of the projection matrix (information unknown in applications) the effect of randomness could be overlooked (undersized).

An important "instrument" useful to define "appropriately" $M(\eta_i)$ could be the specification of the model. Let's suppose that a matrix Z_i (with rank q) of covariates is defined, then, implicitly we assume that $\text{rank}(\Omega_k) \leq q$. In this case there are two possible options for the choice of Q . We could ignore the additional information coming from Z_i and set $Q = I_k$, or define a semi-orthogonal $q \times k$ matrix, Q such that $Z_i = X_i Q'$. If $\text{rank}(\Omega_k) = q$, the projection of Z_i on the kernel of X allows to construct a test statistic very close to F_W and more powerful than a test computed with $Q = I_k$. If the rank of Ω_k is less than q , the matrix Q does not allow to "capture" the "full" randomness. In this case $M(\eta_i)$ is not a "true" arithmetic mean, the number of non zero eigenvalues of the numerator is less than the denominator. Therefore, any attempt to improve the performance of the test goes through the definition of $M(\eta_i)$ as a "true" arithmetic mean. To achieve this goal we propose to define Q as a row vector. This approach produces a set of k test statistics (one for each column of the matrix X_i) that are analyzed within a multiple testing procedure.

5.3 A multiple-testing approach: the statistic T_B

The above analysis of $M(\eta_i)$ suggests a way to improve the power of the test statistic T : define Q so that $M(\eta_i)$ is a "true" arithmetic mean. The consequence of this approach is the computation of k test statistics, each of which tests the same null hypothesis, $H_0 : \Omega = 0$. To emphasize this column-by-column approach, the quantities T and θ will be indicated as T_j , θ_j $j = 1, \dots, k$.

Following, we compute a set of k correlated tests, T_j $j = 1, \dots, k$, which show the following

features,

- i.- The expected value and the shape of T_j depend on the parameter θ_j which is similar (in value) for each individual test. As a consequence the statistics T_j show similar summary statistics (see Table 2).
- ii.- They have the same distribution under the null hypothesis: $T_j \sim F(n-1, N_t - nk)$. Under the alternative hypothesis the functional form is described in section 4.3 and depends on θ_j (the "non-centrality parameter"). These parameters are close from each to the others (see Table 2). Figures 2.1 – 2.5 show the simulated T_j and the exact density functions under H_0 (long dashed lines) and under the alternative hypothesis (solid line).
- iii.- Both statistical powers (and p-values) are very close for each statistic T_j (see Table 1)
- iv.- All individual tests are more powerful than F_W (see Table 1).

Therefore, given $T_j, j = 1, \dots, k$, the problem is how to summarize this set of statistics so that we reject the null hypothesis without losing the statistical power of the individual tests. At first we could proceed by computing the arithmetic mean $\bar{T} = \frac{1}{k} \sum_{i=1}^k T_j$ and then making inference with \bar{T} . This approach is not analyzed, instead we deal with the problem of the "synthesis" within a "multiple testing" procedure.

Let us transform the p-values associated with each individual tests T_j into k realizations of Bernoulli variables and denote with T_B the number of rejections (p-value less than a level α) after performing the k tests individually. If $H_i, i = 0, 1$ is true, then $T_B \sim BB(k, \gamma, \phi)$ where γ denotes the probability parameter and ϕ the over-dispersion parameter. The beta binomial distribution is analyzed by simulation using the HRQoL package of R-program Najera-Zuloaga et al. (2017) to estimate the parameters. We observed the following results:

- i.- All simulations show that the estimate of γ (by method of moments) is always equal to the arithmetic mean of the simulated power of the individual tests T_j . This leads us to state that the "true" probability parameter is equal to the arithmetic mean of "true" statistical powers. Following, we set $\gamma = \frac{1}{k} \sum_{j=1}^k P(T_j > cv | H_j : \theta_j \geq 1)$ where cv is the critical value

equal for each statistic. Under H_0 , γ is known, equal to the probability of Type I error, $\gamma = \alpha$. Under the alternative hypothesis it is unknown but as it is the arithmetic mean of the statistical powers of individual test it does not lose the power of the test.

ii.- An analysis of the estimates of the overdispersion parameter is more complex. Our findings show two aspects: under H_0 we estimate an intraclass correlation coefficient around 0.5, under H_1 the larger is θ_j , the larger is the simulated ϕ with a magnitude around the arithmetic mean of θ_j . These observations induced us to set $\phi = \frac{1}{k} \sum_{j=1}^k \theta_j$. Under H_0 $\phi = 1$ which reflects the intraclass correlation coefficient.

The above observations are highlighted in Table 3 and Figures 2.(i) – 2.(iii).

Therefore, provided that H_0 is true, we reject the null hypothesis when T_B is greater than or equal to the $1 - \alpha$ percentile of the beta binomial distribution $BB(k, 0.05, 1)$. More precisely, with $k = 5$ we reject the null hypothesis if $T_B > 2 = 0.0406$. We can look at the Table 3 and evaluate the statistical power. It is equal to $\sum_{i=3}^5 P(T_B = i | H_1) = 0.482$ which is equal to a mean of the simulated power of the individual tests. Therefore, it is an estimate of the true power $\gamma = 0.503$.

n. replic. =1000	Number of units			
	$n = 10$	$n = 20$	$n = 50$	$n = 100$
Test Statistic				
F_W	0.24	0.37	0.61	0.85
T_1	0.37	0.57	0.83	0.99
T_2	0.34	0.56	0.79	0.98
T_3	0.36	0.59	0.84	0.99
T_4	0.35	0.55	0.82	0.98
T_5	0.36	0.56	0.80	0.99

Table 1: Statistical Power, $\alpha = 0.05$, $\text{rank}(\Omega_k = 3)$, $\tau = 0.18$

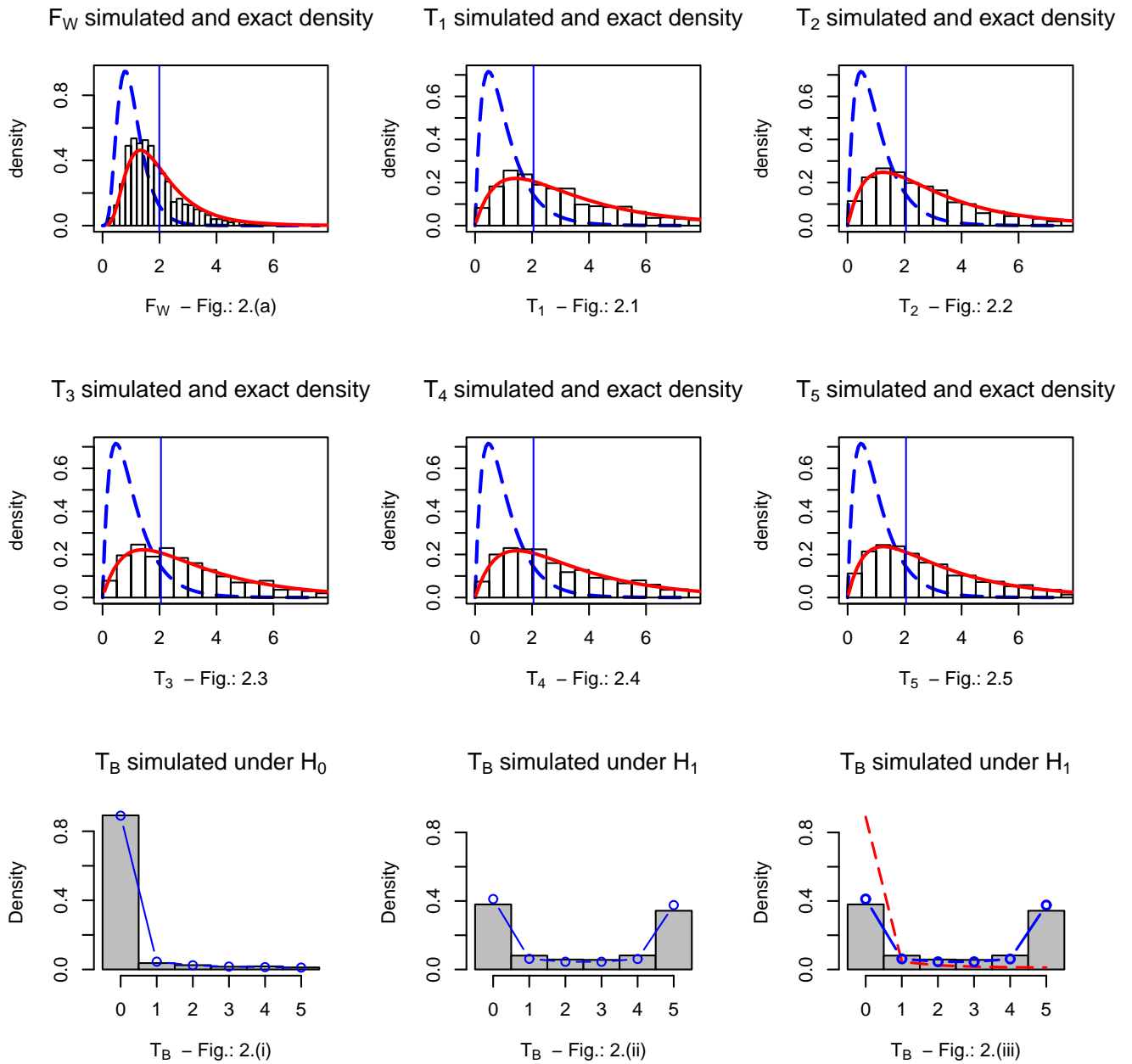


Figure 2: - Fig.: 2.(a) and Fig.: 2.1 – 2.5 show simulated histogram of F_W and T_j , $j = 1, \dots, 5$ under H_1 . The longdash line is the F-Fisher distribution under H_0 . The vertical line is the critical value and the solid line is the exact distribution. - Fig.: 2.(i) and Fig.: 2.(ii) represent simulated barplot of T_B under H_0 and under H_1 . The solid line (with circle) draws the beta binomial distribution. - Fig.: 2.(iii) shows the barplot under H_1 and beta binomial distributions under H_0 and H_1 .

Summary Statistics	test statistics						
	T_1	T_2	T_3	T_4	T_5	F_W	
2.5th percentile	0.367	0.321	0.365	0.3758	0.329	0.5667	
5th percentile	0.5435	0.476	0.539	0.555	0.487	0.688	
Median	2.778	2.456	2.749	2.807	2.568	1.779	
Mean	3.558	3.151	3.52	3.581	3.321	2.05	
95th percentile	9.211	8.18	9.108	9.224	8.706	4.33	
97.5th percentile	11.253	9.989	11.13	11.26	10.65	5.112	
θ_j	3.349	2.966	3.133	3.370	3.126	1.953	$\phi = 2.662$
θ_j^{-1}	0.298	0.337	0.302	0.297	0.319		
True power	0.523	0.463	0.517	0.523	0.485	0.336	$\gamma = 0.503$
Simulated power	0.504	0.446	0.501	0.498	0.459		$\hat{\gamma} = 0.482$

Table 2: Summary statistics of T_j

T_B	T_B			
	Simulated $T_B H_0$	$BB(5, 0.05, 1)$	Simulated $T_B H_1$	$BB(5, \gamma, \phi)$
0	0.885	0.89	0.38	0.411
1	0.056	0.045	0.081	0.0624
2	0.019	0.024	0.058	0.045
3	0.014	0.0167	0.056	0.0446
4	0.014	0.013	0.082	0.0614
5	0.012	0.011	0.343	0.375

Table 3: Beta Binomial distribution of T_B

6 Conclusions

A combination of an F -type test with a multiple testing approach allows for constructing a test statistic which generalizes and significantly improves the power of an F -test. Our analysis is based on a two-stage approach where in the first stage we construct an F -type test we called "repeatable" in the sense that by changing the projected matrix which defines the least squares statistic we can apply the test repeatedly to the same data. This produces a set of statistics associated with a set of p-values used for assessing the significance. In the second stage we transform this set of p-values into a sequence of Bernoulli trials, the sum of which (number of rejections) is analyzed with a beta binomial distribution. Our simulations show an increase in

statistical power by up to 25% – 30% with respect to an F -test. In the light of these results we believe that our two-stage approach based on a combination of a "repeatable" F -type test with a multiple testing approach may suggest a procedure for improving statistical power in linear mixed models. Future work should entail the refining of the second stage of the procedure by exploiting different ways for assessing the Bernoulli trials.

Appendix A Density and moments of the test statistic T

Let consider the following quadratic form $T^* = trV_Q \frac{bb'}{p(n-1)\sigma^2}$. obtained from (11) by replacing s^2 with σ^2 . According to the assumptions of the model, T^* has the same distribution as $\frac{1}{(n-1)p} \sum_{i=1}^{n-1} \left(1 + \frac{\lambda_i}{\sigma^2}\right) Z_i^2$ where Z_i^2 are independent central χ^2 random variables each with one degree of freedom; λ_i , $i = 1, \dots, n-1$ are the eigenvalues of the product $V_Q \left(V_Q^+ + C \left(I_n \otimes \frac{\Omega_k}{\sigma^2} \right) C' \right)$ (Mathai and Provost (1992) section 3.1a.2, *singular case*, p. 35).

When $\Omega = 0$, T^* has a gamma distribution, $G\left(\alpha = \frac{(n-1)p}{2}, \beta = \frac{2}{(n-1)p}\right)$ and $(n-1)pT^*$ is distributed as a χ^2 with $(n-1)p$ degrees of freedom.

When $\Omega \succeq 0$ T^* is a sum of gamma distributions each of which with same shape parameter, $(n-1)p/2$, but r different scale parameters, $2\left(1 + \frac{\lambda_i}{\sigma^2}\right)/((n-1)p)$.

Then, the sample statistic $T = \left(\frac{\sigma^2}{s^2}\right) T^*$ has the same distribution as the random variable

$$W = \frac{T^*}{\frac{df}{df}(s^2/\sigma^2)} = \frac{df T^*}{df (s^2/\sigma^2)}, \quad df = N_t - nk \quad (17)$$

where the numerator is a sum of $(n-1)$ gamma, $G\left(\alpha = 1/2, \beta_i = 2\left(1 + \frac{\lambda_i}{\sigma^2}\right) \frac{df}{(n-1)p}\right)$, the denominator can be seen as a gamma, $G(\alpha = df/2, \beta = 2)$.

When $\Omega = 0$, W is the ratio of two chi-squared variates divided by the corresponding number of degrees of freedom, thus $T \sim F((n-1)p, N_t - nk)$. If $\Omega \succeq 0$ the distribution is more complex. Using the single gamma series representation proposed by Moschopoulos

(1985) we can write the probability density function of W as

$$f_T(w) = \sum_{k=0}^{\infty} p_k \frac{G(\rho + k, \beta_1)}{G\left(\frac{df}{2}, 2\right)} \quad (18)$$

where $p_k = C\delta_k$, $\beta_1 = \min_i\{\beta_i\}$, $C = \prod_{i=1}^{n-1} \left(\frac{\beta_1}{\beta_i}\right)^{\alpha_i}$, $\rho = \sum_{j=1}^{n-1} \alpha_j$ and the coefficients δ_k can be obtained recursively by the formula

$$\begin{cases} \delta_0 & = 1 \\ \delta_{k+1} & = \frac{1}{k+1} \sum_{i=1}^{k+1} \left[\sum_{j=1}^{n-1} \alpha_j \left(1 - \frac{\beta_1}{\beta_j}\right)^i \right] \delta_{k+1-i}, \quad k = 0, 1, 2, \dots \end{cases}$$

The expression (18) is the ratio of two independent gamma random variables then we can re-propose the probability density of W as a generalized F -distribution (GF -distribution) getting,

$$f_T(w) = \sum_{k=0}^{\infty} p_k GF\left(\rho + k, \frac{df}{2}, \frac{\beta_1}{2}\right)$$

where GF is for generalized F -distribution.

Moments of T of order s are given by

$$E(T^s) = \sum_{k=0}^{\infty} p_k E(X_{GF}^s)$$

where

$$E(X_{GF}^s) = (\beta_1/2)^s \frac{\Gamma(\rho + k + s)\Gamma(\gamma - s)}{\Gamma(\rho + k)\Gamma(\gamma)}$$

are moments of order s of a GF -distribution. Simple algebra allows us to write

$$E(T^s) = \frac{(\beta_1/2)^s}{(\gamma - 1) \dots (\gamma - s)} \sum_{k=0}^{\infty} p_k (\rho + k)_s \quad (19)$$

where $(.)_s$ is the Pochhammer symbol for rising factorial, $\gamma = df/2$.

References

- Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. In *Second international symposium on information theory* (eds. H. Akaike, B. N. Petrov and F. Csaki), 267–281. Akadèmiai Kiadó.
- Demidenko, E. (2013) *Mixed Models. Theory and application with R*. 2nd ed. Hoboken, New Jersey, Wiley.
- Demidenko, E., Sargent, J. and Onega, T. (2012) Random effects coefficient of determination for mixed and meta analysis models. *Cooun Stat Theory Methods*, **41(6)**, 953–969.
- Drikvandi, R., Verbeke, G., Khodadadi, A. and Nia, V. (2013) Testing multiple variance components in linear mixed-effects models. *Biostatistics*, **14**, 144–159.
- Fitzmaurice, G., Lipsitz, S. and Ibrahim, J. (2007) A note on permutation tests for variance components in multilevel generalized linear mixed mdels. *Biometrics*, **63**, 942–946.
- Hui, F., Muller, S. and Welsh, A. (2019) Testing random effects in linear mixed models: another look at the f-test (with discussion). *Australian and New Zealand Journal of Statistics*, **61(1)**, 61–84.
- Hunt, D., Cheng, C. and Pounds, S. (2009) The beta-binomial distribution for estimating the number of false rejections in microarray gene expression studies. *Computational Statistics and Data analysis*, **53**, 1688–1700.
- Jiang, J., Rao, J. S., Gu, Z. and Nguyen, T. (2008) Fence methods for mixed model selection. *Ann. Statist.*, **36**, 1669–1692.
- Laird, N. (2004) *Chapter 5: Random effects and the linear mixed model*, vol. Volume 8 of *Regional Conference Series in Probability and Statistics*, 79–95. Beechwood OH and Alexandria VA: Institute of Mathematical Statistics and American Statistical Association. URL<https://projecteuclid.org/euclid.cbms/1462106081>.
- Laird, N. M. and Ware, J. K. (1982) Random effect models for longitudinal data. *Biometrics*, **38**, 963–974.
- Mathai, A. M. and Provost, S. B. (1992) *Quadratic forms in random variables: theory and applications*. Marcel Dekker Inc.
- Moschopoulos, P. G. (1985) The distribution of the sum of independent gamma random variables. *Ann. Inst. Statist. Math.*, **37, Part A**, 541–544.
- Muller, S., Scealy, J. L. and Welsh, A. H. (2013) Model selection in linear mixed models. *Statistical Science*, **28, No. 2**, 135–167.
- Nagao, H. (1973) On some test criteria for covariance matrix. *Annals of Statistics*, **1**, 700–709.
- Najera-Zuloaga, J., Lee, D.-J. and Arostegui, I. (2017) *HRQoL: Health Related Quality of Life Analysis*. URL<https://CRAN.R-project.org/package=HRQoL>. R package version 1.0.
- Pillai, K. C. S. (1955) Some new test criteria in multivariate analysis. *Ann. Mathem. Stat.*, **No. 1, 26**, 117–121.
- Pillai, K. C. S. and Jayachandran, K. (1968) Power comparison of tests of equality of two covariance matrices based on four criteria. *Biometrika*, **55**, 335–342.
- Qiu, W. and Joe., H. (2015) *clusterGeneration: Random Cluster Generation (with Specified Degree of Separation)*. URL<https://CRAN.R-project.org/package=clusterGeneration>. R package version 1.3.4.

- R Core Team (2014) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Rocha, F. M. M. and Singer, J. M. (2017) Selection of terms in random coefficient regression models. *Journal of Applied Statistics*, DOI: **10.1080/02664763.2016.1273884**.
- Roy, S. (1953) On a heuristic method of test construction and its use in multivariate analysis. *Annals of Mathematical Statistics*, **24**, 220–238.
- Samuh, M., Grilli, L., Rampichini, C., Salmaso, L. and Lunardon, F. (2012) The use of permutation tests for variance components in linear mixed models. *Communications in Statistics-Theory and methods*, **41**, 3020–3029.
- Scheipl, F., Greven, S. and Kuechenhoff, H. (2008) Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. *Computational Statistics and Data Analysis*, **52**, 3283–3299.
- Schwarz, G. (1978) Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.
- Searle, S. R. (1971) *Linear Models*. New York: Wiley.
- Seely, J. and El-Bassiouni, Y. (1983) Applying Wald's variance component test. *The Annals of Statistics*, **11**, No 1, 197–201.
- Sinha, S. (2009) Bootstrap tests for variance components in generalized linear mixed models. *The Canadian Journal of Statistics*, **37**, 219–234.
- Soetaert, K. and Herman, P. M. (2009) *A practical Guide to Ecological Modelling. Using R as a Simulation Platform*. Springer.
- Theil, H. (1963) On the Use of Incomplete Prior Information in Regression Analysis. *Journal of America Statistical Association*, Vol. **58**, 401–414.
- Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *Journal of Roy. Statist. Soc. Ser. B*, **58**, 267–288.
- Vaida, F. and Blanchard, S. (2005) Conditional Akaike information for mixed-effects models. *Biometrika*, **92** (2), 351–370.
- Wald, W. (1947) A note on regression analysis. *The Annals of Mathematical Statistics*, **18**, 586–589.

