Asymptotic distribution of likelihood ratio test statistics for variance components in nonlinear mixed effects models

Charlotte Baey, Paul-Henry Cournède, Estelle Kuhn

Version of Record: https://www.sciencedirect.com/science/article/pii/S0167947319300283
Manuscript_e3961f201c7bca72121af3e7fdf942ee

Asymptotic distribution of likelihood ratio test statistics for variance components in nonlinear mixed effects models

Charlotte Baey

Laboratoire Paul Painlevé, Université de Lille, 59655 Villeneuve d'Asq, France

Paul-Henry Cournède

Laboratoire MICS, CentraleSupélec, 92190 Gif sur Yvette, France

Estelle Kuhn*

MaIAGE, INRA, Domaine de Vilvert, 78352 Jouy-en-Josas, France

Abstract

Mixed effects models are widely used to describe heterogeneity in a population. A crucial issue when adjusting such a model to data consists in identifying fixed and random effects. Testing the nullity of the variances of a subset of random effects can help to investigate this issue. Some authors have proposed to use the likelihood ratio test and have established its asymptotic distribution in some particular cases. Extending the existing results, a likelihood ratio test procedure is studied, to test that the variances of any subset of the random effects are equal to zero in nonlinear mixed effects model. More precisely, the asymptotic distribution of the test statistics is shown to

*Corresponding author

Email address: estelle.kuhn@inra.fr ; INRA, MaIAGE, Batiment 210, Domaine de Vilvert, 78352 Jouy-en-Josas, France ; tel: 0033134652701 ; fax: 0033134652901 (Estelle Kuhn)

Preprint submitted to Computational Statistics and Data Analysis *January 18, 2019*

be a chi-bar-square distribution, that is to say a mixture of chi-square distributions, and the corresponding weights are identified. In particular, it is highlighted that the limiting distribution depends strongly on the presence of correlations between the random effects. The finite sample size properties of the test procedure are illustrated through simulation studies and the test procedure is applied to two real datasets of dental growth and of coucal growth.
Keywords: Chi-bar-square distribution, inference under constraints, hypothesis testing, likelihood ratio test, nonlinear mixed effects models, variance components

---

1. Introduction

Mixed effects models have been extensively used in population models in order to account for heterogeneity in populations and to describe the intra and inter-individual variability (see Pinheiro and Bates (2000); Davidian and Giltinan (2003); Lavielle (2014)). In mixed effects models, parameters are of two types: on one side, fixed effects that are common to all the individuals of the population; on the other side, random effects that vary from one individual to the other. The last ones are also called individual parameters.

From a modelling point of view, a key question when adjusting a population model to a dataset is to identify the fixed and random effects of the model. Therefore, being able to compare a simpler model with less random effects to a larger one will help to choose a parsimonious model. From a statistical point of view, it can be rephrased as a test on the nullity of the variances of a subset of all the random effects (Silvapulle and Sen (2011), p14).

Several authors have been interested in likelihood ratio tests (LRT) in this context. We recall here some of the main existing results. Let us denote by Θ the parameter space, by Θ_0 the subset of Θ corresponding to the null hypothesis and by Θ_1 the subset corresponding to the alternative hypothesis, with $\Theta_0 \subset \Theta_1 \subset \Theta$. Chernoff (1954), assuming that Θ is open, treated the case where the true value of the parameter lies on the boundary of Θ_0 and Θ_1, which is assumed to be a proper set of Θ, i.e. strictly contained in Θ. He provided a representation of the asymptotic distribution of the LRT and proved that it is asymptotically equivalent to testing the mean of a multivariate Gaussian distribution based on one single observation. A few years later, Chant (1974) generalized these results by considering the case where the true value lies in a subset of Θ which may not be a proper subset. Shapiro (1985) studied the asymptotic distribution of a larger class of tests when the true value is on the boundary of Θ_0 and an interior point of Θ. He established that the asymptotic distribution is a mixture of chi-square distributions. Simultaneously, Self and Liang (1987) obtained similar results in the case where the true value is on the boundary of Θ. They established a general convergence result and they derived the expression of the asymptotic distribution only in specific cases, assuming in particular that the parameter space Θ is equal to the product of a finite number of either closed, half-open or open intervals of $\mathbb{R}$. They considered several specific cases for applications. They identified in particular the limiting distribution of the LRT for testing that the variance of one single random effect is equal to zero as a mixture $\frac{1}{2}\delta_0 + \frac{1}{2}\chi^2_1$, where δ_0 is the Dirac distribution at zero and χ^2_1 is the chi-square distribution with one degree of freedom (Self and Liang, 1987, case 5). They also identified the limiting distribution of the LRT for testing simultaneously that the variance of one single random effect is equal to zero and that its mean is equal to a constant as a mixture $\frac{1}{2}\chi_1^2 + \frac{1}{2}\chi_2^2$ (Self and Liang, 1987, case 6). Nevertheless, their results do not allow to identify the asymptotic distribution of the LRT statistics in all cases, for example testing that one variance is equal to zero when considering two correlated random effects. Building upon those works, several authors have addressed the issue of variance components testing in mixed effects models. In the context of linear mixed effects models, Stram and Lee (1994, 1995) considered the likelihood ratio test procedure and identified the limiting distribution of the LRT statistics in some particular cases. They suggested also that the limiting distribution might in fact be influenced by the presence of correlations between random effects but they do not investigate this issue. Some authors have also proposed finite sample test procedures for variance components testing in linear mixed models, either by deriving the finite sample distribution of test statistics, or using bootstrap and permutation tests. For example, the finite sample size distribution of the likelihood and restricted likelihood ratio test statistics was studied by Crainiceanu and Ruppert (2004) for linear mixed models with one single random effect, and Greven et al. (2008) extended these results to linear mixed models with more than one random effect.

Other approaches were also inquired. Several years later, Qu et al. (2013) proposed a procedure based on the score test for testing several variance components in linear mixed models, and Wood (2013) studied the finite sample distribution of a test based on the restricted likelihood for testing that one variance is null in generalized linear mixed models. Also in the context of linear mixed models, Sinha (2009) studied a bootstrap test based on the score test for testing several variance components in a generalized linear mixed model, while Fitzmaurice et al. (2007), Samuh et al. (2012) and Drikvandi et al. (2013) considered permutation tests for testing several variance components in the context of linear and generalized mixed effects models. Molenberghs and Verbeke (2007) proposed a review of the existing results for testing variance components in mixed effects models, and studied in particular the equivalence between the LRT, the Score test and the Wald test, based on results by Silvapulle and Silvapulle (1995) or Stram and Lee (1994). They also exhibited the common limiting distribution in some specific cases. However, to the best of our knowledge, there exist no results identifying the limiting distribution of the LRT for general tests on variance components in mixed effects models.

In this paper, we consider the LRT in nonlinear mixed effects models to test that the variances of any subset of the random effects are equal to zero and identify its asymptotic distribution as a mixture of chi-square distributions. In Section 2, we present the framework of nonlinear mixed effects models. Section 3 is devoted to the description of the proposed test and its theoretical properties. Practical implementation guidelines are presented in Section 4. Experimental results illustrate the performances of the procedure through simulation studies and real datasets analysis in Section 5. The paper ends with some discussion in Section 6. The technical proofs are given in Appendix.

2. Nonlinear mixed effects model

2.1. Description

We consider the following nonlinear mixed effects model (Davidian and Giltinan (1995) p98, Pinheiro and Bates (2000) p306, Lavielle (2014) p24):

$$y_i = g(\varphi_i, x_i) + \varepsilon_i,$$

where y_i denotes the vector of observations of the i-th individual of size J, $1 \le i \le N$, g a nonlinear function, φ_i the vector of individual parameters of individual i, x_i a vector of covariates, and ε_i the random error term. The vectors of individual parameters $(\varphi_i)_{1\le i\le N}$ are modeled as:

$$\varphi_i = U_i\beta + V_i b_i \quad , \; 1 \le i \le N,$$

where β is the vector of fixed effects taking values in $\mathbb{R}^b$, b_i is the centered vector of Gaussian random effects with covariance matrix Γ, U_i and V_i are covariates matrices of individual i of sizes $p \times b$ and $p \times p$ respectively, and Γ is a covariance matrix of size $p \times p$. The random vectors (b_i) are assumed to be independent. The vectors $(\varepsilon_i)_{1\le i\le N}$ are assumed to be independent and identically distributed centered Gaussian vectors with covariance matrix Σ, and the sequences (ε_i) and (b_i) are assumed mutually independent.

Let us denote by $\theta = (\beta, \Gamma, \Sigma)$ the vector of all the model parameters and by q its dimension. Thus, the parameter space is defined as $\Theta = \mathbb{R}^p \times \mathbb{S}_+^p \times \mathbb{S}_+^J$, where $\mathbb{S}_+^p$ is the set of symmetric, positive semi-definite $p \times p$ matrices.

2.2. Examples

A special but very common case is the one where the function g is linear. The model can be rewritten in the following usual form (Pinheiro and Bates

(2000), p58):

$$y_i = X_i\beta + Z_i b_i + \varepsilon_i \ ,$$

where y_i is the observation vector for individual i, X_i and Z_i are matrices of known covariates, β is the vector of fixed effects, b_i is the vector of centered random effects for individual i, with $b_i \sim \mathcal{N}(0,\Gamma)$, and ε_i is a random error term, with $\varepsilon_i \sim \mathcal{N}(0,\Sigma)$.

Another famous example of a nonlinear mixed effects model is the logistic growth model, which was studied for example by Pinheiro and Bates (2000) in their well known example of orange trees growth. In this model, a logistic curve is used to model the growth of each individual in the population as a nonlinear function of three individual parameters. Denoting by y_{ij} the variable measured for individual i at age x_j (e.g. the trunk circumference in the orange trees example), for each individual i these three parameters are: the asymptotic value of the trunk circumference φ_{i1}, the age at which the individual reaches half its asymptotic value, φ_{i2}, and the growth scale φ_{i3}. More precisely, we have, for $1 \leq i \leq N, \ 1 \leq j \leq J$:

$$y_{ij} \quad = \quad \frac{\varphi_{i1}}{1+\exp\left(-\frac{x_j-\varphi_{i2}}{\varphi_{i3}}\right)} + \varepsilon_{ij}, \quad \text{with} \quad \varphi_i = \beta + b_i \ ,$$

where $\beta = (\beta_1, \beta_2, \beta_3) \in \mathbb{R}^3$, $b_i \sim \mathcal{N}(0,\Gamma)$ and $\varepsilon_{ij} \sim \mathcal{N}(0,\sigma^2)$.

3. Variance components testing

3.1. Description of the testing procedure

Let $r \in \{1,\ldots,p\}$. We consider general test hypotheses of the following form, to test the nullity of r variances and of the corresponding covariances:

$$H_0 : \theta \in \Theta_0 \quad \text{against} \quad H_1 : \theta \in \Theta, \tag{1}$$

where $\Theta_0 \subset \Theta$. Up to permutations of rows and columns of the covariance matrix Γ, we can assume that we are testing the nullity of the last r variances. We write Γ in blocks as follows:

$$\Gamma = \left(\begin{array}{c|c} \Gamma_1 & \Gamma_{12}^t \\ \hline \Gamma_{12} & \Gamma_2 \end{array} \right),$$

with Γ_1 a $(p-r) \times (p-r)$ matrix, Γ_2 a $r \times r$ matrix, Γ_{12} a $r \times (p-r)$ matrix and where A^t denotes the transposition of matrix A, for any matrix A.

The spaces associated to the null and alternative hypotheses are then:

$$\Theta_0 = \{\theta \in \mathbb{R}^q \mid \beta \in \mathbb{R}^p, \Gamma_1 \in \mathbb{S}_+^{p-r}, \Gamma_{12} = 0, \Gamma_2 = 0, \Sigma \in \mathbb{S}_+^J\}$$
$$\Theta = \{\theta \in \mathbb{R}^q \mid \beta \in \mathbb{R}^p, \Gamma \in \mathbb{S}_+^p, \Sigma \in \mathbb{S}_+^J\}.$$

We emphasize that the parameter space Θ is not open, and that the tested parameter values are on the boundary of Θ.

We recall below the likelihood ratio test procedure. Let us denote by y_1^N the joint vector of a N-sample $(y_1, \ldots, y_N)$, and by $L(y_1^N; \theta)$ the joint likelihood. We then define the likelihood ratio test statistics by:

$$LRT_N = -2 \ \log \left(\frac{\sup_{\theta \in \Theta_0} L(y_1^N; \theta)}{\sup_{\theta \in \Theta} L(y_1^N; \theta)} \right). \tag{2}$$

For a nominal level $0 < \alpha < 1$, the rejection region R_α is defined by $R_\alpha = \{LRT_N \geq q_\alpha\}$, where q_α is the $(1-\alpha)$ quantile of the distribution of LRT_N under the null hypothesis. However in practice the finite sample distribution of LRT_N is generally intractable in the case of nonlinear mixed effects models. Therefore we focus on its asymptotic distribution.

3.2. Asymptotic property of the likelihood ratio test

Let us denote by θ^* the true value of the parameters. We assume that the following condition $\boldsymbol{(C1)}$ is satisfied:

(i) The value θ^* is in Θ_0, i.e. θ^* is of the form $\theta^* = (\beta^*, \Gamma^*, \Sigma^*)$, with $\Gamma^* = \left(\begin{array}{c|c} \Gamma_1^* & \mathbf{0}_{(p-r)\times r} \\ \hline \mathbf{0}_{r\times(p-r)} & \mathbf{0}_{r\times(r)} \end{array}\right)$ where Γ_1^* is a $(p-r)\times(p-r)$ matrix.
(ii) The matrices Γ_1^* and Σ^* are positive definite. In particular, we assume that the variances that are not being tested are strictly positive.

To derive the asymptotic distribution of the likelihood ratio test statistics under the null hypothesis, we need to ensure the consistency of the maximum likelihood estimate (MLE). Therefore, we assume that the following general condition ***(C2)*** is fulfilled (Silvapulle and Sen, 2011):

(i) The function L is injective in θ (to ensure the identifiability of the model),
(ii) The first three derivatives of the log-likelihood w.r.t. θ exist and are bounded by a function whose expectation exists,
(iii) The Fisher information matrix is finite and positive definite.

Remark 1. *Note that the consistency and asymptotic normality of the MLE models in the specific context of nonlinear mixed effects have been studied in Nie (2006, 2007). He exhibited specific assumptions that ensure these theoretical results. However, these assumptions might be difficult to verify in practice.*

Before stating the expression of the asymptotic distribution of the likelihood ratio test statistics, we recall the definition of the chi-bar-square distribution (for more details, see Shapiro (1985); Silvapulle and Sen (2011)).

Definition 1. *Let $\mathcal{C}$ be a closed convex cone of $\mathbb{R}^q$, V a positive definite matrix of size $q \times q$ and $Z \sim \mathcal{N}(0, V)$. The distribution of the random* *variable defined by*

$$\bar{\chi}^2(V,\mathcal{C}) = Z^t V^{-1} Z - \min_{\theta\in\mathcal{C}} (Z-\theta)^t V^{-1}(Z-\theta)$$

is called a chi-bar-square distribution*. It is equal to a mixture of chi-square distributions with different degrees of freedom as follows:*

$$\forall t \geq 0 \;\; P(\bar{\chi}^2(V,\mathcal{C}) \leq t) = \sum_{i=0}^{q} w_i(q,V,\mathcal{C}) P(\chi_i^2 \leq t),$$

where the weights $(w_i(q,V,\mathcal{C}))_{0\leq i\leq q}$ are non-negative numbers summing up to one, and where χ_i^2 is a random variable following the chi-square distribution with i degrees of freedom, with the convention that $\chi_0^2 \equiv 0$.

We can now establish the asymptotic property of the likelihood ratio test statistics.

Theorem 1. *Assume that conditions **(C1)** and **(C2)** are fulfilled. Consider the test defined in (1). We denote by I_* the Fisher information matrix evaluated at the true value $\theta^* \in \Theta_0$. Then:*

$$LRT_N \xrightarrow[N\to\infty]{} \bar{\chi}^2(I_*^{-1}, T(\Theta,\theta^*) \cap T(\Theta_0,\theta^*)^\perp), \tag{3}$$

where $T(\Theta,\theta)$ is the tangent cone to Θ at θ, and $S^\perp$ is the orthogonal complement of S, for any subset S of $\mathbb{R}^q$.

Proof. First we apply Theorem 3 of Self and Liang (1987) to derive a general expression of the asymptotic distribution of the likelihood ratio test statistics. Indeed under conditions **(C1)** and **(C2)**, the conditions stated in Section 1 of Self and Liang (1987) are fulfilled. In particular, condition **(C2)** ensures the consistency of the maximum likelihood estimates on Θ_0 and Θ. Moreover we prove that both Θ and Θ_0 can be approximated by cones, on which the consistency of the MLE also holds. We recall that the sets Θ_0 and Θ are defined by:

$$\begin{aligned}\Theta_0 &= \{\theta \in \mathbb{R}^q \mid \beta \in \mathbb{R}^p, \Gamma_1 \in \mathbb{S}_+^{p-r}, \Gamma_{12} = 0, \Gamma_2 = 0, \Sigma \in \mathbb{S}_+^J\},\\ \Theta &= \{\theta \in \mathbb{R}^q \mid \beta \in \mathbb{R}^p, \Gamma \in \mathbb{S}_+^p, \Sigma \in \mathbb{S}_+^J\},\end{aligned}$$

where the constraints are that Γ_1, of size $(p-r) \times (p-r)$ and Γ, of size $p \times p$, are positive semi-definite. Following Sylvester's criterion (see Gilbert (1991)), a matrix is positive semi-definite if and only if all its leading principal minors are positive. Therefore both Θ_0 and Θ can be written as a set of polynomial equalities and inequalities corresponding to the different determinants. In particular, it means that both Θ_0 and Θ are semi-algebraic sets. They are therefore *Chernoff-regular*, i.e. they admit approximating cones at every point, and in particular at every point $\theta^* \in \Theta_0$ (Drton, 2009). Therefore by applying Theorem 3 of Self and Liang (1987), we get that the asymptotic distribution of the likelihood ratio test is the same as the one of the likelihood ratio test from the null hypothesis "$\theta \in \mathcal{A}(\Theta_0, \theta^*)$" versus the alternative one "$\theta \in \mathcal{A}(\Theta, \theta^*)$", where $\mathcal{A}(\Theta_0, \theta^*)$ and $\mathcal{A}(\Theta, \theta^*)$ are respectively the approximation cones of Θ_0 and Θ at θ^*:

$$LRT_N \to \inf_{\theta \in \mathcal{A}(\Theta_0, \theta^*) \setminus \theta^*} \|Z - \theta\|_{I_*}^2 - \inf_{\theta \in \mathcal{A}(\Theta, \theta^*) \setminus \theta^*} \|Z - \theta\|_{I_*}^2, \tag{4}$$

where Z is a random variable with multivariate Gaussian distribution with mean θ and covariance matrix I_*^{-1}.

We use now the link between approximating and tangent cones, due to Geyer (1994). More precisely, if we denote by $T(\Theta, \theta^*)$ the tangent cone of

Θ at θ^* and by $\mathcal{A}(\Theta, \theta^*)$ the approximating cone of Θ at θ^*, then $\mathcal{A}(\Theta, \theta^*) = \theta^* + T(\Theta, \theta^*)$. Thus, we obtain:

$$LRT_N \to \inf_{\theta \in \mathcal{T}(\Theta_0, \theta^*)} \|Z - \theta\|^2_{I_*} - \inf_{\theta \in \mathcal{T}(\Theta, \theta^*)} \|Z - \theta\|^2_{I_*}. \tag{5}$$

Finally, applying Theorem 3.7.1, page 84 of Silvapulle and Sen (2011), we derive that:

$$LRT_N \xrightarrow[N \to \infty]{} \bar{\chi}^2(I_*^{-1}, T(\Theta, \theta^*) \cap T(\Theta_0, \theta^*)^\perp), \tag{6}$$

which concludes the proof. □

3.3. Analytical expression of the asymptotic distribution

We emphasize that the cone $T(\Theta, \theta^*) \cap T(\Theta_0, \theta^*)^\perp$ always admits an analytical expression in our context. We consider in the following three common cases for the correlation structure of the p-dimensional random effects, namely independent, block-correlated and full correlated, corresponding respectively to a covariance matrix Γ diagonal, block-diagonal and full. The following proposition details the expressions of the cone $T(\Theta, \theta^*) \cap T(\Theta_0, \theta^*)^\perp$ in these three cases.

Proposition 1.
(i) Assume that $\Theta = \{\theta \in \mathbb{R}^q \mid \beta \in \mathbb{R}^p, \Gamma \in \mathbb{S}_+^p, \Gamma$ *diagonal*$, \Sigma \in \mathbb{S}_+^J\}$. *Then*

$$T(\Theta, \theta^*) \cap T(\Theta_0, \theta^*)^\perp = \{0\}^p \times \{0\}^{p-r} \times \mathbb{R}_+^r \times \{0\}^{J(J+1)/2}.$$

(ii) Assume that $\Theta = \{\theta \in \mathbb{R}^q \mid \beta \in \mathbb{R}^p, \Gamma \in \mathbb{S}_+^p, \Gamma$ *block-diagonal*$, \Sigma \in \mathbb{S}_+^J\}$ *with* $\Gamma = \mathrm{diag}(\Gamma_1, \ldots, \Gamma_K)$, *where, for* $k = 1, \ldots, K$, Γ_k *is a full covariance matrix of size* $r_k \times r_k$, *associated with the* k*-th sub-group of correlated random*

effects. Assume here that we want to test that the K-th block Γ_K of variances is null and consider the corresponding subset Θ_0. Then

$$T(\Theta, \theta^*) \cap T(\Theta_0, \theta^*)^\perp = \{0\}^p \times \left(\bigotimes_{k=1}^{K-1} \{0\}^{r_k(r_k+1)/2} \right) \times \mathbb{S}_+^{r_K} \times \{0\}^{J(J+1)/2}.$$

(iii) Assume that $\Theta = \{\theta \in \mathbb{R}^q \mid \beta \in \mathbb{R}^p, \Gamma \in \mathbb{S}_+^p, \Gamma$ full$, \Sigma \in \mathbb{S}_+^J\}$. Then

$$T(\Theta, \theta^*) \cap T(\Theta_0, \theta^*)^\perp = \{0\}^p \times \{0\}^{(p-r)(p-r+1)/2} \times \mathbb{R}^{r(p-r)} \times \mathbb{S}_+^r \times \{0\}^{J(J+1)/2}.$$

The proof of Proposition 1 relies on technical elements from convex analysis (see Hiriart-Urruty and Lemarechal (1996); Hiriart-Urruty and Malick (2012)) and can be easily adapted to general covariance matrix structure using similar tools. It is postponed to the Appendix.

Moreover, thanks to the expressions of the cone $T(\Theta, \theta^*) \cap T(\Theta_0, \theta^*)^\perp$ established in Proposition 1, we can deduce that several weights involved in the chi-bar-square distribution defined in (3) are equal to 0. The following corollary details this result for the two common cases of a diagonal or full covariance matrix Γ, described in Proposition 1.

Corollary 1.

(i) Assume that $\Theta = \{\theta \in \mathbb{R}^q \mid \beta \in \mathbb{R}^p, \Gamma \in \mathbb{S}_+^p, \Gamma$ diagonal$, \Sigma \in \mathbb{S}_+^J\}$. Then the distribution of the random variable $\bar{\chi}^2(I_*^{-1}, T(\Theta, \theta^*) \cap T(\Theta_0, \theta^*)^\perp)$ is a mixture of $(r+1)$ chi-square distributions with degrees of freedom between 0 and r.

(ii) Assume that $\Theta = \{\theta \in \mathbb{R}^q \mid \beta \in \mathbb{R}^p, \Gamma \in \mathbb{S}_+^p, \Gamma$ full$, \Sigma \in \mathbb{S}_+^J\}$. Then the distribution of the random variable $\bar{\chi}^2(I_*^{-1}, T(\Theta, \theta^*) \cap T(\Theta_0, \theta^*)^\perp)$ is a mixture of $(r(r+1)/2 + 1)$ chi-square distributions with degrees of freedom between $r(p-r)$ and $(r(p-r) + r(r+1)/2)$.

The proof is postponed to the Appendix.

3.4. Theoretical result for non identically distributed variables

The theoretical results above can be extended in a natural way to non identically distributed observations, in particular to models involving covariates as defined in Section 2, or to models involving a specific number of observations J_i per individual possibly different for each i.

However Theorem 3 of Self and Liang (1987), which leads to the asymptotic representation of the likelihood ratio test, can not be applied for non identically distributed random variables. In particular, the Fisher information matrix of N independent non identically distributed observations is no more defined as N times the Fisher information matrix of one observation as for a N-sample. Nevertheless, the asymptotic distribution of the LRT can be identified provided that suitable assumptions are fulfilled. We consider here the same conditions as those proposed in Silvapulle and Sen (2011) p156. Therefore we define the condition $\boldsymbol{(C2)(iv)}$ for all θ, there exists some positive definite matrix $\nu(\theta)$ such that

$$N^{-1/2}\nabla_\theta \log L(y_1^N;\theta) \to \mathcal{N}(0,\nu(\theta)) \text{ and } N^{-1}\nabla^2_\theta \log L(y_1^N;\theta) \to -\nu(\theta) \ a.s.,$$

and the function $\theta \to \nu(\theta)$ is continuous. Moreover we define the condition $\boldsymbol{(C3)}$ the maximum likelihood estimates on Θ_0 and Θ are consistent. General assumptions ensuring this condition in mixed effects models are provided in Nie (2006). We can then state the following result:

Theorem 2. *Assume that conditions* $\boldsymbol{(C1)}$, $\boldsymbol{(C2)(i)(ii)(iv)}$ *and* $\boldsymbol{(C3)}$ *are fulfilled. Consider the test defined in (1). Then:*

$$LRT_N \xrightarrow[N\to\infty]{} \bar{\chi}^2(\nu(\theta^*)^{-1}, T(\Theta,\theta^*) \cap T(\Theta_0,\theta^*)^{\perp}), \tag{7}$$

where $T(\Theta,\theta)$ *is the tangent cone to* Θ *at* θ, *and* $S^{\perp}$ *is the orthogonal complement of* S, *for any subset* S *of* $\mathbb{R}^q$.

Proof. We adapt the proof of our theoretical result stated in Subsection 3.2 to the case of non identically distributed variables using suited tools presented in Silvapulle and Sen (2011). We detail all the steps below. Assuming $\boldsymbol{(C2)(i)(ii)(iv)}$, we establish the following fundamental quadratic approximation using standard tools of asymptotic statistics:

$$\begin{aligned}\log L(y_1^N;\theta) &= \log L(y_1^N;\theta^*) + \frac{1}{2N}\nabla_\theta \log L(y_1^N;\theta^*)^t \nu(\theta^*)^{-1} \nabla_\theta \log L(y_1^N;\theta^*) \\ &\quad -\frac{1}{2}(Z_N - u)^t \nu(\theta^*)(Z_N - u) + \delta_N(u)\end{aligned}$$

where $Z_N = N^{-1/2}\nu(\theta^*)^{-1}\nabla_\theta \log L(y_1^N;\theta)$ and $\sup_{\|u\|\leq K} |\delta_N(u)|$ converges toward zero in probability for $K > 0$ given. Then, applying Lemma 4.2.3 in Silvapulle and Sen (2011), we obtain the $\sqrt{N}$-consistency of the maximum likelihood estimates on Θ_0 and on Θ (see Andrews (1999) for the proof of this lemma in the non identically distributed setting). Following the lines of the proof of Theorem 1, we prove that both sets Θ and Θ_0 are Chernoff-regular and can be approximated by cones. By Corollary 4.7.5 in Silvapulle and Sen (2011), we get that the consistency of the maximum likelihood estimates also holds on the corresponding approximation cones. Then applying Proposition

4.8.2 in Silvapulle and Sen (2011), we get

$$LRT_N \to \inf_{\theta \in \mathcal{A}(\Theta_0, \theta^*) \backslash \theta^*} \|Z - \theta\|^2_{\nu(\theta^*)} - \inf_{\theta \in \mathcal{A}(\Theta, \theta^*) \backslash \theta^*} \|Z - \theta\|^2_{\nu(\theta^*)},$$

where Z is a random variable with multivariate Gaussian distribution with mean θ and covariance matrix $\nu(\theta^*)^{-1}$. Finally, we conclude the proof following the same lines as in the proof of Theorem 1. □

3.5. Extension of the test and discussion

Testing simultaneously fixed effects and variance components. The same kind of theoretical results can be established when testing simultaneously fixed effects values and variance components. Indeed, the components of the null hypothesis subset Θ_0 corresponding to the fixed effects values tested are modified as follow: they are equal to the corresponding value of β^* in place of $\mathbb{R}$. Based on this new expression, we calculate the corresponding tangent cone to be able to derive finally the expression of the asymptotic distribution of the LRT. Let us detail the case of a mixed model with p independent random effects where we are interested in testing simultaneously that $0 < s \leq b$ fixed effects are equal to the corresponding values of β^* and that $0 < r \leq p$ variances are equal to zero. We obtain:

$$\Theta = \mathbb{R}^b \times \mathbb{R}_+^p \times \mathbb{S}^J \ ,$$

$$\Theta_0 = \prod_1^s \{\beta_j^*\} \times \mathbb{R}^{b-s} \times \{0\}^r \times (\mathbb{R}_+^*)^{p-r} \times \mathbb{S}^J \ ,$$

and

$$T(\Theta, \theta^*) = \mathbb{R}^b \times \mathbb{R}_+^p \times \mathbb{S}^J \ ,$$

$$T(\Theta_0, \theta^*) = \prod_1^s \{0\} \times \mathbb{R}^{b-s} \times \{0\}^r \times \mathbb{R}^{p-r} \times \mathbb{S}^J \ ,$$

leading finally to

$$T(\Theta, \theta^*) \cap T(\Theta_0, \theta^*)^{\perp} = \mathbb{R}^s \times \{0\}^{b-s} \times \mathbb{R}_+^r \times \{0\}^{p-r} \times \{0\}^{J(J+1)/2} .$$

Choosing $s = r = 1$, we recover in particular the asymptotic distribution $0.5\chi_1^2 + 0.5\chi_2^2$ given in the case 6 of Self and Liang (1987).

Effects of nuisance parameters. Assumption **(*C*1)** ii) ensures that the nuisance variance parameters are not on the boundary of the parameter space. Indeed those nuisance parameters have no action on the asymptotic distribution whatever their number and their type (mean or variance components). Their corresponding components in the subsets Θ and Θ_0 are the same, and consequently also in the tangent cones $T(\Theta, \theta^*)$ and $T(\Theta_0, \theta^*)$. This leads to null components in the cone $T(\Theta, \theta^*) \cap T(\Theta_0, \theta^*)^{\perp}$. This appears also in the asymptotic distributions obtained in cases 5 and 6 of Self and Liang (1987), which do not depend on the number of nuisance parameters. However as illustrated by the case 8 in Self and Liang (1987), the case of a test involving nuisance parameters on the boundary of the parameter space is more intricate and requires further studies.

4. Practical implementation

4.1. Computation of the likelihood ratio test statistic

The computation of the likelihood ratio test requires the computation of the maximum likelihood values under the null and alternative hypotheses, denoted respectively by $\hat{\theta}_0$ and $\hat{\theta}_1$, as well as the values of the likelihood at these two points, $L(y_1^N; \hat{\theta}_0)$ and $L(y_1^N; \hat{\theta}_1)$.

However, in the context of nonlinear mixed effects models the likelihood is not available in a closed form, and we need to resort to stochastic variants of the Expectation-Maximization (EM) algorithm, such as the Stochastic Approximation EM algorithm for example (Kuhn and Lavielle, 2005), to compute $\hat{\theta}_0$ and $\hat{\theta}_1$. For the same reason, $L(y_1^N; \hat{\theta}_0)$ and $L(y_1^N; \hat{\theta}_1)$ cannot be computed explicitly, and should be approximated using appropriate methods such as numerical or stochastic integration.

Since the decision to reject the null hypothesis relies on the value of the test statistics, the approximation of $L(y_1^N; \theta)$ must be computed precisely. Let us denote by $L(y_i; \theta)$ the marginal likelihood of the i-th individual, and by $\ell(y_1^N; \theta)$ the joint log-likelihood. Then:

$$\ell(y_1^N; \theta) = \log\left(\prod_{i=1}^{N} L(y_i; \theta)\right) = \sum_{i=1}^{N} \log\left(\int_{\mathbb{R}^p} f(y_i \mid \varphi_i; \theta) p(\varphi_i; \theta) d\varphi_i\right) ,$$

where $f(\cdot \mid \varphi_i; \theta)$ is the conditional probability density function of y_i given the random effect φ_i, and $p(\cdot; \theta)$ is the probability density function of the random effect φ_i. This quantity can be approximated using classical methods for integral approximations. However in the case of high dimensional random effects, stochastic integration is preferred over numerical approximations, allowing a better approximation for comparable computation times.

In practice, each $L(y_i; \theta)$ can be approximated independently from the others using Monte Carlo methods, and this can be done in parallel to optimize the execution time. Computing the joint log-likelihood using the sum of the marginal log-likelihoods, instead of taking the logarithm of their product can also result in less numerical issues. To further reduce the variability of the LRT statistics estimate, we can compute directly the Monte Carlo estimate of LRT_N based on the same sample of size M rather than using two estimates of $\ell(y_1^N;\hat\theta_0)$ and $\ell(y_1^N;\hat\theta_1)$ based on two different samples. Thus, let us consider:

$$L\hat{R}T_{N,M} = -2\sum_{i=1}^{N}\log\frac{\sum_{m=1}^{M} f(y_i \mid \tilde\varphi_{i,0}^m;\hat\theta_0)}{\sum_{m=1}^{M} f(y_i \mid \tilde\varphi_{i,1}^m;\hat\theta_1)},$$

where $\tilde\varphi_{i,0}^m = \hat\beta_0 + \hat\Gamma_0^{1/2}Z_i^m$, $\tilde\varphi_{i,1}^m = \hat\beta_1 + \hat\Gamma_1^{1/2}Z_i^m$ and $Z_i^m \sim \mathcal{N}(0, I_p)$.

4.2. Computation of chi-bar-square weights when Γ *is diagonal*

In general, the weights involved in the definition of the chi-bar-square distribution are not available in a tractable form. However, in the special case of a diagonal covariance matrix, the cone $\mathcal{C}$ involved in the chi-bar-square distribution is polyhedral of dimension r (see Proposition 1). Indeed, in this case the cone can be written as $\mathcal{C} = \{\theta \in \mathbb{R}^q \mid R\theta \geq 0\}$, with $R = \left(\mathbf{0}_{r\times(p+p-r)} \mid I_r \mid \mathbf{0}_{r\times\frac{J(J+1)}{2}}\right)$, a full-rank matrix of dimension $r \times q$. For polyhedral cones of this type, Shapiro (1985) provided the exact weights expressions for $1 \leq r \leq 3$. The case $r = 2$ is also treated by Self and Liang (1987). Following the notation of Shapiro (1985), we denote by $\rho_{ij} = v_{ij}/(v_{ii}v_{jj})^{1/2}$ and $\rho_{ij.k} = (\rho_{ij} - \rho_{ik}\rho_{jk})/((1-\rho_{ik}^2)(1-\rho_{jk}^2))^{1/2}$, respectively the correlation coefficient, and the partial correlation coefficient associated with the covariance matrix $RI_*^{-1}R^t$, where v_{ij} stands for the element in row i and column j of matrix $RI_*^{-1}R^t$. Using Proposition 3.6.1 of Silvapulle and Sen (2011), we have $w_i(q, I_*^{-1}, \mathcal{C}) = w_i(r, RI_*^{-1}R^t, \mathbb{R}_+^r)$ and denoting by $w_{i,r} = w_i(r, RI_*^{-1}R^t, \mathbb{R}_+^r)$, we have the following expressions: (i) for $r = 1$, we have $w_{0,1} = w_{1,1} = 1/2$, (ii) for $r = 2$, we have $w_{0,2} = 1/2\ \pi^{-1}\cos^{-1}(\rho_{12})$, $w_{1,2} = 1/2$, and $w_{2,2} = 1/2 - 1/2\ \pi^{-1}\cos^{-1}(\rho_{12})$, (iii) for $r = 3$, we have: $w_{3,3} = (4\pi)^{-1}(2\pi - \cos^{-1}(\rho_{12}) - \cos^{-1}(\rho_{13}) - \cos^{-1}(\rho_{23}))$, $w_{2,3} = (4\pi)^{-1}(3\pi -$

$\cos^{-1}(\rho_{12.3}) - \cos^{-1}(\rho_{13.2}) - \cos^{-1}(\rho_{23.1}))$, $w_{1,3} = 1/2 - w_{3,3}$, and $w_{0,3} = 1/2 - w_{2,3}$.

For $r > 3$, and in more general settings, e.g. when $\mathcal{C}$ is not a polyhedral cone, one has to either approximate the weights through numerical integration or Monte Carlo simulations, or to directly compute the tail probability of the chi-bar-square distribution (see Silvapulle and Sen (2011), page 78).

5. Experiments

5.1. Simulation study

Simulation settings. We consider a linear mixed effects model and the logistic mixed effects model described in Section 2.2 with two and three random effects to explore different settings. We denote by θ^* the true parameter value used to generate the data under H_0.

Let us denote by $\mathcal{M}_1$ the linear model with three individual parameters where we set $g(\varphi_i, x_j) = \varphi_{i1} + \varphi_{i2}x_j + \varphi_{i3}x_j^2$ and $\varphi_i = \beta^* + b_i$. We choose $\beta^* = (0, 7, 2)^t$, $\gamma_1^* = 1.3$, $\gamma_2^* = 1$ and $\gamma_{12}^* = 1.04$, corresponding to a correlation coefficient of 0.8 between b_{i1} and b_{i2}. We consider the null hypothesis H_0 defined by $\gamma_3^* = \gamma_{13}^* = \gamma_{23}^* = 0$. In the sub-model with two individual parameters, we set $g(\varphi_i, x_j) = \varphi_{i1} + \varphi_{i2}x_j$, where $\varphi_i = \beta^* + b_i$, $\beta^* = (0, 7)^t$ and $\gamma_1^* = 1.3$. In this case we consider H_0 defined by $\gamma_2^* = \gamma_{12}^* = 0$. In each simulation settings, $x_j = j$, $J = 20$ and $\sigma^* = 1.5$.

Let us denote by $\mathcal{M}_2$ the logistic model with three random effects where we set $g(\varphi_i, x_j) = \varphi_{i1}/(1 + \exp(-(x_j - \varphi_{i2})/\varphi_{i3}))$, where $\varphi_i = \beta^* + b_i$. We set $\beta^* = (200, 500, 150)^t$, $\gamma_2^* = 50$, $\gamma_3^* = 15$ and $\gamma_{23}^* = 375$, corresponding to a correlation coefficient of 0.5 between φ_{i2} and φ_{i3}. We consider here the null hypothesis H_0 defined by $\gamma_1^* = \gamma_{12}^* = \gamma_{13}^* = 0$. In the sub-model with two random effects, we set $\beta^* = (200, 500)^t$ and $\gamma_2^* = 50$. In this case we consider H_0 defined by $\gamma_1^* = \gamma_{12}^* = 0$. The vector of observation times (x_j) is the same for all the individuals, and is defined as a vector of 20 equally spaced values between 50 and 1000, plus 5 equally spaced values between 1100 and 1500. In each simulation settings, $\sigma^* = 10$. When only two random effects are considered in the model, β_3 is fixed to 150 and not estimated by the algorithm.

For the two models $\mathcal{M}_1$ and $\mathcal{M}_2$, we consider several test cases, according to the number of variances being tested, to the total number of random effects in the model, and to the presence of correlations between the random effects. More precisely, we consider the following cases: (i) testing that *one* variance is null in a model with *two* independent (Case 1) or non independent (Case 2) random effects, and in a model with *three* independent (Case 3) or non independent (Case 4) random effects, (ii) testing that *two* variances are null in a model with *three* independent (Case 5) random effects, and (iii) testing that *two* variances are null in a model with *two* independent random effects (Case 6), i.e. testing the absence of random effects. The limiting distribution of the test statistics in each of these cases is respectively $0.5\chi_0^2 + 0.5\chi_1^2$ (Case 1), $0.5\chi_1^2 + 0.5\chi_2^2$ (Case 2), $0.5\chi_0^2 + 0.5\chi_1^2$ (Case 3), $0.5\chi_2^2 + 0.5\chi_3^2$ (Case 4), and $w_{0,2}\chi_0^2 + 0.5\chi_1^2 + (0.5 - w_{0,2})\chi_2^2$ (Case 5 and Case 6) (see Section 4.2). We emphasize here that the limiting distribution strongly depends on the correlation structure of the random effects.

Parameter estimation is performed either using the `lmer` function implemented in the R package `lme4` (Bates et al., 2015), for the linear model, or using the SAEM algorithm implemented in the R package `saemix` (Comets et al., 2011), for the nonlinear model. Others parts of the codes are also developed in R and are available on request.

Empirical level. For each test case, to evaluate the level of the test we generate K datasets $D_{0,1}, \ldots, D_{0,K}$ under the null hypothesis, and we denote by $\hat{\theta}_{0,k}$ (resp. $\hat{\theta}_{1,k}$) the maximum likelihood estimates of θ^* using dataset $D_{0,k}$ under H_0 (resp. H_1). The likelihood ratio test statistics estimate is denoted by $L\hat{R}T_k$. Then, the empirical level of the test for a sample size K is equal to $\hat{\alpha}_K = \frac{1}{K}\sum_{k=1}^{K} \mathbb{1}_{L\hat{R}T_k > c_\alpha}$, where c_α is the $(1-\alpha)$ quantile of the limiting distribution of the LRT statistics. In practice, c_α is not always available in a closed form and may be estimated as mentioned in Section 4.

Note that in the linear mixed model case, the Fisher information matrix I_* is known and is given by:

$$(I_*)_{i,j} = \left(\frac{\partial X\beta}{\partial \theta_i}\right)^t\Bigg|_{\theta=\theta^*} (\Omega^*)^{-1} \left.\frac{\partial X\beta}{\partial \theta_j}\right|_{\theta=\theta^*} + \frac{1}{2}\mathrm{tr}\left((\Omega^*)^{-1} \left.\frac{\partial \Omega}{\partial \theta_i}\right|_{\theta=\theta^*} (\Omega^*)^{-1} \left.\frac{\partial \Omega}{\partial \theta_j}\right|_{\theta=\theta^*} \right),$$

where $\Omega^* = Z\Gamma^* Z^t + (\sigma^*)^2 I_J$, θ_i is the i-th element of vector θ, where $\mathrm{tr}(A)$ denotes the trace of A for any matrix A and where for a matrix A of size $m \times n$, $\partial A/\partial x$ is the matrix of size $m \times n$ whose element (i,j) is given by $(\partial A/\partial x)_{i,j} = \partial A_{ij}/\partial x$.

We first analyze the finite sample size properties of the LRT statistics when performing the test in the linear model $\mathcal{M}_1$ involving two or three random effects, with and without correlations between the random effects. We start by testing that the variance of one random effect is zero. We compute the empirical level as detailed above for nominal level α in $\{0.01, 0.05, 0.10\}$ and for a sample size N varying in $\{100, 500, 800\}$. Results are presented

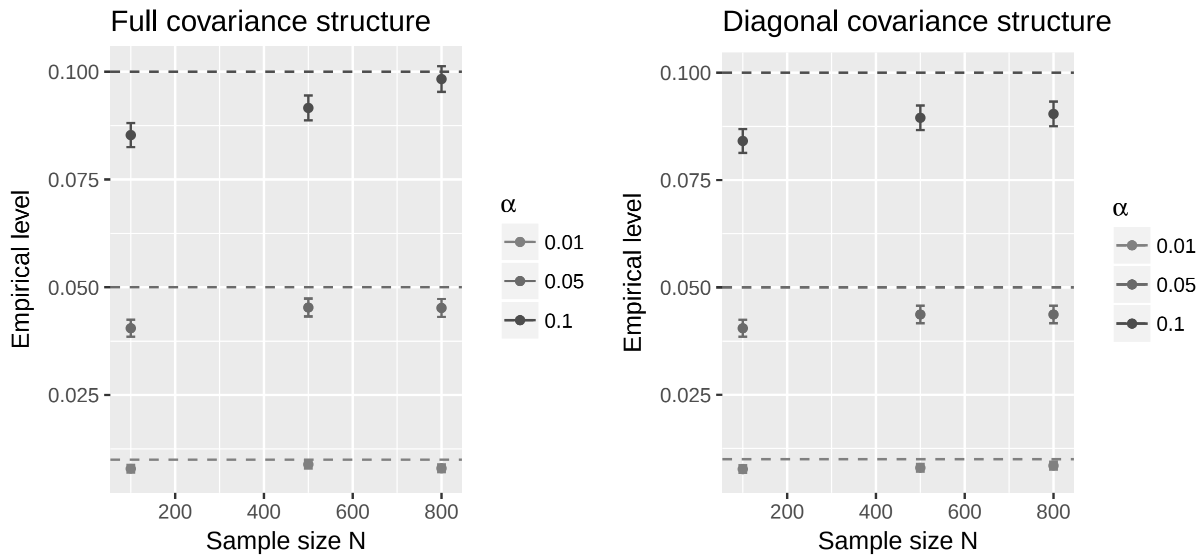


Figure 1: Empirical level and Monte-Carlo error in the **linear model** for a theoretical level $\alpha \in \{0.01, 0.05, 0.10\}$ when testing that one variance is null in a model with 2 random effects (CASES 1-2), evaluated on $K = 10000$ datasets of size $N \in \{100, 500, 800\}$

in Figures 1 and 2. We observe that the empirical levels are closer to the nominal ones when N grows, for random effects involving two and three components. We also observe that the empirical levels are lower than the nominal ones, particularly when N is smaller. This is a known property of chi-bar-square distributions (see for example Fitzmaurice et al. (2007) and Drikvandi et al. (2013)). However, the theoretical results developed in this paper are asymptotic ones. For small values of N, non-asymptotic test procedures such as permutation tests are more appropriate (see Drikvandi et al. (2013)).

Let us now highlight that one can be led to false conclusions when performing the LRT in a model without taking into account the presence of

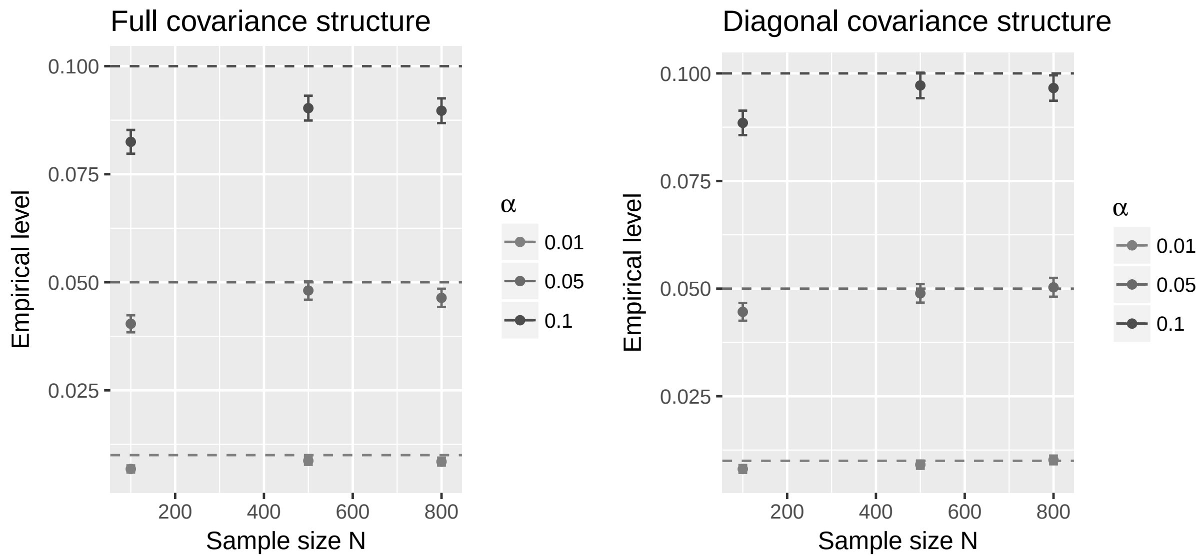


Figure 2: Empirical level and Monte-Carlo error in the **linear model** for a theoretical level $\alpha \in \{0.01, 0.05, 0.10\}$ when testing that one variance is null in a model with 3 random effects (Cases 3-4), evaluated on $K = 10000$ datasets of size $N \in \{100, 500, 800\}$

correlations between random effects. For example, let us consider Case 2, corresponding to a limiting distribution $0.5\chi_1^2 + 0.5\chi_2^2$. If we do not assume a correlation structure between the two random effects although there exists one, then the corresponding limiting distribution is $0.5\chi_0^2 + 0.5\chi_1^2$. We compute the corresponding empirical quantiles. Results are displayed in Table 1. We observe that the empirical levels in column 3 are too large, leading to possibly wrong conclusions. This emphasizes that the presence of correlations between the random effects in the model plays a crucial role when performing a test on variance components in mixed effects models.

We then evaluate the finite sample size properties of the LRT statistics when performing the test of two variances equal to zero in the linear model

Table 1: Empirical level and Monte Carlo standard errors in parenthesis in the **linear model** for a theoretical level $\alpha \in \{0.01, 0.05, 0.10\}$, when testing that one variance is null in a model with two correlated random effects, using the theoretical quantiles of the limiting distribution $0.5\chi_1^2 + 0.5\chi_2^2$ (column 2), and using the quantiles of the limiting distribution $0.5\chi_0^2 + 0.5\chi_1^2$ obtained when considering a model with uncorrelated random effects (column 3), for $N = 500$

α	$\hat{\alpha}_{0.5\chi_1^2+0.5\chi_2^2}$	$\hat{\alpha}_{0.5\chi_0^2+0.5\chi_1^2}$
0.01	0.009	0.050
	(0.0009)	(0.0019)
0.05	0.045	0.174
	(0.0021)	(0.0037)
0.10	0.092	0.311
	(0.0029)	(0.0046)

$\mathcal{M}_1$ involving three independent random effects. Results are detailed in Figure 2. As previously, we observe that the empirical levels converge to the nominal ones when N grows. However the asymptotic effect seems to occur slower when testing that two variances are equal to zero. This may be explained by the fact that more parameters have to be estimated than in the case where one variance is tested equal to zero.

We also study the properties of the LRT statistics when testing that the two variances are equal to zero in the linear model $\mathcal{M}_1$ involving two independent random effects, which amounts to testing the absence of random effects in the model. Results are presented in Table 2. We observe the convergence of the empirical levels towards the nominal ones as N increases.

Let us now focus on the nonlinear model $\mathcal{M}_2$. We analyze the finite sample size properties of the LRT statistics when performing the test in the

Table 2: Empirical level and Monte-Carlo standard errors in parenthesis in the **linear model** for a theoretical level $\alpha \in \{0.01, 0.05, 0.10\}$, when testing that two variances are null in a model with two independent random effects (Case 6), evaluated on $K = 10000$ datasets of size $N \in \{100, 200, 500\}$

α	$N = 100$	$N = 500$
0.01	0.008	0.009
	(0.0009)	(0.0009)
0.05	0.041	0.046
	(0.0020)	(0.0021)
0.10	0.082	0.088
	(0.0027)	(0.0028)

model involving two and three random effects, with and without correlations between random effects. Results are detailed in Table 3. We observe that in the case of this nonlinear mixed effects model, the empirical levels are not as close to the nominal ones as in the linear case and are always lower than the nominal levels in our example. We study the asymptotic behaviour of the empirical levels and notice that they are not improved when N is increased to 1000 (results not presented). These numerical results might be explained by the numerical integrations which have to be performed to compute the LRT statistics in nonlinear mixed effects models, which is not the case in the linear setting. To illustrate this, we use the linear model $\mathcal{M}_1$ where the likelihood can be computed exactly. We compare the exact value (given by the `lmer` function in R) with the approximation obtained by an importance sampling scheme (given by the `saemix` function in R), for $K = 1000$ datasets. It appears on this specific example that the approximation procedure of `saemix` has a tendency to under-estimate the likelihood value, especially under the alternative hypothesis. Consequently, the LRT statistics is in general lower than the exact value when computed by `saemix`, which implies lower empirical levels (see Table 4).

When we consider correlations between the random effects in model $\mathcal{M}_2$, the empirical levels are globally slightly lower than in the models without correlations. This may be explained again by the fact that the models with correlations involved more parameters to estimate. We also observe that the empirical levels in model $\mathcal{M}_2$ with 3 random effects are lower than those in model $\mathcal{M}_2$ involving only 2 random effects. The same argument can be retained in this case.

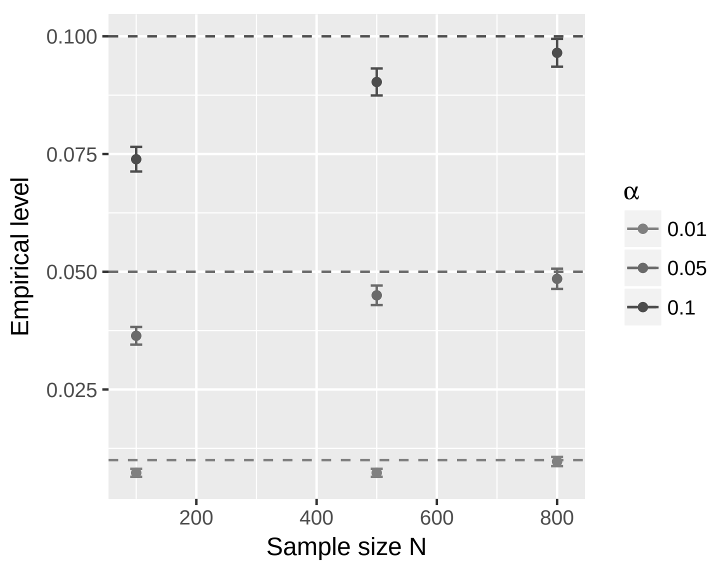


Figure 3: Empirical level and Monte-Carlo error in the **linear model** for a theoretical level $\alpha \in \{0.01, 0.05, 0.10\}$, when testing that the two variances γ_1^2 and γ_2^2 are null (Case 5), evaluated on $K = 10000$ datasets of size $N \in \{100, 500, 800\}$

Table 3: Empirical level in the **nonlinear model** and Monte Carlo standard errors in parenthesis for a theoretical level $\alpha \in \{0.01, 0.05, 0.10\}$ when testing that one variance is null (Cases 1-4), evaluated on $K = 1000$ datasets of size $N = 500$

	2 random effects		3 random effects	
α	Γ diagonal (Case 1)	Γ full (Case 2)	Γ diagonal (Case 3)	Γ full (Case 4)
0.01	0.003 (0.0017)	0.007 (0.0020)	0 (0)	0.003 (0.0033)
0.05	0.038 (0.0060)	0.033 (0.0048)	0.040 (0.0113)	0.033 (0.0104)
0.10	0.082 (0.0087)	0.073 (0.0068)	0.077 (0.0154)	0.073 (0.0151)

Empirical power. We assess the empirical power of the procedure in both models $\mathcal{M}_1$ and $\mathcal{M}_2$, when performing the test of one variance equal to zero in a model with two random effects. Results are given in Tables 5 and 6 for the linear model, and in Table 7 for the nonlinear model. We observe that the empirical power rises rapidly as the true value of the variance being tested γ_2^* increases. As $\beta_2^* = 7$, it corresponds to a relative standard deviation of approximately 1.4%. In the model with correlated random effects, the empirical powers increase with the correlation coefficient between the random effects.

Misspecification of the random effects distribution.. We assess the effect of a misspecification of the random effects distribution on the empirical level. More precisely, we generate $K = 10000$ datasets in the linear model $\mathcal{M}_1$ with two independent random effects, following non-Gaussian distributions. Then, we compute the maximum likelihood estimators under H_0 and H_1 and

Table 4: Comparison of the empirical level and Monte Carlo standard error in parenthesis computed using the exact value of the LRT statistics and an approximation of the LRT statistics using an importance sampling (IS) scheme, in the **linear model**, when testing that one variance is null, and evaluated on $K = 1000$ datasets of size $N = 500$

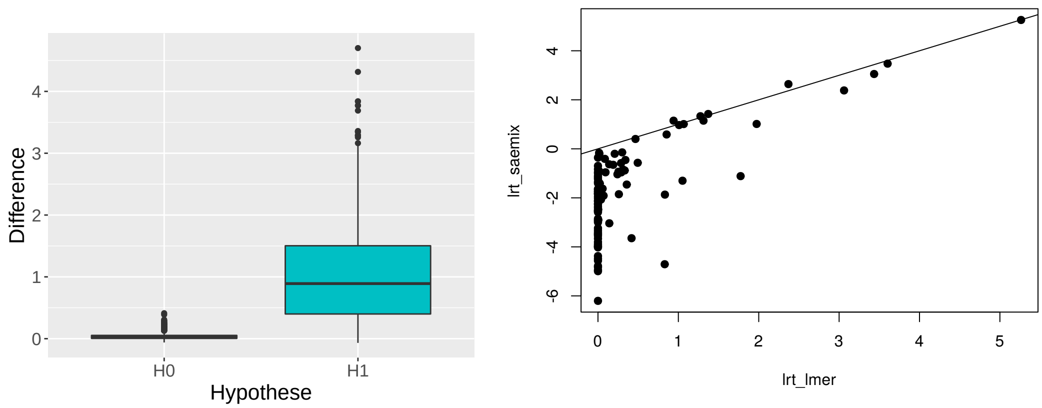
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
Exact value of the LRT	0.004 (0.0018)	0.048 (0.0064)	0.104 (0.0092)
Approximation of LRT by IS	0.002 (0.0013)	0.034 (0.0054)	0.078 (0.0081)

Table 5: Empirical power and Monte Carlo standard error in parenthesis when testing that one variance is null in the **linear model** with two **independent** random effects, evaluated on $K = 10000$ datasets of size N.

	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.10$	
γ_2	$N = 100$	$N = 500$	$N = 100$	$N = 500$	$N = 100$	$N = 500$
0.01	0.0453 (0.0021)	0.0761 (0.0026)	0.1869 (0.0039)	0.2643 (0.0044)	0.3290 (0.0047)	0.4178 (0.0049)
0.02	0.1341 (0.0034)	0.4880 (0.0050)	0.3569 (0.0048)	0.7466 (0.0043)	0.5150 (0.0050)	0.8546 (0.0035)
0.05	0.9803 (0.0004)	1 (0)	0.9972 (0.0019)	1 (0)	0.9991 (0.0033)	1 (0)

Table 6: Empirical power and Monte-Carlo standard error in parenthesis when testing that one variance is null in the **linear model** with two **correlated** random effects, evaluated on $K = 10000$ datasets of size N.

		$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.10$	
γ_2	ρ_{12}	$N = 100$	$N = 500$	$N = 100$	$N = 500$	$N = 100$	$N = 500$
0.01	0.1	0.0135	0.0371	0.0624	0.1250	0.1193	0.2051
		(0.0005)	(0.0019)	(0.0021)	(0.0033)	(0.0035)	(0.0040)
	0.5	0.0385	0.2674	0.1280	0.4957	0.2152	0.6236
		(0.0019)	(0.0044)	(0.0033)	(0.0050)	(0.0041)	(0.0048)
	0.75	0.0737	0.5617	0.2059	0.7827	0.3103	0.8640
		(0.0026)	(0.0049)	(0.0040)	(0.0041)	(0.0046)	(0.0034)
	0.9	0.1010	0.7265	0.2547	0.8857	0.3705	0.9365
		(0.0030)	(0.0044)	(0.0043)	(0.0032)	(0.0048)	(0.0024)
0.02	0.1	0.0694	0.4182	0.1869	0.6585	0.2825	0.7699
		(0.0025)	(0.0049)	(0.0039)	(0.0047)	(0.0045)	(0.0042)
	0.5	0.2556	0.9652	0.4728	0.9912	0.5993	0.9952
		(0.0044)	(0.0018)	(0.0050)	(0.0009)	(0.0049)	(0.0007)
	0.75	0.4825	0.9991	0.7008	0.9999	0.8022	1
		(0.0050)	(0.0003)	(0.0046)	(0.0001)	(0.0040)	(0)
	0.9	0.6025	1	0.8058	1	0.8756	1
		(0.0049)	(0)	(0.0039)	(0)	(0.0033)	(0)
0.05	0.1	0.9644	1	0.9925	1	0.9964	1
		(0.0018)	(0)	(0.0009)	(0)	(0.0006)	(0)



(a) Difference in likelihood evaluation

(b) LRT statistics

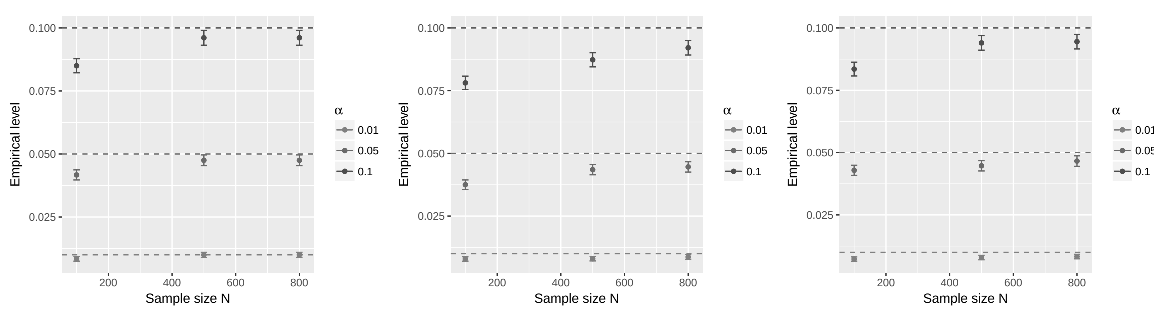
Figure 4: Comparison of the results based on the exact evaluation of the likelihood (computed by the `lmer` function) and on its approximation using importance sampling (computed by the `saemix` function), evaluated on 1000 datasets. Figure 4a : difference between the exact value of the likelihood and its approximation, under each hypotheses. Figure 4b : LRT statistics based on the approximated values of the likelihood, as a function of its exact value.

the corresponding test statistics for each dataset, assuming that the random effects are Gaussian. We consider three types of distribution: a Student, a log-normal and a mixture of two Gaussian distributions. Hence, we cover three different departures from the Gaussian distribution: higher tails with the Student distribution, asymmetry with the log-normal distribution, and bimodality with the mixture of Gaussian distributions. Data are generated under the null hypothesis, with $\gamma_2^* = 0$, and φ_{i1} generated according to one of the three types of distribution mentioned above. For the Student distribution, we choose a degree of freedom of 4.89, leading to a standard deviation of

Table 7: Empirical power Monte-Carlo standard error in parenthesis when testing that one variance is null in the **nonlinear model** with two **independent** random effects, evaluated on $K = 1000$ datasets of size N.

		$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.10$	
γ_2	ρ	$N = 100$	$N = 500$	$N = 100$	$N = 500$	$N = 100$	$N = 500$
2	0	0.30 (0.0145)	0.98 (0.0044)	0.46 (0.0158)	0.995 (0.0022)	0.55 (0.0157)	0.999 (0.001)
2	0.2	0.182 (0.0122)	0.945 (0.0072)	0.310 (0.0146)	0.985 (0.0038)	0.400 (0.0155)	0.991 (0.0030)
2	0.5	0.448 (0.0157)	1 (0)	0.639 (0.0152)	1 (0)	0.717 (0.0142)	1 (0)
2	0.8	0.880 (0.0103)	1 (0)	0.953 (0.0067)	1 (0)	0.970 (0.0054)	1 (0)

1.3. For the log-normal distribution, we choose the parameters such that the variable $\log \varphi_{i1}$ is centered and its standard deviation takes two different values, namely 0.25 and 1.5, corresponding respectively to a distribution which is slightly asymmetric with a mode around 0.95, and to a distribution which is highly asymmetric, with a mode around 0.10. For the mixture of Gaussian distributions, we consider two cases: (i) a symmetric one where the expectations of the Gaussian distributions are set respectively to -1 and 1, both standard deviations to 1, and both weights to 0.5, and (ii) an asymmetric one where the expectations are set respectively to -3 and 1, both standard deviations to 1, and the weights to 0.25 and 0.75. Results for random effects following the Student distribution $t_{4.89}$, the log-normal distribution such that $\log(\varphi_{i1}) \sim \mathcal{N}(0, 1.5^2)$ and the Gaussian mixture $0.5\mathcal{N}(-1, 1) + 0.5\mathcal{N}(1, 1)$ are presented in Figure 5. Results for the two other distributions are very similar and therefore not presented here. We observe on these specific examples that the empirical levels are globally closed to the nominal ones. Moreover, except in the case of the highly asymmetric log-normal distribution, the empirical levels are closer to the nominal ones than the empirical levels obtained in the case where the random effects are Gaussian. This numerical study gives some insight on the effects of misspecification of the random effects distribution on the proposed test procedure. Yet, it can not lead to a general conclusion about the robustness of the proposed test procedure to misspecification of the random effects distribution. However let us mention that there exist diagnostic tools for the normality assumption of the random effects distribution in mixed effects models. Let us quote for example those proposed recently by Drikvandi et al. (2017) and Drikvandi (2017).



(a) *Student* (b) *log-normal* (c) *Gaussian mixture*

Figure 5: Empirical level and Monte-Carlo error in the **linear model** for a theoretical level $\alpha \in \{0.01, 0.05, 0.10\}$, evaluated on $K = 10000$ datasets of size $N \in \{100, 500, 800\}$, when testing that γ_2^2 is null, under different misspecification of the random effects distribution: (a) $\varphi_{i1} \sim t_{4.89}$, (b) $\log(\varphi_{i1}) \sim \mathcal{N}(0, 1.5^2)$, (c) $\varphi_{i1} \sim 0.5\mathcal{N}(-1, 1) + 0.5\mathcal{N}(1, 1)$

5.2. Real data analysis

The method is illustrated on two sets of real data. The first one is the famous dental growth dataset from Potthoff and Roy (1964), in which the distance from the center of the pituitary gland to the pteryomaxillary fissure was measured at 4 different ages for 27 children (16 boys and 11 girls). This dataset is available in the `R` package `mice`. The second dataset comes from a study of coucal growth rates, available as a Dryad package (Goymann et al., 2016). Body masses of 678 nestlings from two species (white-browed coucals and black coucals) were recorded every two days from their hatching date until they left the nests. In this paper, we only consider data from the white-browed coucals species, corresponding to the highest sample size (385 individuals).

A linear model is fitted to the dental growth data using the `lme4` package, with two random effects as described in Section 5.1 (model $\mathcal{M}_1$). More precisely, if we denote by y_{ij}, $1 \leq i \leq 27$, $1 \leq j \leq 4$, the dental measurement of child i of sex x_i, at age t_j, the following model is considered, with a random slope and a random intercept:

$$
\begin{aligned}
& y_{ij} = (\alpha_0 + \alpha_1 x_i + b_{i1}) + (\beta_0 + \beta_1 x_i + b_{i2}) t_j + \varepsilon_{ij} \ , \\
& \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2), \quad (b_{i1}, b_{i2})^t \sim \mathcal{N}(0, \Gamma) \ .
\end{aligned}
$$

The model can be written in the following form :

$$y_i = X_i\beta + Z_ib_i + \varepsilon_i$$

$$X_i = \begin{pmatrix} 1 & x_i & t_1 & x_it_1 \\ 1 & x_i & t_2 & x_it_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_i & t_J & x_it_J \end{pmatrix}, \ \beta = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \beta_0 \\ \beta_1 \end{pmatrix}, \ Z_i = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_J \end{pmatrix}.$$

First, we assess if the random effects are correlated or not using a LRT test, and we obtain a test statistics of 0.3027. In this case, the limiting distribution is a chi-square with one degree of freedom. Thus, at the level of 5%, the null hypothesis is not rejected. Therefore in the sequel we consider a diagonal covariance matrix $\Gamma = \text{diag}(\gamma_1^2, \gamma_2^2)$. We first test $H_0 : \{\gamma_1^2 = 0, \gamma_2^2 = 0\}$ against $H_1 : \{\gamma_1^2 \geq 0, \gamma_2^2 \geq 0\}$, i.e. we test for the absence of randomness in the model. We compute the likelihood ratio test statistics and compare it to the rejection threshold q_α^d associated with the limiting distribution d and corresponding to a level α. The test statistics is equal to 50.13, the limiting distribution is the mixture $w_0\chi_0^2 + 0.5\chi_1^2 + (1 - w_0)\chi_2^2$, with $w_0 = 0.37$, and the rejection threshold is $q_{0.05}^{0.37\chi_0^2+0.5\chi_1^2+0.63\chi_2^2} = 3.61$. Therefore, at the level of 5% we reject the null hypothesis that there is no random effects in the model. Next, we want to test $H_0 : \{\gamma_1^2 \geq 0, \gamma_2^2 = 0\}$ against $H_1 : \{\gamma_1^2 \geq 0, \gamma_2^2 \geq 0\}$, i.e. we want to test if there is a random slope. In this case, the test statistics is equal to 0.5304, the rejection threshold is equal to $q_{0.05}^{0.5\chi_0^2+0.5\chi_1^2} = 2.706$ and the p-value is equal to 0.233. Therefore the null hypothesis is not rejected, which suggests that there is no randomness in the slope. These results are consistent with previous results obtained on this dataset, in particular in Drikvandi et al. (2012).

A nonlinear model is fitted to the white-browed coucals data using the `saemix` package, with three random effects as described in Section 5.1 (model $\mathcal{M}_2$). More precisely, if we denote by y_{ij}, $1 \leq i \leq 385$, $1 \leq j \leq n_i$ the body mass of nestling i at age x_j, we consider the following model:

$$y_{ij} = \frac{\varphi_{i1}}{1 + \exp\left(\frac{x_j - \varphi_{i2}}{\varphi_{i3}}\right)} \ , \quad \varphi_i = \beta + b_i \ , \quad b_i \sim \mathcal{N}(0, \Gamma) \ , \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \ ,$$

where φ_{i1} is the asymptotic body mass of individual i, φ_{i2} the age in days at which individual i reaches half its asymptotic body mass and φ_{i3} the growth rate of individual i.

We test whether the variances of the inflexion point and the growth rate are equal to 0, and thus we consider the CASE 4 described previously. We consider a diagonal covariance matrix $\Gamma = \text{diag}(\gamma_1^2, \gamma_2^2, \gamma_3^2)$ and we test $H_0 : \{\gamma_1^2 > 0, \gamma_2^2 = \gamma_3^2 = 0\}$ against $H_1 : \{\gamma_1^2 > 0, \gamma_2^2 \geq 0, \gamma_3^2 \geq 0\}$. The limiting distribution of the LRT statistics is the mixture $w_{0,2}\chi_0^2 + 0.5\chi_1^2 + (0.5 - w_{0,2})\chi_2^2$, where $w_{0,2}$ is defined in Section 4 and can be computed from the correlation coefficient between parameters γ_2^2 and γ_3^2 obtained from the Fisher information matrix.

The estimated Fisher information matrix $\hat{I}$ is obtained as an output of the `saemix` package when the full model corresponding to H_1 is fitted to the data. From this matrix we can easily compute the covariance matrix $\hat{V} = R\hat{I}^{-1}R^t$, where $R = (\mathbf{0}_{2\times 4} \mid I_2 \mid \mathbf{0}_{2\times 1})$, and the corresponding correlation matrix $\hat{C} = \text{diag}(\hat{V})^{-1/2}\,\hat{V}\,\text{diag}(\hat{V})^{-1/2}$. Then, the correlation coefficient $\hat{\rho}_{12}$ needed to compute $w_{0,2}$, is the element $(1, 2)$ of $\hat{C}$.

In our case, $\hat{\rho}_{12} = -0.644$, leading to the three following weights: $w_{0,2} = 0.139$, $w_{1,2} = 0.5$ and $w_{2,2} = 0.361$, and thus to the limiting distribution

$0.139\chi_0^2 + 0.5\chi_1^2 + 0.361\chi_2^2$. The likelihood ratio test statistics is equal to $LRT = 3.119$, and the rejection threshold is equal to $q_{0.05}^{0.139\chi_0^2+0.5\chi_1^2+0.361\chi_2^2} = 4.682$. The corresponding p-value is evaluated at 0.114 on this dataset. In other words, assuming a diagonal covariance matrix Γ, we do not reject the null hypothesis that both the inflection point and the growth rates are fixed effects and do not vary among individuals, at the asymptotic level of 5%.

6. Discussion

Several perspectives of this work are of great interest, both from a theoretical and a practical point of view. For a wide-range application of these results, adapted tools to compute precisely the test statistics in nonlinear mixed effects models have to be developed. Indeed the reliability of the test is linked to the precise evaluation of the test statistics. Moreover, the procedure opens very promising perspectives in applications with high dimensional random effects. For example, models of plant growth have raised expectations to help improving the understanding of gene by environment interactions by developing a predictive capacity that scales from genotype to phenotype (Letort et al., 2008). In plant ecophysiological models, one genotype should be represented by one unique set of parameters, and reversely, two different genotypes should potentially be characterized by two different sets of parameters (Tardieu, 2003). Such models are often descriptive ones, involving mechanistic parameters. Therefore, considering these parameters as random effects is relevant in order to understand how they vary within a given population (Baey et al., 2016). However this leads to many computational issues, since the number of parameters is high. Therefore, the issue of reducing the model size in a mixed effects model has gained an increasing interest in plant growth modelling. Also more advanced works on the computational methods are still necessary for the computation of the Fisher information matrix. Besides, from a theoretical point of view, it would be interesting to establish results when considering nuisance parameters lying on the boundary of the parameter space, as well as results for the test power. Finally, since the asymptotic regime is not always reached in practice, it would be of a great interest to develop a finite sample-size procedure using for example bootstrap methods or permutation tests in the spirit of the ones developed in the context of linear mixed effects models (Fitzmaurice et al., 2007; Drikvandi et al., 2013).

7. Appendix

7.1. Proof of Proposition 1

The proof relies on technical elements from convex analysis and can be easily adapted to general covariance matrix structure using similar tools.

To calculate the tangent cones to Θ_0 and Θ at θ^*, we can use general results from Hiriart-Urruty and Lemarechal (1996) on the definition of tangent cones, and more recent results by Hiriart-Urruty and Malick (2012) on the tangent cone of the set of symmetric positive semi-definite matrices.

The proof is carried out in the cases where the covariance matrix Γ is diagonal and full. Similar tools can be used in other cases where a more sparse structure is assumed for Γ.

Tangent cone of Θ. We recall that Θ is defined as:

$$\Theta = \{\theta \in \mathbb{R}^q \mid \beta \in \mathbb{R}^p, \Gamma \in \mathbb{S}_+^p, \Sigma \in \mathbb{S}_+^J\} \ ,$$
$$\Theta = \mathbb{R}^p \times \mathbb{S}_+^p \times \mathbb{S}_+^J \ .$$

Now, since each term in the above product is a convex cone, we have that $T(\mathbb{R}^p \times \mathbb{S}_+^p \times \mathbb{S}_+^J, (\beta^*, \Gamma^*, \Sigma^*)) = T(\mathbb{R}^p, \beta^*) \times T(\mathbb{S}_+^p, \Gamma^*) \times T(\mathbb{S}_+^J, \Sigma^*)$ (see for example (Hiriart-Urruty and Lemarechal, 1996, Proposition 5.3.1.)). Therefore, the tangent cone of Θ at θ^* is given by:

$$T(\Theta, \theta^*) = \mathbb{R}^p \times T(\mathbb{S}_+^p, \Gamma^*) \times T(\mathbb{S}_+^J, \Sigma^*) \ ,$$

where $T(\mathbb{S}_+^p, \Gamma^*)$ is the tangent cone of the set of symmetric positive semi-definite matrices of size $p \times p$ at Γ^*, and $T(\mathbb{S}_+^J, \Sigma^*)$ the tangent cone of the set of symmetric positive semi-definite matrices of size $J \times J$ at Σ^*.

To identify $T(\mathbb{S}_+^p, \Gamma^*)$ and $T(\mathbb{S}_+^J, \Sigma^*)$, we can use the result established by Hiriart-Urruty and Malick (2012). According to the authors, the tangent cone of $\mathbb{S}_+^p$ at $A \in \mathbb{S}_+^p$ is given by $T_A = \{M \in \mathbb{S}^p \mid \langle Mu, u \rangle \geq 0 \text{ for all } u \in \ker A\}$, where $\mathbb{S}^p$ is the set of symmetric matrices of size $p \times p$.

In our case, since $\Gamma^* = \left[\begin{smallmatrix} \Gamma_1^* & 0 \\ 0 & 0 \end{smallmatrix}\right]$, we have:

$$\begin{aligned}
T(\mathbb{S}_+^p, \Gamma^*) &= \{M \in \mathbb{S}^p \mid \forall u \in \ker \Gamma^*, \langle Mu, u \rangle \geq 0\} \\
&= \left\{ \left(\begin{array}{c|c} M_{11} & M_{12} \\ \hline M_{12}^t & M_{22} \end{array} \right) \in \mathbb{S}^p \mid \forall u = (\underbrace{0, \ldots, 0}_{p-r}, u_{p-r+1}, \ldots, u_p), u^t M u \geq 0 \right\} \\
&= \left\{ \left(\begin{array}{c|c} M_{11} & M_{12} \\ \hline M_{12}^t & M_{22} \end{array} \right) \in \mathbb{S}^p \mid M_{22} \geq 0 \right\}
\end{aligned}$$
$$T(\mathbb{S}_+^p, \Gamma^*) = \mathbb{R}^{(p-r)(p-r+1)/2} \times \mathbb{R}^{r(p-r)} \times \mathbb{S}_+^r.$$

Similarly, we have:

$$\begin{aligned} T(\mathbb{S}_+^J, \Sigma^*) &= \left\{ M \in \mathbb{S}^J \mid \forall u \in \ker \Sigma^*, \langle Mu, u \rangle \geq 0 \right\} \\ &= \left\{ M \in \mathbb{S}^J \mid \forall u \in \{0\}, \langle Mu, u \rangle \geq 0 \right\} \\ &= \mathbb{S}^J \ . \end{aligned}$$

In the end, we get:

$$T(\Theta, \theta^*) = \mathbb{R}^p \times \mathbb{R}^{(p-r)(p-r+1)/2} \times \mathbb{R}^{r(p-r)} \times \mathbb{S}_+^r \times \mathbb{S}^J \ .$$

In the case where Γ is diagonal, i.e. when the effects are assumed to be independent, we have:

$$T(\Theta, \theta^*) = \mathbb{R}^p \times \mathbb{R}^{p-r} \times \mathbb{R}_+^r \times \mathbb{S}^J \ .$$

Tangent cone of Θ_0. We recall that Θ_0 is defined as:

$$\begin{aligned} \Theta_0 =& \{\theta \in \mathbb{R}^q \mid \beta \in \mathbb{R}^p, \Gamma_1 \in \mathbb{S}_+^{p-r}, \Gamma_{12} = 0, \Gamma_2 = 0, \Sigma \in \mathbb{S}_+^J\} \\ =& \mathbb{R}^p \times \mathbb{S}_+^{p-r} \times \{0\}^{r(p-r)} \times \{0\}^{r(r+1)/2} \times \mathbb{S}_+^J \ . \end{aligned}$$

Using similar tools than for Θ, we get:

$$\begin{aligned} T(\Theta_0, \theta^*) &= \mathbb{R}^p \times \mathbb{S}^{p-r} \times \{0\}^{r(p-r)} \times \{0\}^{r(r+1)/2} \times \mathbb{S}^J \\ &= \mathbb{R}^p \times \mathbb{R}^{(p-r)(p-r+1)/2} \times \{0\}^{r(p-r)+r(r+1)/2} \times \mathbb{R}^{J(J+1)/2}. \end{aligned}$$

Note that in the particular case where Γ is diagonal, i.e. when the effects are supposed to be independent, the parameter space Θ_0 and hence its tangent cone can be simplified, and we have:

$$T(\Theta_0, \theta^*) = \mathbb{R}^p \times \mathbb{R}^{p-r} \times \{0\}^r \times \mathbb{R}^{J(J+1)/2}.$$

Finally, we derive the expressions stated in (i) and (iii).

7.2. Proof of Corollary 1

Let V be a positive-definite matrix and $\mathcal{C}$ a closed convex cone of $\mathbb{R}^q$. We denote by $\mathcal{C}^o = \{x \in \mathbb{R}^q \mid x^t y \leq 0, \ \forall y \in \mathcal{C}\}$ its polar cone. We recall the following properties for the weights of the chi-bar-square distribution $\bar{\chi}^2(V, \mathcal{C})$ (Shapiro, 1985, 1988):

1. for $0 \leq i \leq q$, $w_i(q, V, \mathcal{C}) = w_{q-i}(q, V, \mathcal{C}^o)$;,
2. if $\mathcal{C}$ is included in a linear space of dimension $(q-k)$, for $1 \leq k \leq q$, then the first k weights $\{w_i(q, V, \mathcal{C}^o), i = 0, \ldots, k-1\}$ are zero,
3. if $\mathcal{C}$ contains a linear space of dimension l, for $1 \leq l \leq q$, then the last l weights $\{w_i(q, V, \mathcal{C}^o), i = q-l+1, \ldots, q\}$ are zero.

In our case, $\mathcal{C} = T(\Theta, \theta^*) \cap T(\Theta_0, \theta^*)^\perp$ and $V = I_*^{-1}$, and we have for both cases mentioned in the corollary: (i) $\mathcal{C} = \{0\}^p \times \{0\}^{(p-r)(p-r+1)/2} \times \mathbb{R}^{r(p-r)} \times \mathbb{S}_+^r \times \{0\}^{J(J+1)/2}$ which is included in $\mathbb{R}^{r(p-r)+r(r+1)/2}$, i.e. in a linear space of dimension $q - (p + (p-r)(p-r+1)/2 + J(J+1)/2)$. Therefore, using properties (i) and (ii) above, the weights $w_i(q, I_*^{-1}, \mathcal{C})$, for $i = r(p-r) + r(r+1)/2 + 1, \ldots, q$ are zero. Moreover, $\mathcal{C}$ contains $\mathbb{R}^{r(p-r)}$, i.e. a linear space of dimension $r(p-r)$, which means using properties (i) and (iii) above, that the weights $w_i(q, I_*^{-1}, \mathcal{C})$, for $i = 0, \ldots, r(p-r) - 1$ are zero. (ii) $\mathcal{C} = \{0\}^p \times \{0\}^{p-r} \times \mathbb{R}_+^r \times \{0\}^{J(J+1)/2}$ which is included in $\mathbb{R}^r$, a linear space of dimension $q - p - (p-r) - J(J+1)/2$. It follows that the weights $w_i(q, I_*^{-1}, \mathcal{C})$, for $i = r+1, \ldots, q$ are zero. Then, since $\mathcal{C}$ does not contain any linear space of dimension $k > 0$, all the other weigths are non-zero.

8. References

Andrews, D. W., 1999. Estimation when a parameter is on a boundary. Econometrica, 1341–1383.

Baey, C., Trevezas, S., Cournède, P.-H., 2016. A nonlinear mixed effects model of plant growth and estimation via stochastic variants of the EM algorithm. Communications in Statistics – Theory and Methods 45 (6), 1643–1669.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67 (1), 1–48.

Chant, D., 1974. On Asymptotic Tests of Composite Hypotheses in Nonstandard Conditions. Biometrika 61 (2), 291–298.

Chernoff, H., 1954. On the Distribution of the Likelihood Ratio. The Annals of Mathematical Statistics 25 (3), 573–578.

Comets, E., Lavenu, A., Lavielle, M., 2011. SAEMIX, an R version of the SAEM algorithm. 20^{th} meeting of the Population Approach Group in Europe, Athens, Greece.

Crainiceanu, C. M., Ruppert, D., feb 2004. Likelihood ratio tests in linear mixed models with one variance component. Journal of the Royal Statistical Society. Series B: Statistical Methodology 66 (1), 165–185.

Davidian, M., Giltinan, D. M., 1995. Nonlinear models for repeated measurement data. Chapman & Hall.

Davidian, M., Giltinan, D. M., 2003. Nonlinear models for repeated measurement data: An overview and update. Journal of Agricultural, Biological, and Environmental Statistics 8 (4), 387–419.

Drikvandi, R., 2017. Nonlinear mixed-effects models for pharmacokinetic data analysis: assessment of the random-effects distribution. Journal of pharmacokinetics and pharmacodynamics 44 (3), 223–232.

Drikvandi, R., Khodadadi, A., Verbeke, G., 2012. Testing variance components in balanced linear growth curve models. Journal of Applied Statistics 39 (3), 563–572.

Drikvandi, R., Verbeke, G., Khodadadi, A., Partovi Nia, V., Nia, V. P., 2013. Testing multiple variance components in linear mixed-effects models. Biostatistics 14 (1), 144–159.

Drikvandi, R., Verbeke, G., Molenberghs, G., 2017. Diagnosing misspecification of the random-effects distribution in mixed models. Biometrics 73 (1), 63–71.

Drton, M., 2009. Likelihood ratio tests and singularities. Annals of Statistics 37 (2), 979–1012.

Fitzmaurice, G. M., Lipsitz, S. R., Ibrahim, J. G., 2007. A note on permutation tests for variance components in multilevel generalized linear mixed models. Biometrics 63 (3), 942–946.

Geyer, C. J., 1994. On the asymptotics of constrained m-estimation. The Annals of Statistics 22 (4), 1993–2010.

Gilbert, G. T., 1991. Positive definite matrices and Sylvester's criterion. American Mathematical Monthly 98 (1), 44–46.

Goymann, W., Safari, I., Muck, C., Schwabl, I., 2016. Data from: Sex roles, parental care and offspring growth in two contrasting coucal species. URL http://dx.doi.org/10.5061/dryad.23gt0

Greven, S., Crainiceanu, C. M., Küchenhoff, H., Peters, A., dec 2008. Restricted Likelihood Ratio Testing for Zero Variance Components in Linear Mixed Models. Journal of Computational and Graphical Statistics 17 (4), 870–891.

Hiriart-Urruty, J., Lemarechal, C., 1996. Convex Analysis and Minimization Algorithms I: Fundamentals. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg.

Hiriart-Urruty, J. B., Malick, J., 2012. A Fresh Variational-Analysis Look at the Positive Semidefinite Matrices World. Journal of Optimization Theory and Applications 153 (3), 551–577.

Kuhn, E., Lavielle, M., 2005. Maximum likelihood estimation in nonlinear mixed effects models. Computational Statistics & Data Analysis 49 (4), 1020–1038.

Lavielle, M., 2014. Mixed effects models for the population approach: models, tasks, methods and tools. CRC Press.

Letort, V., Mahe, P., Cournède, P.-H., De Reffye, P., Courtois, B., 2008. Quantitative genetics and functional–structural plant growth models: simulation of quantitative trait loci detection for model parameters and application to potential yield optimization. Annals of Botany 101 (8), 1243–1254.

Molenberghs, G., Verbeke, G., 2007. Likelihood Ratio, Score, and Wald Tests in a Constrained Parameter Space. The American Statistician 61 (1), 22–27.

Nie, L., 2006. Strong consistency of the maximum likelihood estimator in generalized linear and nonlinear mixed-effects models. Metrika 63 (2), 123–143.

Nie, L., 2007. Convergence rate of MLE in generalized linear and nonlinear mixed-effects models: theory and applications. Journal of Statistical Planning and Inference 137 (6), 1787–1804.

Pinheiro, J., Bates, D., 2000. Mixed-Effects Models in S and S-PLUS. Springer.

Potthoff, R. F., Roy, S., 1964. A generalized multivariate analysis of variance model useful especially for growth curve problems. Biometrika 51 (3-4), 313–326.

Qu, L., Guennel, T., Marshall, S. l., 2013. Linear score tests for variance components in linear mixed models and applications to genetic association studies. Biometrics 69 (4), 883–892.

Samuh, M. H., Grilli, L., Rampichini, C., Salmaso, L., Lunardon, N., 2012. The Use of Permutation Tests for Variance Components in Linear Mixed

Models. Communications in Statistics - Theory and Methods 41 (16-17), 3020–3029.

Self, S. G., Liang, K.-Y., 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. Journal of the American Statistical Association 82 (398), 605–610.

Shapiro, A., 1985. Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. Biometrika 72, 133–144.

Shapiro, A., 1988. Towards a Unified Theory of Inequality Constrained Testing in Multivariate Analysis. International Statistical Review / Revue Internationale de Statistique 56 (1), pp. 49–62.

Silvapulle, M. J., Sen, P. K., 2011. Constrained statistical inference: Order, inequality, and shape constraints. Vol. 912. John Wiley & Sons.

Silvapulle, M. J., Silvapulle, P., 1995. A score test against one-sided alternatives. Journal of the American Statistical Association 90 (429), 342–349.

Sinha, S. K., 2009. Bootstrap tests for variance components in generalized linear mixed models. Canadian Journal of Statistics 37 (2), 219–234.

Stram, D. O., Lee, J. W., 1994. Variance components testing in the longitudinal mixed effects model. Biometrics 50 (4), 1171–1177.

Stram, D. O., Lee, J. W., 1995. Corrections to Variance components testing in the longitudinal mixed effects model. Biometrics 51 (3), 1196.

Tardieu, F., 2003. Virtual plants: modelling as a tool for the genomics of tolerance to water deficit. Trends in plant science 8 (1), 9–14.

Wood, S. N., 2013. A simple test for random effects in regression models. Biometrika 100 (4), 1005–1010.