©2022 Ada Academica https://adac.ee

Eur. J. Stat. 2 (2022) 12
doi: 10.28924/ada/stat.2.12

Testing Fixed and Random Terms in Linear Mixed Models

Marco Barnabani

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

Abstract. In linear mixed models the selection of fixed and random effects using a testing hypothesis approach brings up several problems. We deal with the boundary point problem emerging when no randomness is hypotesized and the confounding impact of randomness on the coefficients arising when fixed effects are tested. The test statistics are defined by a ratio of two quadratic forms derived from ordinary least squares, are simple, sufficiently general, easy to compute, with known finite sample properties. The test statistic on randomness has a known exact distribution, the density of the statistic on fixed effect is unknown and is approximated by a noncentral $F-$distribution. The goodness-of-approximation and the selection approach is examined in-depth by simulation. The method proposed in this paper must be seen as complementary to existing selection procedures widening and enriching all information necessary for taking a decision.

1. Introduction

Linear mixed-effect models are widely used to analyze longitudinal and repeated measurements data because of their flexibility and relative simplicity. In particular, they are used in the form of random coefficient regression model for analyzing the specification of the within-unit covariance structure. In this context, deciding which random or fixed coefficient should be included in the model becomes a fundamental problem.

In order to address the issue of which model is more suitable, one might use standard model selection measures based on information criteria. These approaches rely on the choice of models that minimize (an estimate of) a specific criterion which usually involves a trade-off between the closeness of the fit to the data and the complexity of the model; see [1] for a comprehensive review of model selection in linear mixed models. All these methods deal with the problem of selection working simultaneously with both fixed and random component resulting computationally burdensome. To overcoming this computational problem the penalized likelihood methods (dating back to [2]) are proposed. These procedures treat the selection problem via a separate selection approach to avoid the impact of random or fixed effects from one set of coefficient in the other set.

Received: 15 May 2022.
Key words and phrases. generalized F-distribution; hypothesis testing; linear mixed models; selection procedure.

Often the fixed effects are selected by first keeping all the random effects in the models, then the random effects are selected by keeping selected fixed effects from the previous step. The two steps are implemented iteratively until the parameters in the model no longer change. [3], [4] proposed separate penalties for the fixed and random effects that are summed together. [5], [6], [7] proposed two-stage methods where the fixed and random effects selection are performed independently. Note that to remove random effects from a model, entire rows and columns of the covariance matrix must be eliminated to form the final working model. Accounting for these issues, the unknown covariance matrix of the random effects is sometimes replaced with a suitable proxy matrix (see for example an orthogonalization-based approach proposed by [8]).

Although the penalized likelihood methods may avoid the need to search through the entire model space, it may remain computationally intensive. A further complication of these methods is how to define a "good" penalty function (for a discussione see [9]) and how to perform the shrinkage appropriately. Moreover, the results obtained can be interpreted only asymptotically, assigning to simulations the analysis of the behaviour in small samples.

Because the selection of terms is closely related to hypothesis testing, the choice of fixed and random coefficients to be included in the model could be conducted by assessing the significance of appropriate test statistics. As known, this approach brings up several problems. Testing randomness is associated to the so-called boundary point problem. Testing fixed parameters is related to the "confounding" impact of the random effect on the coefficients which can lead to a misleading interpretation of the significance of the statistic. This is the same problem encountered in penalty function approaches. We agree with some authors [10] that none of the proposed procedures should be used as the only procedure to select the fixed and random coefficients in linear mixed models. They should be taken as complementary and the decision should be based on all information available.

The goal of this work is to propose two additional (perhaps useful) tools for selecting terms defining two simple statistics based on a ratio between a statistic which contains the effect and the same statistic without the effect. The test statistic for randomness is constructed by comparing the $h-$th diagonal element of the sample covariance matrix of ordinary least squares, ols, (see [11]) with the same term under the hypothesis of zero random effect. The resulting statistic is simple and has a known exact distribution for any sample size under the null and alternative hypothesis. The test statistic for fixed effects is a ratio between the square of the quadratic mean of $h-$th element of the sample average of ols (which contains both fixed and random effects) and the $h-$th diagonal element of the sample covariance matrix of ols (which captures random effect only). The ratio: $(random\ effect + \ fixed\ effects)/random\ effect$ defines a test statistic with random effect "removed" by division. The distribution of this statistic is unknown ad is approximated by a noncentral $F-$distribution. A selection procedure is conducted through a joint analysis on the significance of these two tests.

The approach proposed in this paper is simple as it uses ols, allows to make inference through point estimates and (approximate) confidence intervals and may be seen as an attempt to overcome the "boundary" and "confounding" problems of testing procedures.

The paper is structured as follows. Section 2 introduces the two stage linear mixed model. Section 3 defines the statistics for testing randomness (subsection 3.1) and fixed effects (subsection 3.2). Section 4 discusses the exact (subsection 4.1) and the approximate (subsection 4.2) density functions of the test statistics. Section 5 is divided into three subsections. Subsection 5.1 defines the settings for simulations, subsection 5.2 analyzes the goodness-of-approximation of the noncentral $F-$distribution, subsection 5.3 deals with the selection of terms of the linear mixed model based on the significance of the two test statistics proposed. Appendix A discusses the approximated distribution.

2. Two-Stage Random Effects: Model and notation

The linear mixed model for longitudinal data can be described as follows: $y_i = X_i^*\beta^* + Z_i d_i^* + u_i$, $i = 1, \dots, n$ 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 $d_i^* \sim N(0, \Omega_q)$, Ω_q is a $q \times q$ positive semidefinite matrix, $\Omega_q \succeq 0$, $u_i \sim N\left(0, \sigma^2 I_{t_i}\right)$. The u_{ij}'s are iid so can be thought of as measurement error. We assume that u_i and d_i^* are independent.

Following [10] we re-express the linear mixed model as a two-stage random coefficients model [12],

$$y_i = X_i\beta_i + u_i, \qquad i = 1, \dots, n \tag{1}$$

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 $j-$th element of β_i is given by $\beta_j^* + d_{ji}^*$ if column j is common to X_i^* and Z_i, by β_j^* if column j is unique to X_i^* or by d_{ji}^* if column j is unique to Z_i. We can therefore write $\beta_i = \beta + d_i$, where null elements may be added to the original β^* and d_i^* vectors so that they have the same dimension.

Regarding (1) as a two stage model, it follows that $y_i|d_i^* \sim N(X_i\beta_i; \sigma^2 I_{t_i})$ 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 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 $(t_i - k)s_i^2 = y_i'\left(I_{t_i} - X_i(X_i'X_i)^{-1}X_i'\right)y_i$ 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 Ω_k consists of Ω_q augmented with null rows and/or columns corresponding to the null elements in the random vectors d_i. Let $\beta_{hi} = \beta_h + d_{hi}$ be the h-th component of the vector β_i where β_h is the h-th component of β and d_{hi} is the h-th element of d_i such that $d_{hi} \sim N(0, \omega_{hh})$, ω_{hh} the h-th diagonal element of Ω_k. Setting $\omega_{hh} = 0$ is equivalent to setting all the elements in the $h-$th column and $h-$th row of the matrix Ω_k to zero. This means that a single parameter controls the inclusion/exclusion of the random effects in the model.

3. Test statistics

The test statistics defined in this section are based on ols, $b_i \sim \left(\beta, \sigma^2(X_i'X_i)^{-1} + \Omega_k\right)$. Let denote with b_{hi} the h-th component of the vector b_i. The sample average of ols estimators, $\overline{b} = \frac{1}{n}\sum_{i=1}^n b_i$, is normally distributed with expected value β and variance $var(\overline{b}) = \frac{\sigma^2}{n}\overline{V} + \frac{1}{n}\Omega_k$ where $\overline{V} = n^{-1}\sum_{i=1}^n (X_i'X_i)^{-1}$. Let $\overline{b}_h$ be the h-th component of $\overline{b}$, and $\overline{v}_{hh}$ the h-th main diagonal element of $\overline{V}$.

According to the assumptions of the model, $(b_i - \overline{b}) \sim N(\,0\,,\, \sigma^2 V_{ii} + \frac{n-1}{n}\Omega_k)$ with $V_{ii} = \frac{1}{n}\overline{V} + \frac{n-2}{n}(X_i'X_i)^{-1}$, $E(b_i - \overline{b})(b_j - \overline{b})' = \sigma^2 V_{ij} + h_{ij}\Omega_k$, $V_{ij} = \frac{1}{n}\overline{V} - \frac{1}{n}(X_i'X_i)^{-1} - \frac{1}{n}(X_j'X_j)^{-1}$ and $h_{ij} = \dfrac{n-1}{n}$ 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 with (i,j)-th block V_{ij}. V is a positive semidefinite and symmetric matrix with rank $(n-1)k$. We define two statistics, one for testing randomness "removing" the fixed effect by a difference, the other for testing the nullity of the fixed effect "removing" randomness from β_{hi} by a ratio.

3.1. Test statistic T_h, $H_0 : \omega_{hh} = 0 \bigcap \beta_h \in \mathbb{R}$; $H_1 : \omega_{hh} > 0 \bigcap \beta_h \in \mathbb{R}$.

In this section we define a statistic for testing randomness of $\beta_{hi} = \beta_h + d_{hi}$.

Observe that $\omega_{hh} = 0$ implies $d_{hi} = 0$ with probability 1. We develop a test statistic based on $S_{b_h} = (n-1)^{-1}\sum_{i=1}^n \left(b_{hi} - \overline{b}_h\right)^2$ which is the h-th diagonal element of the sample covariance matrix of ols proposed by [11], $S_b = (n-1)^{-1}\sum_{i=1}^n \left(b_i - \overline{b}\right)\left(b_i - \overline{b}\right)'$. We recall that $E(S_b) = \sigma^2\overline{V} + \Omega_k$ which implies that $E\left(S_{b_h}|H_1\right) = \sigma^2\overline{v}_{hh} + \omega_{hh}$ and $E\left(S_{b_h}|H_0\right) = \sigma^2\overline{v}_{hh}$. The ratio of these two expected values is

$$\frac{E\left[\frac{1}{n-1}\sum_{i=1}^n (b_{hi} - \overline{b}_h)^2 | H_1\right]}{E\left[\frac{1}{n-1}\sum_{i=1}^n (b_{hi} - \overline{b}_h)^2 | H_0\right]} = \frac{\sigma^2\overline{v}_{hh} + \omega_{hh}}{\sigma^2\overline{v}_{hh}} = 1 + \frac{\omega_{hh}}{\sigma^2\overline{v}_{hh}} = \theta_h$$

Observe that the difference, $b_{hi} - \overline{b}_h$ allows us to "remove" the fixed effect. The statistic we propose is an estimate of θ_h, defined by a ratio between S_{b_h} and $s^2\overline{v}_{hh}$ where s^2 is the sample variance of σ^2. We have,

$$T_h = \frac{1}{n-1}\frac{\sum_{i=1}^n (b_{hi} - \overline{b}_h)^2}{s^2\ \overline{v}_{hh}} \tag{2}$$

The expected value of T_h is given by

$$E(T_h) = df\ E\left(\frac{S_{b_h}}{\sigma^2 \overline{v}_{hh}}\right) E\left(\frac{\sigma^2}{df\ s^2}\right) = \frac{df}{df-2}\,\theta_h \quad with \quad df = \sum_{i=1}^{n}(t_i - k)$$

Since $E\left(S_{b_h}|H_1\right) = \sigma^2\overline{v}_{hh} + \omega_{hh}$, the difference $\widehat{\omega}_{hh} = S_{b_h} - s^2\overline{v}_{hh}$ gives an unbiased estimate of ω_{hh} and $\widehat{\theta}_h = 1 + \frac{\widehat{\omega}_{hh}}{s^2\overline{v}_{hh}} = T_h$. $\frac{df-2}{df}\ T_h$ is an unbiased estimator of θ_h.

The parameter θ_h can be interpreded as a measure of the relative change of the "total" variability" of the $h-th$ coefficient, $\sigma^2\overline{v}_{hh} + \omega_{hh}$, with respect to the residual variance $\sigma^2\overline{v}_{hh}$. Given finite $\sigma^2 > 0$ when $\omega_{hh} = 0$, $\theta_h = 1$ and T_h takes values around $E(T_h) = \frac{df}{df-2}$. If $\omega_{hh} > 0$ then $\frac{\omega_{hh}}{\sigma^2\overline{v}_{hh}} > 0$, θ_h is greater than 1 and T_h deviates from its expected value. The farther $\frac{\omega_{hh}}{\sigma^2\overline{v}_{hh}}$ is from zero, the greater θ_h and T_h, everything else being equal. The greater T_h the stronger the evidence against H_0. We call $rp_h = \frac{\omega_{hh}}{\sigma^2\overline{v}_{hh}}$ randomness parameter.

The parameter, rp_h, plays the same role as the noncentrality parameter of an F-distribution. As we shall see, if rp_h increases, the shape of the distribution of T_h shifts to the right and a larger percentage of the curve moves to the right of the critical value. θ_h can be seen as the unknown parameter of the model to be tested and estimated. Testing randomness is equivalent to testing θ_h. We can restate the null and alternative hypotheses as follows: $H_0 : \theta_h = 1 (H_0 : \theta_h \leq 1)$ and $H_1 : \theta_h > 1$. H_0 is rejected if the test statistic T_h is "much" greater than one.

3.2. **Test statistic** F_h, $H_0 : \beta_h = 0 \bigcap \omega_{hh} \geq 0$, $H_1 : \beta_h > 0 \bigcap \omega_{hh} \geq 0$. For testing the fixed effec, we develop a test statistic based on $\overline{b}_h$. The quadratic mean under H_1, $E(n\ \overline{b}_h^2|H_1)$, is compared with $E(n\ \overline{b}_h^2|H_0) = E\left(S_{b_h}|H_1\right)$ by a ratio. We have,

$$\frac{E(n\ \overline{b}_h^2 \mid H_1)}{E(n\ \overline{b}_h^2 \mid H_0)} = \frac{\sigma^2\overline{v}_{hh} + \omega_{hh} + n\ \beta_h^2}{\sigma^2\overline{v}_{hh} + \omega_{hh}} = 1 + \frac{n\beta_h^2}{\sigma^2\overline{v}_{hh} + \omega_{hh}} = 1 + \frac{n\beta_h^2}{\sigma^2\overline{v}_{hh}}\ \theta_h^{-1} = 1 + ncp_h$$

where ncp_h is a noncentrality parameter, $\theta_h^{-1} = \frac{\sigma^2\overline{v}_{hh}}{\sigma^2\overline{v}_{hh}+\omega_{hh}}$ is the reciprocal of $1 + rp_h$, ranges in the interval $[0, 1]$ and can be interpreted as the share of "residual" variance on the "total" variability. The numerator of $1 + ncp_h$ incorporates both random and fixed effect, the ratio allows us to "remove" the random effect. The statistic we propose is an estimate of the above ratio,

$$F_h = \frac{n\ \overline{b}_h^2}{S_{b_h}} = \frac{n\ \overline{b}_h^2}{(n-1)^{-1}\sum_{i=1}^{n}\left(b_{hi} - \overline{b}_h\right)^2} \tag{3}$$

When $\beta_h = 0$ the test statistic F_h takes values around its expected value. If $\beta_h \neq 0$ then F_h deviates from its expected value. As we shall see, the greater β_h^2 the further away the peak of the distribution from zero.The bigger the ncp_h, the more the alternative sampling distribution moves to the right and the more power we have. The null hypothesis is rejeted for large value of F_h. The test statistic (3) is similar to the one proposed in [10].

4. Density functions of T_h and F_h

In this section we develop and define the exact and approximate distributions respectively of T_h and F_h both under the null and the alternative hypotheses.

4.1. **Exact density function of T_h.** Let v_{hi}, $h = 1, \ldots, k$ be the h-th diagonal element of the block matrix V_{ii} and denote $\overline{V}_D^{-1/2} = diag(1/\sqrt(\overline{v}_{11}), \ldots, 1/\sqrt(\overline{v}_{hh}), \ldots, 1/\sqrt(\overline{v}_{kk}))$ where $\overline{v}_{hh}$ is the h-th main diagonal element of $\overline{V}$. Let's define $W = R + G$ where $R = \left(I_n \otimes \overline{V}_D^{-1/2}\right) V \left(I_n \otimes \overline{V}_D^{-1/2}\right)$ is the $nk \times nk$ covariance matrix when $\Omega_k = 0$, $G = \left(I_n \otimes \overline{V}_D^{-1/2}\right)\left(H_n \otimes \frac{\Omega_h}{\sigma^2}\right)\left(I_n \otimes \overline{V}_D^{-1/2}\right)$ the $nk \times nk$ covariance matrix of random components and $H_n = [h_{ij}]$.

Let $W_h = R_h + G_h$ be the $n \times n$ matrix of rank $n-1$ obtained from W dropping the rows and columns that do not refer to the h-th element. According to the hypotheses of the model $(b_{hi} - \overline{b}_h)/(\sigma\sqrt{\overline{v}_{hh}})$, is $N(0, \theta_{hi})$ where $\theta_{hi} = \frac{v_{hi}}{\overline{v}_{hh}} + \left(\frac{n-1}{n}\right)\frac{\omega_{hh}}{\sigma^2 \overline{v}_{hh}}$ is the i-th diagonal element of W_h and $\theta_h = \frac{1}{n-1}\sum_{i=1}^{n}\theta_{hi}$. The square $\left(b_{hi} - \overline{b}_h\right)^2/(\sigma^2\overline{v}_{hh})$, is a gamma with shape parameter $1/2$ and scale parameter $2\theta_{hi}$. For $i = 1, \ldots, n$ we have a set of correlated gamma with same shape parameter and different scale parameter. The density function of $S_{b_h}/(\sigma^2\overline{v}_{hh})$ is defined with the gamma-series representation of [13] (see also [14]). The statistic T_h obtained by replacing σ^2 with s^2, can be seen as the ratio of two random variables where the numerator is a sum of gamma, $\sum_{i=1}^{n-1} G(\frac{1}{2}, \beta_i)$ with $\beta_i = 2\, df\, \phi_i/(n-1)$ where ϕ_i's are the eigenvalues of W_h and the denominator is a $G(\frac{df}{2}, 2)$ (si veda [15]). It can be shown that $\theta_h = \frac{1}{n-1}\sum_{i=1}^{n-1}\phi_i$ and when H_0 is true $\sum_{i=1}^{n-1}\dfrac{\phi_i}{n-1} = 1$.

The ratio of these two gamma is a generalized F-distribution denoted with GF. By expressing the numerator as a single gamma-series representation [13], the density function of T_h can be written as

$$f_{T_h}(x) = \sum_{k=0}^{\infty} p_k\; GF\left(\rho + k, \frac{df}{2}, \frac{\beta_1}{2}\right) \tag{4}$$

where $p_k = C\delta_k$, $\beta_1 = \min_i\{\beta_i\}$, $C = \prod_{i=1}^{n-1}\left(\frac{\beta_1}{\beta_i}\right)^{\alpha_i}$, $\rho = \sum_{j=1}^{n-1}\alpha_j$, $\alpha_j = 1/2\ \forall j$ and the coefficients δ_k are obtained recursively by the formula

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

$(k, p_k)_{k=0,1,2\ldots}$ is a discrete probability distribution. Since the gamma-series representation of [13] is CPU-time intensive when the shape parameters are small and the scale parameters have large variation, [16] proposed to approximate the probability distribution $(k, p_k)_{k=0,1,2\ldots}$ with a generalized negative binomial distribution.

The function (4) is uniform convergent [13]. This property justifies the interchange of the integration and summation and allows us to compute the distribution function and quantiles.

4.2. **Approximate density function of F$_h$.** Let divide and multiply expression F_h (formula (3)) by $\sigma^2\overline{v}_{hh} + \omega_{hh}$ and, individually, discuss numerator and denominator. According to the assumptions of the model,

$$\frac{n\,\overline{b}_h^2}{\sigma^2\overline{v}_{hh} + \omega_{hh}} \sim \chi^2\left(1, ncp_h = \frac{n\beta_h^2}{\sigma^2\overline{v}_{hh} + \omega_{hh}}\right) \quad for\ any\ \ \omega_{hh} \geq 0 \tag{5}$$

and

$$Q_h = \frac{S_{b_h}}{\sigma^2\overline{v}_{hh} + \omega_{hh}} = \frac{1}{n-1}\frac{\sum_{i=1}^{n}\left(b_{hi} - \overline{b}_h\right)^2}{\sigma^2\overline{v}_{hh} + \omega_{hh}} \sim \frac{1}{n-1}\sum_{i=1}^{n-1}\tau_i\,\chi^2(1) \tag{6}$$

where $\tau_i = \phi_i/\theta_h$. Q_h is distributed as a linear combination of $\chi^2(1)$ the exact distribution of which can be obtained, for example, through the gamma-series representation of [13]. However, the knowledge of the exact distribution of Q_h is not useful for defining a "simple" distribution of the statistic F_h, so following [17] we approximate the distribution of Q_h by an adjusted chi-square distribution as in $Q_h \approx a\,\chi^2(b)$ where a and b are determined by matching the first two moments of Q_h with those of $a\chi^2(b)$ (see Appendix A).

The ratio between the exact chi-square distribution (expression (5)) and the approximate chi-square distribution of Q_h each divided by their degrees of freedom gives the following approximation

$$F_h = \frac{n\,\overline{b}_h^2}{(n-1)^{-1}\sum_{i=1}^{n}\left(b_{hi} - \overline{b}_h\right)^2} \approx F\left(1, b, ncp_h\right) \tag{7}$$

We recall (see (Appendix A) that b depends on the random component and ranges between b_0 when $\omega_{hh} = 0$ and $(n-1)$ when ω_{hh} is large (with respect to σ^2), then (7) defines a family of density functions. According to the critical discussion of Appendix A, we propose setting $b = n-1$ for any $\omega_{hh} \geq 0$.

5. Simulation

This section is divided into three subsections. The first defines the settings for simulations which are valid unless otherwise specified (settings by default). The second subsection analyzes the goodness-of-approximation of the statistic F_h. The test statistic T_h is based on the works of [16] and [18] where the power function and consistency of the test is partly analyzed and discussed. The results are not reproposed here but are available in a supplementary material. The third subsection discusses the selection of terms of the linear mixed model based on the significativity of the test statistics T_h and F_h.

5.1. **"Base" Scenario for simulations.** To allow the maximum of generality and flexibility, we define the following scenario for all simulations unless otherwise specified.

(i) The number of parameters and units are respectively $k = 6$ and $n = 10$. The number of observations per units, t_i, $i = 1, \ldots n$, are drawn randomly from a uniform distribution, $U(k+4, 3\,k)$.

(ii) The vector of regression coefficients, β, is generated randomly from a $N(0, 2)$.

(iii) For each units, the columns of X_i are drawn from an $N(mean, sqrt)$ 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.

(iv) We define first a positive definite matrix, Ψ, by extracting elements from a standard normal distribution then the covariance matrix Ω_k is obtained by selecting q columns and rows from Ψ and zero elsewhere. This allows us to define (indirectly) the random and fixed parameters of the model.

(v) The index of the tested parameter is drawn randomly from a uniform distribution, $U(1, k)$.

(vi) The variance, σ^2, is fixed proportionally to the maximum entry of the main diagonal of Ω_k.

5.2. **Goodness-of-approximation.** This study is based on a set of matrices, $M_1, \dots, M_l, \dots, M_k$ defined as follows.

Let ncp_l, $l = 1, \dots, h, \dots k$ be one specific value of the noncentrality parameter in the arbitrary set $A = \{0, 1, 2, 3, 5, 8\}$ and consider $nrepl = 100$ different values of θ_h, θ_{hj}, $j = 1, \dots, nrepl$ drawn randomly from a uniform distribution on the interval $[1, 10]$. Given ncp_l, for each different parameter combination (ncp_l, θ_{hj}), we compute the test statistic F_h on $N = 1000$ simulated samples of size $n = 10$. This yields an $N \times nrepl$ matrix, M_l, of statistics F_h computed with the same noncentrality parameter but different parameter θ_{hj}. The matrix M_l is defined for each value of $ncp_l \in A$. The set of M_l matrices, $l = 1, \dots, k$ is the basis for our analysis on the goodness-of-approximation of the test statistic F_h.

We proceed following two steps. First we extract one column from the matrix M_l. We have $N = 1000$ simulated samples, replicated by a model defined by the pair (ncp_l, θ_{hj}). With these data we compute different goodness-of-fit statistics and estimate the parameters of the approximating $F-$distribution by maximizing the likelihood function. Subsequently we extend the analysis to all the column of the matrix M_l so that we can evaluate the impact of the randomness parameter on the test statistic.

Given the pair $(ncp_l = 3, \theta_{hj} = 3.85)$, the empirical distribution function of the N simulated value of the statistic F_h is shown in Fig.: 1.a where the solid line is the noncentral $F-$distribution and the dotted line is the central $F-$distribution. The deviation between the two curves is the effect of the noncentrality parameter. Fig.: 1.b shows the empirical cdf of F_h with the 95% Kolmogorov-Smirnov (K.S.) confidence bands for the unknown cumulative distribution function. Fig.: 1.c shows the Q-Q plot between the quantiles of the noncentral distribution, $F(1, n-1, ncp)$ and the empirical quantiles. The points of both sets of quantiles form a line that's roughly straight. Fig.: 1.d is a P-P plot computed as follows: on the abscissa there is the set of probabilities: $p = \{0.025, 0.05, 0.075, \dots, 0.975\}$ on the vertical axis there is the empirical probability, $\widehat{P}_p = \frac{1}{N}\sum_{i=1}^{N} \delta(F_h < q_p)$ where $\delta(true) = 1$, $\delta(false) = 0$ and q_p is the quantile of the non central $F-$ distribution. The points close to the $0-1$ line highlights goodness of the approximation.

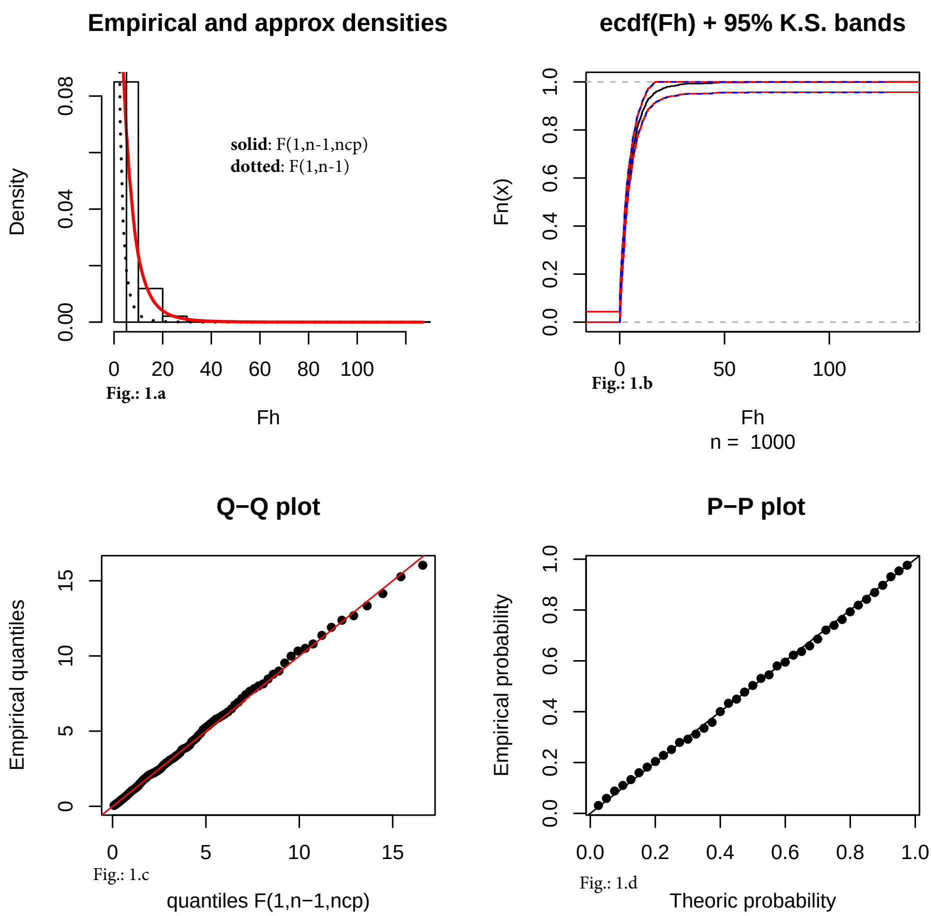

Figure 1. – Fig.: 1.a shows simulated histogram of F_h, the central F (dotted line) and noncentral F (solid line). The displacement of the solid line from the dotted line is due to the ncp with $\theta_h \geq 1$. The parameters are: $h = 3$, $ncp = 3$ $\theta_h = 3.85$. – Fig.: 1.b show the graph of the empirical cdf of F_h with K.S. confidence bands at 95%. – Fig.: 1.c represents Q-Q plot plot between empirical quantiles and quantiles of noncentral $F-$distribution functions. . Fig.: 1.d shows the empirical probabilities plottted against theoretical quantiles of $F(1, n-1, ncp)$.

The Kolmogorov-Smirnov method is used to test the null hypothesis that the hypotesized distribution is $F(1, 9, ncp = 3)$ against the alternative that the "exact" cdf does not equal the $F(1, 9, ncp = 3)$. The result is a statistic $ks = 0.02130233$ with a $pvalue = 0.75$. (The chi-square goodness of fit test gives similar results).

The method of maximum likelihood is used to estimate the parameters of a noncentral $F-$distribution. We expect that the estimates are "close" to the parameters $(1, 9, ncp = 3)$. The fitdistrplus package of R produces the result of Tab.: 1. Tab.: 2 shows the confidence intervals obtained with the basic bootstrap procedure. All the results are quite satisfactory.

Table 1. Maximum likelihood estimation

Parameters	Estimate.	Std. Error
$df1 = 1$	1.018527	0.05633942
$df2 = 9$	8.633183:	1.27773298
$ncp = 3$	2.951871:	0.14260946
Loglik: -2567.894	AIC: 5141.787	BIC: 5156.51

Table 2. Parametric bootstrap medians and 95% percentile CI

	Median	2.5%	97.5%
$df1 = 1$	0.9613134	0.8688267	1.074114
$df2 = 9$	8.7609549	6.8027519	11.999008
$ncp = 3$	3.0368	2.69518	3.08946

The above analysis is carried out on $N = 1000$ simulated samples. To evaluate the "stability" of the results we keep fixed the noncentrality parameter and repeat ($nrepl = 100$) the simulations drawing randomly θ_{hj} from a uniform distribution on the interval $[0, 10]$. This means that we work on the whole matrix M_l.

For each column of the matrix M_l we compute the empirical vigintiles of F_h. Fig.: 2.a shows the bundle of lines "close" to each others which envelop the vigintiles (black points) of the approximating noncentral distribution. Fig.: 2.b shows the boxplots of vigintiles and the points of the approximating distribution. The approximation which collocates points of the replicated simulation inside the box or within the whiskers of the boxplot can be defined "good" ("excellent"). Fig.: 2.c reproposes part of Fig.: 2.b focusing on the first and third quartiles.

Some other results concerning the "goodness-of-approximation" of the F_h test statistic can be found in [15].

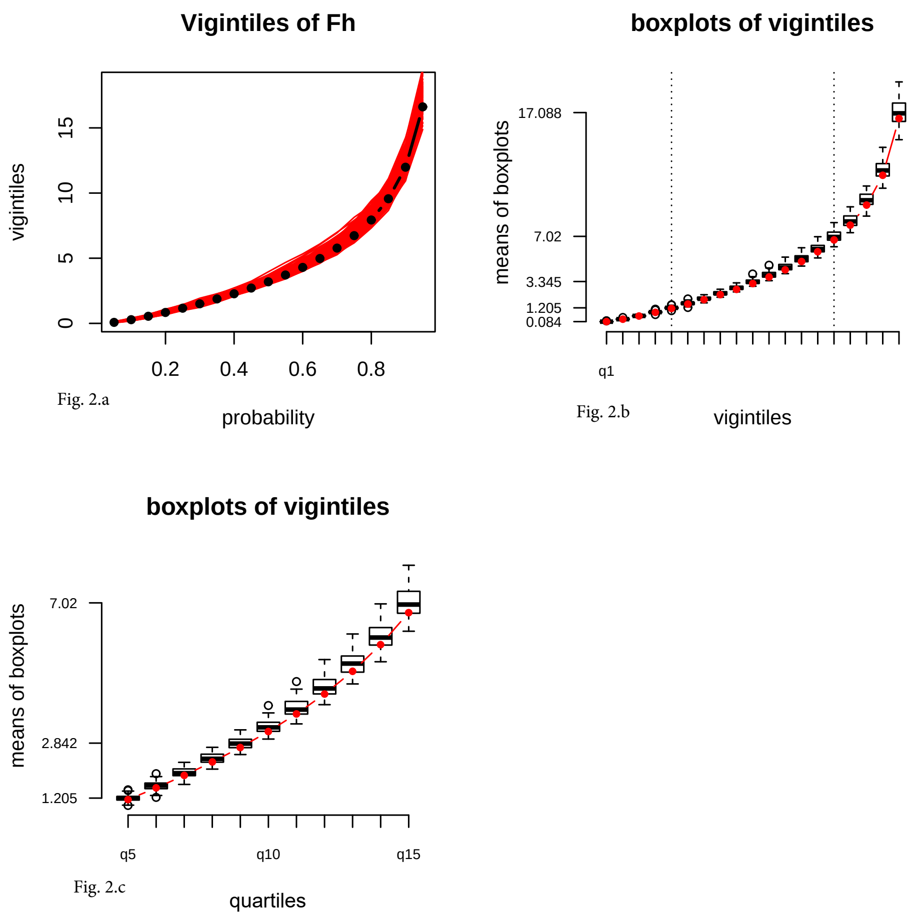

FIGURE 2. – Fig.:2.a shows the bundle of lines of empirical vigintiles and the points of the approximating noncentral distribution. – Fig.: Fig.:2.b shows the boxplots of empirical vigintiles and the points of the approximating distribution. – Fig.: Fig.:2.c shows the graph of Fig.:2.b limited to the quartiles.

5.3. **The selection procedure.** In this section we discuss a selection procedure of the $h-$th term based on the significance of the test statistics T_h and F_h and following the decision-making scheme of Tab.: 3.

TABLE 3. Selection of terms in a linear mixed model

	Significance of F_h	
Significance of T_h	**Yes = 1**	**No=0**
Yes = 1	$\beta_{hi} = \beta_h + d_{hi}$	$\beta_{hi} = d_{hi}$
No=0	$\beta_{hi} = \beta_h$	$\beta_{hi} = 0$

The Table can be read by column and/or by row. Let consider the pair ($yes = 1,\ yes = 1$). By row, $yes = 1$ means that T_h is statistically significant, $\theta_h >> 1\ \ \forall \beta_h$. Likely, the term β_{hi} has a random component. By column, $yes = 1$ implies that presumably $\beta_h \neq 0$ for any $\theta_h \geq 1$. The joint significance ($yes = 1,\ yes = 1$) leads us to claim that β_{hi} could be composed of both a fixed and a random component, $\beta_{hi} = \beta_h + d_{hi}$. The "goodness" of the selection is evaluated by simulating a table of marginal and joint empirical significance measured by the power of T_h and F_h. That is we calculate the percentage of $(1,1)$ that occur on 10000 replications under different settings.

We recall that the noncentrality parameter, ncp_h, and the randomness parameter, rp_h, are a measure of the degree to which the null hypothesis is false and then, they tell us something about the significance of the two test statistics. We saw that $ncp_h = \dfrac{n\beta_h^2}{\sigma^2\overline{v}_{hh}}\,\theta_h^{-1}$ then, n, β_h, σ^2, $\overline{v}_{hh}$ and ω_{hh} are all factors that influence the "goodness" of the selection approach. In this section we assume that n, β_h, σ^2 and $\overline{v}_{hh}$ are given and we discuss the power of T_h and F_h by varying the parameter $0 \leq \theta_h^{-1} \leq 1$. This means that the analysis is restricted to the discussion of the pairs $(1,0)$ $(0,1)$ of Tab.: 3. Simulations based on different settings of the ratio $\dfrac{n\beta_h^2}{\sigma^2\overline{v}_{hh}}$ and on θ_h^{-1} are not considered in this paper.

Table 4. Simulated power of T_h and F_h for different values of the randomness parameter

$\sigma^2 = 3.27,\ \overline{v}_{hh} = 2.9,\ \ \beta_h = 3.05,\ \ n = 10$			
Share: θ_h^{-1}	ncp_h	power of F_h	power of T_h
0	0	0.057	1
0.016	0.15	0.0647	1
0.028	0.26	0.0689	0.996
0.1047	0.97	0.1446	0.9636
0.208	1.93	0.2477	0.7461
0.32	2.958	0.3572	0.4435
0.5138	4.762	0.5443	0.1551
0.81	7.51	0.7324	0.0652
1	9.26	0.8113	0.0545

Table 4 shows simulated power of T_h and F_h for different value of θ_h^{-1}.

(1) The power of T_h depends inversely on θ_h^{-1}. The larger θ_h^{-1} ($\omega_{hh} \longrightarrow 0$) the lower the power of T_h. When $\theta_h^{-1} = 1$ ($\omega_{hh} = 0$) the power of T_h is equal to the level of significance. The smaller θ_h^{-1} ($\omega_{hh} \longrightarrow \infty$) the greater the power of T_h. As $\theta_h^{-1} \to 0$ the power of T_h tends to one.

(2) The power of F_h is directly related to θ_h^{-1}. The larger θ_h^{-1} the greater the power of F_h. When $\theta_h^{-1} = 1$ ($\omega_{hh} = 0$) the power of F_h depends on $\dfrac{n\beta_h^2}{\sigma^2\overline{v}_{hh}}$ which increases if the number

of unit n and/or the magnitude of β_h increases. The lower θ_h^{-1} ($\omega_{hh} \longrightarrow \infty$) the smaller the power of F_h.

The last two columns of Tab.: 4 can be taken as marginal probabilities and tell us the percentage the significance (yes/no) of the two individual test statistics occurs on 10000 replications. Tab.: 5 ($(a)-(h)$) shows the empirical percentage the pairs $(1,1)$, $(1,0)$, $(0,1)$, $(0,0)$ occur on 10000 replications for different values of θ_h^{-1}.

Tab.: 4 shows that with a low value of θ_h^{-1} (large power of T_h and low power of F_h) likely we observe the pair $(1,0)$. In this situation ω_{hh} is large compared to the "residual" variance $\sigma^2\overline{v}_{hh}$ and the fixed effect is dominated by the "randomness". In this case the significance of the fixed effect plays a minor role in a selection approach. Tab.: 5 (subtable $(a)-(d)$) shows the empirical "joint probability" of selecting the term $\beta_{hi} = d_{hi}$. This "probability" decreases from 0.934 when $\theta_h^{-1} \simeq 0.16$ to 0.6087 when $\theta_h^{-1} \simeq 0.208$. Of course other factors such as n or the magnitude of β_h which influence the power of F_h could address towards the selection of $\beta_{hi} = \beta_h + d_{hi}$ instead of $\beta_{hi} = d_{hi}$.

When θ_h^{-1} is large (low power of T_h and large power of F_h) presumably we observe the pair $(0,1)$. In this case the random component is dominated by the fixed effect and the selection of the $h-th$ terms is based on the significance of F_h ignoring the possible presence of ("irrelevant") randomness. Tab.: 5 (subtable $(f)-(h)$) shows the "joint probability" of selecting the term $\beta_{hi} = \beta_h$. It is greater than 0.70 when $\theta_h^{-1} > 0.8$.

(a) $\theta_h^{-1} \approxeq 0.016$, $ncp_h \approxeq 0.15$

	F_h	
T_h	Yes = 1	No = 0
Yes = 1	0.066	0.934
No = 0	0	0

(b) $\theta_h^{-1} \approxeq 0.028$, $ncp_h \approxeq 0.26$

	F_h	
T_h	Yes = 1	No = 0
Yes = 1	0.07	0.93
No = 0	0	0

(c) $\theta_h^{-1} \approxeq 0.1047$, $ncp_h \approxeq 0.97$

	F_h	
T_h	Yes = 1	No = 0
Yes = 1	0.128	0.8356
No = 0	0.0166	0.0198

(d) $\theta_h^{-1} \approxeq 0.208$, $ncp_h \approxeq 1.93$

	F_h	
T_h	Yes = 1	No = 0
Yes = 1	0.1374	0.6087
No = 0	0.1103	0.1436

(e) $\theta_h^{-1} \approxeq 0.32$, $ncp_h \approxeq 2.958$

	F_h	
T_h	Yes = 1	No = 0
Yes = 1	0.0898	0.3537
No = 0	0.2674	0.2891

(f) $\theta_h^{-1} \approxeq 0.5138$, $ncp_h \approxeq 4.7618$

	F_h	
T_h	Yes = 1	No = 0
Yes = 1	0.032	0.1231
No = 0	0.5123	0.3328

6. Conclusions

A hypothesis testing approach designed for selecting fixed and random coefficients to be included in a linear mixed model brings up several complications. The two test statistics proposed in this

(g) $\theta_h^{-1} \approxeq 0.81$, $ncp_h \approxeq 7.51$

	F_h	
T_h	**Yes = 1**	**No = 0**
Yes = 1	0.0201	0.0451
No = 0	0.7123	0.2225

(h) $\theta_h^{-1} \approxeq 1.00$, $ncp_h \approxeq 9.26$

	F_h	
T_h	**Yes = 1**	**No = 0**
Yes = 1	0.0170	0.0315
No = 0	0.7943	0.1572

Table 5. Settings: $\sigma^2 = 3.27$, $\overline{v}_{hh} = 2.9$, $\beta_h = 3.05$, $n = 10$. Each subtable shows the empirical percentage the pairs $(1,1)$, $(1,0)$, $(0,1)$, $(0,0)$ occur on 10000 replications for different values of θ_h^{-1} and ncp_h. For example, in subtable (a) in 6.6% of cases both T_h and F_h are significant, in 93.4% of cases T_h is significant and F_h is not.

paper are developed trying to solve the so called "boundary" and "confounding" problems which are crucial for evaluating the significance of the tests.

In our opinion the approach based on a ratio between an "appropriatre" statistic which contains the effect and another (or the same) that does not contain the same effect, may be a good method to overcome the above problems. Since the statistics used in the work are based on ordinary least squares, they are easy to compute, do not need any estimate of covariance matrices, allow to investigate exact and approximate density function in small samples.

By using ordinary regression, the selection method based on the joint significance of the two test statistics maintains great simplicity. However this approach must be taken as complementary, point estimates and (approximate) confidence interval of randomness and noncentrality parameters widen and enrich information needed to take a decision.

Appendix A. Approximation

According to the hypotheses of the linear mixed model, the ratio $Q_h = \dfrac{S_{b_h}}{\sigma^2 \overline{v}_{hh}} \theta_h^{-1} = \frac{1}{n-1} \dfrac{\sum_{i=1}^n \left(b_{hi} - \overline{b}_h\right)^2}{\sigma^2 \overline{v}_{hh} + \omega_{hh}}$ is distributed as $\frac{1}{n-1} \sum_{i=1}^n \tau_i \chi^2(1)$ where $\tau_i = \phi_i / \theta_h$, with $\frac{1}{n-1} \sum_{i=1}^n \tau_i = 1$. The exact distribution of Q_h can be derived using the gamma-series representation of [13]. $Fig.: 3.a$ shows a simulated histogram of $(n-1)Q_h$ and its exact density function.

Following [17] we develop the approximations $(n-1)\, Q_h \approx a\, \chi^2(b)$ where a and b are determined by matching the first two moments of $(n-1)Q_h$ with those of $a\chi^2(b)$. Straightforward calculation leads to $a = \dfrac{M(\phi_i^2)}{M(\phi_i)^2}$ and $b = \dfrac{n-1}{a}$ where $M(.)$ is for arithmetic mean and ϕ_i's are the eigenvalues of W_h (see subsection 4.1). Observe that a and b depend both on the unknown ω_{hh} and can be can be computed only under the null hypothesis. Let consider the following approximation,

$$\frac{(n-1)Q_h}{a} \approx \chi^2(b) \tag{8}$$

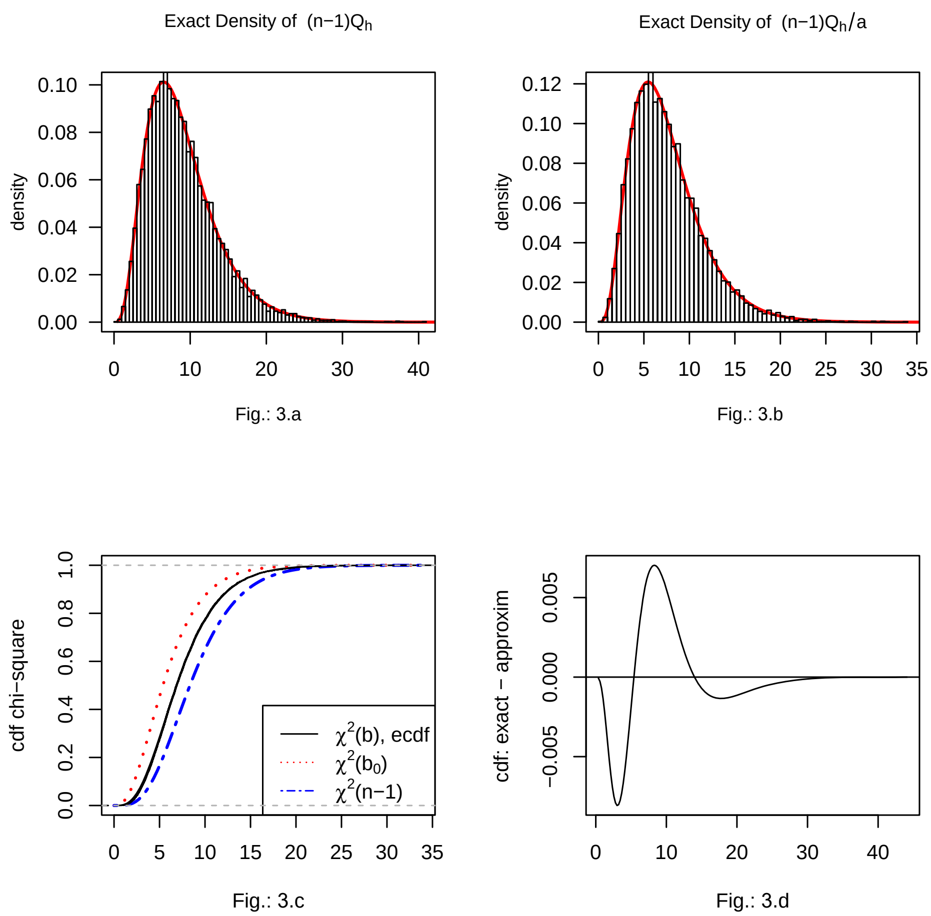

FIGURE 3. - Fig.: 3.(a) and Fig.: 3.(b) show simulated histogram and exact density of $(n-1)Q_h$ and $(n-1)Q_h/a$ for the third element ($h = 3$). - Fig.: 3.(c) represents the distribution functions (cdf) of $\chi^2(b_0)$ (dotted line), $\chi^2(n-1)$ (twodashed line) and $\chi^2(b)$ (solid black line). Fig.: 3.(d) shows the diffferfences between the cdf of the approximated $\chi^2(b)$ and the cdf of the exact distribution.

$Fig.$: 3.b shows a simulated histogram of $(n-1)Q_h/a$ and the chi-square approximation (8) which depends on the relative sizes of the τ_i's, on their variabilities and on the degrees of freedom. Let's make some comments:

(1) If $\tau_1 = \tau_2 = \ldots = \tau_{n-1} = \tau$, then $a = 1$, $b = n-1$, $\tau = \theta_h$ and the approximation (8) is exact, $(n-1)Q_h \sim \chi(n-1)$. The equality of τ_i's occurs when given $\overline{v}_{hh}$, ω_{hh} is (very) large with respect to σ^2, that is, the parameter θ_h is "much larger" than one. The greater ω_{hh} (with respect to σ^2) the farther θ_h is from one, the less the variability of the eigenvalues. From a practical point of view we may capture this "limit" situation through the *pvalue* of the test statistic T_h. Our evidences show that if *pvalue* < 0.001 then the variability of

eigenvalues is (approximately) zero, $b = n-1$ and $(n-1)Q_h \sim \chi(n-1)$. We can show algebraically that when $\omega_{hh} \to \infty$, b reaches the maximum value at $n-1$ [17]. Therefore, b is always less or equal to $n-1$

(2) The maximum variability of τ_i is reached when $\omega_{hh} = 0$. In this case we can compute the minimum value of b, b_0, and the maximum value of a, a_0. Therefore, as $\omega_{hh} = 0$ ranges between zero and infinity, $b_0 \leq b \leq n-1$ and $1 \leq a \leq a_0$.

Starting from the zero variability of eigenvalues, as the τ_i's depart from each other, b decreases towards b_0 and a increases towards a_0. In Figure 3 the bottom left graph ($Fig.:3.c$), shows the distribution functions (cdf) of $\chi^2(b_0)$, $\chi^2(n-1)$ and the empirical distribution function (ecdf) of (8) which collocates between the two curves. The ecdf of (8) is "well" approximated by a $\chi^2(b)$. $Fig.:3.c$ shows the difference between the two curves which is less than 0.5%.

(3) The approximation depends on the number of unit, n. As n increases, according to the central limit teorem, the exact distribution of $(n-1)Q_h$ may be approximately described by a normal distribution, and so may be $a\chi(b)$. Thus, we may expect that the approximation will improve as n increases.

References

[1] S. Müller, J.L. Scealy, A.H. Welsh, Model selection in linear mixed models, Stat. Sci. 28 (2013) 135-167. https://doi.org/10.1214/12-sts410.

[2] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc.: Ser. B (Methodol.) 58 (1996) 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

[3] H.D. Bondell, A. Krishna, S.K. Ghosh, Joint variable selection for fixed and random effects in linear mixed-effects models, Biometrics. 66 (2010) 1069-1077. https://doi.org/10.1111/j.1541-0420.2010.01391.x.

[4] J.G. Ibrahim, H. Zhu, R.I. Garcia, R. Guo, Fixed and random effects selection in mixed effects models, Biometrics. 67 (2010) 495-503. https://doi.org/10.1111/j.1541-0420.2010.01463.x.

[5] J. Fan, R. Li, variable selection in linear mixed effects models, Ann. Stat. 40 (2012) 2043-2068, https://www.jstor.org/stable/41806528.

[6] H. Peng, Y. Lu, Model selection in linear mixed effect models, J. Multivar. Anal. 109 (2012) 109-129. https://doi.org/10.1016/j.jmva.2012.02.005.

[7] B. Lin, Z. Pang, J. Jiang, Fixed and random effects selection by reml and pathwise coordinate optimization, J. Comput. Graph. Stat. 22 (2013) 341-355. https://www.jstor.org/stable/43304834.

[8] P. Wu, X. Luo, P. Xu, L. Zhu, New variable selection for linear mixed-effects models, Ann. Inst. Stat. Math. 69 (2017) 627-646. https://doi.org/10.1007/s10463-016-0555-z.

[9] J. Fan, R. Li, Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties, J. Amer. Stat. Assoc. 96 (2001) 1348-1360. https://doi.org/10.1198/016214501753382273.

[10] F.M.M. Rocha, J.M. Singer, Selection of terms in random coefficient regression models, J. Appl. Stat. 45 (2017) 225-242. https://doi.org/10.1080/02664763.2016.1273884.

[11] M. Gumpertz, S.G. Pantula, A simple approach to inference in random coefficient model, Amer. Stat. 43 (1989) 203-210. https://www.jstor.org/stable/2685362.

[12] N. Laird, Chapter 5: Random effects and the linear mixed model, in: NSF-CBMS Regional Conference Series in Probability and Statistics, Institute of Mathematical Statistics and American Statistical Association, Beechwood OH and Alexandria VA, 2004: pp. 79-95. https://doi.org/10.1214/cbms/1462106081.

[13] P.G. Moschopoulos, The distribution of the sum of independent gamma random variables, Ann. Inst. Stat. Math. 37 (1985) 541-544. https://doi.org/10.1007/bf02481123.

[14] S. Kotz, J.W. Adams, Distribution of sum of identically distributed exponentially correlated gamma-variables, Ann. Math. Stat. 35 (1964) 277-283. https://www.jstor.org/stable/2238038.

[15] M. Barnabani, Testing fixed and random effects in linear mixed models, Working papers DISIA - University of Florence 9 (2020) 1-20.

[16] M. Barnabani, A parametric test to discriminate between a linear regression model and a linear latent growth model, Int. J. Stat. Probab. 6 (2017) 157. https://doi.org/10.5539/ijsp.v6n3p157.

[17] K.H. Yuan, P.M. Bentler, Two simple approximations to the distributions of quadratic forms, Br. J. Math. Stat. Psychol. 63 (2010) 273-291. https://doi.org/10.1348/000711009x449771.

[18] M. Barnabani, An F-type multiple testing approach for assessing randomness of linear mixed models, Amer. Rev. Math. Stat. 7 (2019) 15-28. https://doi.org/10.15640/arms.v7n2a4.