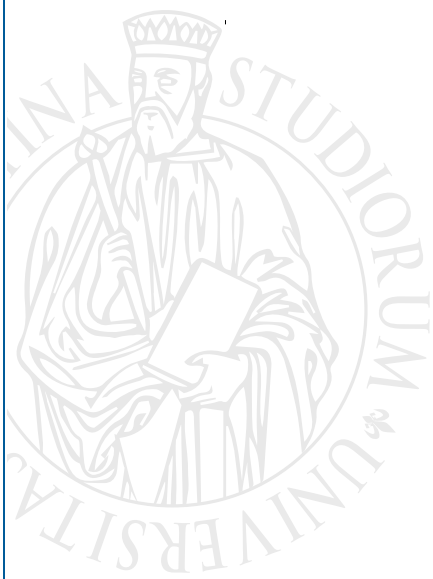
UNIVERSITÀ DEGLI STUDI FIRENZE

DISIA
DIPARTIMENTO DI STATISTICA, INFORMATICA, APPLICAZIONI "GIUSEPPE PARENTI"

Testing fixed and random effects in linear mixed models

Marco Barnabani



DISIA WORKING PAPER
2020/09

© Copyright is held by the author(s).

Testing fixed and random effects in 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 selection of fixed and random effects using a testing hypothesis approach brings up several problems. In this paper, we consider the so called boundary problem and the confounding impact of effects from one set of coefficient in the other set. These problems are addressed by defining two test statistics based on ordinary least squares obtained by dividing two quadratic forms, one that contains the effect and another that does not. As a result, the test statistics are sufficiently general, easy to compute, with known finite sample properties. The test on randomness has a known exact distribution under the null and alternative hypothesis, the test on fixed effect is approximated by a noncentral $F-$distribution. Because of its importance in the selection variable approach, the goodness-of-approximation is examined in-depth in final simulations.

keywords: Selection procedure; Hypothesis testing; Linear Mixed Models; Generalized F-distribution;

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 such as the widely used Akaike Information

Criteria (*AIC*; Akaike (1973)), the Bayesian Information Criteria (*BIC*; Schwarz (1978)) the conditional Akaike Information Criterion (*cAIC*, Vaida and Blanchard (2005)). These approaches are based 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 Muller et al. (2013) for a comprehensive review of model selection in linear mixed models. All these methods deals with the problem of selection working simultaneously with both fixed and random component resulting computationally burdensome. One approach to overcoming this computational problem is penalized likelihood methods (dating back to Tibshirani (1996)). These procedures treat the selection problem via a separate selection approach to avoid the impact of effects from one set of coefficient in the other set. 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. Bondell et al. (2010), Ibrahim et al. (2011) proposed separate penalties for the fixed and random effects that are summed together. Fan and Li (2012), Peng and Lu (2012), Lin et al. (2013) 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 usually replaced with a suitable proxy matrix (see for example an orthogonalization-based approach proposed by Wu et al. (2017)).

Although the penalized likelihood methods may avoid the need to search through the entire model space, it may remain computationally intensive. A Bayesian method was proposed by Chen and Dunson (2003) by selecting a prior with mass at zero for the random effect variances. A further complication of these methods is how to define a "good" penalty function (for a discussione see Fan and Li (2001)) and how to perform the shrinkage appropriately. Finally, 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. This approach brings up several problems. Testing randomness is associated to the fact that the null hypothesis places the parameter on the border of the parametric

space. Testing fixed parameters is related to the confounding impact on the statistic of the random effect that must be "removed" to avoid a misleading interpretation of the significance. We face the same type of problem encountered in penalty function approaches.

We agree with some authors (Rocha and Singer, 2017) that none of the proposed procedures should be used as the only procedure to select the fixed and random coefficients in linear mixed models. In fact they should be taken as complementary and the decision should be based on all available information. The goal of this work is to propose an additional (hopefully useful) instrument for such purposes. Specifically, we show how to choose fixed and random coefficients through ordinary least squares estimators, defining two simple statistics based on a ratio between a statistic which contains the effect and another that does not. More precisely, the test for randomness is constructed by comparing the sample covariance matrix of ordinary least squares, ols, (see Gumpertz and Pantula (1989)) with the same matrix under the hypothesis of zero random effect. As a result, the statistic has a known distribution for any sample size and captures randomness indirectly avoiding the boundary problem. The 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: $(random\ effect + \ fixed\ effect)/random\ effect$ defines a test statistic with random effect removed. The distribution of this statistic is approximated with a noncentral $F-$distribution. A selection procedure may be conducted through a joint analysis on the significance of these two tests.

The use of ordinary least squares makes the approach simple and may be seen as an attempt to overcome the boundary and confounding problems of testing procedures. Furthermore, point estimates and (approximate) confidence intervals can be constructed as complementary information useful for choosing variables.

Section 2 introduces some notations and defines the two stage linear mixed model. Section 3 defines the statistics for testing randomness and fixed effects. Section 4 discusses the density function of the test statistics. Section 5 introduces the simulations outlining the limits of the analysis. This section is divided into two subsections. Subsection 5.1 describes the base scenarios for all simulations. Subsection 5.2 study and discuss the goodness-of-approximation of the noncentral $F-$distribution. Appendix A examines and defines the exact density function of the test statistic for randomness. Appendix B deals with the approximated distribution.

2 Two-Stage Random Effects: Model and notations

The linear mixed model for longitudinal data can be described as follows: $y_i = X_i^* \beta^* + Z_i d_i^* + u_i$, $i = 1, \ldots, n$ 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 Rocha and Singer (2017) we re-express the linear mixed model as a two-stage random coefficients model Laird (2004),

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

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 element 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 element 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 recall that the $h-$th element of the vector β_i is $\beta_{hi} = \beta_h + d_{hi}$ that is $\beta_{hi} = fixed\ +\ random$. Then, we define two statistics, one for testing randomness "removing" the fixed effect. Is $E(\beta_{hi} - \beta_h) = 0$ for any $\beta_h \in \mathbb{R}$? The other statistic is defined for testing the nullity of the fixed effect "removing" randomness from β_{hi}.

1. Is β_{hi} a random parameter? Hypotheses: $H_0 : \omega_{hh} = 0 \bigcap \beta_h \in \mathbb{R}\ ,\ H_1 : \omega_{hh} > 0 \bigcap \beta_h \in \mathbb{R}$.
   Observe that $\omega_{hh} = 0$ implies $d_{hi} = 0$ with probability 1 and $\beta_{hi} = \beta_h \in \mathbb{R}$.
   We construct 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 (Gumpertz and Pantula, 1989), $S_b = (n-1)^{-1}\sum_{i=1}^{n}\left(b_{hi} - \overline{b}_h\right)\left(b_{hi} - \overline{b}_h\right)'$. We recall that $E(S_b) = \sigma^2\overline{V} + \Omega_k$ then $E\left(S_{b_h}|H_1\right) = \sigma^2\overline{v}_{hh} + \omega_{hh}$ and $E\left(S_{b_h}|H_0\right) = \sigma^2\overline{v}_{hh}$. By comparing these two expected values we capture the randomness, by working on the difference, $b_{hi} - \overline{b}_h$ we remove the fixed effect from the statistic. The test statistic developed in this work is an estimate of the ratio of this two variances

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

   The statistic is otained from the ratio $S_h = \frac{1}{n-1}\sum_{i=1}^{n}(b_{hi} - \overline{b}_h)^2/(\sigma^2\overline{v}_{hh})$ by replacing σ^2 with the sample variance s^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 parameter θ_h may be interpreded as a measure of the relative change of the total variance of b_{hi} with respect to the variance of b_{hi} under H_0. In this sense it is a relative measure of randomness. We call $rp = \dfrac{\omega_{hh}}{\sigma^2 \overline{v}_{hh}}$ randomness parameter.

The expected value of T_h is given by

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

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

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 $\frac{\omega_{hh}}{\sigma^2} > 0$, θ_h is greater than 1 and T_h deviates from its expected value. The farther $\frac{\omega_{hh}}{\sigma^2} > 0$ is from zero, the greater are θ_h and T_h, everything else being equal. The greater T_h the stronger the evidence against H_0. The parameter, *rp*, plays the same role as a noncentrality parameter of an F-distribution. As we shall see, if *rp* increases, θ_h increases too, 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 $H_1 : \theta_h > 1$ with T_h taken as a test statistic. H_0 is rejected if T_h is "much" greater than one.

2. Is $\beta_h = 0$? Hypotheses: $H_0 : \beta_h = 0 \bigcap \omega_{hh} \geq 0$, $H_1 : \beta_h > 0 \bigcap \omega_{hh} \geq 0$.

   We develop a test based on $\overline{b}_h$. The quadratic mean under H_1, $E(n\,\overline{b}_h^2|H_1)$ is compared by a ratio with $E(n\,\overline{b}_h^2|H_0) = E\,(S_{b_h})$,

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

   where $ncp = \dfrac{n\,\beta_h^2}{\sigma^2 \overline{v}_{hh} + \omega_{hh}}$ is a noncentrality parameter. The numerator incorporates both random and fixed effect, the ratio removes the random effect. The statistic we propose is an estimate of the above ratio,

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

When $\beta_h = 0$ the test statistic F_h takes values around the expected value of an $F-$distribution. If $\beta_h \neq 0$ then F_h deviates from its expected value. The greater β_h^2 the further away the peak of the distribution from zero.The bigger the ncp, the more the alternative sampling distribution moves to the right and the more power we have. The larger the noncentrality parameter the greater the power. The null hypothesis is rejeted for large value of F_h.

The test statistics T_h and F_h can be jointly used for a selection variable approach. Tab.: 1 shows by row the significane of T_h and by column the significance of F_h,

Table 1: Selection of fixed and random coefficients
in a linear mixed model

	Significance of F_h		
Significance of T_h	*yes*	*No*	
Yes	$\beta_{hi} = \beta_h + d_{hi}$	$\beta_{hi} = d_{hi}$	$\theta_h > 1,\ \forall\, \beta_h$
No	$\beta_{hi} = \beta_h$	$\beta_{hi} = 0$	$\theta = 1,\ \forall\, \beta_h$
	$\beta_h \neq 0,\ \theta_h \geq 1$	$\beta_h = 0,\ \theta_h \geq 1$	

The Table can be read by column or by row. Let consider the cell $(yes,\ yes)$. Reading by row, *yes* means that T_h is significative: we reject the hypothesis, H_0, "no randomness" $(\theta_h > 1,\ \forall\, \beta_h)$ then likely β_{hi} is random. By column, *yes* implies the significance of F_h $(\beta_h \neq 0,\ \theta_h \geq 1)$. The cell $(yes,\ yes)$, (significance of both tests) means that presumably β_{hi} consists of both a fixed and a random component: $\beta_{hi} = \beta_h + d_{hi}$.

4 Density functions

Appendix A describes the exact density function of T_h both under the null and the alternative hypotheses. In this section we discuss and develop the density function of F_h.

Let us divide and multiply expression F_h (formula (3)) by $\sigma^2 \overline{v}_{hh} + \omega_{hh}$ and analyze the numerator and the denominator. According to the assumptions of the model,

$$\frac{n\,\overline{b}_h^2}{\sigma^2 \overline{v}_{hh} + \omega_{hh}} \sim \chi^2\left(1, ncp = \frac{n\,\beta_h^2}{\sigma^2 \overline{v}_{hh} + \omega_{hh}}\right) \quad for\ any\ \ \omega_{hh} \geq 0 \tag{4}$$

and

$$Q_h = \frac{S_{b_h}}{\sigma^2 \overline{v}_{hh} + \omega_{hh}} = \frac{1}{n-1} \frac{\sum_{i=1}^{n} \left(b_{hi} - \overline{b}_h\right)^2}{\sigma^2 \overline{v}_{hh} + \omega_{hh}} \sim \frac{1}{n-1} \sum_{i=1}^{n-1} \tau_i\, \chi^2(1) \approx\ a\, \chi^2(b) \tag{5}$$

where τ_i, a and b are defined in Appendix B.

Q_h is distributed as a linear combination of $\chi^2(1)$ the exact distribution of which is derived in Appendix B. The knowledge of the exact distribution of Q_h is not very useful for defining a "simple" distribution of the statistic F_h, so following (Yuan and Bentler, 2010) we approximate the distribution of Q_h by $a\chi^2(b)$ (we write $Q_h \approx a\chi^2(b)$) where a and b are determined by matching the first two moments of Q_h with those of $a\chi^2(b)$ (see Appendix B). The ratio between the exact Chi-square distribution given by expression (4) and the approximate Chi-square distribution given by (5) each divided by its degrees of freedom gives the following approximate distribution of F_h,

$$F_h = \frac{n\,\overline{b}^2}{(n-1)^{-1}\sum_{i=1}^{n}\left(b_{hi}-\overline{b}_h\right)^2} \approx F\left(1, b=n-1, ncp\right) \quad for\ any\ \ \omega_{hh} \geq 0 \tag{6}$$

We recall (see (Appendix B) that the degrees of freedom, b, of the approximated Chi-square distribution depend on the random component and ranges between b_0 when $\omega_{hh}=0$ and $(n-1)$ when ω_{hh} is large. We work setting $b=n-1$ mainly for the following reasons

- Given $\bar{v}_{hh}$, if ω_{hh} is large with respect to σ^2 (θ_h "much larger" than one), then the statistic F_h has an exact $F-$distribution, $F_h \sim F(1, n-1, ncp)$. We recall that the farther $\frac{\omega_{hh}}{\sigma^2} > 0$ from zero, the greater θ_h and T_h, everything else being equal. The greater T_h the stronger the evidence in favour of the presence of randomness. Simulations show that if the *pvalue* of T_h is less than 0.001 the distribution may be considered "exact". In applications $T_h >>$ *critical value* or *pvalue* $<< 0.05$ justify the choice $b=n-1$.
- If the number of observations, n, is large then the shape of the noncentral $F-$distribution is similar to the noncentral chi squared distribution with 1 degrees of freedom,
  $F_h \sim \chi^2(1, ncp)$ as $n \to \infty$.
- The approximating $F-$distribution depends on the number of unit, n. As n increases, according to the central limit teorem, the exact distribution of $Q_h = S_{b_h}$ may be approximately described by a normal distribution, and so may $a\chi(b)$. Thus, we expect that the approximation will improve as n increases.
- We recall that as $\theta_h \to 1$ ($\theta_h = 1$), $b \to b_0$ (see Appendix B). Then, if $ncp = 0$ the quantiles of $F(1, b_0)$ are greater than the quantiles of $F(1, n-1)$ (so is the critical value) this implies that, given α, the rejection region defined by $F(1, b_0)$, R_0, is a subset of the rejection region defined by $F(1, n-1)$, R. This has two consequences, the first is $P(F_h \in R_0 | ncp > 0) \leq P(F_h \in R | ncp > 0)$ then the approximation with b_0 degrees of freedom is less powerful than

the approximation with $n-1$ degrees of freedom, the second is that the test based on b_0 is conservative when randomness is different from zero.

5 Simulation

The selection procedure proposed in table 1 is powerful if the two test statistics defined for the selection, T_h and F_h are powerful. Any simulation for evaluating the performance of the approach proposed requires a preliminary analysis of T_h and F_h. As a consequence an exhaustive analysis of the selection procedure produces a huge amount of results which cannot all be shown in this paper and a choice is made.

The statistic T_h is based on the works of Barnabani (2017) and Barnabani (2019) 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 statistic F_h needs more attention. We expect that given the noncentrality parameter the distribution of F_h is not influenced by the random effect, that is $F_h \approx F(1, n-1, ncp)$ for any $\theta_h \geq 1$. In this sense we say that the test statistic, F_h, is "stable" with respect to the randomness parameter. Therefore, a simulation study on the goodness-of-approximation focusing on the effect of θ_h (randomness) and ncp (noncentrality parameter) on F_h plays a crucial role for evaluating power, consistency and "stability" of the test statistic. This study is conducted by analyzing a set of matrices, $M_1, \ldots, M_l, \ldots, M_s$ constructed as follows.

Let denote with ncp_l, $l = 1, \ldots, s = 6$ one specific value of ncp in $A = \{0, 1, 2, 3, 5, 8\}$. We consider $nrepl = 100$ different values of θ_{hj}, $j = 1, \ldots, nrepl$ 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 all 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, $l = 1, \ldots, s$ is the basis of our analysis on the goodness-of-approximation of the test statistic.

5.1 "Base" Scenario for simulations

To allow the maximum of generality and arbitrariness, 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, $U(k+4, 3k)$.

(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

We have the scheme: $F_h \sim$ "*exact*" *cdf* $\approx$ "*approx*" *cdf*. Since we are not able to evaluate the goodness of the approximating *cdf* to the "exact" *cdf* we by-pass the second step and use methods for analysing the fitting of the $F-$distribution to the test statistic F_h.

To this purpose we choose a specific combination of (ncp_l, θ_{hj}) and extract one column from M_l. With these data ($N = 1000$ simulated samples) we show and comment some usual graphs, 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 "stability" of the results obtained.

Given the parameter combination $(ncp_l = 3, \theta_{hj} = 3.85)$, 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: $p = \{0.025, 0.05, 0.075, \ldots, 0.975\}$ 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.

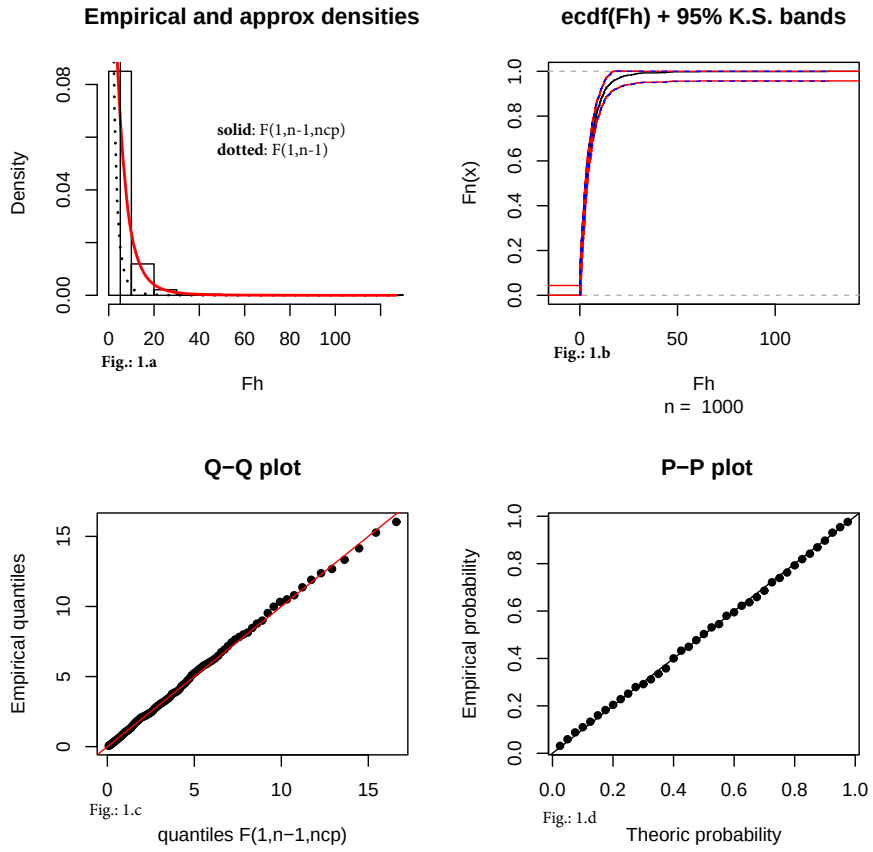


Figure 1: - Fig.: 1.a shows simulated histogram of F_h, the central F (dotted line) and noncentral F (solid line). The displacement of the solid line from the dotted line is due to the ncp with $\theta_h \geq 1$. The parameters are: $h = 3$, $ncp = 3$ $\theta_h = 3.85$. - Fig.: 1.b show the graph of 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.: 2. Tab.: 3 shows the confidence intervals obtained with the basic bootstrap procedure. All the results are quite satisfactory.

Table 2: 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 3: 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 $[0,10]$. Moreover, working on the whole matrix M_l, we can analyze the goodness of fit of the $F-$distribution to the "exact" unknown distribution function of F_h.

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.

Let define the empirical probability of rejecting F_h under H_0 and under H_1,

$$\widehat{\alpha}_j=\frac{1}{N}\sum_{i=1}^{N}\delta(F_{ij}>q_{0.95}|H_0)\qquad\widehat{\gamma}_j=\frac{1}{N}\sum_{i=1}^{N}\delta(F_{ij}>q_{0.95}|H_1)$$

where $\delta(true)=1$ and $\delta(false)=0$, $P(F(1,n-1)\leq q_{0.95})=0.95$ and F_{ij} is the element (i,j) of the matrix M_l, that is, the value of the statistic F_h computed on the $i-$th simulated sample $(i=1,\ldots,N)$ and a specific value of θ_{hj}, $j=1,\ldots,nrepl$.

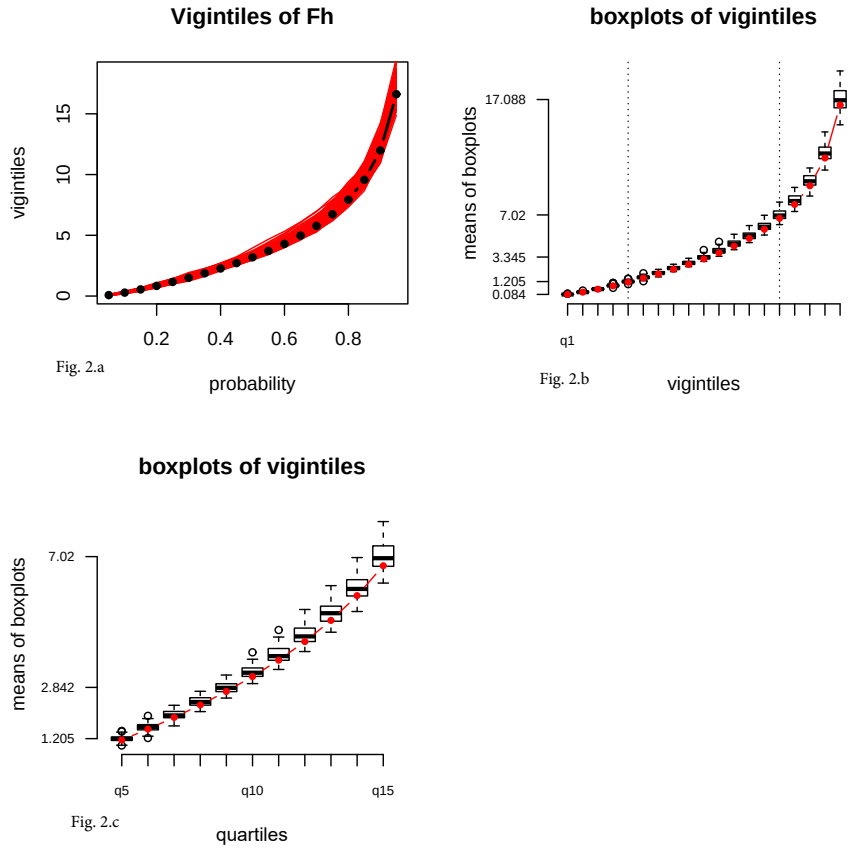


Figure 2: - Fig.:2.a shows the bundle of lines of empirical vigintiles and the points of the approximating noncentral distribution. - Fig.: Fig.:2.b shows the boxplots of empirical vigintiles and the points of the approximating distribution. - Fig.: Fig.:2.c shows the graph of Fig.:2.b limited to the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

We expect that $\widehat{\alpha}_j$ and $\widehat{\gamma}_j$ does not change "significantly" as $\theta_{hj}A$ changes. Tab.: 4 shows summary statistics of $\widehat{\alpha}_j$ (first row) and $\widehat{\gamma}_j$ (for different value of ncp) the interquartile difference can be taken as a measure of the variability.

Table 4: Summary statistics of $\widehat{\alpha}_j$ and $\widehat{\gamma}_j$, $N = 1000$, $n = 10$, $nrepl = 50$,

Hypothesis		Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
$H_0 : ncp = 0 \bigcap \theta_h \geq 1$	$\widehat{\alpha}_j$:	0.036	0.045	0.05	0.0498	0.05475	0.076
$H_1 : ncp = 2 \bigcap \theta_h \geq 1$	$\widehat{\gamma}_j$:	0.1060	0.2572	0.2685	0.2612	0.2775	0.307
$H_1 : ncp = 5 \bigcap \theta_h \geq 1$	$\widehat{\gamma}_j$:	0.1860	0.441	0.553	0.504	0.572	0.599
$H_1 : ncp = 8 \bigcap \theta_h \geq 1$	$\widehat{\gamma}_j$:	0.353	0.69	0.75	0.705	0.77	0.81

Let f_{oss} be a realized value of the test statistic F_h. The cumulative distribution function of the pvalue (written as $cdf\ [P(F_h \geq f_{oss} \mid "true" df)]$, is a 45° line. Then, if the (empirical) cumulative probability of $P(F_h \geq f_{oss} \mid "approx" df)$ that is the cdf of the pvalue computed with the approximating $F-$distribution, is "close" to the 45 degrees line we face to a "good" approximation.

Fig.:3.a shows the ”true” cdf of the pvalue (black points) and the empirical cdf of 10 replications of the approximated pvalues computed with an $F(1,9,ncp=3)$.

The comparison between the ”exact” and ”approx” pvalue allows us also to detect the ”approx” power function of the test (see Fig.:3.b). In terms of pvalue we recall that the (empirical) power function of the test is given by $cdf(pvalue < p|H_1)$ varying p on the interval $[0,1]$ and H_1 defined for a specific combination $(ncp_l, \theta_h > 1)$. The 45° line can be seen as the power function computed when $H_0 = H_1$, the pvalue always equals power. The graph of Fig.:3.b is generated by varying f_{oss}. The upper right hand of the graph corresponds to $f_{oss} = 0$. Both pvalue and power are 1 at this point. The lower left-hand corner corresponds to a very large f_{oss}, so large that the test statistic will never exceed it. Both pvalue and power are 0 at this point. Normally, we would expect the power of the test to exceed its pvalue for any given f_{oss},except when pvalue and power are both equal to 0 or 1. The curves in Fig.:3.b are examples of tests for which this is the case.

power tradeoff curve given by the 45° line. A test for which size always equals power has a size-è If the DGP for which the tradeoff curve is constructed actually satisfies the null hypothesis. The graph of Fig.:3.b is generated by varying f_{oss}.

The size-power tradeoff curve is generated by varying f_{oss}. The upper right hand of the graph corresponds to $f_{oss} = 0$. Both size (pvalue) and power are 1 at this point. The lower left-hand corner corresponds toa very large $f_{oss} = 0$, so large that the test statistic will never exceed it. Both size and power are 0 at this point.

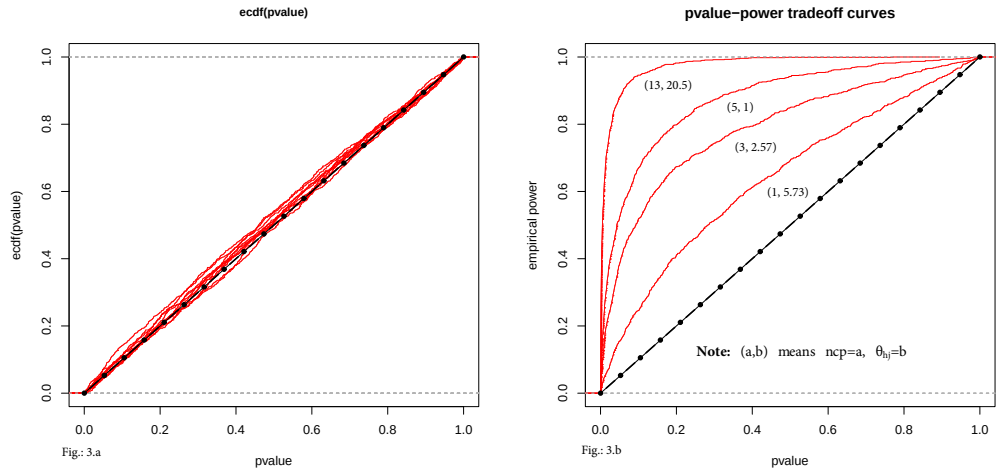


Figure 3: - Fig.:3.a shows the points of $cdf\ [P(F_h \geq f_{oss} \mid \text{”}true\text{”}df)]$ (45° line) and 10 replications of the (empirical) cumulative probability of $P(F_h \geq f_{oss} \mid \text{”}approx\text{”}df)$ (solid lines). The approximating distribution is an $F(1,9,ncp=3)$. - Fig.: Fig.:3.b shows empirical power functions for different combinations $(ncp_l, \theta hj)$. The power increases as ncp_l increases for any $\theta_{hj} \geq 1$.

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 the work are developed for investigating the "boundary" and "confounding" problems that are crucial for evaluating the significance of the tests.

In our opinion an indirect approach based on a ratio between two quadratic forms one that contains the effect and another that does not, may be a good method to tackle the above-mentioned 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 the exact or approximative density function and allow an analysis of the tests in small samples.

By using only ordinary regression results, the analysis maintains great simplicity and in addition to the analysis on the significance, the approach can be developed by improving the power of the two test statistics, T_h and F_h and by introducing other instruments such as point estimates and (approximative) confidence intervals of randomness and noncentrality parameter.

Appendix A Exact probability density function of $\mathbf{T}_h$

Consider first the distribution of $(b_i - \overline{b})$ when $\Omega_k \succeq 0$ and σ^2 is known. The vector has a normal distribution with mean zero and variance covariance matrix $\sigma^2 V_{ii} + \dfrac{n-1}{n}\Omega_k$. Let v_{hi}, $h = 1,\ldots,k$ be the h-th diagonal element of V_{ii} and denote with $D^{-1/2} = I_n \otimes \overline{V}_D^{-1/2}$ the block diagonal matrix with $\overline{V}_D^{-1/2} = diag(1/\sqrt{(}\overline{v}_{11}),\ldots,1/\sqrt{(}\overline{v}_{hh}),\ldots,1/\sqrt{(}\overline{v}_{kk}))$. Then, the vector $\overline{V}_D^{-1/2}(b_i - \overline{b})/\sigma$ is multivariate normal with mean zero and covariance matrix $\overline{V}_D^{-1/2}\left(V_{ii} + \frac{n-1}{n}\frac{\Omega_k}{\sigma^2}\right)\overline{V}_D^{-1/2}$. The h-th component, $s_{hi} = (b_{hi} - \overline{b}_h)/(\sigma\sqrt{\overline{v}_{hh}})$, is $N(0,\theta_{hi})$ where $\theta_{hi} = \frac{v_{hi}}{\overline{v}_{hh}} + \left(\frac{n-1}{n}\right)\frac{\omega_{hh}}{\sigma^2\overline{v}_{hh}}$ and ω_{hh} denotes the h-th element of the main diagonal of Ω_k.

The square, s_{hi}^2, has a gamma distribution with shape parameter $\alpha = 1/2$, scale parameter $2\,\theta_{hi} = 2\left[\frac{v_{hi}}{\overline{v}_{hh}} + \left(\frac{n-1}{n}\right)\frac{\omega_{hh}}{\sigma^2\overline{v}_{hh}}\right]$ and $E(s_{hi}^2) = \theta_{hi}$

Let's denote $W = R + G$ where $R = D^{-1/2} V D^{-1/2}$ is the $nk \times nk$ covariance matrix when $\Omega_k = 0$ (H_0 is true) and $G = D^{-1/2} H_n \otimes \frac{\Omega_h}{\sigma^2} D^{-1/2}$ the $nk \times nk$ covariance matrix of random components. Moreover, 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. The i-th diagonal element of W_h is θ_i, the other

elements are covariances. The column vector $[s_{h1}, s_{h2}, \ldots, s_{hn}]$ is multivariate normal with zero mean and covariance matrix W_h, $\{s_{h1}^2, \ldots, s_{hi}^2, \ldots, s_{hn}^2\}$ is a set of correlated gamma.

The average, $S_h = \frac{1}{n-1}\sum_{i=1}^{n} s_{hi}^2$ is a sum (scaled by $n-1$) of correlated gamma with same shape parameter, $\alpha = 1/2$, and different scale parameters. The expected value of S_h is equal to $\theta_h = \frac{1}{n-1}\sum_{i=1}^{n}\theta_{hi} = 1 + \frac{\omega_{hh}}{\sigma^2 \bar{v}_{hh}}$. When the h-th coefficient is not random, $\omega_{hh} = 0$ and $E(S_h) = 1$. The arithmetic mean S_h has the same distribution as a sum of chi-square random variables (Mathai and Provost, 1992) (section $3.2a.2$, p. 36, Eq. $(3.2a.8)$),

$$S_h \sim \sum_{i=1}^{n-1} \frac{\phi_i}{n-1} G\left(\alpha = \frac{1}{2}, \beta = 2\right) \tag{7}$$

where the ϕ_i's are the eigenvalues of W_h and capture the presence of randomness. Observe that $E(S_h) = \theta_h = \frac{1}{n-1}\sum_{i=1}^{n-1}\phi_i$. When H_0 is true $\sum_{i=1}^{n-1}\frac{\phi_i}{n-1} = 1$.

If in $\{s_{hi}^2\}$ we replace σ^2 with the sample variance s^2 we get

$$t_{hi}^2 = \frac{(b_{hi} - \bar{b}_h)^2}{s^2\,\bar{v}_{hh}} = \frac{s_{hi}^2}{(df\,s^2)/(df\,\sigma^2)}, \qquad with \quad df = \sum_{i=1}^{n}(t_i - k) \tag{8}$$

which can be seen as the ratio of two independent gamma distributions: the numerator is $G(\frac{1}{2}, 2\,df\,\theta_{hi})$, the denominator is $G(\frac{df}{2}, 2)$. Therefore, $t_{hi}^2 \sim GF(\frac{1}{2}, \frac{df}{2}, 1, \frac{2\,df\,\theta_{hi}}{2}) = GF(\frac{1}{2}, \frac{df}{2}, df\,\theta_{hi})$ where GF denotes generalized F-distribution and the argument 1 is omitted.

The arithmetic mean $T_h = \frac{1}{n-1}\sum_{i=1}^{n} t_{hi}^2$ has the same distribution as the ratio of two random variables where the numerator is a sum of gamma. By (7) we have,

$$T_h \sim \sum_{i=1}^{n-1} \frac{\phi_i}{n-1}\,\frac{G(\,\frac{1}{2}\,, 2\,df)}{G(\frac{df}{2}, 2)} \tag{9}$$

Following Moschopoulos (1985), by expressing the numerator of (9) as a single gamma-series representation, the density function of T_h is written as

$$f_{T_h}(x) = \sum_{k=0}^{\infty} p_k \frac{G(\rho + k, \beta_1)}{G\left(\frac{df}{2}, 2\right)} = \sum_{k=0}^{\infty} p_k\, GF\left(\rho + k, \frac{df}{2}, \frac{\beta_1}{2}\right) \tag{10}$$

where $p_k = C\delta_k$, $\beta_1 = \min_i\{\beta_i\}$, $C = \prod_{i=1}^{n-1}\left(\frac{\beta_1}{\beta_i}\right)^{\alpha_i}$, $\rho = \sum_{j=1}^{n-1}\alpha_j$, $\alpha_j = 1/2\ \forall j$, $\beta_i = 2\phi_i\frac{df}{n-1}$, and

the coefficients δ_k obtained recursively by the formula

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

Since the series representation of Moschopoulos (1985) is CPU-time intensive when the shape parameters are small and the scale parameters have large variation, Barnabani (2017) proposed to approximate the probability distribution $(k,p_k)_{k=0,1,2\ldots}$ with a generalized negative binomial distribution.

The function (10) is uniform convergent (Moschopoulos, 1985). This property justifies the interchange of the integration and summation and allows us to compute the distribution function and quantiles.

Appendix B Approximation

Following the describtion of the exact density distribution of S_h of App.A, the ratio $Q_h = S_h/\theta_h$, $Q_h = \frac{1}{n-1}\frac{\sum_{i=1}^{n}\left(b_{hi}-\overline{b}_h\right)^2}{\sigma^2\overline{v}_{hh}+\omega_{hh}}$ is distributed as $\frac{1}{n-1}\sum_{i=1}^{n-1}\tau_i\chi^2(1)$ where $\tau_i = \phi_i/\theta_h$. We can derive the exact distribution of Q_h using the single gamma-series representation proposed by Moschopoulos (1985). *Fig.* : 1.*a* shows a simulated histogram of $(n-1)Q_h$ and its exact density function. The expected value of $(n-1)\,Q_h$ is $n-1$, the variance is $Var\left[\left((n-1)\,Q_h\right] = K(\tau)\right.$ where $K(\tau)$ denotes an unknown expression that depends on the random component (ω_{hh}) through θ_h.

Following Yuan and Bentler (2010) we discuss 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 $a = \dfrac{(n-1)^2}{K(\tau)}$ and $b = \dfrac{n-1}{a}$. We discuss the following approximation,

$$\frac{(n-1)Q_h}{a} = \frac{1}{n-1}\frac{\sum_{i=1}^{n}\left(b_{hi}-\overline{b}_h\right)^2}{\sigma^2\overline{v}_{hh}+\omega_{hh}}\,\frac{K(\tau)}{n-1} \;\approx\; \chi^2(b) \tag{11}$$

observe that if (11) is divided by the degree of freedom, b, we remove $K(\tau)$ from the quadratic form.

As known (Yuan and Bentler, 2010) the chi-square approximation (11) depends on the relative sizes of the τ_i's, on their variabilities and on the degrees of freedom. In this section we analyze these effects on the approximation.

1. When $\tau_1 = \tau_2 = \ldots = \tau_{n-1} = \tau$, then $a = 1$, $b = n-1$, $\tau = \theta_h$ and the approximation in (11)

is exact, $(n-1)Q_h \sim \chi(n-1)$. The equality of τ_i's occurs when given $\bar{v}_{hh}$, ω_{hh} is (very) large with respect to σ^2, or when 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. Let t_{oss} be the observed value of T_h, if $P(T_h > t_{oss}|H_0) < 0.001$ that is, if $pvalue < 0.001$ then the variability of eigenvalues is (approximately) zero, $b = n-1$ and $(n-1)Q_h \sim \chi(n-1)$.

Yuan and Bentler (2010), shows algebraically that when $\omega_{hh} = 0$, b reaches the maximum value at $n-1$.

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, $b_0 \leq b \leq n-1$ and $1 \leq a \leq a_0$. We recall that b depends on ω_{hh} and can not be calculate while b_0 and a_0 can be. 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 4 the bottom left graph ($Fig.: 1.c$), shows the distribution functions (cdf) of $\chi^2(b_0)$, $\chi^2(n-1)$ and the empirical distribution function (ecdf) of (11) which collocates between the two curves. The ecdf of (11) is "well" approximated by a $\chi^2(b)$. $Fig.: 1.c$ 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 $a\chi(b)$. Thus, we may expect that the approximation will improve as b increases.

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 Kiadò.

Barnabani, M. (2017) A parametric test to discriminate between a linear regression model and a linear latent growth model. *International Journal of Statistics and Probability*, **6(3)**, 157–167.

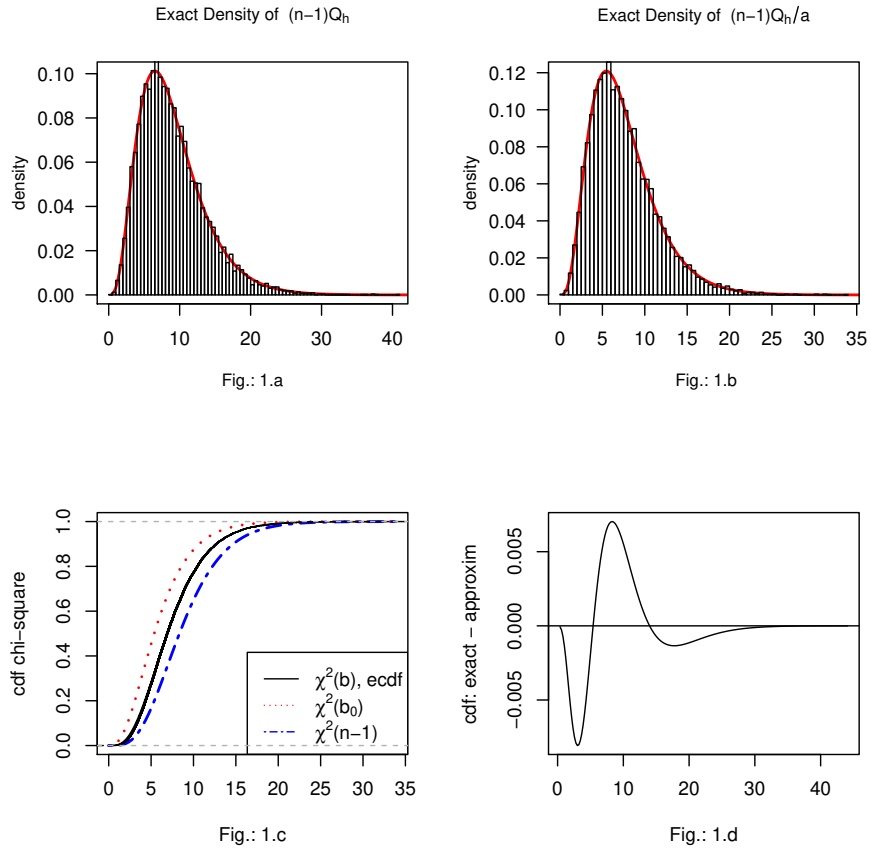


Figure 4: - Fig.: $1.(a)$ and Fig.: $1.(b)$ show simulated histogram and exact density of $(n-1)Q_h$ and $(n-1)Q_h/a$ for the third element $(h=3)$. - Fig.: $1.(c)$ represents the distribution functions (cdf) of $\chi^2(b_0)$ (dotted line), $\chi^2(n-1)$ (twodashed line) and $\chi^2(b)$ (solid black line). Fig.: $1.(d)$ shows the diffferfences between the cdf of the approximated $\chi^2(b)$ and the cdf of the exact distribution.

— (2019) An f-type multiple testing approach for assessing randomness of linear mixed models. *American Review of Mathematics and Statistics*, **7(2)**, 15–28.

Bondell, H. D., Krishna, A. and Ghosh, S. K. (2010) Joint variable selection for fixed and random effects in linear mixed-effects models. *Biometrics*, **66**, 1069–1077.

Chen, Z. and Dunson, D. B. (2003) Random effects selection in linear mixed models. *Biometrics*, **59**, 762–769.

Fan, J. and Li, R. (2001) variable selection via concave penalized likelihood and its oracle properties. *American Stat. Assoc.*, **96(456)**, 1348–1360.

— (2012) variable selection in linear mixed effects models. *The Annals of Statistics*, **40**, 2043–2068.

Gumpertz, M. and Pantula, S. G. (1989) A simple approach to inference in random coefficient model. *The American Statistician*, **43, No. 4**, 203–210.

Ibrahim, J. G., Zhu, H., Garcia, R. I. and Guo, R. (2011) Fixed and random effects selection in mixed effects models. *Biometrics*, **67**, 495–503.

Laird, N. M. (2004) *Chapter 5: Random effects and the linear mixed model*, vol. 8 of *Regional Conference Series in Probability and Statistics*, 79–95. Beechwood OH and Alexandria VA: Institute of Mathematical Statistics and American Statistical Association. URLhttps://projecteuclid.org/euclid.cbms/1462106081.

Lin, B., Pang, Z. and Jiang, J. (2013) Fixed and random effects selection by reml and pathwise coordinate optimization. *Journal of Computational and Graphical Statistics*, **22**, 341–355.

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.

Peng, H. and Lu, Y. (2012) Model selection in linear mixed effect models. *Journal of Multivariate Analysis*, **109**, 109–129.

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**.

Schwarz, G. (1978) Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.

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.

Wu, P., Luo, X., Xu, P. and Zhu, L. (2017) New variable selection for linear mixed-effects models. *Ann Inst Stat Math*, **69**, 627–646.

Yuan, K. H. and Bentler, P. M. (2010) Two simple approximations to the distributions of quadratic forms. *British Journal of Mathematical and Statistical Psychology*, **63, No. 2**, 273–291.