Ann Inst Stat Math (2015) 67:885–895
DOI 10.1007/s10463-014-0478-5

On local power properties of the LR, Wald, score and gradient tests in nonlinear mixed-effects models

Artur J. Lemonte

Received: 22 August 2013 / Revised: 20 June 2014 / Published online: 10 August 2014

Abstract The local powers of some tests under the presence of a parameter vector, $\boldsymbol{\omega}$ say, that is orthogonal to the remaining parameters are studied in this paper. We show that some of the coefficients that define the local powers of the tests remain unchanged regardless of whether $\boldsymbol{\omega}$ is known or needs to be estimated, whereas the others can be written as the sum of two terms, the first of which being the corresponding term obtained as if $\boldsymbol{\omega}$ were known, and the second, an additional term yielded by the fact that $\boldsymbol{\omega}$ is unknown. We apply our general result in the class of nonlinear mixed-effects models and compare the local powers of the tests in this class of models.

Keywords Asymptotic expansions · Gradient test · Likelihood ratio test · Nonlinear mixed-effects models · Score test · Wald test

1 Introduction

The most common tests in parametric models are the likelihood ratio (LR), Wald, and Rao score tests (Wilks 1938; Wald 1943; Rao 1948). These tests are widely used in areas such as economics, biology, and engineering, among others, since exact tests are not always available. A new criterion for testing hypotheses, referred to as the *gradient statistic*, was proposed in Terrell (2002), which shares the same first order asymptotic properties with the LR, Wald and score statistics. An advantage of the gradient statistic over the Wald and score statistics is that it does not involve knowledge of the information matrix, neither expected nor observed.

A. J. Lemonte (✉)
Departamento de Estatística, Universidade Federal de Pernambuco,
Cidade Universitária, Recife, PE 50740-540, Brazil
e-mail: arturlemonte@gmail.com

Let $\ell(\boldsymbol{\theta})$, $\boldsymbol{U}_{\boldsymbol{\theta}} = \partial\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ and $\boldsymbol{K}_{\boldsymbol{\theta}} = \mathbb{E}(\boldsymbol{U}_{\boldsymbol{\theta}}\boldsymbol{U}_{\boldsymbol{\theta}}^{\top})$ be the log-likelihood function for the parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)^{\top}$ of dimension k, the score vector and the Fisher information matrix, respectively. Consider the partition $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\top}, \boldsymbol{\omega}^{\top})^{\top}$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\top}$ and $\boldsymbol{\omega} = (\omega_{p+1}, \ldots, \omega_k)^{\top}$ are vectors of dimensions p and $k - p$, respectively. Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^{\top}, \boldsymbol{\beta}_2^{\top})^{\top}$, where the dimensions of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are q and $p - q$, respectively. Suppose the interest lies in testing the composite null hypothesis $\mathcal{H}_0 : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20}$ against the two-sided alternative hypothesis $\mathcal{H}_a : \boldsymbol{\beta}_2 \neq \boldsymbol{\beta}_{20}$, where $\boldsymbol{\beta}_{20}$ is a specified vector of dimension $p - q$, and $\boldsymbol{\beta}_1$ and $\boldsymbol{\omega}$ act as vectors of nuisance parameters. From the partition of $\boldsymbol{\theta}$, we have the corresponding partitions: $\boldsymbol{U}_{\boldsymbol{\theta}} = (\boldsymbol{U}_{\boldsymbol{\beta}_1}^{\top}, \boldsymbol{U}_{\boldsymbol{\beta}_2}^{\top}, \boldsymbol{U}_{\boldsymbol{\omega}}^{\top})^{\top}$,

$$\boldsymbol{K}_{\boldsymbol{\theta}} = \begin{bmatrix} \boldsymbol{K}_{\boldsymbol{\beta}} & \boldsymbol{K}_{\boldsymbol{\beta}\boldsymbol{\omega}} \\ \boldsymbol{K}_{\boldsymbol{\omega}\boldsymbol{\beta}} & \boldsymbol{K}_{\boldsymbol{\omega}} \end{bmatrix}, \quad \boldsymbol{K}_{\boldsymbol{\beta}} = \begin{bmatrix} \boldsymbol{K}_{11} & \boldsymbol{K}_{12} \\ \boldsymbol{K}_{21} & \boldsymbol{K}_{22} \end{bmatrix}, \quad \boldsymbol{K}_{\boldsymbol{\beta}}^{-1} = \begin{bmatrix} \boldsymbol{K}^{11} & \boldsymbol{K}^{12} \\ \boldsymbol{K}^{21} & \boldsymbol{K}^{22} \end{bmatrix}.$$

In this paper, we shall assume that $\boldsymbol{\beta}$ is globally orthogonal to $\boldsymbol{\omega}$ in the sense of Cox and Reid (1987). In other words, the Fisher information matrix for $\boldsymbol{\theta}$ and its inverse are block-diagonal: $\boldsymbol{K}_{\boldsymbol{\theta}} = \mathrm{diag}\{\boldsymbol{K}_{\boldsymbol{\beta}}, \boldsymbol{K}_{\boldsymbol{\omega}}\}$ and $\boldsymbol{K}_{\boldsymbol{\theta}}^{-1} = \mathrm{diag}\{\boldsymbol{K}_{\boldsymbol{\beta}}^{-1}, \boldsymbol{K}_{\boldsymbol{\omega}}^{-1}\}$. There are numerous statistical models like the nonlinear mixed-effects model (NLMM) for which global orthogonality holds. We will show an interesting decomposition of the $n^{-1/2}$ term of the expansions of the nonnull distribution functions of the LR, Wald, score and gradient statistics.

The LR, Wald, Rao score and gradient statistics for testing $\mathcal{H}_0$ versus $\mathcal{H}_a$ are defined by $S_{\mathrm{LR}} = 2\big[\ell(\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2, \widehat{\boldsymbol{\omega}}) - \ell(\widetilde{\boldsymbol{\beta}}_1, \boldsymbol{\beta}_{20}, \widetilde{\boldsymbol{\omega}})\big]$, $S_{\mathrm{W}} = (\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_{20})^{\top}(\widehat{\boldsymbol{K}}^{22})^{-1}(\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_{20})$, $S_{\mathrm{R}} = \widetilde{\boldsymbol{U}}_{\boldsymbol{\beta}_2}^{\top}\widetilde{\boldsymbol{K}}^{22}\widetilde{\boldsymbol{U}}_{\boldsymbol{\beta}_2}$ and $S_{\mathrm{T}} = \widetilde{\boldsymbol{U}}_{\boldsymbol{\beta}_2}^{\top}(\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_{20})$, respectively. Here, $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}_1^{\top}, \widehat{\boldsymbol{\beta}}_2^{\top}, \widehat{\boldsymbol{\omega}}^{\top})^{\top}$ and $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\beta}}_1^{\top}, \boldsymbol{\beta}_{20}^{\top}, \widetilde{\boldsymbol{\omega}}^{\top})^{\top}$ denote the unrestricted and restricted (under $\mathcal{H}_0$) maximum likelihood estimators of $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^{\top}, \boldsymbol{\beta}_2^{\top}, \boldsymbol{\omega}^{\top})^{\top}$, respectively, $\widehat{\boldsymbol{K}}^{22} = \boldsymbol{K}^{22}(\widehat{\boldsymbol{\theta}})$, $\widetilde{\boldsymbol{K}}^{22} = \boldsymbol{K}^{22}(\widetilde{\boldsymbol{\theta}})$ and $\widetilde{\boldsymbol{U}}_{\boldsymbol{\beta}_2} = \boldsymbol{U}_{\boldsymbol{\beta}_2}(\widetilde{\boldsymbol{\theta}})$. The null hypothesis is rejected for a given nominal level, γ say, if the test statistic exceeds the upper $100(1-\gamma)\%$ quantile of the χ^2_{p-q} distribution, denoted by $\chi^2_{p-q}(\gamma)$.

One of the aims of this paper is to study the local powers of the LR, Wald, score and gradient tests for testing the null hypothesis $\mathcal{H}_0 : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20}$ (under a sequence of local alternatives) in general continuous parametric models, when global orthogonality between $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$ holds. The nonnull distribution function of the statistic S_i under local alternatives for testing $\mathcal{H}_0 : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20}$ takes the form

$$\Pr(S_i \leq x) = G_{p-q,\lambda}(x) + \sum_{j=0}^{3} b_{ij} G_{p-q+2j,\lambda}(x) + O(n^{-1}), \quad i = \mathrm{LR}, \mathrm{W}, \mathrm{R}, \mathrm{T}, \tag{1}$$

where $G_{\nu,\lambda}(x)$ is the cumulative distribution function of a non-central chi-square variate with ν degrees of freedom and an appropriate non-centrality parameter λ. Clearly, the local powers (up to order $O(n^{-1/2})$) of the four corresponding tests are given by $\Pi_i = 1 - \Pr(S_i \leq x)$, where x is replaced by $\chi^2_{p-q}(\gamma)$. The coefficients

b_{ij} (i = LR, W, R, T and $j = 0, 1, 2, 3$) and λ are given in Hayakawa (1975), Harris and Peers (1980) and Lemonte and Ferrari (2012a). See also Mukerjee (1993) and Kakizawa (2012).

Firstly, we will show that the coefficients b_{i2} and b_{i3} (for i = LR, W, R, T) in (1) remain unchanged regardless of whether $\boldsymbol{\omega}$ is known or needs to be estimated, whereas the coefficients b_{i1} (for i = LR, W, R, T) can be written as the sum of two terms, the first of which being the corresponding term obtained as if $\boldsymbol{\omega}$ were known, and the second, an additional term yielded by the fact that $\boldsymbol{\omega}$ is unknown. A sufficient condition under which this additional term is zero will be given. Second, we will apply the general result in the class of NLMMs. In particular, we derive closed-form expressions in matrix notation for the coefficients that define the nonnull asymptotic expansions of these statistics in this class of models.

2 Main result

In order to describe our main result, we shall assume that the local alternative hypothesis is $\mathcal{H}_{an} : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} = (\epsilon_{q+1}, \ldots, \epsilon_p)^\top$ is of order $O(n^{-1/2})$. We define the quantities

$$\boldsymbol{\epsilon}^* = \begin{bmatrix} \boldsymbol{K}_{11}^{-1}\boldsymbol{K}_{12} \\ -\boldsymbol{I}_{p-q} \\ \boldsymbol{0}_{k-p,p-q} \end{bmatrix}\boldsymbol{\epsilon}, \qquad \boldsymbol{A} = (a_{rs})_{r,s=1,\ldots,k} = \begin{bmatrix} \boldsymbol{A}_{\boldsymbol{\beta}} & \boldsymbol{0}_{p,k-p} \\ \boldsymbol{0}_{k-p,p} & \boldsymbol{K}_{\boldsymbol{\omega}}^{-1} \end{bmatrix},$$

$$\boldsymbol{A}_{\boldsymbol{\beta}} = \begin{bmatrix} \boldsymbol{K}_{11}^{-1} & \boldsymbol{0}_{q,p-q} \\ \boldsymbol{0}_{p-q,q} & \boldsymbol{0}_{p-q,p-q} \end{bmatrix}, \qquad \boldsymbol{M} = (m_{rs})_{r,s=1,\ldots,k} = \begin{bmatrix} \boldsymbol{M}_{\boldsymbol{\beta}} & \boldsymbol{0}_{p,k-p} \\ \boldsymbol{0}_{k-p,p} & \boldsymbol{0}_{k-p,k-p} \end{bmatrix},$$

where $\boldsymbol{M}_{\boldsymbol{\beta}} = \boldsymbol{K}_{\boldsymbol{\beta}}^{-1} - \boldsymbol{A}_{\boldsymbol{\beta}}$, and $\boldsymbol{I}_z$ and $\boldsymbol{0}_{h,u}$ denote an identity matrix of order z and a $h \times u$ matrix of zeros, respectively. In the following, we use the standard notation for the log-likelihood derivatives: $U_r = \partial\ell(\boldsymbol{\theta})/\partial\theta_r$, $U_{rs} = \partial^2\ell(\boldsymbol{\theta})/\partial\theta_r\partial\theta_s$, $U_{rst} = \partial^3\ell(\boldsymbol{\theta})/\partial\theta_r\partial\theta_s\partial\theta_t$ etc. for $r, s, t = 1, \ldots, k$. Then, we write $\kappa_{rs} = \mathbb{E}(U_{rs})$, $\kappa_{rst} = \mathbb{E}(U_{rst})$, $\kappa_{r,st} = \mathbb{E}(U_r U_{st})$, $\kappa_{r,s,t} = \mathbb{E}(U_r U_s U_t)$, etc. All moments κ's refer to a total over the sample and are, in general, of order $O(n)$.

In what follows, assume that the indices r, s and t vary from 1 to p (i.e. on the elements of the parameter vector $\boldsymbol{\beta}$) and the indices R and S vary from $p+1$ to k (i.e. on the elements of the parameter vector $\boldsymbol{\omega}$). We arrive, after long and tedious algebraic manipulations, at the following general result.

Theorem 1 *Let $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \boldsymbol{\omega}^\top)^\top$ be the parameter vector of dimension k, where the dimensions of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are q and $p-q$, respectively, and $\boldsymbol{\omega}$ is a $(k-p)$-dimensional vector of parameters. Assume that $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$ and $\boldsymbol{\omega}$ are globally orthogonal. The nonnull asymptotic expansions of the distribution functions of the LR, Wald, score and gradient statistics for testing the null hypothesis $\mathcal{H}_0 : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20}$ under a sequence of local alternatives are given by (1) with $\lambda = \boldsymbol{\epsilon}^\top(\boldsymbol{K}_{22} - \boldsymbol{K}_{21}\boldsymbol{K}_{11}^{-1}\boldsymbol{K}_{12})\boldsymbol{\epsilon}$, $b_{\text{LR}1} = b_{\text{LR}1}^0 + \xi$, $b_{\text{W}1} = b_{\text{W}1}^0 + \xi$, $b_{\text{R}1} = b_{\text{R}1}^0 + \xi$, $b_{\text{T}1} = b_{\text{T}1}^0 + \xi$, $b_{\text{LR}3} = 0$,*

$$b_{\mathrm{LR}2} = b_{\mathrm{R}3} = -\frac{1}{6}\sum_{r,s,t=1}^{p} \kappa_{r,s,t}\epsilon_r^*\epsilon_s^*\epsilon_t^*, \quad b_{\mathrm{W}3} = -2b_{\mathrm{T}3} = \frac{1}{6}\sum_{r,s,t=1}^{p} \kappa_{rst}\epsilon_r^*\epsilon_s^*\epsilon_t^*,$$

$$b_{\mathrm{W}2} = \frac{1}{2}\sum_{r,s,t=1}^{p} \kappa_{r,st}\epsilon_r^*\epsilon_s^*\epsilon_t^* + \frac{1}{2}\sum_{r,s,t=1}^{p} \kappa_{rst}m_{rs}\epsilon_t^*, \quad b_{\mathrm{R}2} = -\frac{1}{2}\sum_{r,s,t=1}^{p} \kappa_{r,s,t}m_{rs}\epsilon_t^*,$$

$$b_{\mathrm{T}2} = -\frac{1}{4}\sum_{r,s,t=1}^{p} \kappa_{rst}m_{rs}\epsilon_t^* + \frac{1}{4}\sum_{r,s,t=1}^{p} (\kappa_{rst} + 2\kappa_{r,st})\epsilon_r^*\epsilon_s^*\epsilon_t^*,$$

$$b_{\mathrm{LR}1}^0 = -\frac{1}{6}\sum_{r,s,t=1}^{p} (\kappa_{rst} - 2\kappa_{r,s,t})\epsilon_r^*\epsilon_s^*\epsilon_t^* - \frac{1}{2}\sum_{r,s,t=1}^{p} (\kappa_{rst} + 2\kappa_{r,st})a_{rs}\epsilon_t^*$$
$$-\frac{1}{2}\sum_{r=q+1}^{p}\sum_{s,t=1}^{p} (\kappa_{rst} + \kappa_{r,st})\epsilon_r\epsilon_s^*\epsilon_t^*,$$

$$b_{\mathrm{W}1}^0 = -\frac{1}{2}\sum_{r,s,t=1}^{p} (\kappa_{rst} + 2\kappa_{r,st})\epsilon_r^*\epsilon_s^*\epsilon_t^* + \sum_{r,s,t=1}^{p} \kappa_{r,st}m_{rs}\epsilon_t^*$$
$$-\frac{1}{2}\sum_{r,s,t=1}^{p} (\kappa_{rst} + 2\kappa_{r,st})\kappa^{r,s}\epsilon_t^* - \frac{1}{2}\sum_{r=q+1}^{p}\sum_{s,t=1}^{p} (\kappa_{rst} + \kappa_{r,st})\epsilon_r\epsilon_s^*\epsilon_t^*,$$

$$b_{\mathrm{R}1}^0 = -\frac{1}{6}\sum_{r,s,t=1}^{p} (\kappa_{rst} - 2\kappa_{r,s,t})\epsilon_r^*\epsilon_s^*\epsilon_t^* + \frac{1}{2}\sum_{r,s,t=1}^{p} \kappa_{r,s,t}m_{rs}\epsilon_t^*$$
$$-\frac{1}{2}\sum_{r,s,t=1}^{p} (\kappa_{rst} + 2\kappa_{r,st})a_{rs}\epsilon_t^* - \frac{1}{2}\sum_{r=q+1}^{p}\sum_{s,t=1}^{p} (\kappa_{rst} + \kappa_{r,st})\epsilon_r\epsilon_s^*\epsilon_t^*,$$

$$b_{\mathrm{T}1}^0 = \frac{1}{4}\sum_{r,s,t=1}^{p} \kappa_{rst}\kappa^{r,s}\epsilon_t^* - \frac{1}{2}\sum_{r,s,t=1}^{p} (\kappa_{rst} + 2\kappa_{r,st})\epsilon_r^*\epsilon_s^*\epsilon_t^*$$
$$-\frac{1}{4}\sum_{r,s,t=1}^{p} (4\kappa_{r,st} + 3\kappa_{rst})a_{rs}\epsilon_t^* - \frac{1}{2}\sum_{r=q+1}^{p}\sum_{s,t=1}^{p} (\kappa_{rst} + \kappa_{r,st})\epsilon_r\epsilon_s^*\epsilon_t^*,$$

where $\boldsymbol{K}_{\boldsymbol{\beta}}^{-1} = (\kappa^{r,s})_{r,s=1,\ldots,p}$, *and*

$$\xi = \frac{1}{2}\sum_{t=1}^{p}\sum_{R,S=p+1}^{k} \kappa_{RSt}\kappa^{R,S}\epsilon_t^*, \tag{2}$$

with $\boldsymbol{K}_{\boldsymbol{\omega}}^{-1} = (\kappa^{R,S})_{R,S=p+1,\ldots,k}$, $\kappa_{RSt} = \mathbb{E}(\partial^3\ell(\boldsymbol{\theta})/\partial\omega_R\partial\omega_S\partial\beta_t)$, *and* $b_{i0} = -(b_{i1} + b_{i2} + b_{i3})$, *for* $i = \mathrm{LR, W, R, T}$.

Proof The proof is provided in Appendix. □

Remark 1 As pointed out by an anonymous referee, the finding for the S_{LR}, S_{W} and S_{R} statistics in Theorem 1 has been discussed by Eguchi (1991), who adopts the differential geometric framework. Additionally, the local power function (1) for the S_{LR}, S_{W}, S_{R}, and S_{T} statistics can be described in terms of C^* function as given in

Kakizawa (2012, § 3); see also Mukerjee (1993). It should be noticed that ξ in Theorem 1 can be found for a broad class of asymptotic chi-squared tests, which includes the LR, Wald, score and gradient tests, in Kakizawa (2012). Finally, Theorem 1 generalizes the results of Lemonte and Ferrari (2012b), which holds only for the scalar nuisance parameter case.

From Theorem 1, notice that b_{i1}^0 ($i=$ LR, W, R, T) and b_{ij} ($i=$ LR, W, R, T and $j=2,3$) represent the contribution of the parameter vector $\boldsymbol{\beta}$ to the local powers of the LR, Wald, score and gradient tests for testing the null hypothesis $\mathcal{H}_0 : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20}$, since these expressions are only obtained over the components of $\boldsymbol{\beta}$, i.e. as if $\boldsymbol{\omega}$ were known. On the other hand, the quantity ξ, which depends on the moments involving $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$, can be regarded as the contribution of the parameter vector $\boldsymbol{\omega}$ to the local powers of the LR, Wald, score and gradient tests when it is unknown, that is, when it needs to be estimated. It is interesting to note that the contribution yielded by the fact that $\boldsymbol{\omega}$ is unknown is the same for the four tests. Additionally, the contribution of the parameter vector $\boldsymbol{\omega}$ to the local powers of the tests only appears in the coefficient b_{i1} ($i=$ LR, W, R, T) and, of course, in b_{i0} ($i=$ LR, W, R, T).

Theorem 1 has a practical application when the goal is to obtain explicit formulas to the nonnull distribution functions of any of the four test statistics for special models in which orthogonality holds. It suggests that the coefficients b_{ij}'s should be obtained as if the orthogonal parameter vector $\boldsymbol{\omega}$ were known, and the extra contribution due to the estimation of $\boldsymbol{\omega}$ should be obtained from (2). We will use this general result to derive explicit formulas for the nonnull distribution functions of the LR, Wald, score and gradient statistics in the class of NLMMs.

Now, let Π_i^0 and Π_i, for $i=$ LR, W, R, T, be the local powers (ignoring terms of order smaller than $n^{-1/2}$) of the test that uses the statistic S_i when $\boldsymbol{\omega}$ is known and when $\boldsymbol{\omega}$ is unknown, respectively. It is well known that

$$G_{m,\lambda}(x) - G_{m+2,\lambda}(x) = 2g_{m+2,\lambda}(x), \tag{3}$$

where $g_{\nu,\lambda}(x)$ is the probability density function of a non-central chi-square variate with ν degrees of freedom and non-centrality parameter λ. We can then write $\Pi_i - \Pi_i^0 = z\xi$ for $i=$ LR, W, R, T, where $z = 2g_{p-q+2,\lambda}(x) > 0$ and x is replaced by $\chi^2_{p-q}(\gamma)$. Therefore, the difference between the local powers can be zero, or it can increase or decrease when $\boldsymbol{\omega}$ needs to be estimated, depending on the sign of the components of $\boldsymbol{\epsilon}$. If $\kappa_{RSt} = 0$ for $t = 1, \ldots, p$ and $R, S = p+1, \ldots, k$, we have $\xi = 0$ and hence the nonnull asymptotic expansions up to order $O(n^{-1/2})$ for the nonnull distribution functions of the LR, Wald, score and gradient statistics do not change when the parameter vector $\boldsymbol{\omega}$, which is globally orthogonal to the remaining parameters, is included in the model specification.

3 Nonlinear mixed-effects model

The NLMM can be expressed as

$$\boldsymbol{y}_i = \boldsymbol{f}_i(\boldsymbol{X}_i, \boldsymbol{\beta}) + \boldsymbol{Z}_i \boldsymbol{b}_i + \boldsymbol{\varepsilon}_i, \qquad i = 1, \ldots, N, \tag{4}$$

where $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$ is a $n_i \times 1$ vector of responses on the ith experimental unit, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is a p-vector of fixed effects parameters, $\boldsymbol{b}_i = (b_{i1}, \ldots, b_{im})^\top$ is a m-vector of random effects, $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{in_i})^\top$ with $\boldsymbol{x}_{ij} = (x_{ij1}, \ldots, x_{ijp})^\top$ and $\boldsymbol{Z}_i = (z_{i1}, \ldots, z_{in_i})^\top$ with $z_{ij} = (z_{ij1}, \ldots, z_{ijm})^\top$ are $n_i \times p$ and $n_i \times m$ known matrices of full rank, respectively, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{in_i})^\top$ is a $n_i \times 1$ random (within-subject) vector of measurement errors, and $\boldsymbol{f}_i(\boldsymbol{X}_i, \boldsymbol{\beta}) = (f_{i1}(\boldsymbol{X}_i, \boldsymbol{\beta}), \ldots, f_{in_i}(\boldsymbol{X}_i, \boldsymbol{\beta}))^\top$ is a $n_i \times 1$ vector of nonlinear functions of $\boldsymbol{\beta}$. It is often assumed that $\boldsymbol{b}_i \sim \mathcal{N}_m(\boldsymbol{0}_m, \boldsymbol{G})$ and $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_{n_i}(\boldsymbol{0}_{n_i}, \boldsymbol{R}_i)$ for $i = 1, \ldots, N$, with $\boldsymbol{b}_i$ and $\boldsymbol{\varepsilon}_i$ independent (for $i = 1, \ldots, N$), where $\boldsymbol{G} = \boldsymbol{G}(\boldsymbol{\tau})$ and $\boldsymbol{R}_i = \boldsymbol{R}_i(\boldsymbol{\phi})$ are $m \times m$ and $n_i \times n_i$ positive definite matrices whose elements are expressed as functions of vectors of covariance parameters $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_{c_1})^\top$ and $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_{c_2})^\top$ of dimensions $c_1 \times 1$ and $c_2 \times 2$, respectively, not functionally related to $\boldsymbol{\beta}$. Here, $\boldsymbol{0}_u$ denotes a u-dimensional vector of zeros.

A hierarchical formulation is behind the model (4) when it is assumed that $\boldsymbol{y}_i|\boldsymbol{b}_i \sim \mathcal{N}_{n_i}(\boldsymbol{f}_i(\boldsymbol{X}_i, \boldsymbol{\beta}) + \boldsymbol{Z}_i\boldsymbol{b}_i, \boldsymbol{R}_i)$, $\boldsymbol{b}_i \sim \mathcal{N}_m(\boldsymbol{0}_m, \boldsymbol{G})$ and $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_{n_i}(\boldsymbol{0}_{n_i}, \boldsymbol{R}_i)$, with $\boldsymbol{b}_i$ and $\boldsymbol{\varepsilon}_i$ independent (for $i = 1, \ldots, N$). So, the joint distribution of $(\boldsymbol{y}_i^\top, \boldsymbol{b}_i^\top)^\top$ becomes

$$\begin{pmatrix} \boldsymbol{y}_i \\ \boldsymbol{b}_i \end{pmatrix} \sim \mathcal{N}_{n_i+m}\left(\begin{pmatrix} \boldsymbol{f}_i(\boldsymbol{X}_i, \boldsymbol{\beta}) \\ \boldsymbol{0}_m \end{pmatrix}, \begin{bmatrix} \boldsymbol{Z}_i\boldsymbol{G}\boldsymbol{Z}_i^\top + \boldsymbol{R}_i & \boldsymbol{Z}_i\boldsymbol{G} \\ \boldsymbol{G}\boldsymbol{Z}_i^\top & \boldsymbol{G} \end{bmatrix}\right).$$

Hence, classical inference may be based on the likelihood function of the marginal model $\boldsymbol{y}_i \sim \mathcal{N}_{n_i}(\boldsymbol{f}_i(\boldsymbol{X}_i, \boldsymbol{\beta}), \boldsymbol{\Sigma}_i)$, where $\boldsymbol{\Sigma}_i = \boldsymbol{Z}_i\boldsymbol{G}\boldsymbol{Z}_i^\top + \boldsymbol{R}_i$. Let $n = \sum_{i=1}^N n_i$. Model (4) can be written in matrix form as $\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\beta}) + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{\varepsilon}$, where $\boldsymbol{Y} = (\boldsymbol{y}_1^\top, \ldots, \boldsymbol{y}_N^\top)^\top$ is a $n \times 1$ vector, $\boldsymbol{X} = (\boldsymbol{X}_1^\top, \ldots, \boldsymbol{X}_N^\top)^\top$ is a $n \times p$ matrix, $\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\beta}) = (\boldsymbol{f}_1(\boldsymbol{X}_1, \boldsymbol{\beta})^\top, \ldots, \boldsymbol{f}_N(\boldsymbol{X}_N, \boldsymbol{\beta})^\top)^\top$ is a $n \times 1$ vector of nonlinear functions of $\boldsymbol{\beta}$, $\boldsymbol{Z}$ is a $n \times Nm$ block-diagonal matrix given by $\boldsymbol{Z} = \mathrm{diag}\{\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_N\}$, $\boldsymbol{b} = (\boldsymbol{b}_1^\top, \ldots, \boldsymbol{b}_N^\top)^\top$ is a Nm-vector and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^\top, \ldots, \boldsymbol{\varepsilon}_N^\top)^\top$ is $n \times 1$. Thus, $\boldsymbol{b} \sim \mathcal{N}_{Nm}(\boldsymbol{0}_{Nm}, \boldsymbol{I}_N \otimes \boldsymbol{G})$ and $\boldsymbol{\varepsilon} \sim \mathcal{N}_n(\boldsymbol{0}_n, \boldsymbol{R})$, where $\boldsymbol{R}$ is a $n \times n$ block-diagonal matrix given by $\boldsymbol{R} = \mathrm{diag}\{\boldsymbol{R}_1, \ldots, \boldsymbol{R}_N\}$, and "$\otimes$" denotes the Kronecker product. It is also possible to express the model (4) as $\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\beta}) + \boldsymbol{e}$, where $\boldsymbol{e} = \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{\varepsilon} \sim \mathcal{N}_n(\boldsymbol{0}_n, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\omega}) = \boldsymbol{Z}(\boldsymbol{I}_N \otimes \boldsymbol{G})\boldsymbol{Z}^\top + \boldsymbol{R}$ and $\boldsymbol{\omega} = (\boldsymbol{\tau}^\top, \boldsymbol{\phi}^\top)^\top$ is a $(c_1 + c_2) \times 1$ vector of unknown parameters.

Let $\boldsymbol{\mu} = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\beta})$ and $\boldsymbol{u} = \boldsymbol{Y} - \boldsymbol{\mu}$. The log-likelihood function for $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\omega}^\top)^\top$ can be expressed as $\ell(\boldsymbol{\theta}) = -(n/2)\log(2\pi) - (1/2)\log|\boldsymbol{\Sigma}| - (1/2)\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{u}\boldsymbol{u}^\top\}$, where $|\boldsymbol{\Sigma}|$ denotes the determinant of the matrix $\boldsymbol{\Sigma}$, and $\mathrm{tr}\{\cdot\}$ is the trace operator. Some additional notation is in order. Let $\boldsymbol{d}_r = \partial\boldsymbol{\mu}/\partial\beta_r$, $\boldsymbol{d}_{rs} = \partial^2\boldsymbol{\mu}/\partial\beta_r\partial\beta_s$, $\dot{\boldsymbol{\Sigma}}_R = \partial\boldsymbol{\Sigma}/\partial\omega_R$, with $r, s = 1, \ldots, p$, and $R = 1, \ldots, c_1 + c_2$. Also, define $\boldsymbol{D} = [\boldsymbol{d}_1, \ldots, \boldsymbol{d}_p]$ and $\boldsymbol{V} = [\mathrm{vec}\{\dot{\boldsymbol{\Sigma}}_1\}, \ldots, \mathrm{vec}\{\dot{\boldsymbol{\Sigma}}_{c_1+c_2}\}]$, where $\mathrm{vec}\{\cdot\}$ is the vec operator, which transforms a matrix into a vector by stacking the columns of the matrix one underneath the other. Let

$$\boldsymbol{F} = \begin{bmatrix} \boldsymbol{D} \\ \boldsymbol{V} \end{bmatrix}, \qquad \boldsymbol{H} = \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{0}_{n,n^2} \\ \boldsymbol{0}_{n^2,n} & 2(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) \end{bmatrix}^{-1}, \qquad \boldsymbol{v} = \begin{pmatrix} \boldsymbol{u} \\ -\mathrm{vec}\{\boldsymbol{\Sigma} - \boldsymbol{u}\boldsymbol{u}^\top\} \end{pmatrix}.$$

So, after some straightforward matrix algebra, the score function and the Fisher information matrix for $\boldsymbol{\theta}$ can be written, respectively, as $\boldsymbol{U}_{\boldsymbol{\theta}} = \boldsymbol{F}^\top\boldsymbol{H}\boldsymbol{v}$ and $\boldsymbol{K}_{\boldsymbol{\theta}} = \boldsymbol{F}^\top\boldsymbol{H}\boldsymbol{F}$.

The maximum likelihood estimator $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^\top, \widehat{\boldsymbol{\omega}}^\top)^\top$ of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\omega}^\top)^\top$ satisfies the equation $\boldsymbol{U}_{\boldsymbol{\theta}}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = \mathbf{0}_k$, with $k = p + c_1 + c_2$. The Fisher scoring method can be used to estimate $\boldsymbol{\theta}$ by iteratively solving $(\boldsymbol{F}^{(w)\top}\boldsymbol{H}^{(w)}\boldsymbol{F}^{(w)})\boldsymbol{\theta}^{(w+1)} = \boldsymbol{F}^{(w)\top}\boldsymbol{H}^{(w)}\boldsymbol{v}^{*(w)}$, where $\boldsymbol{v}^{*(w)} = \boldsymbol{F}^{(w)}\boldsymbol{\theta}^{(w)} + \boldsymbol{v}^{(w)}$ and $w = 0, 1, \ldots$ is the iteration counter. Each loop, through the above iterative scheme, consists of an iterative re-weighted least squares algorithm to optimize the log-likelihood $\ell(\boldsymbol{\theta})$. A recent discussion about estimation in NLMMs can be found in Meza et al. (2012).

4 Nonnull asymptotic expansions in NLMMs

In what follows, we shall consider the LR, Wald, Rao score and gradient statistics for testing a composite null hypothesis on the fixed effects in the class of NLMMs, i.e. tests on the parameter vector $\boldsymbol{\beta}$. Tests on the variance components (i.e. on the parameter vector $\boldsymbol{\omega} = (\boldsymbol{\tau}^\top, \boldsymbol{\phi}^\top)^\top$) will not be the subject of this paper. The reader is referred to Nobre et al. (2013) and references therein for tests on the variance components. The null hypothesis of interest is $\mathcal{H}_0 : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20}$, which will be tested against the alternative hypothesis $\mathcal{H}_a : \boldsymbol{\beta}_2 \neq \boldsymbol{\beta}_{20}$, where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$ with $\boldsymbol{\beta}_1 = (\beta_1, \ldots, \beta_q)^\top$ and $\boldsymbol{\beta}_2 = (\beta_{q+1}, \ldots, \beta_p)^\top$. Here, $\boldsymbol{\beta}_{20}$ is a fixed column vector of dimension $p - q$, and $\boldsymbol{\beta}_1$ and $\boldsymbol{\omega}$ act as nuisance parameter vectors. Let $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}_1^\top, \widehat{\boldsymbol{\beta}}_2^\top, \widehat{\boldsymbol{\omega}}^\top)^\top$ and $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\beta}}_1^\top, \boldsymbol{\beta}_{20}^\top, \widetilde{\boldsymbol{\omega}}^\top)^\top$ be the unrestricted and restricted (under $\mathcal{H}_0$) maximum likelihood estimators of $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \boldsymbol{\omega}^\top)^\top$, respectively. The LR, Wald, score and gradient statistics for testing $\mathcal{H}_0$ are defined by $S_{\mathrm{LR}} = \log(|\widetilde{\boldsymbol{\Sigma}}||\widehat{\boldsymbol{\Sigma}}|^{-1}) + \mathrm{tr}\{\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\boldsymbol{u}}\widetilde{\boldsymbol{u}}^\top - \widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{u}}\widehat{\boldsymbol{u}}^\top\}$, $S_{\mathrm{W}} = (\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_{20})^\top\widehat{\boldsymbol{D}}_2^\top\widehat{\boldsymbol{\Sigma}}^{-1}(\widehat{\boldsymbol{D}}_2 - \widehat{\boldsymbol{D}}_1\widehat{\boldsymbol{\Upsilon}})(\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_{20})$, $S_{\mathrm{R}} = \widetilde{\boldsymbol{u}}^\top\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\boldsymbol{D}}_2[\widetilde{\boldsymbol{D}}_2^\top\widetilde{\boldsymbol{\Sigma}}^{-1}(\widetilde{\boldsymbol{D}}_2 - \widetilde{\boldsymbol{D}}_1\widetilde{\boldsymbol{\Upsilon}})]^{-1}\widetilde{\boldsymbol{D}}_2^\top\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\boldsymbol{u}}$ and $S_{\mathrm{T}} = \widetilde{\boldsymbol{u}}^\top\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\boldsymbol{D}}_2(\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_{20})$, respectively, where $\boldsymbol{\Upsilon} = (\boldsymbol{D}_1^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_1)^{-1}\boldsymbol{D}_1^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_2$ and $\boldsymbol{D} = \begin{bmatrix}\boldsymbol{D}_1 & \boldsymbol{D}_2\end{bmatrix}$, $\boldsymbol{D}_1$ being $n \times q$ and $\boldsymbol{D}_2$ being $n \times (p - q)$. Also, tildes and hats indicate evaluation at the restricted and unrestricted maximum likelihood estimates, respectively.

We assume the local alternative hypothesis $\mathcal{H}_{an} : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} = \boldsymbol{\beta}_2 - \boldsymbol{\beta}_{20} = (\epsilon_{q+1}, \ldots, \epsilon_p)^\top$ is of order $O(n^{-1/2})$. From Theorem 1 and after some algebra, it can be shown that $\xi = 0$ and therefore the nonnull distribution functions (up to order $O(n^{-1/2})$) of the four statistics do not change when the parameter vector $\boldsymbol{\omega} = (\boldsymbol{\tau}^\top, \boldsymbol{\phi}^\top)^\top$ is included in the model specification. Let $\lambda = \boldsymbol{\epsilon}^\top\boldsymbol{D}_2^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{D}_2 - \boldsymbol{D}_1\boldsymbol{\Upsilon})\boldsymbol{\epsilon}$. The coefficients b_{ij}'s that define the nonnull expansions are obtained from Theorem 1 and, after extensive algebra, they can be expressed in matrix notation as follows: $b_{\mathrm{LR2}} = b_{\mathrm{LR3}} = b_{\mathrm{R2}} = b_{\mathrm{R3}} = 0$, $b_{\mathrm{W3}} = -(1/2)\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}(\boldsymbol{\epsilon}^* \otimes \boldsymbol{I}_p)\boldsymbol{\epsilon}^*\boldsymbol{\epsilon}^{*\top}\boldsymbol{D}^\top\}$, $b_{\mathrm{T3}} = -(1/2)b_{\mathrm{W3}}$,

$$\begin{aligned}
b_{\mathrm{LR1}} = b_{\mathrm{R1}} = &\ \frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}(\boldsymbol{\epsilon}^* \otimes \boldsymbol{I}_p)\boldsymbol{\epsilon}^*\boldsymbol{\epsilon}^{*\top}\boldsymbol{D}^\top\} \\
&+ \mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}_{(2)}(\boldsymbol{\epsilon}^* \otimes \boldsymbol{I}_{p-q})\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{*\top}\boldsymbol{D}^\top\} \\
&+ \frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}^{(1)}\mathrm{vec}\{(\boldsymbol{D}_1^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_1)^{-1}\}\boldsymbol{\epsilon}^{*\top}\boldsymbol{D}^\top\},
\end{aligned}$$

$$
\begin{aligned}
b_{\mathrm{W}1} = {} & \frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}(\boldsymbol{\epsilon}^{*}\otimes \boldsymbol{I}_{p})\boldsymbol{\epsilon}^{*}\boldsymbol{\epsilon}^{*\top}\boldsymbol{D}^{\top}\}\\
& + \mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}(\boldsymbol{\epsilon}^{*}\otimes \boldsymbol{I}_{p})(\boldsymbol{D}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{D})^{-1}\boldsymbol{D}^{\top}\}\\
& - \mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}_{(1)}(\boldsymbol{\epsilon}^{*}\otimes \boldsymbol{I}_{q})(\boldsymbol{D}_{1}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{1})^{-1}\boldsymbol{D}_{1}^{\top}\}\\
& + \frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}\mathrm{vec}\{(\boldsymbol{D}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{D})^{-1}\}\boldsymbol{\epsilon}^{*\top}\boldsymbol{D}^{\top}\}\\
& + \mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}_{(2)}(\boldsymbol{\epsilon}^{*}\otimes \boldsymbol{I}_{p-q})\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{*\top}\boldsymbol{D}^{\top}\},\\
b_{\mathrm{W}2} = {} & \frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}(\boldsymbol{\epsilon}^{*}\otimes \boldsymbol{I}_{p})\boldsymbol{\epsilon}^{*}\boldsymbol{\epsilon}^{*\top}\boldsymbol{D}^{\top}\}\\
& - \frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}\mathrm{vec}\{(\boldsymbol{D}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{D})^{-1}\}\boldsymbol{\epsilon}^{*\top}\boldsymbol{D}^{\top}\}\\
& - \mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}(\boldsymbol{\epsilon}^{*}\otimes \boldsymbol{I}_{p})(\boldsymbol{D}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{D})^{-1}\boldsymbol{D}^{\top}\}\\
& + \frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}^{(1)}\mathrm{vec}\{(\boldsymbol{D}_{1}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{1})^{-1}\}\boldsymbol{\epsilon}^{*\top}\boldsymbol{D}^{\top}\}\\
& + \mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}_{(1)}(\boldsymbol{\epsilon}^{*}\otimes \boldsymbol{I}_{q})(\boldsymbol{D}_{1}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{1})^{-1}\boldsymbol{D}_{1}^{\top}\},\\
b_{\mathrm{T}1} = {} & -\frac{1}{4}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}\mathrm{vec}\{(\boldsymbol{D}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{D})^{-1}\}\boldsymbol{\epsilon}^{*\top}\boldsymbol{D}^{\top}\}\\
& - \frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}(\boldsymbol{\epsilon}^{*}\otimes \boldsymbol{I}_{p})(\boldsymbol{D}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{D})^{-1}\boldsymbol{D}^{\top}\}\\
& + \frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}(\boldsymbol{\epsilon}^{*}\otimes \boldsymbol{I}_{p})\boldsymbol{\epsilon}^{*}\boldsymbol{\epsilon}^{*\top}\boldsymbol{D}^{\top}\}\\
& + \frac{3}{4}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}^{(1)}\mathrm{vec}\{(\boldsymbol{D}_{1}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{1})^{-1}\}\boldsymbol{\epsilon}^{*\top}\boldsymbol{D}^{\top}\}\\
& + \frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}_{(1)}(\boldsymbol{\epsilon}^{*}\otimes \boldsymbol{I}_{q})(\boldsymbol{D}_{1}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{1})^{-1}\boldsymbol{D}_{1}^{\top}\}\\
& + \mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}_{(2)}(\boldsymbol{\epsilon}^{*}\otimes \boldsymbol{I}_{p-q})\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{*\top}\boldsymbol{D}^{\top}\},\\
b_{\mathrm{T}2} = {} & \frac{1}{4}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}\mathrm{vec}\{(\boldsymbol{D}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{D})^{-1}\}\boldsymbol{\epsilon}^{*\top}\boldsymbol{D}^{\top}\}\\
& + \frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}(\boldsymbol{\epsilon}^{*}\otimes \boldsymbol{I}_{p})(\boldsymbol{D}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{D})^{-1}\boldsymbol{D}^{\top}\}\\
& - \frac{1}{4}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}^{(1)}\mathrm{vec}\{(\boldsymbol{D}_{1}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{1})^{-1}\}\boldsymbol{\epsilon}^{*\top}\boldsymbol{D}^{\top}\}\\
& - \frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}_{(1)}(\boldsymbol{\epsilon}^{*}\otimes \boldsymbol{I}_{q})(\boldsymbol{D}_{1}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{1})^{-1}\boldsymbol{D}_{1}^{\top}\}\\
& - \frac{1}{4}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}(\boldsymbol{\epsilon}^{*}\otimes \boldsymbol{I}_{p})\boldsymbol{\epsilon}^{*}\boldsymbol{\epsilon}^{*\top}\boldsymbol{D}^{\top}\},
\end{aligned}
$$

where $\boldsymbol{\epsilon}^{*} = \left[\boldsymbol{D}_{2}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{1}(\boldsymbol{D}_{1}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{1})^{-1} \quad - \boldsymbol{I}_{p-q}\right]^{\top}\boldsymbol{\epsilon}$, $\boldsymbol{\Phi}_{r} = [\boldsymbol{\Phi}_{r}^{(1)}, \boldsymbol{\Phi}_{r}^{(2)}]$, $\boldsymbol{\Phi}_{r}^{(1)} = [\boldsymbol{d}_{1r}, \ldots, \boldsymbol{d}_{qr}]$, $\boldsymbol{\Phi}_{r}^{(2)} = [\boldsymbol{d}_{(q+1)r}, \ldots, \boldsymbol{d}_{pr}]$, $\boldsymbol{\Delta}^{(1)} = [\boldsymbol{\Phi}_{1}^{(1)}, \ldots, \boldsymbol{\Phi}_{q}^{(1)}]$, $\boldsymbol{\Phi} = [\boldsymbol{\Phi}_{(1)}, \boldsymbol{\Phi}_{(2)}]$, $\boldsymbol{\Phi}_{(1)} = [\boldsymbol{\Phi}_{1}, \ldots, \boldsymbol{\Phi}_{q}]$ and $\boldsymbol{\Phi}_{(2)} = [\boldsymbol{\Phi}_{q+1}, \ldots, \boldsymbol{\Phi}_{p}]$, with $r = 1, \ldots, p$. The coefficients b_{i0} are obtained from $b_{i0} = -(b_{i1} + b_{i2} + b_{i3})$ for $i =$ LR, W, R, T. The b_{ij}'s are of order $O(n^{-1/2})$ and all quantities except $\boldsymbol{\epsilon}$ are evaluated under the null hypothesis $\mathcal{H}_0$. The detailed derivation of the above expressions is long and very tedious, and hence will be omitted to save space.

A brief commentary on the coefficients that define the nonnull asymptotic expansions of the distribution functions of the LR, Wald, score and gradient statistics in NLMMs is in order. Note that they depend heavily on the derivative matrix $\boldsymbol{D}$. They also involve the second derivative of the (possibly nonlinear) function $\boldsymbol{f}_i(\boldsymbol{X}_i;\boldsymbol{\beta})$. Additionally, they depend on the variance-covariance matrix $\boldsymbol{\Sigma}$ through its inverse. Unfortunately, they are not easy to interpret in generality and provide no indication as to what structural aspects of the model contribute significantly to their magnitude. The matrix $\boldsymbol{\Phi}_r$ $(r = 1, \ldots, p)$ may be considered as the amount of nonlinearity in the LR, Wald, score and gradient statistics induced by the (possibly nonlinear) function $\boldsymbol{f}_i(\boldsymbol{X}_i;\boldsymbol{\beta})$, since it vanishes for linear models. It is interesting to note that $b_{\mathrm{LR}j} = b_{\mathrm{R}j}$ $(j = 0, 1, 2, 3)$ and hence the nonnull asymptotic expansions of the distribution functions of the LR and score statistics (up to order $O(n^{-1/2})$) are equal; see also Kakizawa (2012) with $\kappa_{r,s,t} = 0$ for $r, s, t = 1, \ldots, p$.

Next, we shall compare the local powers of the rival tests based on the general nonnull asymptotic expansions derived above for testing the null hypothesis $\mathcal{H}_0$: $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20}$ in the class of NLMMs. Let Π_i be the power function, up to order $O(n^{-1/2})$, of the test that uses the statistic S_i for $i = \mathrm{LR}, \mathrm{W}, \mathrm{R}, \mathrm{T}$. We have

$$\Pi_i - \Pi_l = \sum_{j=0}^{3}(b_{lj} - b_{ij})G_{p-q+2j,\lambda}(x), \qquad i \neq l, \tag{5}$$

where x is replaced by $\chi^2_{p-q}(\gamma)$. From (3) and (5), after some algebra, it follows that

$$\begin{aligned} \Pi_{\mathrm{LR}} - \Pi_{\mathrm{W}} &= \Pi_{\mathrm{R}} - \Pi_{\mathrm{W}} = -2[\vartheta_1 g_{p-q+4,\lambda}(x) + \vartheta_2 g_{p-q+6,\lambda}(x)], \\ \Pi_{\mathrm{LR}} - \Pi_{\mathrm{T}} &= \Pi_{\mathrm{R}} - \Pi_{\mathrm{T}} = \vartheta_1 g_{p-q+4,\lambda}(x) + \vartheta_2 g_{p-q+6,\lambda}(x), \\ \Pi_{\mathrm{W}} - \Pi_{\mathrm{T}} &= 3[\vartheta_1 g_{p-q+4,\lambda}(x) + \vartheta_2 g_{p-q+6,\lambda}(x)], \end{aligned} \tag{6}$$

where $\vartheta_2 = -(1/2)\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}(\boldsymbol{\epsilon}^* \otimes \boldsymbol{I}_p)\boldsymbol{\epsilon}^*\boldsymbol{\epsilon}^{*\top}\boldsymbol{D}^\top\}$ and

$$\begin{aligned} \vartheta_1 = &-\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}(\boldsymbol{\epsilon}^* \otimes \boldsymbol{I}_p)(\boldsymbol{D}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{D})^{-1}\boldsymbol{D}^\top\} \\ &+ \mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}_{(1)}(\boldsymbol{\epsilon}^* \otimes \boldsymbol{I}_q)(\boldsymbol{D}_1^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_1)^{-1}\boldsymbol{D}_1^\top\} \\ &- \frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}\mathrm{vec}\{(\boldsymbol{D}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{D})^{-1}\}\boldsymbol{\epsilon}^{*\top}\boldsymbol{D}^\top\} \\ &+ \frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}^{(1)}\mathrm{vec}\{(\boldsymbol{D}_1^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_1)^{-1}\}\boldsymbol{\epsilon}^{*\top}\boldsymbol{D}^\top\}. \end{aligned}$$

As earlier noted, we arrive at the following general conclusions from equations (6). We have that $\Pi_{\mathrm{LR}} = \Pi_{\mathrm{R}}$ and hence the LR and score tests have the same local power (up to order $O(n^{-1/2})$) for testing hypotheses on the fixed effects in NLMMs. Also, if $\vartheta_1 \geq 0$ and $\vartheta_2 \geq 0$ with $\vartheta_1 + \vartheta_2 > 0$, we have $\Pi_{\mathrm{W}} > \Pi_{\mathrm{LR}} = \Pi_{\mathrm{R}} > \Pi_{\mathrm{T}}$. On the other hand, if $\vartheta_1 \leq 0$ and $\vartheta_2 \leq 0$ with $\vartheta_1 + \vartheta_2 < 0$, we have $\Pi_{\mathrm{T}} > \Pi_{\mathrm{LR}} = \Pi_{\mathrm{R}} > \Pi_{\mathrm{W}}$. For the linear mixed model, we have that $\Pi_{\mathrm{LR}} = \Pi_{\mathrm{W}} = \Pi_{\mathrm{R}} = \Pi_{\mathrm{T}}$, as expected.

The general conclusions above are very interesting and can be used to choose the most powerful test to make inference on the fixed effects parameters in the class of NLMMs; that is, if $\vartheta_1 \geq 0$ and $\vartheta_2 \geq 0$ with $\vartheta_1 + \vartheta_2 > 0$, then the Wald test will be the most powerful test, on the other hand, the gradient test will be the most powerful test if $\vartheta_1 \leq 0$ and $\vartheta_2 \leq 0$ with $\vartheta_1 + \vartheta_2 < 0$. It should be noticed that these conditions can be easily verified numerically in practical applications after fitting the NLMM to the data.

Appendix: Proof of Theorem 1

We provide the sketch of the proof for the LR statistic, since for the other ones (Harris and Peers 1980; Lemonte and Ferrari 2012a) the proof is obtained in a similar fashion. With the abuse of notation, let $r, s, t = 1, \ldots, k$, where k is the total number of parameters. From Hayakawa (1975), we have that $b_{\mathrm{LR}3} = 0$, $b_{\mathrm{LR}0} = -(b_{\mathrm{LR}1} + b_{\mathrm{LR}2} + b_{\mathrm{LR}3})$, $b_{\mathrm{LR}2} = -(1/6)\sum_{r,s,t=1}^{k} \kappa_{r,s,t}\epsilon_r^*\epsilon_s^*\epsilon_t^*$ and $b_{\mathrm{LR}1} = -(1/6)\sum_{r,s,t=1}^{k}(\kappa_{rst} - 2\kappa_{r,s,t})\epsilon_r^*\epsilon_s^*\epsilon_t^* - (1/2)\sum_{r,s,t=1}^{k}(\kappa_{rst} + 2\kappa_{r,st})a_{rs}\epsilon_t^* - (1/2)\sum_{r=q+1}^{p}\sum_{s,t=1}^{k}(\kappa_{rst} + \kappa_{r,st})\epsilon_r\epsilon_s^*\epsilon_t^*$. First, note that

$$b_{\mathrm{LR}2} = -\frac{1}{6}\sum_{r,s,t=1}^{p} \kappa_{r,s,t}\epsilon_r^*\epsilon_s^*\epsilon_t^*,$$

since $\epsilon_r^* = \epsilon_s^* = \epsilon_t^* = 0$ for $r, s, t = p+1, \ldots, k$. Also,

$$-\frac{1}{6}\sum_{r,s,t=1}^{k}(\kappa_{rst} - 2\kappa_{r,s,t})\epsilon_r^*\epsilon_s^*\epsilon_t^* = -\frac{1}{6}\sum_{r,s,t=1}^{p}(\kappa_{rst} - 2\kappa_{r,s,t})\epsilon_r^*\epsilon_s^*\epsilon_t^*,$$

$$-\frac{1}{2}\sum_{r=q+1}^{p}\sum_{s,t=1}^{k}(\kappa_{rst} + \kappa_{r,st})\epsilon_r\epsilon_s^*\epsilon_t^* = -\frac{1}{2}\sum_{r=q+1}^{p}\sum_{s,t=1}^{p}(\kappa_{rst} + \kappa_{r,st})\epsilon_r\epsilon_s^*\epsilon_t^*,$$

$$-\frac{1}{2}\sum_{r,s,t=1}^{k}(\kappa_{rst} + 2\kappa_{r,st})a_{rs}\epsilon_t^* = -\frac{1}{2}\sum_{t=1}^{p}\sum_{r,s=1}^{k}(\kappa_{rst} + 2\kappa_{r,st})a_{rs}\epsilon_t^*.$$

We have that $a_{rs} = 0$ for $r = 1, \ldots, p$ and $s = p+1, \ldots, k$, and $a_{rs} = 0$ for $r = p+1, \ldots, k$ and $s = 1, \ldots, p$. Also, $a_{rs} = \kappa^{r,s}$ for $r, s = p+1, \ldots, k$. Hence,

$$-\frac{1}{2}\sum_{t=1}^{p}\sum_{r,s=1}^{k}(\kappa_{rst} + 2\kappa_{r,st})a_{rs}\epsilon_t^* = -\frac{1}{2}\sum_{r,s,t=1}^{p}(\kappa_{rst} + 2\kappa_{r,st})a_{rs}\epsilon_t^* - \frac{1}{2}\sum_{t=1}^{p}\sum_{r,s=p+1}^{k}(\kappa_{rst} + 2\kappa_{r,st})\kappa^{r,s}\epsilon_t^*.$$

Now, according to the notation of Theorem 1, it follows that

$$\xi = -\frac{1}{2}\sum_{t=1}^{p}\sum_{r,s=p+1}^{k}(\kappa_{rst} + 2\kappa_{r,st})\kappa^{r,s}\epsilon_t^* = -\frac{1}{2}\sum_{t=1}^{p}\sum_{R,S=p+1}^{k}(\kappa_{RSt} + 2\kappa_{R,St})\kappa^{R,S}\epsilon_t^*,$$

where $\kappa_{R,St} = \mathbb{E}[(\partial\ell(\boldsymbol{\theta})/\partial\omega_R)(\partial^2\ell(\boldsymbol{\theta})/\partial\omega_S\partial\beta_t)]$. By using the Bartlett identity $\kappa_{R,St} = \kappa_{St}^{(R)} - \kappa_{RSt}$ for $t = 1, \ldots, p$ and $R, S = p+1, \ldots, k$, where $\kappa_{St}^{(R)} = \partial\kappa_{St}/\partial\omega_R$, we have $\kappa_{R,St} = -\kappa_{RSt}$ since the orthogonality between $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$ implies that $\kappa_{St} = 0$. Therefore, we can express

$$\xi = \frac{1}{2}\sum_{t=1}^{p}\sum_{R,S=p+1}^{k}\kappa_{RSt}\kappa^{R,S}\epsilon_t^*,$$

thus $b_{\mathrm{LR1}} = b_{\mathrm{LR1}}^0 + \xi$.

Acknowledgments I am very grateful to the Associate Editor and two anonymous referees for the valuable comments and suggestions which have improved considerably the first version of the manuscript. The author gratefully acknowledges the financial support of CNPq (Brazil) and FACEPE (Pernambuco, Brazil).

References

Cox, D. R., Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society B*, *40*, 1–39.

Eguchi, S. (1991). A geometric look at nuisance parameter effect of local powers in testing hypothesis. *Annals of the Institute of Statistical Mathematics*, *43*, 245–260.

Harris, P., Peers, H. W. (1980). The local power of the efficient score test statistic. *Biometrika*, *67*, 525–529.

Hayakawa, T. (1975). The likelihood ratio criterion for a composite hypothesis under a local alternative. *Biometrika*, *62*, 451–460.

Kakizawa, Y. (2012). Second-order powers of a class of tests in the presence of a nuisance parameter. *Communications in Statistics Theory and Methods*, *41*, 3676–3691.

Lemonte, A. J., Ferrari, S. L. P. (2012a). The local power of the gradient test. *Annals of the Institute of Statistical Mathematics*, *64*, 373–381.

Lemonte, A. J., Ferrari, S. L. P. (2012b). A note on the local power of the LR, Wald, score and gradient tests. *Electronic Journal of Statistics*, *6*, 421–434.

Meza, C., Osorio, F., De la Cruz, R. (2012). Estimation in nonlinear mixed-effects models using heavy-tailed distributions. *Statistics and Computing*, *22*, 121–139.

Mukerjee, R. (1993). An extension of the conditional likelihood ratio test to the general multiparameter case. *Annals of the Institute of Statistical Mathematics*, *45*, 759–771.

Nobre, J. S., Singer, J. M., Sen, P. K. (2013). U-tests for variance components in linear mixed models. *Test*, *22*, 580–605.

Rao, C. R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problens of estimation. *Proceedings of the Cambridge Philosophical Society*, *44*, 50–57.

Terrell, G. R. (2002). The gradient statistic. *Computing Science and Statistics*, *34*, 206–215.

Wald, A. (1943). Tests of statistical hypothesis concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, *54*, 426–482.

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypothesis. *Annals of Mathematical Statistics*, *9*, 60–62.