Computational Statistics and Data Analysis 53 (2009) 2573–2582

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Improved testing inference in mixed linear models

Tatiane F.N. Melo [a], Silvia L.P. Ferrari [a,*], Francisco Cribari-Neto [b]

[a] Departamento de Estatística, Universidade de São Paulo, Rua do Matão, 1010, São Paulo/SP, 05508-090, Brazil

[b] Departamento de Estatística, Universidade Federal de Pernambuco, Cidade Universitária, Recife/PE, 50740-540, Brazil

ARTICLE INFO

Article history:
Received 18 June 2008
Received in revised form 10 November 2008
Accepted 18 December 2008
Available online 30 December 2008

ABSTRACT

Mixed linear models are commonly used in repeated measures studies. They account for the dependence amongst observations obtained from the same experimental unit. Often, the number of observations is small, and it is thus important to use inference strategies that incorporate small sample corrections. In this paper, we develop modified versions of the likelihood ratio test for fixed effects inference in mixed linear models. In particular, we derive a Bartlett correction to such a test, and also to a test obtained from a modified profile likelihood function. Our results generalize those in [Zucker, D.M., Lieberman, O., Manor, O., 2000. Improved small sample inference in the mixed linear model: Bartlett correction and adjusted likelihood. Journal of the Royal Statistical Society B, 62, 827–838] by allowing the parameter of interest to be vector-valued. Additionally, our Bartlett corrections allow for random effects nonlinear covariance matrix structure. We report simulation results which show that the proposed tests display superior finite sample behavior relative to the standard likelihood ratio test. An application is also presented and discussed.

1. Introduction

In recent years, repeated measures of data have been widely analyzed in many fields, including biology and medicine. In such studies, the observations are obtained from different experimental units, each unit being observed more than once (Brown and Prescott, 2006). In particular, some of these studies use longitudinal data (Verbeke and Molenberghs, 2000), in which the observations are collected over time. Mixed linear models have been extensively used by practitioners to analyze repeated measures, since they account for within units correlation . It is also noteworthy that there is available software, specifically designed for the estimation of such models; see Pinheiro and Bates (2000) and Littel et al. (2006).

A common shortcoming lies in the fact that, in many studies, the sample size is small, which renders approximate inferential procedures unreliable. Improved inference may be based on the theory of higher order asymptotics. Practical applications of such theory may be found in Brazzale et al. (2007). The likelihood ratio test, which is commonly used to make inference on the fixed effects parameters, quite often displays large size distortions when the sample size is small. This happens because its null distribution is poorly approximated by the limiting χ^2 distribution, from which critical values are obtained. It is possible to obtain a Bartlett correction factor and use it to modify the likelihood ratio test statistic in such a way as to bring its null distribution closer to its limiting counterpart; the approximation error is reduced from $O(n^{-1})$ to $O(n^{-2})$, where n is the sample size, thus making any size distortion vanish at a faster rate.

Another shortcoming relates to the effect of the nuisance parameters on the resulting inference on the parameters of interest. Different modifications to the profile likelihood function have been proposed with the aim of reducing such effect. For a review, see Severini (2000, Chapter 9); see also Sartori et al. (1999) and Sartori (2003). The adjustment proposed by

* Corresponding author. Tel.: +55 11 3091 6129; fax: +55 11 3091 6130.
E-mail address: silviaferrari@usp.br (S.L.P. Ferrari).

0167-9473/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2008.12.007

Cox and Reid (1987) can be used whenever the nuisance and interest parameters are orthogonal. DiCiccio and Stern (1994) have shown that the Cox–Reid test statistic can be Bartlett-corrected, just as the likelihood ratio test statistic. The combined use of modified profile likelihoods and Bartlett correction can deliver accurate and reliable inference in small samples, as evidenced by the results in Ferrari et al. (2004, 2005) and Cysneiros and Ferrari (2006).

Zucker et al. (2000) obtained improved likelihood ratio testing inference by deriving Bartlett corrections to the profile, and modified (Cox–Reid) profile likelihood ratio tests on the fixed effects parameters in mixed linear models. Their results, however, are only applicable for testing one parameter at a time, since they only allow for a scalar parameter of interest. In many studies, nonetheless, practitioners wish to perform joint testing inference on a set of parameters, especially when comparing three or more treatments in medical trials. Also, they derived the Bartlett correction to the profile likelihood ratio test only for the situation where the covariance matrix for the random effects has a linear structure. Hence, their results are not fully applicable in many situations of interest (e.g. when the responses of a single subject are measured sequentially and the errors are assumed to be autocorrelated). Our chief goal is to generalize their results so that they are valid in situations where the parameter of interest is vector-valued and the covariance matrix for the random effects is allowed to have a non-linear structure. We obtain the Cox–Reid profile likelihood adjustment, and also Bartlett correction factors for the profile and adjusted profile likelihood ratio test statistics.

The paper unfolds as follows. Section 2 introduces the mixed linear model, Section 3 contains the three improved tests (Cox–Reid and Bartlett-corrected tests), and Section 4 presents a simulation study on the finite sample behavior of the standard likelihood ratio test and its modified counterparts. An application that uses real data is presented and discussed in Section 5. Finally, Section 6 concludes the paper. Technical details are collected in two appendices.

2. Mixed linear models

The mixed linear model is given by

$$\mathbf{y}_i = X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, N, \tag{1}$$

where $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{i\tau_i})^\top$ is a $\tau_i \times 1$ vector of responses on the ith experimental unit, $\boldsymbol{\beta}$ is an n-vector of fixed effects parameters, X_i is a $\tau_i \times n$ known matrix, $\mathbf{b}_i$ is a random effects vector ($q \times 1$), Z_i is a known $\tau_i \times q$ matrix, and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \epsilon_{i2}, \ldots, \epsilon_{i\tau_i})^\top$ is a $\tau_i \times 1$ vector of random errors. It is often assumed that $\epsilon_i \sim \mathcal{N}_{\tau_i}(\mathbf{0}, \sigma^2 I_{\tau_i})$, where I_{τ_i} denotes the $\tau_i \times \tau_i$ identity matrix and $\mathbf{0}$ is a vector of zeros. It is also assumed that $\mathbf{b}_i \sim \mathcal{N}_q(\mathbf{0}, G)$, where $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_N, \boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \ldots, \boldsymbol{\epsilon}_N$ are independent and $G = G(\boldsymbol{\varrho})$ is a $q \times q$ positive definite matrix, $\boldsymbol{\varrho}$ being an $m \times 1$ vector of unknown parameters. Model (1) can be written in matrix form as

$$\mathbf{Y} = X\boldsymbol{\beta} + Z\mathbf{b} + \boldsymbol{\epsilon}, \tag{2}$$

where $\mathbf{Y} = (\mathbf{y}_1^\top, \mathbf{y}_2^\top, \ldots, \mathbf{y}_N^\top)^\top$ is $T \times 1$, with $T = \sum_{i=1}^N \tau_i$, $X = (\mathbf{X}_1^\top, \mathbf{X}_2^\top, \ldots, \mathbf{X}_N^\top)^\top$ is a $T \times n$ matrix, Z is a $T \times Nq$ diagonal matrix given by $Z = \mathrm{diag}(Z_1, Z_2, \ldots, Z_N)$, $\mathbf{b} = (\mathbf{b}_1^\top, \mathbf{b}_2^\top, \ldots, \mathbf{b}_N^\top)^\top$ is an Nq-vector and $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1^\top, \boldsymbol{\epsilon}_2^\top, \ldots, \boldsymbol{\epsilon}_N^\top)^\top$ is $T \times 1$. Thus, $\mathbf{b} \sim \mathcal{N}_{Nq}(\mathbf{0}, I_N \otimes G)$, where $\otimes$ denotes the Kronecker product and $\boldsymbol{\epsilon} \sim \mathcal{N}_T(\mathbf{0}, \sigma^2 I_T)$; $\mathbf{b}$ and $\boldsymbol{\epsilon}$ are independent.

It is possible to write model (2) as

$$\mathbf{Y} = X\boldsymbol{\beta} + \mathbf{e}, \tag{3}$$

where $\mathbf{e} = Z\mathbf{b} + \boldsymbol{\epsilon}$. Hence, $\mathbf{e} \sim \mathcal{N}_T(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\omega}) = Z(I_N \otimes G)Z^\top + \sigma^2 I_T$, $\boldsymbol{\omega} = (\boldsymbol{\varrho}^\top, \sigma^2)^\top$ being an $(m+1) \times 1$ vector of unknown parameters. Hence, the log-likelihood function for model (3) can be expressed as

$$\ell(\boldsymbol{\beta}, \boldsymbol{\omega}; \mathbf{Y}) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}(\mathbf{Y} - X\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Y} - X\boldsymbol{\beta}), \tag{4}$$

where $|\cdot|$ denotes matrix determinant.

Let $\theta = (\boldsymbol{\psi}^\top, \boldsymbol{\varsigma}^\top, \boldsymbol{\omega}^\top)^\top$ be the $(n+m+1)$-vector of parameters, where $\boldsymbol{\psi} = (\beta_1, \beta_2, \ldots, \beta_p)^\top$ is the p-vector ($p \leq n$) containing the first p elements of $\boldsymbol{\beta}$ and $(\boldsymbol{\varsigma}^\top, \boldsymbol{\omega}^\top)^\top$ is the $(n-p+m+1) \times 1$ vector of nuisance parameters with $\boldsymbol{\varsigma} = (\beta_{p+1}, \beta_{p+2}, \ldots, \beta_n)^\top$. In what follows, we shall focus on fixed effects inference. In particular, we wish to test $\mathcal{H}_0 : \boldsymbol{\psi} = \boldsymbol{\psi}^{(0)}$ against $\mathcal{H}_1 : \boldsymbol{\psi} \neq \boldsymbol{\psi}^{(0)}$, where $\boldsymbol{\psi}^{(0)}$ is a given p-vector.

We follow Zucker et al. (2000) and use a reparameterization in which the nuisance ($(\boldsymbol{\varsigma}^\top, \boldsymbol{\omega}^\top)^\top$) and interest ($\boldsymbol{\psi}$) parameters are orthogonal. In particular, we transform $\theta = (\boldsymbol{\psi}^\top, \boldsymbol{\varsigma}^\top, \boldsymbol{\omega}^\top)^\top$ into $\vartheta = (\boldsymbol{\psi}^\top, \boldsymbol{\xi}^\top, \boldsymbol{\omega}^\top)^\top$, with

$$\boldsymbol{\xi} = \boldsymbol{\varsigma} + (\widetilde{X}_{n-p}^\top \boldsymbol{\Sigma}^{-1}\widetilde{X}_{n-p})^{-1}\widetilde{X}_{n-p}^\top \boldsymbol{\Sigma}^{-1}\widetilde{X}_p\boldsymbol{\psi}, \tag{5}$$

where $\widetilde{X}_p$ denotes the matrix formed out of the first p columns of X and $\widetilde{X}_{n-p}$ contains the remaining $(n-p)$ columns of X. It is easy to show that $\boldsymbol{\psi}$ is orthogonal to $\boldsymbol{\phi} = (\boldsymbol{\xi}^\top, \boldsymbol{\omega}^\top)^\top$, i.e., the expected values of $\partial^2\ell(\vartheta; \mathbf{Y})/\partial\boldsymbol{\psi}\partial\boldsymbol{\xi}^\top$ and $\partial^2\ell(\vartheta; \mathbf{Y})/\partial\boldsymbol{\psi}\partial\omega_j$, for $j = 1, 2, \ldots, m+1$, are matrices of zeros. By partitioning X as $(\widetilde{X}_p, \widetilde{X}_{n-p})$ and $\boldsymbol{\beta}$ as $(\boldsymbol{\psi}^\top, \boldsymbol{\varsigma}^\top)^\top$, we can write $X\boldsymbol{\beta} = \widetilde{X}_p\boldsymbol{\psi} + \widetilde{X}_{n-p}\boldsymbol{\varsigma}$. Using (5) we obtain

$$X\boldsymbol{\beta} = \widetilde{X}_p'\boldsymbol{\psi} + \widetilde{X}_{n-p}\boldsymbol{\xi},$$

where $\widetilde{X}'_p = [I_T - \widetilde{X}_{n-p}(\widetilde{X}^\top_{n-p}\boldsymbol{\Sigma}^{-1}\widetilde{X}_{n-p})^{-1}\widetilde{X}^\top_{n-p}\boldsymbol{\Sigma}^{-1}]\widetilde{X}_p$. It follows that the log-likelihood function in (4) can be written as

$$\ell = \ell(\vartheta; \mathbf{Y}) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\mathbf{z}^\top\boldsymbol{\Sigma}^{-1}\mathbf{z}, \tag{6}$$

where $\mathbf{z} = \mathbf{z}(\mathbf{Y}, X, \vartheta) = \mathbf{Y} - \widetilde{X}'_p\boldsymbol{\psi} - \widetilde{X}_{n-p}\boldsymbol{\xi}$.

3. Improved likelihood ratio tests

3.1. Bartlett correction

The profile likelihood function, which only involves the vector of parameters of interest, is defined as $\ell_p(\boldsymbol{\psi}) = \ell(\boldsymbol{\psi}, \widehat{\boldsymbol{\phi}}(\boldsymbol{\psi}))$, where $\widehat{\boldsymbol{\phi}}(\boldsymbol{\psi})$ is the maximum likelihood estimator of $\boldsymbol{\phi}$ for a fixed value of $\boldsymbol{\psi}$. The likelihood ratio statistic for testing $\mathcal{H}_0$ is

$$LR = LR(\boldsymbol{\psi}^{(0)}) = 2\left\{\ell_p(\widehat{\boldsymbol{\psi}}) - \ell_p(\boldsymbol{\psi}^{(0)})\right\},$$

where $\widehat{\boldsymbol{\psi}}$ denotes the maximum likelihood estimator of $\boldsymbol{\psi}$. Under the standard regularity conditions and under $\mathcal{H}_0$, LR converges in distribution to χ^2_p. This first order approximation may not work well in small samples, however. In order to achieve more accuracy, Bartlett (1937) proposed multiplying LR by a constant, $(1 + C/p)^{-1}$, thus obtaining what is now known as the Bartlett-corrected test statistic:

$$LR^* = \frac{LR}{1 + C/p},$$

where C is a constant of order n^{-1} chosen such that, under $\mathcal{H}_0$, $\mathrm{E}(LR^*) = p + O(n^{-3/2})$. In regular problems, and under the null hypothesis LR^* is χ^2_p distributed up to an error of order n^{-2}; see Barndorff-Nielsen and Hall (1988). A general expression for C in terms of log-likelihood cumulants up to the fourth order was obtained by Lawley (1956).

One of our goals is to obtain the Bartlett correction term C for testing $\mathcal{H}_0 : \psi = \psi^{(0)}$ against $\mathcal{H}_1 : \psi \neq \psi^{(0)}$ for mixed linear models. This is done in Appendix A using Lawley's results; see (A.1). For simplicity, here we only give the expression for C when the $\psi^{(0)} = \mathbf{0}$, which is common in practical applications:

$$C = \mathrm{tr}\left(D^{-1}\left(-\frac{1}{2}M + \frac{1}{4}P - \frac{1}{2}(\gamma + \nu)\tau^\top\right)\right), \tag{7}$$

where $\mathrm{tr}(\cdot)$ is the trace operator. Here, D, M and P are $(m+1)\times(m+1)$ matrices given by

$$D = \{(1/2)\,\mathrm{tr}(\dot{\boldsymbol{\Sigma}}^j\dot{\boldsymbol{\Sigma}}_k)\},$$
$$M = \{\mathrm{tr}((\widetilde{X}'^\top_p\boldsymbol{\Sigma}^{-1}\widetilde{X}'_p)^{-1}(\widetilde{X}'^\top_p\ddot{\boldsymbol{\Sigma}}^{jk}\widetilde{X}'_p + 2\dot{X}'^\top_k\dot{\boldsymbol{\Sigma}}^j\widetilde{X}'_p))\},$$
$$P = \{\mathrm{tr}((\widetilde{X}'^\top_p\dot{\boldsymbol{\Sigma}}^j\widetilde{X}'_p)(\widetilde{X}'^\top_p\boldsymbol{\Sigma}^{-1}\widetilde{X}'_p)^{-1}(\widetilde{X}'^\top_p\dot{\boldsymbol{\Sigma}}^k\widetilde{X}'_p)(\widetilde{X}'^\top_p\boldsymbol{\Sigma}^{-1}\widetilde{X}'_p)^{-1})\}$$

and τ, γ and ν are $(m+1)$-vectors whose jth elements are $\mathrm{tr}((\widetilde{X}'^\top_p\boldsymbol{\Sigma}^{-1}\widetilde{X}'_p)^{-1}(\widetilde{X}'^\top_p\dot{\boldsymbol{\Sigma}}^j\widetilde{X}'_p))$, $\mathrm{tr}(D^{-1}A^{(j)})$ and $\mathrm{tr}((\widetilde{X}^\top_{n-p}\boldsymbol{\Sigma}^{-1}\widetilde{X}_{n-p})^{-1}(\widetilde{X}^\top_{n-p}\dot{\boldsymbol{\Sigma}}^j\widetilde{X}_{n-p}))$, respectively. Note that we give the (j, k) element of each matrix. In our notation, $A^{(j)}$ is the $(m+1)\times(m+1)$ matrix given by

$$A^{(j)} = \{(1/2)\,\mathrm{tr}(\dot{\boldsymbol{\Sigma}}^l\ddot{\boldsymbol{\Sigma}}_{jk}) - (1/2)\,\mathrm{tr}(\dot{\boldsymbol{\Sigma}}^k\ddot{\boldsymbol{\Sigma}}_{jl}) - (1/2)\,\mathrm{tr}(\dot{\boldsymbol{\Sigma}}^j\ddot{\boldsymbol{\Sigma}}_{lk})\}.$$

Also, $\dot{\boldsymbol{\Sigma}}_j = \partial\boldsymbol{\Sigma}/\partial\boldsymbol{\omega}_j$, $\dot{\boldsymbol{\Sigma}}^j = \partial\boldsymbol{\Sigma}^{-1}/\partial\boldsymbol{\omega}_j = -\boldsymbol{\Sigma}^{-1}\dot{\boldsymbol{\Sigma}}_j\boldsymbol{\Sigma}^{-1}$, $\ddot{\boldsymbol{\Sigma}}_{jk} = \partial^2\boldsymbol{\Sigma}/\partial\boldsymbol{\omega}_j\partial\boldsymbol{\omega}_k$, $\ddot{\boldsymbol{\Sigma}}^{jk} = \partial^2\boldsymbol{\Sigma}^{-1}/\partial\boldsymbol{\omega}_j\partial\boldsymbol{\omega}_k = -2\dot{\boldsymbol{\Sigma}}^k\dot{\boldsymbol{\Sigma}}_j\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\ddot{\boldsymbol{\Sigma}}_{jk}\boldsymbol{\Sigma}^{-1}$ and $\dot{X}'_j = \partial\widetilde{X}'_p/\partial\boldsymbol{\omega}_j = -\widetilde{X}_{n-p}(\widetilde{X}^\top_{n-p}\boldsymbol{\Sigma}^{-1}\widetilde{X}_{n-p})^{-1}\widetilde{X}^\top_{n-p}\dot{\boldsymbol{\Sigma}}^j\widetilde{X}'_p$.

It is noteworthy that (7) generalizes the result in Zucker et al. (2000, Eq. (3)). Their expression is only valid when the parameter under test is scalar and the covariance matrix for the random effects has a linear structure and so does $\boldsymbol{\Sigma}$, i.e., $\boldsymbol{\Sigma} = \sum\boldsymbol{\omega}_jQ_j$, where Q_j are known matrices. Note that when $\boldsymbol{\Sigma}$ has a linear structure, we have $\dot{\boldsymbol{\Sigma}}_j = Q_j$, $\forall j$, $\ddot{\boldsymbol{\Sigma}}_{jk} = 0$, $\forall j, k$, and Eq. (7) becomes

$$C = \mathrm{tr}\left(D^{-1}\left(-\frac{1}{2}M + \frac{1}{4}P - \frac{1}{2}\nu\tau^\top\right)\right). \tag{8}$$

Additionally, when ψ is scalar, our expression (8) reduces to Eq. (3) in Zucker et al. (2000). Also, when the null hypothesis is $\mathcal{H}_0 : \boldsymbol{\beta} = \boldsymbol{\beta}^{(0)}$, (8) reduces to

$$C = \mathrm{tr}\left(D^{-1}\left(-\frac{1}{2}M_1 + \frac{1}{4}P_1\right)\right),$$

where

$$M_1 = \{\mathrm{tr}((X^\top \mathbf{\Sigma}^{-1} X)^{-1}(X^\top \ddot{\mathbf{\Sigma}}^{jk} X))\}$$

and

$$P_1 = \{\mathrm{tr}((X^\top \dot{\mathbf{\Sigma}}^{j} X)(X^\top \mathbf{\Sigma}^{-1} X)^{-1}(X^\top \dot{\mathbf{\Sigma}}^{k} X)(X^\top \mathbf{\Sigma}^{-1} X)^{-1})\}.$$

3.2. Cox–Reid profile likelihood adjustment

Cox and Reid (1987) proposed an adjustment to the profile likelihood function which can be used when the nuisance and interest parameters are orthogonal. The Cox–Reid adjusted profile log-likelihood function is given by

$$\ell_{pa}(\boldsymbol{\psi}) = \ell_p(\boldsymbol{\psi}) - \frac{1}{2}\log\left\{\left|-\ell_{\boldsymbol{\phi\phi}}\left(\widehat{\phi}(\boldsymbol{\psi})\right)\right|\right\},$$

where $\ell_{\boldsymbol{\phi\phi}}$ is the matrix of second derivatives of ℓ with respect to $\boldsymbol{\phi}$. The corresponding likelihood ratio test statistic is

$$LR_{CR}(\boldsymbol{\psi}^{(0)}) = 2\left\{\ell_{pa}(\widetilde{\boldsymbol{\psi}}) - \ell_{pa}(\boldsymbol{\psi}^{(0)})\right\},$$

where $\widetilde{\psi}$ is the maximizer of $\ell_{pa}(\psi)$.

The Cox–Reid test statistic is χ^2_p distributed under $\mathcal{H}_0$ up to an error of order n^{-1}, just like the standard likelihood ratio test statistic. DiCiccio and Stern (1994) defined a Bartlett correction to this test statistic which reduces the order of the approximation error to $O(n^{-2})$. The corrected test statistic is

$$LR^*_{CR} = \frac{LR_{CR}}{1 + C^*/p},$$

where C^* is a constant of order n^{-1} such that, under $\mathcal{H}_0$, $\mathrm{E}(LR^*_{CR}) = p + O(n^{-3/2})$. A general expression for C^* can be found in DiCiccio and Stern (1994, Eq. (25)). In Appendix B, we obtain C^* for testing $\mathcal{H}_0$ in mixed linear models; see (B.1). Here, we give the expression for C^* for the case where $\boldsymbol{\psi}^{(0)} = \mathbf{0}$:

$$C^* = \mathrm{tr}\left(D^{-1}\left\{-M + \frac{1}{4}P + \gamma^*\tau^\top\right\}\right), \tag{9}$$

where D, M, P and τ were given above and the jth element of the vector γ^* is $\mathrm{tr}(D^{-1}C^{(j)})$, with $C^{(j)}$ being an $(m+1)\times(m+1)$ matrix given by

$$C^{(j)} = \{-\mathrm{tr}(\dot{\mathbf{\Sigma}}^{k}\dot{\mathbf{\Sigma}}_j\mathbf{\Sigma}^{-1}\dot{\mathbf{\Sigma}}_l) + (1/2)\ \mathrm{tr}(\dot{\mathbf{\Sigma}}^{j}\ddot{\mathbf{\Sigma}}_{kl}) + (1/2)\ \mathrm{tr}(\dot{\mathbf{\Sigma}}^{k}\ddot{\mathbf{\Sigma}}_{jl})\}.$$

Our expression for C^* generalizes the result in Zucker et al. (2000, Eq. (4)), since their formula is only valid when $p = 1$. We notice that their formula remains valid when the covariance matrix for the random effects has a nonlinear structure. As expected, (9) reduces to equation (4) of Zucker et al. (2000) when $p = 1$. Also, for testing $\mathcal{H}_0 : \boldsymbol{\beta} = \boldsymbol{\beta}^{(0)}$ against $\mathcal{H}_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}^{(0)}$, C^* reduces to (9) with M and P replaced by M_1 and P_1, respectively, and τ replaced by τ_1, with τ_1 being the $(m+1)$-vector whose jth element is $\mathrm{tr}((X^\top\mathbf{\Sigma}^{-1}X)^{-1}(X^\top\dot{\mathbf{\Sigma}}^{j}X))$.

The expressions we give for C and C^* in (7) and (9), respectively, only involve simple operations on vectors and matrices. Therefore, they can be easily computed with the aid of a programming language or software which can perform such operations, e.g. `Ox` (Cribari-Neto and Zarkos, 2003; Doornik, 2006) and `R` (Ihaka and Gentleman, 1996). We note that C and C^* only depend on X, on the inverse covariance matrix $\mathbf{\Sigma}^{-1}$, on the covariance matrix $\mathbf{\Sigma}$ and its first two derivatives with respect to ω.

4. Simulation study

In this section we shall present the results of Monte Carlo simulation experiments in which we evaluate the finite sample performances of the likelihood ratio test (LR), its Bartlett-corrected version (LR^*), the adjusted profile likelihood ratio test (LR_{CR}) and its Bartlett-corrected counterpart (LR^*_{CR}).

The simulations were based on the following mixed linear model:

$$\mathbf{y}_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 x_{1i} + \beta_3 x_{2i} + b_{0i} + b_{1i}t_{ij} + \boldsymbol{\epsilon}_{ij},$$

for $j = 1, 2, \ldots, \tau_i$ with $\tau_i \in \{2, 3, 4, 5, 6, 7, 8, 9\}$ and $i = 1, 2, \ldots, N$. The values of t_{ij} were obtained as random draws from the standard uniform distribution $\mathcal{U}(0, 1)$; x_{1i} and x_{2i} are dummy variables. The fixed effects parameters are $\beta_0, \beta_1, \beta_2, \beta_3$. Also, $\mathbf{b}_i = (b_{0i} \quad b_{1i})^\top \sim \mathcal{N}_2(\mathbf{0}, G)$ with

$$G = \begin{bmatrix} \omega_1 & \omega_2 \\ \omega_2 & \omega_3 \end{bmatrix}. \tag{10}$$

Table 1
Null rejection rates of the tests of $\mathcal{H}_0 : \psi = \mathbf{0}$; entries are percentages.

N	ω_2	ω_3	$\alpha = 5\%$				$\alpha = 10\%$			
			LR	LR^*	LR_{CR}	LR^*_{CR}	LR	LR^*	LR_{CR}	LR^*_{CR}
12	0	0.50	13.0	7.6	4.5	5.3	20.8	13.1	9.2	10.2
	0	1	13.4	7.8	4.8	5.9	21.7	13.5	9.6	10.8
	0.25	0.50	11.2	6.0	3.4	4.1	19.0	11.2	7.5	8.5
	0.25	1	13.8	7.9	5.1	5.8	21.9	13.9	9.6	10.7
24	0	0.50	8.3	5.6	4.7	5.0	14.6	10.9	9.5	10.0
	0	1	8.5	5.8	4.9	5.1	14.6	11.1	10.1	10.5
	0.25	0.50	8.6	5.7	4.8	5.1	14.8	11.1	9.6	10.2
	0.25	1	8.7	6.0	4.8	5.1	15.0	11.4	10.1	10.6
36	0	0.50	6.4	4.6	4.2	4.4	12.8	10.1	9.5	9.8
	0	1	6.1	4.9	4.4	4.7	12.6	9.8	9.0	9.4
	0.25	0.50	6.7	4.8	4.3	4.6	12.4	10.0	9.3	9.6
	0.25	1	6.4	4.7	4.3	4.4	12.6	9.8	9.1	9.4

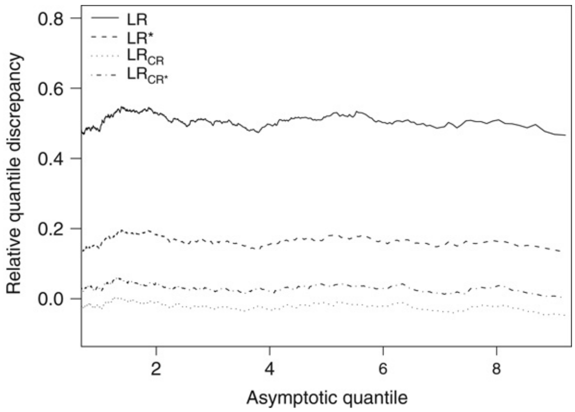

Fig. 1. Relative quantile discrepancies plot: $N = 12$, $\omega_2 = 0$ and $\omega_3 = 0.50$.

Additionally, the ϵ_{ij}'s are independent from the b_i's, and $\epsilon_i \sim \mathcal{N}_{\tau_i}(0, \omega_4 I_{\tau_i})$. We test $\mathcal{H}_0 : \psi = \mathbf{0}$ against $\mathcal{H}_1 : \psi \neq \mathbf{0}$, where $\psi = (\beta_2 \quad \beta_3)^\top$.

All simulations were performed using the Ox matrix programming language (Cribari-Neto and Zarkos, 2003; Doornik, 2006). The number of Monte Carlo replications was 5000 and the sample sizes considered were $N = 12, 24$ and 36. The parameter values are $\beta_0 = 0$, $\beta_1 = 0.2$, $\beta_2 = 0$, $\beta_3 = 0$, $\omega_1 = 1$, $\omega_2 = 0$ and 0.25, $\omega_3 = 0.5$ and 1, and $\omega_4 = 0.05$. All tests were carried out at the following nominal levels: $\alpha = 5\%$ and $\alpha = 10\%$.

The null rejection rates of the four tests under evaluation are displayed in Table 1. We note that the likelihood ratio test is liberal. For instance, when $\omega_2 = 0$, $\omega_3 = 0.50$, $N = 12$ and $\alpha = 10\%$, its rejection rate exceeds 20%. It is noteworthy that the three alternative tests outperform the standard likelihood ratio test. For $N = 12$ and $N = 24$, the two best performing tests are LR_{CR} and LR^*_{CR}; LR^* is slightly oversized. For example, when $\omega_2 = 0$, $\omega_3 = 0.50$, $N = 12$ and $\alpha = 5\%$, the null rejection rates of LR_{CR}, LR^*_{CR} and LR^* are, respectively, 4.5%, 5.3% and 7.6% (LR: 13.0%). It is not possible to single out a global winner between LR_{CR} and LR^*_{CR}. When $N = 36$, the Cox–Reid and the two Bartlett-corrected tests still outperform LR; here, LR^* slightly outperforms the other two alternative tests, LR^*_{CR} being the second best performing test.

Fig. 1 plots the relative quantile discrepancies against the asymptotic quantiles for $N = 12$, the smallest sample size, where the corrections are mostly needed. Relative quantile discrepancies are defined as differences between exact and asymptotic (χ^2_2) quantiles divided by the latter. The closer to zero these discrepancies, the better the approximation used in the test. We note that the test statistics with the smallest relative quantile discrepancies are LR_{CR} and LR^*_{CR}. We also note that quantiles of LR are approximately 50% larger than the respective asymptotic (χ^2_2) quantiles.

Note that the simulated model and the hypothesis under test have practical applications, for instance, when the practitioner wishes to compare two different treatments and the experimental units are observed in different points in time. Here, we assume that the time horizon of the study is limited. This is why we used a bounded distribution for choosing values for t_{ij}. We performed simulations under other situations. We varied the values of all the parameters and considered a gamma distribution with mean 3 and variance 1.5 for choosing values for t_{ij}. Also, we considered an extended model in which interactions between t_{ij} and the dummy variables were included. In this case, we tested the interactions effects. For the sake of brevity, the results are not shown. In short, the Cox–Reid and the two Bartlett-corrected tests outperformed LR. For instance, our simulation experiment with $\beta_0 = 0.2$, $\beta_1 = 0.4$, $\beta_2 = \beta_3 = 0$, $\omega_1 = 1.5$, $\omega_2 = 0.05$, $\omega_3 = 1.2$, $\omega_4 = 0.10$ and $N = 24$ yielded the following null rejection rates at the 10% nominal level: 14.7% (LR), 10.6% (LR^*), 9.4% (LR_{CR}) and

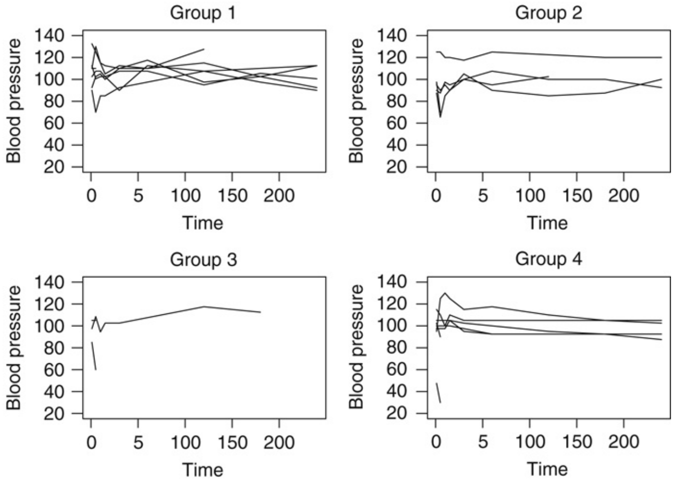

Fig. 2. Blood pressure against time for each rat.

9.8% (LR^*_{CR}). Also, for an extended model which includes two parameters, β_4 and β_5, representing interactions between t_{ij} and the dummy variables, we obtained 7.1% (LR), 5.4% (LR^*), 3.3% (LR_{CR}) and 5.0% (LR^*_{CR}) for $\alpha = 5\%$ and $N = 24$. Here, $\beta_0 = 0.2$, $\beta_1 = 0.4$, $\beta_2 = 0.3$, $\beta_3 = 0.5$, $\beta_4 = \beta_5 = 0$ and the same values for $\omega_1, \ldots, \omega_4$ as before.

5. Blood pressure data

We shall now present an application that uses a real data set. The data consist of a randomly selected subset of the data used by Crepeau et al. (1985). Heart attacks were induced in rats exposed to four different low concentrations of halothane; group 1: 0% (control), group 2: 0.25%, group 3: 0.50% and group 4: 1.0%. Our sample consists of 23 rats. The blood pressure of each rat (in mm Hg) is recorded over different points in time, from 1 to 9 recordings, after the induced heart attack. The main goal is to investigate the effect of halothane on the blood pressure.

Fig. 2 shows plots of blood pressure versus time for each rat. Clearly, the profiles differ on the intercept. However, the slopes are not markedly different. At the outset, we consider a model where blood pressure varies linearly with time, possibly with different intercepts and slopes for each concentration of halothane, and with intercept and slope random effects to account for animal-to-animal variation. As we will see later, the usual likelihood ratio test rejects the null hypothesis of common slope at the 10% nominal level, unlike the modified tests.

The mixed linear model considered here is

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \gamma_{02} G_{2i} + \gamma_{03} G_{3i} + \gamma_{04} G_{4i} + \gamma_{12} G_{2i} t_{ij} + \gamma_{13} G_{3i} t_{ij} + \gamma_{14} G_{4i} t_{ij} + b_{0i} + b_{1i} t_{ij} + \epsilon_{ij}, \tag{11}$$

with $i = 1, 2, \ldots, 23$ and $j = 1, 2, \ldots, \tau_i$, where y_{ij} is the blood pressure of the ith rat at time j, t_{ij} is the jth point in time (in minutes) in which the ith rat blood pressure was recorded, and G_{2i} is a dummy variable that equals 1 if the ith rat belongs to group 2 and 0 otherwise. Also, G_{3i} and G_{4i} equal 1 for groups 3 and 4, respectively. We assume that $b_i = (b_{0i} \quad b_{1i})^\top \overset{\text{i.i.d.}}{\sim} \mathcal{N}_2(\mathbf{0}, G)$, where G is given in (10). Additionally, $\epsilon_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \omega_4)$, the ϵ_{ij}'s being independent of the b_i's.

The maximum likelihood estimates of the fixed effects parameters are $\widehat{\beta}_0 = 104.360$, $\widehat{\beta}_1 = 0.004$, $\widehat{\gamma}_{02} = -0.719$, $\widehat{\gamma}_{03} = 0.203$, $\widehat{\gamma}_{04} = -15.211$, $\widehat{\gamma}_{12} = 0.022$, $\widehat{\gamma}_{13} = 0.109$ and $\widehat{\gamma}_{14} = -0.019$. We wish to make inference on γ_{12}, γ_{13} and γ_{14}. More specifically, we wish to test $\mathcal{H}_0 : \psi = \mathbf{0}$ against $\mathcal{H}_1 : \psi \neq \mathbf{0}$, where $\psi = (\gamma_{12}, \gamma_{13}, \gamma_{14})^\top$. Note that under the null hypothesis, the mean slopes are equal for the different halothane concentrations. The adjusted profile maximum likelihood estimates of γ_{12}, γ_{13} and γ_{14} are $\widetilde{\gamma}_{12} = 0.020$, $\widetilde{\gamma}_{13} = 0.101$ and $\widetilde{\gamma}_{14} = -0.030$, respectively. The test statistics assume the following values: $LR = 6.522$ (p-value: 0.089), $LR^* = 5.678$ (p-value: 0.128), $LR_a = 5.287$ (p-value: 0.152) and $LR^*_{CR} = 6.168$ (p-value: 0.104). The standard likelihood ratio test rejects the null hypothesis at the 10% nominal level, i.e., it suggests that there are differences in mean slopes for different dosages. The three modified tests, however, suggest otherwise, i.e., the null hypothesis is not rejected by these tests at the same nominal level.

We now consider the following reduced model:

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \gamma_{02} G_{2i} + \gamma_{03} G_{3i} + \gamma_{04} G_{4i} + b_{0i} + b_{1i} t_{ij} + \epsilon_{ij},$$

with $i = 1, 2, \ldots, 23$ and $j = 1, 2, \ldots, \tau_i$. We wish to test $\mathcal{H}_0^* : \psi^* = \mathbf{0}$ against $\mathcal{H}_1^* : \psi^* \neq \mathbf{0}$, where $\psi^* = (\gamma_{02}, \gamma_{03}, \gamma_{04})^\top$. Note that we are testing whether the mean blood pressures are equal across the different dosages. The fixed effects maximum likelihood estimates are $\widehat{\beta}_0 = 99.531$, $\widehat{\beta}_1 = 0.006$, $\widehat{\gamma}_{02} = -0.525$, $\widehat{\gamma}_{03} = 2.318$ and $\widehat{\gamma}_{04} = -13.357$. The adjusted profile maximum likelihood estimates of γ_{02}, γ_{03} and γ_{04} are, respectively, $\widetilde{\gamma}_{02} = -0.823$, $\widetilde{\gamma}_{03} = 2.079$ and $\widetilde{\gamma}_{04} = -12.573$.

We now have $LR = 6.143$ (p-value: 0.105), $LR^* = 5.174$ (p-value: 0.159), $LR_a = 4.002$ (p-value: 0.261) and $LR^*_{CR} = 4.167$ (p-value: 0.244). All tests yield the same inference, namely: the null hypothesis is not rejected at the 10% nominal level.

Therefore, we conclude that there is no group effect. In other words, the analysis carried out using the modified tests suggests that the blood pressure is not affected by the administration of halothane at the concentrations considered in the experiment. This conclusion agrees with the findings of Crepeau et al. (1985).

6. Concluding remarks

We addressed the issue of performing likelihood-based testing inference on the fixed effects parameters of mixed linear models when the sample contains a small number of observations. The standard likelihood ratio test is liberal, as evidenced by our Monte Carlo results. We obtained three alternative tests, namely: an adjusted profile likelihood ratio test, its Bartlett-corrected version and also the Bartlett-corrected likelihood ratio test. Our results generalize those in Zucker et al. (2000) in two directions. First, we allow practitioners to test joint restrictions on one or more fixed effects parameters, whereas their results only hold for tests on a parameter at a time. Second, unlike Zucker et al. (2000), we do not assume that the covariance matrix of the random effects is linear when deriving the Bartlett correction to the profile likelihood ratio test. Our main results are stated through closed-form formulas that only involve simple operations on vectors and matrices, and hence they can be easily implemented in matrix programming languages and statistical software. The simulation study we report clearly show that the proposed tests outperform the standard likelihood ratio test, especially when the sample size is small. It shows that the three alternative tests yield reliable inferences even for unbalanced data. In particular, the adjusted profile likelihood ratio test and its Bartlett-corrected version improve the type I error rate, especially when the number of observations is small.

Acknowledgments

We gratefully acknowledge financial support from FAPESP and CNPq. We also thank three anonymous referees for helpful suggestions.

Appendix A. Derivation of C

We use the following tensor notation for log-likelihood cumulants:

$$\kappa_{rs} = \mathrm{E}\left(\frac{\partial^2 \ell}{\partial \boldsymbol{\vartheta}_r \partial \boldsymbol{\vartheta}_s}\right), \qquad \kappa_{rst} = \mathrm{E}\left(\frac{\partial^3 \ell}{\partial \boldsymbol{\vartheta}_r \partial \boldsymbol{\vartheta}_s \partial \boldsymbol{\vartheta}_t}\right) \quad \text{and} \quad \kappa_{rstu} = \mathrm{E}\left(\frac{\partial^4 \ell}{\partial \boldsymbol{\vartheta}_r \partial \boldsymbol{\vartheta}_s \partial \boldsymbol{\vartheta}_t \partial \boldsymbol{\vartheta}_u}\right),$$

$\boldsymbol{\vartheta}_r$ being the rth element of ϑ. The notation used for derivatives of cumulants is the following:

$$(\kappa_{rs})_t = \frac{\partial \kappa_{rs}}{\partial \boldsymbol{\vartheta}_t}, \qquad (\kappa_{rst})_u = \frac{\partial \kappa_{rst}}{\partial \boldsymbol{\vartheta}_u} \quad \text{and} \quad (\kappa_{rs})_{tu} = \frac{\partial \kappa_{rs}}{\partial \boldsymbol{\vartheta}_t \partial \boldsymbol{\vartheta}_u}.$$

In what follows, we shall use similar notation for derivatives of matrices formed out of cumulants. Note that $-\kappa_{rs}$ is the (r, s) element of Fisher's information matrix; the (r, s) element of its inverse is denoted by $-\kappa^{rs}$.

Lawley's (1956) formula for C is

$$C = \sum_{\psi,\xi,\omega} (l_{rstu} - l_{rstuvw}) - \sum_{\xi,\omega} (l_{rstu} - l_{rstuvw}) = C_1 - C_2,$$

where $C_1 = \sum_{\psi,\xi,\omega} l_{rstu} - \sum_{\xi,\omega} l_{rstu}$ and $C_2 = \sum_{\psi,\xi,\omega} l_{rstuvw} - \sum_{\xi,\omega} l_{rstuvw}$ with

$$l_{rstu} = \kappa^{rs}\kappa^{tu}\left\{\frac{1}{4}\kappa_{rstu} - (\kappa_{rst})_u - (\kappa_{rt})_{su}\right\}$$

and

$$l_{rstuvw} = \kappa^{rs}\kappa^{tu}\kappa^{vw}\left\{\kappa_{rtv}\left(\frac{1}{6}\kappa_{suw} - (\kappa_{sw})_u\right) + \kappa_{rtu}\left(\frac{1}{4}\kappa_{svw} - (\kappa_{sw})_v\right) + (\kappa_{rt})_v\,(\kappa_{sw})_u + (\kappa_{rt})_u\,(\kappa_{sw})_v\right\},$$

where the indices r, s, t, u, v, w refer to the components of $\vartheta = (\boldsymbol{\psi}^\top, \boldsymbol{\xi}^\top, \boldsymbol{\omega}^\top)^\top$. Here, $\sum_{\psi,\xi,\omega}$ denotes summation over all possible combinations of the $n + m + 1$ parameters in ϑ, and $\sum_{\xi,\omega}$ denotes summation over the combinations of the $n - p + m + 1$ parameters in $(\xi^\top, \omega^\top)^\top$. We use indices a, b, c, d in reference to the components of ψ, indices f, g for the components of ξ, and indices j, k, l, o for the elements of ω. Further notation used here is given in Sections 2 and 3.

The first-order derivatives of the log-likelihood function in (6) are

$$\frac{\partial \ell(\vartheta; \mathbf{Y})}{\partial \boldsymbol{\psi}} = \widetilde{X}_p'^\top \boldsymbol{\Sigma}^{-1}\mathbf{z}, \qquad \frac{\partial \ell(\vartheta; \mathbf{Y})}{\partial \boldsymbol{\xi}} = \widetilde{X}_{n-p}^\top \boldsymbol{\Sigma}^{-1}\mathbf{z},$$

$$\frac{\partial \ell(\vartheta; \mathbf{Y})}{\partial \boldsymbol{\omega}_j} = -\frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\dot{\boldsymbol{\Sigma}}_j) - \frac{1}{2}\mathbf{z}^\top \dot{\boldsymbol{\Sigma}}^j \mathbf{z} + \boldsymbol{\psi}^\top \dot{X}_j'^\top \boldsymbol{\Sigma}^{-1}\mathbf{z}.$$

The second-order derivatives are

$$\frac{\partial^2\ell(\vartheta;\mathbf{Y})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\top}=-\widetilde{X}_p'^\top\boldsymbol{\Sigma}^{-1}\widetilde{X}_p',\qquad \frac{\partial^2\ell(\vartheta;\mathbf{Y})}{\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}^\top}=-\widetilde{X}_{n-p}^\top\boldsymbol{\Sigma}^{-1}\widetilde{X}_{n-p},\qquad \frac{\partial^2\ell(\vartheta;\mathbf{Y})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\xi}^\top}=\mathbf{0},$$

$$\frac{\partial^2\ell(\vartheta;\mathbf{Y})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\omega}_j}=(\dot{X}_j'^\top\boldsymbol{\Sigma}^{-1}+\widetilde{X}_p'^\top\dot{\boldsymbol{\Sigma}}^j)\mathbf{z},\qquad \frac{\partial^2\ell(\vartheta;\mathbf{Y})}{\partial\boldsymbol{\xi}\partial\boldsymbol{\omega}_j}=\widetilde{X}_{n-p}^\top\dot{\boldsymbol{\Sigma}}^j(\mathbf{Y}-\widetilde{X}_{n-p}\boldsymbol{\xi}),$$

$$\frac{\partial^2\ell(\vartheta;\mathbf{Y})}{\partial\boldsymbol{\omega}_j\partial\boldsymbol{\omega}_k}=-\frac{1}{2}\operatorname{tr}(\dot{\boldsymbol{\Sigma}}^j\dot{\boldsymbol{\Sigma}}_k)-\frac{1}{2}\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\ddot{\boldsymbol{\Sigma}}_{jk})-\boldsymbol{\psi}^\top\dot{X}_k'^\top\boldsymbol{\Sigma}^{-1}\dot{X}_j'\boldsymbol{\psi}+\boldsymbol{\psi}^\top(\ddot{X}_{jk}'^\top\boldsymbol{\Sigma}^{-1}+\dot{X}_k'^\top\dot{\boldsymbol{\Sigma}}^j+\dot{X}_j'^\top\dot{\boldsymbol{\Sigma}}^k)\mathbf{z}-\frac{1}{2}\mathbf{z}^\top\ddot{\boldsymbol{\Sigma}}^{jk}\mathbf{z},$$

where

$$\ddot{X}_{jk}'=\frac{\partial\dot{X}_j'}{\partial\boldsymbol{\omega}_k}=2\widetilde{X}_{n-p}(\widetilde{X}_{n-p}^\top\boldsymbol{\Sigma}^{-1}\widetilde{X}_{n-p})^{-1}\widetilde{X}_{n-p}^\top\dot{\boldsymbol{\Sigma}}^k\widetilde{X}_{n-p}(\widetilde{X}_{n-p}^\top\boldsymbol{\Sigma}^{-1}\widetilde{X}_{n-p})^{-1}\widetilde{X}_{n-p}^\top\dot{\boldsymbol{\Sigma}}^j\widetilde{X}_p'-\widetilde{X}_{n-p}(\widetilde{X}_{n-p}^\top\boldsymbol{\Sigma}^{-1}\widetilde{X}_{n-p})^{-1}\widetilde{X}_{n-p}^\top\ddot{\boldsymbol{\Sigma}}^{jk}\widetilde{X}_p'.$$

Additionally, the third-order derivatives are

$$\frac{\partial^3\ell(\vartheta;\mathbf{Y})}{\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}^\top\partial\boldsymbol{\omega}_j}=-\widetilde{X}_{n-p}^\top\dot{\boldsymbol{\Sigma}}^j\widetilde{X}_{n-p},\qquad \frac{\partial^3\ell(\vartheta;\mathbf{Y})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\top\partial\boldsymbol{\omega}_j}=-\widetilde{X}_p'^\top\dot{\boldsymbol{\Sigma}}^j\widetilde{X}_p',$$

$$\frac{\partial^3\ell(\vartheta;\mathbf{Y})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\xi}^\top\partial\boldsymbol{\xi}_f}=\frac{\partial^3\ell(\vartheta;\mathbf{Y})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\xi}^\top\partial\boldsymbol{\omega}_j}=\mathbf{0},\qquad \frac{\partial^3\ell(\vartheta;\mathbf{Y})}{\partial\boldsymbol{\xi}\partial\boldsymbol{\omega}_j\partial\boldsymbol{\omega}_k}=\widetilde{X}_{n-p}^\top\ddot{\boldsymbol{\Sigma}}^{jk}(\mathbf{Y}-\widetilde{X}_{n-p}\boldsymbol{\xi}),$$

$$\frac{\partial^3\ell(\vartheta;\mathbf{Y})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\omega}_j\partial\boldsymbol{\omega}_k}=(\ddot{X}_{jk}'^\top\boldsymbol{\Sigma}^{-1}+\dot{X}_k'^\top\dot{\boldsymbol{\Sigma}}^j+\dot{X}_j'^\top\dot{\boldsymbol{\Sigma}}^k+\widetilde{X}_p'^\top\ddot{\boldsymbol{\Sigma}}^{jk})\mathbf{z},$$

$$\begin{aligned}\frac{\partial^3\ell(\vartheta;\mathbf{Y})}{\partial\boldsymbol{\omega}_j\partial\boldsymbol{\omega}_k\partial\boldsymbol{\omega}_l}\ =\ &-\frac{1}{2}(\operatorname{tr}(\ddot{\boldsymbol{\Sigma}}^{lk}\dot{\boldsymbol{\Sigma}}_j)+\operatorname{tr}(\dot{\boldsymbol{\Sigma}}^k\ddot{\boldsymbol{\Sigma}}_{lj})+\operatorname{tr}(\dot{\boldsymbol{\Sigma}}^l\ddot{\boldsymbol{\Sigma}}_{jk})+\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\ddot{\boldsymbol{\Sigma}}_{jkl})+\mathbf{z}^\top\ddot{\boldsymbol{\Sigma}}_{jkl}\mathbf{z})\\&+\boldsymbol{\psi}^\top(\ddot{X}_{lk}'^\top\dot{\boldsymbol{\Sigma}}^j+\dot{X}_k'^\top\ddot{\boldsymbol{\Sigma}}^{lj}+\ddot{X}_{jk}'^\top\dot{\boldsymbol{\Sigma}}^l+\ddot{X}_{lj}'^\top\dot{\boldsymbol{\Sigma}}^k+\dot{X}_j'^\top\ddot{\boldsymbol{\Sigma}}^{lk}+\dot{X}_l'^\top\ddot{\boldsymbol{\Sigma}}^{kj}+\ddot{X}_{jkl}'^\top\boldsymbol{\Sigma}^{-1})\mathbf{z},\end{aligned}$$

where $\ddot{\boldsymbol{\Sigma}}_{jkl}=\partial\ddot{\boldsymbol{\Sigma}}_{jk}/\partial\boldsymbol{\omega}_l$ and $\ddot{X}_{jkl}'=\partial\ddot{X}_{jk}'/\partial\boldsymbol{\omega}_l$. Finally, the fourth-order derivatives can be shown to be

$$\frac{\partial^4\ell(\vartheta;\mathbf{Y})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\top\partial\boldsymbol{\omega}_j\partial\boldsymbol{\omega}_k}=-2\dot{X}_k'^\top\dot{\boldsymbol{\Sigma}}^j\widetilde{X}_p'-\widetilde{X}_p'^\top\ddot{\boldsymbol{\Sigma}}^{jk}\widetilde{X}_p',$$

$$\frac{\partial^4\ell(\vartheta;\mathbf{Y})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\top\partial\boldsymbol{\xi}_f\partial\boldsymbol{\xi}_g}=\frac{\partial^4\ell(\vartheta;\mathbf{Y})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\top\partial\boldsymbol{\xi}_f\partial\boldsymbol{\omega}_j}=\frac{\partial^3\ell(\vartheta;\mathbf{Y})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\top\partial\boldsymbol{\xi}_f}=\mathbf{0}.$$

Taking expected values of second, third and fourth derivatives, we obtain

$$K_{\boldsymbol{\psi}\boldsymbol{\psi}}=\mathrm{E}\left(\frac{\partial^2\ell(\vartheta;\mathbf{Y})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\top}\right)=-\widetilde{X}_p'^\top\boldsymbol{\Sigma}^{-1}\widetilde{X}_p',$$

$$K_{\boldsymbol{\xi}\boldsymbol{\xi}\boldsymbol{\omega}_j}=\mathrm{E}\left(\frac{\partial^3\ell(\vartheta;\mathbf{Y})}{\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}^\top\partial\boldsymbol{\omega}_j}\right)=-\widetilde{X}_{n-p}^\top\dot{\boldsymbol{\Sigma}}^j\widetilde{X}_{n-p},$$

$$K_{\boldsymbol{\psi}\boldsymbol{\psi}\boldsymbol{\omega}_j\boldsymbol{\omega}_k}=\mathrm{E}\left(\frac{\partial^4\ell(\vartheta;\mathbf{Y})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\top\partial\boldsymbol{\omega}_j\partial\boldsymbol{\omega}_k}\right)=-2\dot{X}_k'^\top\dot{\boldsymbol{\Sigma}}^j\widetilde{X}_p'-\widetilde{X}_p'^\top\ddot{\boldsymbol{\Sigma}}^{jk}\widetilde{X}_p'.$$

In similar fashion, it follows that

$$K_{\boldsymbol{\xi}\boldsymbol{\xi}}=-\widetilde{X}_{n-p}^\top\boldsymbol{\Sigma}^{-1}\widetilde{X}_{n-p},\qquad K_{\boldsymbol{\xi}\boldsymbol{\omega}_j}=\widetilde{X}_{n-p}^\top\dot{\boldsymbol{\Sigma}}^j\widetilde{X}_p'\boldsymbol{\psi},\qquad K_{\boldsymbol{\psi}\boldsymbol{\omega}_j}=\mathbf{0},$$

$$K_{\boldsymbol{\psi}\boldsymbol{\psi}\boldsymbol{\omega}_j}=-\widetilde{X}_p'^\top\dot{\boldsymbol{\Sigma}}^j\widetilde{X}_p',\qquad K_{\boldsymbol{\psi}\boldsymbol{\xi}\boldsymbol{\xi}_f}=K_{\boldsymbol{\psi}\boldsymbol{\xi}\boldsymbol{\omega}_j}=\mathbf{0},$$

$$K_{\boldsymbol{\psi}\boldsymbol{\omega}_j\boldsymbol{\omega}_k}=\mathbf{0},\qquad K_{\boldsymbol{\xi}\boldsymbol{\omega}_j\boldsymbol{\omega}_k}=\widetilde{X}_{n-p}^\top\ddot{\boldsymbol{\Sigma}}^{jk}\widetilde{X}_p'\boldsymbol{\psi},\qquad K_{\boldsymbol{\psi}\boldsymbol{\psi}\boldsymbol{\xi}_f\boldsymbol{\xi}_g}=K_{\boldsymbol{\psi}\boldsymbol{\psi}\boldsymbol{\xi}_f\boldsymbol{\omega}_j}=\mathbf{0}.$$

Additionally,

$$\kappa_{jk}=\frac{1}{2}\operatorname{tr}(\dot{\boldsymbol{\Sigma}}^j\dot{\boldsymbol{\Sigma}}_k)-\boldsymbol{\psi}^\top\dot{X}_j'^\top\boldsymbol{\Sigma}^{-1}\dot{X}_k'\boldsymbol{\psi},$$

$$\kappa_{ljk}=-2\operatorname{tr}(\dot{\boldsymbol{\Sigma}}^l\dot{\boldsymbol{\Sigma}}_k\boldsymbol{\Sigma}^{-1}\dot{\boldsymbol{\Sigma}}_j)+\frac{1}{2}\operatorname{tr}(\dot{\boldsymbol{\Sigma}}^j\ddot{\boldsymbol{\Sigma}}_{lk})+\frac{1}{2}\operatorname{tr}(\dot{\boldsymbol{\Sigma}}^k\ddot{\boldsymbol{\Sigma}}_{lj})+\frac{1}{2}\operatorname{tr}(\dot{\boldsymbol{\Sigma}}^l\ddot{\boldsymbol{\Sigma}}_{jk}).$$

Consider the following matrices formed out of minus Fisher's information inverse: $K^{\psi\psi}=K_{\psi\psi}^{-1}$, $K^{\omega\omega}=(K_{\omega\omega}-K_{\xi\omega}^\top K_{\xi\xi}^{-1}K_{\xi\omega})^{-1}$, $K^{\xi\xi}=K_{\xi\xi}^{-1}+K_{\xi\xi}^{-1}K_{\xi\omega}K^{\omega\omega}K_{\xi\omega}^\top K_{\xi\xi}^{-1}$ and $K^{\xi\omega}=K^{\omega\xi^\top}=-K_{\xi\xi}^{-1}K_{\xi\omega}K^{\omega\omega\top}$, where the jth column of $K_{\xi\omega}$ is

$K_{\boldsymbol{\xi}\boldsymbol{\omega}_j}$ and the (j,k)th element of $K_{\omega\omega}$ is κ_{jk}. It can be shown that

$$\left(K_{\boldsymbol{\psi}\boldsymbol{\psi}}\right)_j = -\widetilde{X}_p'^{\top}\dot{\boldsymbol{\Sigma}}^j\widetilde{X}_p', \qquad \left(K_{\boldsymbol{\psi}\boldsymbol{\psi}}\right)_{jk} = -2\dot{X}_k'^{\top}\dot{\boldsymbol{\Sigma}}^j\widetilde{X}_p' - \widetilde{X}_p'^{\top}\ddot{\boldsymbol{\Sigma}}^{jk}\widetilde{X}_p',$$
$$\left(K_{\boldsymbol{\xi}\boldsymbol{\xi}}\right)_j = -\widetilde{X}_{n-p}^{\top}\dot{\boldsymbol{\Sigma}}^j\widetilde{X}_{n-p}, \qquad (K_{\boldsymbol{\xi}\boldsymbol{\omega}_j})_k = \widetilde{X}_{n-p}^{\top}\ddot{\boldsymbol{\Sigma}}^{jk}\widetilde{X}_p'\boldsymbol{\psi} + \widetilde{X}_{n-p}^{\top}\dot{\boldsymbol{\Sigma}}^j\dot{X}_k'\boldsymbol{\psi},$$
$$(\kappa_{jl})_k = -\mathrm{tr}(\dot{\boldsymbol{\Sigma}}^l\dot{\boldsymbol{\Sigma}}_j\boldsymbol{\Sigma}^{-1}\dot{\boldsymbol{\Sigma}}_k) + \frac{1}{2}\mathrm{tr}(\dot{\boldsymbol{\Sigma}}^j\ddot{\boldsymbol{\Sigma}}_{lk}) + \frac{1}{2}\mathrm{tr}(\dot{\boldsymbol{\Sigma}}^l\ddot{\boldsymbol{\Sigma}}_{jk}).$$

It follows from the orthogonality between ψ and $(\xi^{\top},\omega^{\top})^{\top}$ that $\kappa^{af} = \kappa^{aj} = (\kappa_{af})_{jb} = (\kappa_{aj})_{fb} = 0$. Also, $\kappa_{jfa} = \kappa_{jfab} = 0$. Hence,

$$C_1 = \sum\left(l_{abcd} + l_{abfg} + l_{abfj} + l_{abjf} + l_{abjk} + l_{fgab} + l_{jkab}\right),$$

where $\sum$ ranges over all parameter combinations induced by the indices a,b,c,d,f,g,j,k. It is possible to show that $l_{abcd} = l_{abfg} = l_{abfj} = l_{abjf} = l_{fgab} = 0$. Thus,

$$C_1 = \sum\left(l_{abjk} + l_{jkab}\right) = \sum\left\{\kappa^{ab}\kappa^{jk}\left(\frac{1}{4}\kappa_{abjk} - \left(\kappa_{abj}\right)_k\right) + \frac{1}{4}\kappa^{jk}\kappa^{ab}\kappa_{jkab}\right\}.$$

Since $\kappa_{abjk} = \left(\kappa_{abj}\right)_k = \kappa_{jkab}$, C_1 reduces to

$$C_1 = -\frac{1}{2}\sum\kappa^{ab}\kappa^{jk}\kappa_{abjk}.$$

As for C_2, we have that

$$\begin{aligned}
C_2 &= \sum(l_{abcdjk} + l_{abjkcd} + l_{jkloab} + l_{jkablo} + l_{jkabcd} + l_{fjgkab} + l_{fjkgab} + l_{fjklab} \\
&\quad + l_{jkfgab} + l_{jkflab} + l_{jklfab} + l_{jfabgk} + l_{jfabkg} + l_{jfabkl} + l_{jkabfg} + l_{jkabfl} + l_{jkablf}) \\
&= \sum\left\{-\frac{1}{4}\kappa^{ab}\kappa^{cd}\kappa^{jk}\kappa_{abj}\kappa_{cdk} + \frac{1}{2}\kappa^{ab}\kappa^{fg}\kappa^{jk}\kappa_{abj}\kappa_{fgk}\right. \\
&\quad \left. -\kappa^{ab}\kappa^{fj}\kappa^{kl}\kappa_{abj}\left(2(\kappa_{fk})_l - \frac{3}{2}\kappa_{fkl}\right) + \frac{1}{2}\kappa^{ab}\kappa^{jk}\kappa^{lo}\kappa_{abj}\left(\kappa_{klo} - 2(\kappa_{kl})_o\right)\right\}.
\end{aligned}$$

Therefore, C reduces to

$$\begin{aligned}
C &= \sum\left\{-\frac{1}{2}\kappa^{ab}\kappa^{jk}\kappa_{abjk} + \frac{1}{4}\kappa^{ab}\kappa^{cd}\kappa^{jk}\kappa_{abj}\kappa_{cdk} - \frac{1}{2}\kappa^{ab}\kappa^{jk}\kappa^{lo}\kappa_{abj}\left(\kappa_{lok} - 2(\kappa_{lo})_k\right)\right. \\
&\quad \left. +\kappa^{ab}\kappa^{fk}\kappa^{jl}\kappa_{abj}\left(2(\kappa_{fk})_l - \frac{3}{2}\kappa_{fkl}\right) - \frac{1}{2}\kappa^{ab}\kappa^{fg}\kappa^{jk}\kappa_{abj}\kappa_{fgk}\right\}.
\end{aligned}$$

We now arrive at the matrix expression given by

$$C = \mathrm{tr}\left(K^{\omega\omega}\left\{-\frac{1}{2}M + \frac{1}{4}P - \left(\frac{1}{2}\rho - \delta + \frac{1}{2}\eta\right)\tau^{\top}\right\}\right). \tag{A.1}$$

Here, ρ, δ and η are $(m+1)$-vectors whose jth elements are, respectively, $\mathrm{tr}(K^{\omega\omega}A^{(j)})$, $\mathrm{tr}(K^{\xi\omega^{\top}}B^{(j)})$ and $\mathrm{tr}(-K^{\xi\xi}(\widetilde{X}_{n-p}^{\top}\dot{\boldsymbol{\Sigma}}^j\widetilde{X}_{n-p}))$. In our notation, $B^{(j)}$ is a matrix that contains the $m+1$ column vectors $(1/2\widetilde{X}_{n-p}^{\top}\ddot{\boldsymbol{\Sigma}}^{jk}\widetilde{X}_p' + 2\widetilde{X}_{n-p}^{\top}\dot{\boldsymbol{\Sigma}}^j\dot{X}_k')\psi$ and $A^{(j)}$ is defined in Section 3.1. For testing $\mathcal{H}_0 : \psi = \mathbf{0}$, C reduces to Eq. (7).

Appendix B. Derivation of C^*

We shall now obtain C^*, which is used to Bartlett-correct the adjusted profile likelihood ratio test statistic. DiCiccio and Stern (1994, Eq. (25)) give the following general expression:

$$\begin{aligned}
C^* &= \sum_{\psi,\xi,\omega}\left\{\frac{1}{4}\tau^{ru}\tau^{st}\kappa_{rstu} - \kappa^{ru}\tau^{st}(\kappa_{rst})_u + \left(\kappa^{ru}\kappa^{st} - \nu^{ru}\nu^{st}\right)(\kappa_{rs})_{tu}\right. \\
&\quad - \left(\frac{1}{4}\kappa^{ru}\tau^{st}\tau^{vw} + \frac{1}{2}\kappa^{ru}\tau^{sw}\tau^{tv} - \frac{1}{3}\tau^{ru}\tau^{sw}\tau^{tv}\right)\kappa_{rst}\kappa_{uvw} + \left(\kappa^{ru}\tau^{st}\kappa^{vw} + \kappa^{ru}\kappa^{sw}\kappa^{tv} - \nu^{ru}\kappa^{sw}\nu^{tv}\right)\kappa_{rst}(\kappa_{uv})_w \\
&\quad \left. - \left(\kappa^{ru}\kappa^{st}\kappa^{vw} - \nu^{ru}\nu^{st}\nu^{vw}\right)(\kappa_{rs})_t(\kappa_{uv})_w - \left(\kappa^{ru}\kappa^{sw}\kappa^{tv} - \nu^{ru}\nu^{sw}\nu^{tv}\right)(\kappa_{rs})_t(\kappa_{uv})_w\right\},
\end{aligned}$$

where $\nu^{rs} = \kappa^{rs} - \tau^{rs}$, $\tau^{rs} = \kappa^{rb}\kappa^{sa}\sigma_{ab}$, σ_{ab} being the (a, b) element of the inverse of $K^{\psi\psi}$. From the orthogonality between ψ and ϕ we have that $\tau^{fg} = \tau^{jk} = \tau^{fj} = \tau^{af} = \tau^{aj} = 0$. Also, $\tau^{ab} = \kappa^{ab}$. Thus,

$$C^* = \sum \left\{ \frac{1}{4}\kappa^{ad}\kappa^{bc}\kappa_{abcd} - \kappa^{ru}\kappa^{ab}(\kappa_{rab})_u + \left(\kappa^{ru}\kappa^{st} - \nu^{ru}\nu^{st}\right)(\kappa_{rs})_{tu} - \left(\frac{1}{4}\kappa^{ru}\kappa^{ab}\kappa^{cd} + \frac{1}{2}\kappa^{ru}\kappa^{ad}\kappa^{bc}\right)\kappa_{rab}\kappa_{ucd} \right.$$
$$\left. + \kappa^{ru}\kappa^{ab}\kappa^{vw}\kappa_{rab}(\kappa_{uv})_w + \left(\kappa^{ru}\kappa^{tv} - \nu^{ru}\nu^{tv}\right)\kappa^{sw}\kappa_{rst}(\kappa_{uv})_w \right\}.$$

We have that $\kappa^{ru}\kappa^{tv} - \nu^{ru}\nu^{tv} = \kappa^{ru}\tau^{tv} + \kappa^{tv}\tau^{ru} - \tau^{ru}\tau^{tv}$ and $(\kappa_{bd})_k = \kappa_{bdk}$. Hence,

$$\sum \left(\kappa^{ru}\kappa^{tv} - \nu^{ru}\nu^{tv}\right)\kappa^{sw}\kappa_{rst}(\kappa_{uv})_w = \sum \kappa^{ab}\kappa^{cd}\kappa^{jk}\kappa_{ajc}\kappa_{bdk}.$$

Since $\kappa_{abcd} = \kappa_{abc} = \kappa_{fab} = (\kappa_{ac})_{bu} = (\kappa_{af})_{tu} = (\kappa_{aj})_{tu} = 0$, it follows that C^* reduces to

$$C^* = \sum \left\{ -\kappa^{ab}\kappa^{jk}\kappa_{abjk} + \frac{1}{4}\kappa^{ab}\kappa^{cd}\kappa^{jk}\kappa_{abj}\kappa_{cdk} + \kappa^{ab}\kappa^{jk}\kappa^{lo}\kappa_{abj}(\kappa_{kl})_o + \kappa^{ab}\kappa^{fj}\kappa^{gk}\kappa_{abj}\kappa_{fgk} + 2\kappa^{ab}\kappa^{fj}\kappa^{kl}\kappa_{abj}(\kappa_{fk})_l \right\}.$$

We then arrive at the matrix expression

$$C^* = \mathrm{tr}\left(K^{\omega\omega}\left\{-M + \frac{1}{4}P + (\rho^* + 2\delta^*)\tau^\top\right\}\right) + \tau^\top K^{\omega\xi}\eta^*, \tag{B.1}$$

where the jth elements of the vectors ρ^* and δ^* are, respectively, $\mathrm{tr}(K^{\omega\omega}C^{(j)})$ and $\mathrm{tr}(K^{\xi\omega\top}F^{(j)})$, and the fth element of the vector η^* is $\mathrm{tr}(K^{\omega\xi}G^{(f)})$. Also, $C^{(j)}$ is defined in Section 3.2, $F^{(j)}$ is a matrix that contains the $m+1$ column vectors $\left(\widetilde{X}_{n-p}^\top \ddot{\boldsymbol{\Sigma}}^{jk}\widetilde{X}_p' + \widetilde{X}_{n-p}^\top \dot{\boldsymbol{\Sigma}}^j \dot{X}_k'\right)\psi$ and $G^{(f)}$ is the $(n-p)\times(m+1)$ matrix whose jth column is the fth column of $-\widetilde{X}_{n-p}^\top \dot{\boldsymbol{\Sigma}}^j \widetilde{X}_{n-p}$. For testing $\mathcal{H}_0 : \psi = \mathbf{0}$, C^* reduces to Eq. (8).

References

Barndorff-Nielsen, O.E., Hall, P., 1988. On the level-error after Bartlett adjustment of the likelihood ratio statistic. Biometrika 75, 374–378.

Bartlett, M.S., 1937. Properties of sufficiency and statistical test. Proceeding of the Royal Society A 160, 268–282.

Brazzale, A.R., Davison, A., Reid, N., 2007. Applied Asymptotics: Case Studies in Small-Sample Statistics. Cambridge University Press.

Brown, H., Prescott, R., 2006. Applied Mixed Models in Medicine. Wiley, New York.

Cox, D.R., Reid, N., 1987. Parameter orthogonality and approximate conditional inference. Journal of the Royal Statistical Society B 49, 1–39.

Crepeau, H., Koziol, J., Reid, N., Yuh, Y.S., 1985. Analysis of incomplete multivariate data from repeated measurement experiments. Biometrics 41, 505–514.

Cribari-Neto, F., Zarkos, S.G., 2003. Econometric and Statistical Computing Using Ox. Computational Economics 21, 277–295.

Cysneiros, A.H.M.A., Ferrari, S.L.P., 2006. An improved likelihood ratio test for varying dispersion in exponential family nonlinear models. Statistics and Probability Letters 76, 255–265.

DiCiccio, T.J., Stern, S.E., 1994. Frequentist and Bayesian Bartlett correction of test statistics based on adjusted profile likelihoods. Journal of the Royal Statistical Society B 56, 397–408.

Doornik, J.A., 2006. An Object-oriented Matrix Programming Language — Ox 4. Timberlake Consultants, London.

Ferrari, S.L.P., Cysneiros, A.H.M.A., Cribari-Neto, F., 2004. An improved test for heteroskedasticity using adjusted modified profile likelihood inference. Journal of Statistical Planning and Inference 124, 423–437; Journal of Statistical Planning and Inference 131 207–208 (corrigendum).

Ferrari, S.L.P., Lucambio, F., Cribari–Neto, F., 2005. Improved profile likelihood inference. Journal of Statistical Planning and Inference 134, 373–391.

Ihaka, R., Gentleman, R., 1996. R: A language for data analysis and graphics. Journal of Computational Graphics and Statistics 5, 299–314.

Lawley, D.N., 1956. A general method for approximating to the distribution of the likelihood ratio criteria. Biometrika 71, 233–244.

Littel, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D., Schabenberger, O., 2006. SAS for Mixed Models. SAS Publishing.

Pinheiro, J.C., Bates, D.M., 2000. Mixed-Effects Models in *S* and *S*-PLUS. Springer, New York.

Sartori, N., 2003. Modified profile likelihood in models with stratum nuisance parameters. Biometrika 90, 533–549.

Sartori, N., Bellio, R., Salvan, A., Pace, L., 1999. The directed modified profile likelihood in models with many nuisance parameters. Biometrika 86, 735–742.

Severini, T.A., 2000. Likelihood Methods in Statistics. Oxford University Press, Oxford.

Verbeke, G., Molenberghs, G., 2000. Linear Mixed Models for Longitudinal Data. Springer, New York.

Zucker, D.M., Lieberman, O., Manor, O., 2000. Improved small sample inference in the mixed linear model: Bartlett correction and adjusted likelihood. Journal of the Royal Statistical Society B 62, 827–838.