International Statistical Review

International Statistical Review (2017), 85, 2, 290–324 doi:10.1111/insr.12178

Graphical Tools for Detecting Departures from Linear Mixed Model Assumptions and Some Remedial Measures

Julio M. Singer[1], Francisco M.M. Rocha[2] and Juvêncio S. Nobre[3]

[1]*Departamento de Estatística, Universidade de São Paulo, São Paulo, Brazil*
E-mail: jmsinger@ime.usp.br
[2]*Escola Paulista de Política, Economia e Negócios, Universidade Federal de São Paulo, São Paulo, Brazil*
E-mail: f.marcelo.rocha@gmail.com
[3]*Departamento de Estatística e Matemática Aplicada, Universidade Federal do Ceará, Fortaleza, Brazil*
E-mail: juvencio@ufc.br

Summary

We review some results on the analysis of longitudinal data or, more generally, of repeated measures via linear mixed models starting with some exploratory statistical tools that may be employed to specify a tentative model. We follow with a summary of inferential procedures under a Gaussian set-up and then discuss different diagnostic methods focusing on residual analysis but also addressing global and local influence. Based on the interpretation of diagnostic plots related to three types of residuals (marginal, conditional and predicted random effects) as well as on other tools, we proceed to identify remedial measures for possible violations of the proposed model assumptions, ranging from fine-tuning of the model to the use of elliptically symmetric or skew-elliptical linear mixed models as well as of robust estimation methods. We specify many results available in the literature in a unified notation and highlight those with greater practical appeal. In each case, we discuss the availability of model diagnostics as well as of software and give general guidelines for model selection. We conclude with analyses of three practical examples and suggest further directions for research.

Key words: Leverage; local influence; mixed models; random coefficients; random effects; residual analysis.

1 Introduction

Repeated measures and longitudinal data from diverse fields such as Agriculture, Biology, Economy, Genetics, Medicine and Sociology usually present some degree of correlation and regression models for their analysis have attracted the attention of investigators for a long time. In particular, we mention the seemingly unrelated regression models considered by Zellner (1962) and the growth curve models proposed by Potthoff and Roy (1964) as some of the first attempts to address this problem. The thrust in the development of appropriate statistical methodology designed to analyse this type of data probably stems from the seminal paper

by Laird and Ware (1982). In this context, a myriad of multivariate models (MM), linear mixed models (LMM), generalised linear mixed models (GLMM), generalised estimating equations (GEE)-based models, non-linear mixed models (NLMM) and transition models (TM) have been proposed in the literature. The variety of texts recently published on the subject clearly emphasises its importance. Vonesh and Chinchilli (1997), Singer and Andrade (2000), Verbeke and Molenberghs (2000), Diggle *et al.* (2002), Molenberghs and Verbeke (2005), Hedecker and Gibbons (2006), Fitzmaurice *et al.* (2008), Fitzmaurice *et al.* (2011) or Demidenko (2013), among many others, are good examples.

Gaussian linear mixed models are those most commonly used for such purposes, not only because of their simplicity but also because they may serve as approximations for models in other classes and are computationally easy to work with. They are extremely flexible and useful to model the covariance structure of dependent data and allow both unit-specific or population-averaged analyses as indicated in Verbeke and Molenberghs (2000), for example. Even working within this particular class of models, there are so many alternatives for data analysis that statisticians often have difficulties in selecting an appropriate approach for their applications. This is where model selection and diagnostic tools play an important role. Many alternatives have been developed for such purposes, but they are scattered in the vast literature on linear mixed models. An attempt to gather such tools and provide some guidance to practitioners is made in Loy and Hofman (2013). Our objective is to provide a broader review of the many results in this field, to specify them in an unified notation, to point to those with greater practical appeal and to propose some guidelines for the post-diagnostic analysis.

In Section 2, we introduce the linear mixed model and present essential estimation/prediction results under a Gauss–Markov set-up. In Section 3, we review the basic inferential results under a classical Gaussian set-up. Diagnostic procedures (residual, leverage, case deletion and local influence analyses) under such a framework are presented in Section 4; slight modifications of the existing tools to facilitate comparison of different models are suggested. In Section 5, we describe remedial measures to deal with cases where the diagnostic tools suggest that a proposed model may not be satisfactory; in particular, we consider fine tuning of the Gaussian LMM, or alternatively, fitting elliptically symmetric, skew-elliptical LMM or using robust estimation methods. In each case, we comment on available diagnostic techniques. Computational aspects are outlined in Section 6 along with some exploratory tools that may help in the specification of a tentative model. Some analysis strategies are illustrated with practical examples. We conclude with a brief discussion and an overview of research problems in Section 7.

2 The Linear Mixed Model

The linear mixed model may be expressed as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i, \quad i = 1, \ldots, n, \tag{1}$$

where $\mathbf{y}_i = (y_{i1}, \ldots, y_{im_i})^\top$ is an $m_i \times 1$ vector of observations (*response profile*) for the i-th unit, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is a $p \times 1$ vector of unknown population parameters (*fixed effects*), $\mathbf{X}_i$ is a $m_i \times p$ known specification matrix corresponding to the fixed effects, $\mathbf{b}_i = (b_{i1}, \ldots, b_{iq})^\top$ is a $q \times 1$ vector of unobservable random elements (*random effects*), $\mathbf{Z}_i$ is an $m_i \times q$ known specification matrix corresponding to the random effects and $\mathbf{e}_i = (e_{i1}, \ldots, e_{im_i})^\top$ is an $m_i \times 1$ vector of random errors. Usually, we assume that $\mathbf{b}_i$ and $\mathbf{e}_i$ are all uncorrelated and are such that $\mathbb{E}(\mathbf{b}_i) = \mathbf{0}$, $\mathbb{V}(\mathbf{b}_i) = \mathbf{G} = \mathbf{G}(\boldsymbol{\theta})$, $\mathbb{E}(\mathbf{e}_i) = \mathbf{0}$, $\mathbb{V}(\mathbf{e}_i) = \mathbf{R}_i = \mathbf{R}_i(\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is a $t \times 1$ covariance parameter vector not functionally related to $\boldsymbol{\beta}$ and $\mathbf{G}$ and $\mathbf{R}_i$ are, respectively, $q \times q$ and $m_i \times m_i$ appropriate covariance matrices.

The marginal distribution of the vector of observations from the i-th unit has mean vector $\mathbf{X}_i\boldsymbol{\beta}$ and covariance matrix

$$\mathbb{V}(\mathbf{y}_i) = \boldsymbol{\Omega}_i(\boldsymbol{\theta}) = \mathbf{Z}_i\mathbf{G}(\boldsymbol{\theta})\mathbf{Z}_i^\top + \mathbf{R}_i(\boldsymbol{\theta}). \tag{2}$$

The first component of the right-hand side of (2) models the dispersion of the individual response profiles ($\mathbf{y}_i$) around the mean response profile, and the second component is related to the conditional dispersion of the response within the individual profile.

It is common to consider $\mathbf{R}_i = \sigma^2\mathbf{I}_{m_i}$, with $\mathbf{I}_r$ denoting an r-dimensional identity matrix. In this case, if a single random effect b_i such that $\mathbb{V}(b_i) = \sigma_b^2$ is included, it follows that $\boldsymbol{\theta} = (\sigma_b^2, \sigma^2)^\top$, and a uniform (within-unit) covariance matrix (i.e. with equal variances, $\sigma_b^2 + \sigma^2$ and equal covariances, σ_b^2) is induced for $\mathbb{V}(\mathbf{y}_i)$. When the maximum number of possible random effects is included in the model, an unstructured (i.e. with possibly different variances and covariances) covariance matrix is obtained. On the one hand, the simplicity of the uniform structure may not be adequate to model real data, but on the other, the excessive number of parameters associated to the unstructured covariance matrix may decrease the efficiency of estimators, specially for small or moderate sample sizes. Models for $\mathbf{G}$ and $\mathbf{R}_i$ with a small number (e.g. 2–4) of covariance parameters (e.g. autoregressive, Toeplitz or ante-dependence) are thus interesting alternatives to the two extreme cases mentioned previously; they relax the strong restrictions on the covariance structure imposed by the single random intercept model and avoid the over-parametrisation required by the more general case (see Diggle *et al.* (2002) or Demidenko (2013) for details).

Letting $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{X} = (\mathbf{X}_1^\top, \ldots, \mathbf{X}_n^\top)^\top$, $\mathbf{Z} = \oplus_{i=1}^n \mathbf{Z}_i$, $\mathbf{b} = (\mathbf{b}_1^\top, \ldots, \mathbf{b}_n^\top)^\top$ and $\mathbf{e} = (\mathbf{e}_1^\top, \ldots, \mathbf{e}_n^\top)^\top$, we can write model (1) more compactly as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \mathbf{e}. \tag{3}$$

This implies that $\mathbb{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $\mathbb{V}(\mathbf{y}) = \boldsymbol{\Omega} = \mathbf{Z}\boldsymbol{\Gamma}\mathbf{Z}^\top + \mathbf{R}$, where $\boldsymbol{\Gamma} = \mathbf{I}_n \otimes \mathbf{G}$ and $\mathbf{R} = \oplus_{i=1}^n \mathbf{R}_i$.

Given the model specification and assuming that the covariance matrices $\boldsymbol{\Gamma}$ and $\mathbf{R}$ are known, best linear unbiased estimators (BLUE) of the fixed effects $\boldsymbol{\beta}$ and best linear predictors (BLUP) of the random effects $\mathbf{b}_i$ may be obtained as the solutions to the mixed model equations (MME) [see Henderson (1975) for details].

$$\begin{pmatrix} \mathbf{X}^\top\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}^\top\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}^\top\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}^\top\mathbf{R}^{-1}\mathbf{Z} + \boldsymbol{\Gamma}^{-1} \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\mathbf{b}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^\top\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}^\top\mathbf{R}^{-1}\mathbf{y} \end{pmatrix}, \tag{4}$$

namely

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &= (\mathbf{X}^\top\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\Omega}^{-1}\mathbf{y} \\ \widehat{\mathbf{b}} &= \boldsymbol{\Gamma}\mathbf{Z}^\top\boldsymbol{\Omega}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) = \boldsymbol{\Gamma}\mathbf{Z}^\top\mathbf{Q}\mathbf{y}, \end{aligned} \tag{5}$$

where

$$\mathbf{Q} = \boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}\mathbf{X}(\mathbf{X}^\top\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\Omega}^{-1}.$$

This implies that

$$\mathbb{E}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta},\ \mathbb{V}(\widehat{\boldsymbol{\beta}}) = (\mathbf{X}^\top\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}, \tag{6}$$

$$\mathbb{E}(\widehat{\mathbf{b}}) = \mathbf{0},\ \mathbb{V}(\widehat{\mathbf{b}}) = \boldsymbol{\Gamma}\mathbf{Z}^\top\mathbf{Q}\mathbf{Z}\boldsymbol{\Gamma}. \tag{7}$$

Although some derivations of (5) require assumptions on the form of the distribution of $\mathbf{y}$, they may be obtained by simply specifying the mean and covariance structure of the underlying random variables as in (3). Such results are useful to estimate/predict unit-specific linear combinations of the form $\mathbf{L}_i = \mathbf{K}_1^\top \boldsymbol{\beta} + \mathbf{K}_2^\top \mathbf{b}_i$, where $\mathbf{K}_1$ and $\mathbf{K}_2$ are known $p \times k$ and $q \times k$ matrices, respectively. The reader is referred to Harville (1976) or to Robinson (1991) for excellent accounts on the subject.

In practice, the covariance matrices $\boldsymbol{\Gamma}$ and $\mathbf{R}$ are unknown, and empirical BLUE or BLUP of $\mathbf{L}_i$ may be obtained by using consistent estimates. Inferential procedures, however, depend on further assumptions on the elements of model (3). Mainly because of its mathematical tractability, the classical approach to this problem is to assume Gaussian distributions for $\mathbf{b}$ and $\mathbf{e}$.

3 Inference in a Gaussian Set-up

The Gaussian LMM is given by (1) coupled with

$$\mathbf{b}_i \overset{\text{iid}}{\sim} N_q(\mathbf{0}, \mathbf{G}) \text{ and } \mathbf{e}_i \overset{\text{iid}}{\sim} N_{m_i}(\mathbf{0}, \mathbf{R}_i), \quad i = 1, ..., n, \tag{8}$$

or equivalently with

$$\begin{bmatrix} \mathbf{b} \\ \mathbf{e} \end{bmatrix} \sim N_{M+N}\left(\begin{bmatrix} \mathbf{0}_M \\ \mathbf{0}_N \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Gamma} & \mathbf{0}_{M\times N} \\ \mathbf{0}_{N\times M} & \mathbf{R} \end{bmatrix} \right)$$

where $N = \sum_{i=1}^n m_i$ and $M = nq$.

The corresponding likelihood is

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\theta}) &= \prod_{i=1}^n (2\pi)^{-m_i/2} |\boldsymbol{\Omega}_i(\boldsymbol{\theta})|^{-1/2} \exp\left\{ -\frac{1}{2} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top [\boldsymbol{\Omega}_i(\boldsymbol{\theta})]^{-1} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) \right\} \\ &= (2\pi)^{-N/2} |\boldsymbol{\Omega}(\boldsymbol{\theta})|^{-n/2} \exp\left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top [\boldsymbol{\Omega}(\boldsymbol{\theta})]^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\}. \end{aligned} \tag{9}$$

If $\boldsymbol{\Omega}(\boldsymbol{\theta})$ were known, the maximum likelihood estimator (MLE) of $\boldsymbol{\beta}$ would be

$$\begin{aligned} \widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) &= \left(\sum_{i=1}^n \mathbf{X}_i^\top [\boldsymbol{\Omega}_i(\boldsymbol{\theta})]^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^n \mathbf{X}_i^\top [\boldsymbol{\Omega}_i(\boldsymbol{\theta})]^{-1} \mathbf{y}_i \\ &= \left(\mathbf{X}^\top [\boldsymbol{\Omega}(\boldsymbol{\theta})]^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^\top [\boldsymbol{\Omega}(\boldsymbol{\theta})]^{-1} \mathbf{y}. \end{aligned} \tag{10}$$

In general, $\boldsymbol{\Omega}(\boldsymbol{\theta})$ is not known and must be estimated. For such purposes, we may maximise (9) with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ simultaneously. Given the non-linear nature of the likelihood function, with the exception of some particular situations, it is necessary to use iterative procedures like those based on the Newton–Raphson, Fisher Scoring or EM algorithms to obtain the MLE of $\boldsymbol{\theta}$.

It is well-known (for example, Diggle *et al.*, 2002) that the MLE of variance components are, in general, biased. To reduce the bias, Patterson and Thompson (1971) suggest that we should maximise the likelihood function of a transformation of the response that does not depend on the fixed effects. This method may be implemented by considering the density function of the marginal residuals $\tilde{\mathbf{e}} = \mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}$, where $\tilde{\boldsymbol{\beta}}$ is the ordinary least squares estimator of $\boldsymbol{\beta}$ under the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$. The resulting log-likelihood function is proportional to

$$l(\tilde{\mathbf{e}}, \boldsymbol{\theta}) = -\frac{1}{2}\left\{\log|\mathbf{X}^\top \boldsymbol{\Omega}(\boldsymbol{\theta})^{-1}\mathbf{X}| + \log|\boldsymbol{\Omega}(\boldsymbol{\theta})| + \tilde{\mathbf{e}}^\top \tilde{\mathbf{e}}\right\}. \tag{11}$$

The estimator of $\boldsymbol{\theta}$ obtained via the maximisation of (11), say $\widehat{\boldsymbol{\theta}}_R$, is known as the restricted (residual) maximum likelihood estimator (REMLE). As in the earlier case, we have a non-linear constrained maximisation problem. Both the MLE and REMLE of $\boldsymbol{\theta}$ are asymptotically equivalent; however, in small/moderate samples, the bias of the REMLE is smaller, and for this reason, it is commonly used in practice. The corresponding REMLE of $\boldsymbol{\beta}$ is obtained by replacing $\boldsymbol{\Omega}(\boldsymbol{\theta})$ with its REMLE in (10). For details, the reader is referred to Dey *et al.* (2000), McCulloch *et al.* (2008) and Demidenko (2013), among others.

Under model (1), with the assumptions of normality and known $\boldsymbol{\theta}$, we may show that

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) \sim N_p\left\{\boldsymbol{\beta}, \left(\mathbf{X}^\top[\boldsymbol{\Omega}(\boldsymbol{\theta})]^{-1}\mathbf{X}\right)^{-1}\right\}. \tag{12}$$

When $\boldsymbol{\theta}$ is unknown, if we assume that

$$\lim_{n\to\infty} n^{-1}\left(\mathbf{X}^\top[\Omega(\boldsymbol{\theta})]^{-1}\mathbf{X}\right) = \mathbf{A},$$

where $\mathbf{A}$ is a positive definite matrix, we may use the consistency of the MLE ($\widehat{\boldsymbol{\theta}}$) and Sverdrup's theorem [e.g. Sen *et al.* (2009)] to show that

$$\sqrt{n}[\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}) - \boldsymbol{\beta}] \to N_p\left(0, \mathbf{A}^{-1}\right), \tag{13}$$

so that for large n, the distribution of $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}})$ may be approximated by (12) with $\widehat{\boldsymbol{\theta}}$ in lieu of $\boldsymbol{\theta}$. Given that the MLE and REMLE are asymptotically equivalent, it follows that the same approximate distribution remains valid if we substitute $\widehat{\boldsymbol{\theta}}_R$ for $\widehat{\boldsymbol{\theta}}$ in (10) and (12). For more details, see, for example, Jiang (1996) and Demidenko (2013) or Sen *et al.* (2009).[1]

Other methods of estimation, like the method of moments or minimum norm quadratic unbiased estimation (MINQUE) are presented and discussed in Demidenko (2013), for example. Bayesian methods have also been developed by many authors, like Dey *et al.* (2000), Wu and Zhang (2006), Daniels and Hogan (2008) or Demidenko (2013).

4 Diagnostic in the Gaussian setup

Residual and sensitivity analyses constitute important tools for evaluating the fit of any statistical model to a given data, for checking the validity of its assumptions, and consequently, for evaluating the reliability of statistical inference based on it. In standard Gaussian linear models, residuals are used to verify linearity of effects, normality, independence, homoskedasticity of the errors as well as the presence of outliers and influential observations. Sensitivity analysis is based on a set of tools designed to evaluate changes in the fitted model when some perturbation is imposed on the data or on the model assumptions. They include leverage analysis, based on the effect of the explanatory variables via the 'hat' matrix, case-deletion analysis or local influence analysis, where the effect of a small (infinitesimal) perturbation of the data or of the model is evaluated.

Many alternatives to evaluate the fit of ordinary linear models are available in the literature, as described in Hoaglin and Welsch (1978), Belsley *et al.* (1980), Cook and Weisberg (1982), Cook (1986), Chatterjee and Hadi (1988), Wei *et al.* (1998) and Atkinson and Riani (2000), among others. Generalisations to Gaussian LMM are more recent. For example,

Beckman *et al.* (1987), Christensen *et al.* (1992), Banerjee and Frees (1997), Lesaffre and Verbeke (1998), Tan *et al.* (2001), Fung *et al.* (2002), Demidenko and Stukel (2005), Zewotir and Galpin (2005), Zewotir (2008), Nobre and Singer (2007, 2011), Gumedze *et al.* (2010), Schützenmeister and Piepho (2012) or Demidenko (2013) present a series of such diagnostic methods that we summarise and update in the following sections.

A distinctive feature of diagnostic analysis of LMM relates to the evaluation of the impact of an entire unit or of some within-unit observations on certain features of the model. For example, units identified as possible outliers may be so because of the effect of a single (within-unit) observation.

4.1 Residual Analysis

Diagnostic analysis is more complex in LMM than in the standard case because the associated sources of variation (**e** and **b**) generate the following three types of residuals:

(i) Marginal residuals, $\widehat{\boldsymbol{\xi}} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}$, that predict the marginal errors $\boldsymbol{\xi} = \mathbf{y} - \mathbb{E}[\mathbf{y}] = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$.
(ii) Conditional residuals, $\widehat{\mathbf{e}} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}} - \mathbf{Z}\widehat{\mathbf{b}}$, that predict the conditional errors $\mathbf{e} = \mathbf{y} - \mathbb{E}[\mathbf{y}|\mathbf{b}] = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b}$.
(iii) Random effects residuals, $\mathbf{Z}\widehat{\mathbf{b}}$, that predict the random effects, $\mathbf{Z}\mathbf{b} = \mathbb{E}[\mathbf{y}|\mathbf{b}] - \mathbb{E}[\mathbf{y}]$.

A further complication is related to the concept of confounding. As discussed by Hilden-Minton (1995), a residual is considered confounded for a specific type of error if it depends on other errors than the one that it is supposed to predict. Given that $\widehat{\mathbf{e}} = \mathbf{RQe} + \mathbf{RQZb}$, conditional and random effects residuals may be confounded (except in cases where the columns of $\mathbf{Z}_i$ belong to the space generated by the columns of $\mathbf{X}_i$, $i = 1, \ldots, n$). This implies, for example, that $\widehat{\mathbf{e}}$ may not be adequate to check for the normality of **e** because when **b** is grossly non-Gaussian, $\widehat{\mathbf{e}}$ may not present a Gaussian behaviour even when **e** is Gaussian.

4.1.1 Marginal residuals

Lesaffre and Verbeke (1998) comment that when the within-unit covariance structure is adequate, $\mathcal{V}_i = ||\mathbf{I}_{m_i} - \boldsymbol{\mathcal{E}}_i \boldsymbol{\mathcal{E}}_i^\top||^2$, where $\boldsymbol{\mathcal{E}}_i = \widehat{\boldsymbol{\Omega}}_i^{-1/2}\widehat{\boldsymbol{\xi}}_i$ with $\boldsymbol{\Omega}_i = \boldsymbol{\Omega}_i(\boldsymbol{\theta})$ should be close to zero. Units with large values of $\mathcal{V}_i$ are those for which the proposed covariance structure might not be adequate. Given that the true variance of $\widehat{\boldsymbol{\xi}}_i$ is $\mathbb{V}(\widehat{\boldsymbol{\xi}}_i) = \boldsymbol{\Omega}_i - \mathbf{X}_i(\mathbf{X}_i^\top \boldsymbol{\Omega}_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i^\top$ and not $\boldsymbol{\Omega}_i$, we consider replacing $\boldsymbol{\mathcal{E}}_i$ in $\mathcal{V}_i$ with the standardised marginal residuals $\widehat{\boldsymbol{\xi}}_i^* = [\widehat{\mathbb{V}}(\widehat{\boldsymbol{\xi}}_i)]^{-1/2}\widehat{\boldsymbol{\xi}}_i$, where $\mathbb{V}(\widehat{\boldsymbol{\xi}}_i)$ corresponds to the diagonal block of $\boldsymbol{\Omega} - \mathbf{X}(\mathbf{X}^\top\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top$ associated to the i-unit. Furthermore, to avoid giving much weight to units with many observations, we consider taking $\mathcal{V}_i^* = \sqrt{\mathcal{V}_i}/m_i$ as a standardised measure of adequacy of the within-unit covariance structure. Plots of the $\mathcal{V}_i^*$ versus unit indices, i, (generally known as unit index plots) may help to identify units for which the covariance structure should be modified.

To evaluate the linearity of the fixed effects in model (1), we consider plotting the elements of the standardised marginal residuals $\widehat{\boldsymbol{\xi}}_{ij}^* = \widehat{\boldsymbol{\xi}}_{ij}/[diag_j(\widehat{\mathbb{V}}(\widehat{\boldsymbol{\xi}}_i))]^{1/2}$, where $diag_j(\mathbb{V}(\widehat{\boldsymbol{\xi}}_i))$ is the j-th element of the main diagonal of $\mathbb{V}(\widehat{\boldsymbol{\xi}}_i)$, versus the values of each explanatory variable as well as versus the fitted values. We also recommend plotting $\widehat{\boldsymbol{\xi}}_{ij}^*$ versus the observation indices as a tool to detect outlying observations.

4.1.2 Conditional residuals

Given that $\mathbb{V}(\widehat{\mathbf{e}}) = \mathbf{R}[\boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}\mathbf{X}\left(\mathbf{X}^\top\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^\top\boldsymbol{\Omega}^{-1}]\mathbf{R} = \mathbf{RQR}$, Nobre and Singer (2007) observe that the conditional residuals may have different variances. They suggest plots

of standardised conditional residuals, $\widehat{e}^{*}_{ij} = \widehat{e}_{ij}/diag_{ij}(\widehat{\mathbf{RQ}}\widehat{\mathbf{QR}})^{1/2}$, with $diag_{ij}(\mathbf{RQR})$ denoting the main diagonal element of $\mathbf{RQR}$ corresponding to the j-th observation of the i-th unit versus fitted values to check for homoskedasticity of the conditional errors (in the case of homoskedastic models) or versus observation indices to check for outlying observations.

As indicated in Hilden-Minton (1995), the ability to check for normality of the conditional errors increases as we minimise the fraction of confounding for the k-th conditional residual, namely

$$0 \leq \frac{\mathbf{u}_k^{\top}\mathbf{RQZ\Gamma Z}^{\top}\mathbf{QRu}_k}{\mathbf{u}_k^{\top}\mathbf{RQRu}_k} = 1 - \frac{\mathbf{u}_k^{\top}\mathbf{RQRQRu}_k}{\mathbf{u}_k^{\top}\mathbf{RQRu}_k} \leq 1 \tag{14}$$

where $\mathbf{u}_k$ is the k-th column of $\mathbf{I}_N$. With this in mind, he advocates the use of *least confounded conditional residuals*, that is, of a linear transformation of the conditional residuals that minimises the fraction of confounding. The least confounded residuals are given by

$$\mathbf{c}_k^{\top}\widehat{\mathbf{e}} = \lambda_k^{-1/2}\boldsymbol{\ell}_k^{\top}\mathbf{R}^{-1/2}\widehat{\mathbf{e}} = \lambda_k^{1/2}\boldsymbol{\ell}_k^{\top}\mathbf{R}^{-1/2}\mathbf{y},\ k = 1, \ldots, N-p$$

where $1 \geq \lambda_1 \geq \ldots \geq \lambda_{N-p} > 0$ are the ordered values of $\boldsymbol{\Lambda}$, obtained from the spectral decomposition

$$\mathbf{R}^{1/2}\mathbf{QR}^{1/2} = \mathbf{L\Lambda L}^{\top},\ \mathbf{L}^{\top}\mathbf{L} = \mathbf{I}_{N-p}$$

and $\boldsymbol{\ell}_k$ represents the k-th column of $\mathbf{L}$. Standardised least confounded residuals, $\mathbf{c}_k^{\top}\widehat{\mathbf{e}}^{*}$, may be obtained by dividing $\mathbf{c}_k^{\top}\widehat{\mathbf{e}}$ by the square root of the corresponding element in $\mathbf{C}\widehat{\mathbf{RQ}}\widehat{\mathbf{RC}}^{\top}$, where $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_{N-p}]^{\top}$. QQ plots of the standardised least confounded conditional residuals, $\mathbf{c}_k^{\top}\widehat{\mathbf{e}}^{*}$, may be employed to check for normality.

More recently, Schützenmeister and Piepho (2012) comment that linear transformation of residuals may not be so useful as diagnostic tools because, first, linearly transformed residuals do not correspond to individual observations, and second, they may amplify the super-normality effect, that is, may tend to look more normal than the underlying effects really are. As an alternative, they propose to analyse the LMM as if it were a fixed effects linear model and to use the corresponding studentised residuals as diagnostic tools. They also propose a simulation approach to construct tolerance intervals for the corresponding QQ plots. Their proposal is computer-intensive and may not be reasonable as an exploratory tool in cases where many models are under investigation.

4.1.3 Random effects residuals

When there is no confounding and the random effects follow a q-dimensional Gaussian distribution, $\mathcal{M}_i = \widehat{\mathbf{b}}_i^{\top}\{\mathbb{V}[\widehat{\mathbf{b}}_i - \mathbf{b}_i]\}^{-1}\widehat{\mathbf{b}}_i$ (the Mahalanobis's distance between $\widehat{\mathbf{b}}_i$ and $\mathbb{E}(\mathbf{b}_i) = \mathbf{0}$) should have a chi-squared distribution with q degrees of freedom. Therefore, an χ^2_q QQ plot for $\mathcal{M}_i$ may be used to verify whether the random effects follow a (q-variate) Gaussian distribution. Unit index plots of $\mathcal{M}_i$ may also be employed to detect outliers.

The different uses for the three types of LMM residuals are summarised in Table 1, adapted from Nobre and Singer (2007).

4.2 Global Influence Analysis

4.2.1 Leverage analysis

In a standard linear model, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, the *leverage* of the i-th observation is defined as $h_{ii} = \partial\widehat{y}_i/\partial y_i$ with $\widehat{y}_i$ denoting the i-th fitted value (i.e. the i-th element of $\widehat{\mathbf{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$) and

Table 1. *Uses of residuals for diagnostic purposes.*

Diagnostic for	Type of residual	Plot
Linearity of effects fixed ($\mathbb{E}[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta}$)	Marginal	$\widehat{\boldsymbol{\xi}}^*_{ij}$ versus fitted values or explanatory variables
Presence of outlying observations ($\mathbf{y}_{ij}$)	Marginal	$\widehat{\boldsymbol{\xi}}^*_{ij}$ versus observation indices
Within-units covariance matrix ($\boldsymbol{\Omega}_i$)	Marginal	$\mathcal{V}^*_i$ versus unit indices and
Presence of outlying observations ($\mathbf{y}_{ij}$)	Conditional	$\widehat{e}^*_{ij}$ versus observation indices
Homoskedasticity of conditional errors ($\mathbf{e}_{ij}$)	Conditional	$\widehat{e}^*_{ij}$ versus fitted values
Normality of conditional errors ($\mathbf{e}_{ij}$)	Conditional	Gaussian QQ plot for $\mathbf{c}_k^\top \widehat{\mathbf{e}}^*$
Presence of outlying subjects ($\mathbf{b}_i$)	Random effects	$\mathcal{M}_i$ versus unit indices
Normality of the random effects ($\mathbf{b}_i$)	Random effects	χ^2_q QQ plot for $\mathcal{M}_i$

corresponds to the i-th element of the main diagonal of the orthogonal projection matrix or 'hat' matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$. Outlying observations in the vector space spanned by the columns of $\mathbf{X}$ are said to have *high leverage* with respect to the fitted values $\widehat{y}_i$, as pointed in Cook and Weisberg (1982), Chatterjee and Hadi (1988) and Wei *et al.* (1998), for example. For the linear mixed model (3), Fung *et al.* (2002) define the *generalised marginal leverage matrix* as

$$\begin{aligned}\mathbf{L}_1 = \frac{\partial\widehat{\mathbf{y}}}{\partial\mathbf{y}^\top} &= \frac{\partial\mathbf{X}\left(\mathbf{X}^\top\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^\top\boldsymbol{\Omega}^{-1}\mathbf{y}}{\partial\mathbf{y}^\top} \\ &= \mathbf{X}\left(\mathbf{X}^\top\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^\top\boldsymbol{\Omega}^{-1}.\end{aligned} \tag{15}$$

This is essentially the oblique projection matrix in the space generated by the columns of $\mathbf{X}$. Outlying observations in the vector space spanned by the columns of $\mathbf{X}$ are said to have *high leverage* with respect to the fitted values $\widehat{y}_i$.

Letting $\mathbf{L}_{1i} = \mathbf{X}_i(\mathbf{X}_i^\top\boldsymbol{\Omega}_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i^\top\boldsymbol{\Omega}_i^{-1}$, that is, the i-th diagonal block of $\mathbf{L}_1$, and $\mathbf{L}_{1i(jj)}$ denote the j-th element of the main diagonal of $\mathbf{L}_{1i}$, we have

$$\operatorname{tr}(\mathbf{L}_1) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\mathbf{L}_{1i(jj)} = p.$$

Then, we may consider unit i to have high leverage with respect to the marginal fitted values $\widehat{\mathbf{y}}$ if $\operatorname{tr}(\mathbf{L}_{1i})/m_i$ is larger than some arbitrary value, say $2p/n$ and, similarly, observation j within unit i to have high leverage with respect to the marginal fitted values if $\mathbf{L}_{1i(jj)} \geq 2p/n$, although it is more common to use subjective criteria (based on visual inspection of leverage plots) for such purposes.

Given that in model (3), a unit or a (within-unit) observation can affect both marginal and conditional fitted values, it seems reasonable to evaluate the joint influence of each unit or (within-unit) observation on both. Thus, $\widehat{\mathbf{y}}^* = \mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{Z}\widehat{\mathbf{b}}$ is used instead of $\widehat{\mathbf{y}}$ to define the *generalised joint leverage matrix* as

$$\mathbf{L} = \frac{\partial\widehat{\mathbf{y}}^*}{\partial\mathbf{y}^\top} = \frac{\partial\widehat{\mathbf{y}}}{\partial\mathbf{y}^\top} + \frac{\partial\mathbf{Z}\widehat{\mathbf{b}}}{\partial\mathbf{y}^\top} = \mathbf{L}_1 + \mathbf{Z}\boldsymbol{\Gamma}\mathbf{Z}^\top\mathbf{Q} = \mathbf{L}_1 + \mathbf{L}_2\mathbf{Q}, \tag{16}$$

where $\mathbf{L}_1$ is the generalised marginal leverage matrix (15) and

$$\mathbf{L}_2 = \mathbf{Z}\boldsymbol{\Gamma}\mathbf{Z}^\top. \tag{17}$$

The matrix $\mathbf{L}_2$ represents the portion of the within-unit variability explained by the presence of the random effects. Outlying observations in the vector space spanned by the columns of $\mathbf{Z}$

should affect the random effect component of the conditional fitted values and consequently affect the estimate of the within-unit variance explained by the random effects as advocated by Tan *et al.* (2001). Therefore, estimates of the components of variance of the model should be affected by observations with high leverage with respect to the random effect component of the conditional fitted values. Demidenko and Stukel (2005) suggest that

$$\mathbf{H}_2 = \mathbf{Z}\boldsymbol{\Gamma}\mathbf{Z}^\top\mathbf{Q} \tag{18}$$

could be used as a generalised leverage matrix for such purposes. Nobre and Singer (2011), on the other hand, observe that

$$\mathbf{H}_2 = \mathbf{L}_2\boldsymbol{\Omega}^{-1}\left[\mathbf{I}_n - \mathbf{L}_1\right]$$

also depends on the generalised marginal leverage matrix $\mathbf{L}_1$ and argue that the leverage with respect to the random effects of the conditional fitted values may be confounded by the leverage with respect to the marginal fitted values. Thus, given (16), they suggest that a better alternative to measure the leverage of the observations and units with respect to the random effect components of the conditional fitted values is to use the *generalised random component leverage matrix* $\mathbf{L}_2$ instead of $\mathbf{H}_2$.

Along the lines considered in the case of marginal fitted values, the i-th unit generalised random component leverage matrix is defined as $\mathbf{L}_{2i} = \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i^\top$. The corresponding generalised random component leverage for the j-th observation of the i-th unit is $\mathbf{L}_{2i(jj)}$, that is, the j-th element of the main diagonal of $\mathbf{L}_{2i}$. Similarly, the decision regarding what should be considered as high leverage units or (within-unit) observations is usually based on a subjective assessment.

4.2.2 Case deletion analysis

A simple and easily interpretable tool to evaluate the impact of an observation or of a set of observations on some characteristic of interest (a parameter estimate or the corresponding standard error, for example) is based on *case deletion*, an idea introduced by Cook (1977) and studied by many authors, like Belsley *et al.* (1980), Cook and Weisberg (1982) or Chatterjee and Hadi (1988). In LMM like (3), it seems reasonable to consider the repeated measurements on a unit (case) or on a set of units as the set of observations to be deleted. In this context, part of the effort lies in obtaining BLUEs or BLUPs without having to refit the model when each unit or observation is deleted.

Assuming that $\boldsymbol{\Gamma}$ and $\mathbf{R}$ are known, and that we are interested in examining the impact of the set of observations $I = \{i_1, i_2, ..., i_k\}$ $(1 \leq i_1 \leq i_2 \leq ... \leq i_k \leq N)$, we may consider the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \mathbf{U}_I\boldsymbol{\phi}_{(I)} + \mathbf{e}, \tag{19}$$

where $\boldsymbol{\phi}_I$ represents a k-dimensional (fixed) parameter vector and $\mathbf{U}_I = [\mathbf{u}_{i_1}, \mathbf{u}_{i_2}, ..., \mathbf{u}_{i_k}]$ with $\mathbf{u}_i$ denoting the i-th column of $\mathbf{I}_N$. In general, the set I contains all the observations related to a unit or to a set of units. Hilden-Minton (1995) and Fung *et al.* (2002) show that the BLUE of $\boldsymbol{\beta}$ and the BLUP of $\mathbf{b}$ in model (19) are the same as those obtained from model (3) when the observations in the set I are deleted and denote them by $\widehat{\boldsymbol{\beta}}_{(I)}$ and $\widehat{\mathbf{b}}_{(I)}$, respectively. They also show that the BLUE of $\boldsymbol{\phi}_{(I)}$ is

$$\widehat{\boldsymbol{\phi}}_{(I)} = \left(\mathbf{U}_I^\top\mathbf{Q}\mathbf{U}_I\right)^{-1}\mathbf{U}_I^\top\mathbf{Q}\mathbf{y} \tag{20}$$

and that

$$\widehat{\boldsymbol{\beta}}_{(I)} = \widehat{\boldsymbol{\beta}} - \left(\mathbf{X}^\top \boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^\top \boldsymbol{\Omega}^{-1}\mathbf{U}_I\widehat{\boldsymbol{\phi}}_{(I)} \tag{21}$$

$$\mathbf{b}_{(I)} = \widehat{\mathbf{b}} - \boldsymbol{\Gamma}\mathbf{Z}^\top \mathbf{Q}\mathbf{U}_I\widehat{\boldsymbol{\phi}}_{(I)}. \tag{22}$$

Furthermore, we have

$$\mathbb{V}\left(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(I)}\right) = \left(\mathbf{X}^\top \boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^\top \boldsymbol{\Omega}^{-1}\mathbf{U}_I\left(\mathbf{U}_I^\top \mathbf{Q}\mathbf{U}_I\right)^{-1}\mathbf{U}_I^\top \boldsymbol{\Omega}^{-1}\mathbf{X}\left(\mathbf{X}^\top \boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1} \tag{23}$$

$$\mathbb{V}\left(\widehat{\mathbf{b}} - \widehat{\mathbf{b}}_{(I)}\right) = \boldsymbol{\Gamma}\mathbf{Z}^\top \mathbf{Q}\mathbf{U}_I(\mathbf{U}_I\mathbf{Q}\mathbf{U}_I)^{-1}\mathbf{U}_I\mathbf{Q}\mathbf{Z}\boldsymbol{\Gamma} \tag{24}$$

and

$$\mathbb{V}\left(\widehat{\boldsymbol{\beta}}_{(I)}\right) = \mathbb{V}(\widehat{\boldsymbol{\beta}}) - \mathbb{V}\left(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(I)}\right). \tag{25}$$

With these ingredients, we may consider Cook's distance

$$D_I = \frac{\left(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(I)}\right)^\top \left(\mathbf{X}^\top \widehat{\boldsymbol{\Omega}}^{-1}\mathbf{X}\right)\left(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(I)}\right)}{p} = \frac{\left(\widehat{\mathbf{y}} - \widehat{\mathbf{y}}_{(I)}\right)^\top \widehat{\boldsymbol{\Omega}}^{-1}\left(\widehat{\mathbf{y}} - \widehat{\mathbf{y}}_{(I)}\right)}{p}, \tag{26}$$

as a measure of influence of the observations in the set I on the estimate of the fixed parameter $\boldsymbol{\beta}$.

Tan *et al.* (2001), among others, comment on the limited efficiency of Cook's distance (26) as a measure of influence in the context of LMM. In particular, they stress that unit-oriented measures may not be convenient to detect influential units in view of the relative position of the observations within and across subjects and propose a conditional approach based on observation-oriented influence measures. They assume that the covariance matrix of $\mathbf{e}_i$ in (1) is $\mathbf{R}_i = \sigma^2\mathbf{I}_{n_i}$, but we may generalise their results to cases where $\mathbf{R}_i$ is more general and consider the conditional model

$$\mathbf{y} = \mathbf{X}^*\boldsymbol{\beta}^* + \mathbf{e}$$

where $\mathbf{X}^* = [\mathbf{X}\ \mathbf{Z}]$ and $\boldsymbol{\beta}^* = (\boldsymbol{\beta}^\top, \mathbf{b}^\top)^\top$. The *conditional Cook distance* is defined as

$$D_{i(j)}^{cond} = \sum_{i=1}^{n} \frac{\left(\widehat{\mathbf{y}}_i^* - \widehat{\mathbf{y}}_{i(j)}^*\right)^\top \widehat{\mathbb{V}}[\mathbf{y}_i|\mathbf{b}_i]^{-1}\left(\widehat{\mathbf{y}}_i^* - \widehat{\mathbf{y}}_{i(j)}^*\right)}{(n-1)q + p} = \sum_{i=1}^{n} \frac{\left(\widehat{\mathbf{y}}_i^* - \widehat{\mathbf{y}}_{i(j)}^*\right)^\top \widehat{\mathbf{R}}_i^{-1}\left(\widehat{\mathbf{y}}_i^* - \widehat{\mathbf{y}}_{i(j)}^*\right)}{(n-1)q + p}. \tag{27}$$

where $\widehat{\mathbf{y}}_i^* = \mathbf{X}_i\widehat{\boldsymbol{\beta}} + \mathbf{Z}_i\widehat{\mathbf{b}}$, $\widehat{\mathbf{y}}_{i(j)}^* = \mathbf{X}_i\widehat{\boldsymbol{\beta}}_{(i(j))} + \mathbf{Z}_i\widehat{\mathbf{b}}_{(i(j))}$ and $\widehat{\boldsymbol{\beta}}_{(i(j))}$ and $\widehat{\mathbf{b}}_{(i(j))}$ denote, respectively, the BLUEs of $\boldsymbol{\beta}$ and $\mathbf{b}$ obtained with the elimination of the j-th observation from the i-th unit.

As suggested by Tan *et al.* (2001), (27) may be decomposed as $D_{i(j)}^{cond} = D_{1i(j)}^{cond} + D_{2i(j)}^{cond} + D_{3i(j)}^{cond}$, where

$$D_{1i(j)}^{cond} = [(n-1)q + p]^{-1}\left(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i(j))}\right)^\top \mathbf{X}^\top \widehat{\mathbf{R}}^{-1}\mathbf{X}\left(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i(j))}\right)$$

is a useful measure to evaluate the influence of the j-th observation from the i-th unit on the estimate of $\boldsymbol{\beta}$

$$D_{2i(j)}^{cond} = [(n-1)q+p]^{-1} \sum_{i=1}^{n} \left(\widehat{\mathbf{b}} - \widehat{\mathbf{b}}_{(i(j))}\right)^{\top} \mathbf{Z}_i^{\top} \widehat{\mathbf{R}}_i^{-1} \mathbf{Z}_i \left(\widehat{\mathbf{b}} - \widehat{\mathbf{b}}_{(i(j))}\right)$$

is convenient to evaluate the influence of the j-th observation from the i-th unit on the estimate of $\mathbf{b}$ and

$$D_{3i(j)}^{cond} = [(n-1)q+p]^{-1} 2 \left(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i(j))}\right)^{\top} \sum_{i=1}^{n} \mathbf{X}_i^{\top} \widehat{\mathbf{R}}_i^{-1} \mathbf{Z}_i \left(\widehat{\mathbf{b}} - \widehat{\mathbf{b}}_{(i(j))}\right)$$

is a measure of the covariation between a change in the average profile and a change in the position of the unit-specific profiles relative to the average profile when the j-th observation from the i-th unit is eliminated. They also show that when the interest lies in evaluating the influence of an entire unit on the estimates of the parameters, it suffices to sum (27) over all the observations corresponding to that unit.

A different approach to (conditional) case-deletion diagnostics is considered in Fei and Pan (2003). Zewotir (2008), on the other hand, considers some computational aspects related to unit-deletion diagnostics.

The ratio of the variance ellipsoids

$$\begin{aligned} \rho_{(I)} = \frac{\left|\mathbb{V}(\widehat{\boldsymbol{\beta}}_{(I)})\right|}{|\mathbb{V}(\widehat{\boldsymbol{\beta}})|} &= \frac{\left|\left(\mathbf{X}^{\top}\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{I}_p + \mathbf{X}^{\top}\boldsymbol{\Omega}^{-1}\mathbf{U}_I\left(\mathbf{U}_I^{\top}\mathbf{Q}\mathbf{U}_I\right)^{-1}\mathbf{U}_I^{\top}\boldsymbol{\Omega}^{-1}\mathbf{X}\left(\mathbf{X}^{\top}\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1}\right)\right|}{\left|\left(\mathbf{X}^{\top}\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1}\right|} \\ &= \left|\mathbf{I}_p + \mathbf{X}^{\top}\boldsymbol{\Omega}^{-1}\mathbf{U}_I\left(\mathbf{U}_I^{\top}\mathbf{Q}\mathbf{U}_I\right)^{-1}\mathbf{U}_I^{\top}\boldsymbol{\Omega}^{-1}\mathbf{X}\left(\mathbf{X}^{\top}\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1}\right| \end{aligned} \tag{28}$$

can be used to evaluate the influence of the units in set I on the covariance matrix of $\widehat{\boldsymbol{\beta}}$ as highlighted by Hilden-Minton (1995). In practice, global influence may be evaluated by the index plots summarised in Table 2.

For homoskedastic conditional independence models, we may also consider the decomposition of the conditional Cook distance (27) proposed by Tan *et al.* (2001) as summarised in Table 3.

More recently, Mun and Lindstrom (2013) propose a non-deletion method based on studentized residual sums of squares (TRSS) plots, claiming that it is more efficient and flexible than Cook distance-based methods to identify outlying units or observations.

Table 2. *Global influence plots for observations or units.*

Diagnostic for effect on	Global influence measure	Index plot of
Fixed portion of fitted value ($\mathbf{X}\widehat{\boldsymbol{\beta}}$)	Generalised marginal leverage matrix $\mathbf{L}_1$ (15)	$\mathbf{L}_{1i(jj)}$ $[tr(\mathbf{L}_{1i})/m_i]$ versus observations (units)
Random portion of fitted value ($\mathbf{Z}\widehat{\mathbf{b}}$)	Generalised random component marginal leverage matrix $\mathbf{L}_2$ (17)	$\mathbf{L}_{2i(jj)}$ $[tr(\mathbf{L}_{2i})/m_i]$ versus observations (units)
Regression coefficients ($\widehat{\boldsymbol{\beta}}$)	Cook's distance D_I (26)	$D_{i(j)}$ $[D_i]$ versus observations (units)
Covariance matrix of regression coefficients $[\mathbb{V}(\widehat{\boldsymbol{\beta}})]$	Ratio of variance ellipsoids $\rho_{(I)}$ (24)	$\rho_{i(j)}$ $[\rho_i]$ versus observations (units)

Table 3. *Global influence plots based on decomposition of the conditional Cook distance for homoskedastic conditional independence models.*

Diagnostic for effect on	Index plot of
Regression coefficients ($\widehat{\boldsymbol{\beta}}$)	$D^{cond}_{1i(j)}$ versus observations
Random effects ($\widehat{\mathbf{b}}$)	$D^{cond}_{2i(j)}$ versus observations
Changes in covariance between $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{b}}$	$D^{cond}_{3i(j)}$ versus observations

4.3 Local Influence Analysis

The concept of *local influence* was introduced by Cook (1986) with the objective of evaluating the changes in the analysis resulting from 'small' perturbations on the data or on some element of the model. Essentially, the idea is to investigate the behaviour of the *likelihood displacement* defined as

$$LD(\boldsymbol{\omega}) = 2\left\{L(\widehat{\boldsymbol{\psi}}) - L\left(\widehat{\boldsymbol{\psi}}_{\boldsymbol{\omega}}|\boldsymbol{\omega}\right)\right\}, \tag{29}$$

where L denotes the likelihood for the proposed model, $\boldsymbol{\psi}$ is a p-dimensional parameter vector, $\boldsymbol{\omega} \in \Omega \subset IR^q$ is a q-dimensional vector of 'perturbations' and $\widehat{\boldsymbol{\psi}}$ and $\widehat{\boldsymbol{\psi}}_{\boldsymbol{\omega}}$ are, respectively, the MLEs of $\boldsymbol{\psi}$ based on $L(\boldsymbol{\psi})$ and $L(\boldsymbol{\psi}|\boldsymbol{\omega})$. We assume that there exists some $\boldsymbol{\omega}_0 \in \Omega$ such that $\boldsymbol{\psi}_{\boldsymbol{\omega}_0} = \boldsymbol{\psi}$ and that $L(\boldsymbol{\psi}|\boldsymbol{\omega})$ has continuous first-order and second-order derivatives in a neighbourhood of $\boldsymbol{\omega}_0$. In this context, the normal curvature of the graph $G(\boldsymbol{\omega}) = [\boldsymbol{\omega}^\top, LD(\boldsymbol{\omega})]$ in the direction of the q-dimensional vector $\mathbf{d}$ with unit norm at the point $\boldsymbol{\omega}_0$ is

$$C_{\mathbf{d}} = 2\left|\mathbf{d}^\top \mathbf{H}^\top \ddot{\mathbf{F}}^{-1} \mathbf{H}\mathbf{d}\right|, \tag{30}$$

where $\ddot{\mathbf{F}} = [\partial^2 L(\boldsymbol{\psi})/\partial\boldsymbol{\psi}\,\partial\boldsymbol{\psi}^\top]_{\boldsymbol{\psi}=\widehat{\boldsymbol{\psi}}}$ and $\mathbf{H} = [\partial^2 L(\boldsymbol{\psi}|\boldsymbol{\omega})/\partial\boldsymbol{\psi}\,\partial\boldsymbol{\omega}^\top]_{\boldsymbol{\omega}=\boldsymbol{\omega}_0;\boldsymbol{\psi}=\widehat{\boldsymbol{\psi}}}$. Letting $C_{\min}$ and $C_{\max}$, respectively, denote the smallest and the largest eigenvalues of $-\mathbf{H}^\top \ddot{\mathbf{F}}^{-1}\mathbf{H}$, it is possible to show that $C_{\min} \leq C_{\mathbf{d}} \leq C_{\max}$. The normalised eigenvector $\mathbf{d}_{\max}$ corresponding to $C_{\max}$ may be used to identify what linear combination of the elements of $\boldsymbol{\omega}$ is more influential on the curvature of $LD(\boldsymbol{\omega})$. This is usually employed to evaluate the effect of perturbations on the response variable, that is, $\mathbf{y}_i(\boldsymbol{\omega}_i) = \mathbf{y}_i + \boldsymbol{\omega}_i$, on the explanatory variables, that is, $\mathbf{X}_i(\mathbf{W}_i) = \mathbf{X}_i + \mathbf{W}_i$, where $\mathbf{W}_i = [\boldsymbol{\omega}_{i1}, \ldots, \boldsymbol{\omega}_{ip}]$ with $\boldsymbol{\omega}_{ij} = (\omega_{ij1}, \ldots, \boldsymbol{\omega}_{ijm_i})^\top$, on the variance of the random effects, that is, $\mathbf{G}(\omega) = \omega\mathbf{G}$ or on the variance of the errors, $\mathbf{R}_i(\omega_i) = \omega_i \mathbf{R}_i$. In each case, plots of the absolute values of the elements of $\mathbf{d}_{\max}$ (obtained for the specific $\mathbf{H}$ and $\ddot{\mathbf{F}}$) versus the unit (or within-unit observations) indices may then be used to identify those with more impact on the likelihood displacement.

The selection of the appropriate perturbation scheme is not straightforward, mainly because 'arbitrarily perturbing a model may lead to inappropriate inference about the cause (e.g. influential observations) of a large effect' and 'the components of a perturbation vector may not be orthogonal to each other', which may lead to difficulties in the interpretation, as discussed in Zhu *et al.* (2007). A slightly different and more practical approach for local influence diagnostics in LMM is considered in Beckman *et al.* (1987) and Lesaffre and Verbeke (1998). These authors consider the marginal likelihood of $\boldsymbol{\psi} = (\boldsymbol{\beta}^\top, \boldsymbol{\theta}^\top)^\top$, namely

$$L(\boldsymbol{\psi}) = \sum_{i=1}^{n} L_i(\boldsymbol{\psi}) = \sum_{i=1}^{n}(-1/2)\left\{\log|\boldsymbol{\Omega}_i| + (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top \boldsymbol{\Omega}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right\}, \tag{31}$$

and include perturbations of its individual terms by taking

$$L_i(\boldsymbol{\psi}_{\boldsymbol{\omega}}) = \sum_{i=1}^{n} \omega_i L_i(\boldsymbol{\psi}), \tag{32}$$

where L_i denotes the likelihood for the i-th unit and $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$. Here, $\boldsymbol{\omega}_0 = \mathbf{1}_n$, where $\mathbf{1}_n$ denotes an n-dimensional vector with all elements equal to 1. Letting $\mathbf{u}_i$ denote the i-th column of an n-dimensional identity matrix, the normal curvature (30) computed in the direction $\mathbf{u}_i$ is

$$C_{\mathbf{u}_i} = 2\left|\mathbf{u}_i^\top \mathbf{H}^\top \ddot{\mathbf{F}}^{-1}\mathbf{H}\mathbf{u}_i\right| = 2\left|\mathbf{H}_i^\top \ddot{\mathbf{F}}^{-1}\mathbf{H}_i\right|, \tag{33}$$

where $\mathbf{H}_i$ denotes the i-th column of $\mathbf{H}$. Large values of (33) indicate a large impact of the i-th unit on the likelihood displacement in a local sense, that is, a slight perturbation in the corresponding weight vector ω_i induces a large displacement in the corresponding log-likelihood. Lesaffre and Verbeke (1998) also show that $C_i = C_{\mathbf{u}_i} = -2(\widehat{\boldsymbol{\psi}} - \widehat{\boldsymbol{\psi}}^1_{(i)})^\top \ddot{\mathbf{F}}(\widehat{\boldsymbol{\psi}} - \widehat{\boldsymbol{\psi}}^1_{(i)})$, where $\widehat{\boldsymbol{\psi}}^1_{(i)}$ is referred to as the one-step approximation to $\widehat{\boldsymbol{\psi}}_{(i)}$ and that for large n, it is equivalent to Pregibon's (1981) proposal to measure the *global local influence* of the i-th unit. They show that the computation of the normal curvature (30) in the direction $\mathbf{d}_{\max}$ corresponding to the largest eigenvalue of $-\mathbf{H}^\top \ddot{\mathbf{F}}^{-1}\mathbf{H}$ suggests the perturbation of the model leading to the largest changes in the likelihood displacement.

Because C_i relates to the impact of the i-th unit on the estimator of the parameter vector $\boldsymbol{\psi}$, Lesaffre and Verbeke (1998) discuss a decomposition of this measure that may be used to suggest which part of the model is most affected by a specific unit. In particular, they recommend index plots of $||\mathbf{I}_{m_i} - \boldsymbol{\mathcal{E}}_i\boldsymbol{\mathcal{E}}_i^\top||^2$, $||\boldsymbol{\mathcal{E}}_i\boldsymbol{\mathcal{E}}_i^\top||^2$ $||\boldsymbol{\mathcal{X}}_i\boldsymbol{\mathcal{X}}_i^\top||^2$, $||\boldsymbol{\mathcal{Z}}_i\boldsymbol{\mathcal{Z}}_i^\top||^2$ and $||\boldsymbol{\Omega}_i^\top||^2$, where $\boldsymbol{\mathcal{E}}_i = \widehat{\boldsymbol{\Omega}}_i^{-1/2}\widehat{\boldsymbol{\xi}}_i$, $\boldsymbol{\mathcal{X}}_i = \widehat{\boldsymbol{\Omega}}_i^{-1/2}\mathbf{X}_i$ and $\boldsymbol{\mathcal{Z}}_i = \widehat{\boldsymbol{\Omega}}_i^{-1/2}\mathbf{Z}_i$ as additional diagnostic tools when influential units are highlighted by C_i. The analysis based on $\mathcal{V}_i = ||\mathbf{I}_{m_i} - \boldsymbol{\mathcal{E}}_i\boldsymbol{\mathcal{E}}_i^\top||^2$ is closely related to residual analysis as already pointed out in Table 1. Finally, these authors mention that their results were derived assuming MLEs for the covariance matrix parameters, and that they may not follow when a restricted maximum likelihood approach is considered. However, because the proposed methods should only be viewed with an exploratory spirit, their use in either case may bring more benefits than losses.

5 Some Remedial Measures

Adequate modelling of data necessarily depends on diagnostic tools, because, in many cases, little or no theoretical basis is available to suggest the specific form of the inter-relationships of the intervening (response and explanatory) variables. Diagnostic tools are undoubtedly useful to highlight the inadequacy of some characteristics of the tentative models. The question is how to proceed with the analysis when these tools suggest that the proposed models do not seem to accommodate one or more features of the data. In this section, we indicate a few alternatives to cope with this problem.

5.1 Fine Tuning of the Model

Diagnostic tools like those summarised in the previous section rely on the correct specification of the within-unit covariance structure of the model as pointed out by Zewotir and Galpin (2007), for example. Although Christensen *et al.* (1992) note that good estimates of the covariance parameters might serve the (exploratory) purpose of a diagnostic, some effort in the refinement of the associated model should be employed to prevent or at least to reduce erroneous indication of influential or outlying units or (within-unit) observations.

For such purposes, Rocha and Singer (2016) suggest some guidelines in the case of random polynomial coefficient models, that is, when the columns of $\mathbf{X}_i$ and $\mathbf{Z}_i$ in (1) correspond to values of powers of the time variable; in particular, they propose (i) simple t-tests based on the estimated coefficients of standard linear regression models fitted to each unit's data as a tool for selecting fixed effects and (ii) Bonferroni-corrected reference intervals for selecting random effects. They also note that when $\mathbf{R}_i = \sigma^2\mathbf{I}_{m_i}$, the structure of the columns of the within-unit covariance matrix is similar to that of the individual profiles. More explicitly, letting $\mathbf{z}_{is}^\top$ denote the s-th row of $\mathbf{Z}_i$, it follows that the random effects contribution to the s-th column of the marginal dispersion matrix (2) is $\mathbf{Z}_i\mathbf{G}\mathbf{z}_{is}$, which has the same form as the individual profile component $\mathbf{Z}_i\mathbf{b}_i$. When the data for all subjects are collected at the same time points, we have $\mathbf{Z}_i = \mathbf{Z}^*$ so that $\mathbf{V} = \mathbf{Z}^*\mathbf{G}\mathbf{Z}^{*\top}$ may be estimated by

$$\widehat{\mathbf{V}} = n^{-1}\sum_{i=1}^{n}(\mathbf{y}_i - \overline{\mathbf{y}})(\mathbf{y}_i - \overline{\mathbf{y}})^\top - \widehat{\sigma}^2\mathbf{I}_m$$

where m is the common number of observations per unit, $\overline{\mathbf{y}} = n^{-1}\sum_{i=1}^{n}\mathbf{y}_i$ and $\widehat{\sigma}^2$ is an estimate of σ^2. Rochar and Singer (2016) suggest fitting standard linear regression models to the rows of $\widehat{\mathbf{V}}$ and using Bonferroni-corrected reference intervals for the coefficients to decide what random effects should be considered in the model. Based on a simulation study, they show that the procedure is reasonably efficient even for moderate sample sizes, for example, 25 units with five repeated measures each, and for non-Gaussian random effects or error terms.

Grady and Helms (1995) suggest that plots of covariances and correlations versus time between measurements (lags) may be used as a tool for identifying possible auto-regressive covariance patterns and propose a strategy to compare different longitudinal data models. Autocorrelation plots may also be used for such purposes. Singer and Cúri (2006) consider an experimental data example in which the relations between the estimated variances and covariances obtained under a saturated model for the fixed parameters and an explanatory variable are explored to suggest the form of the $\mathbf{Z}_i$ and $\mathbf{G}$ matrices in (1).

5.2 Elliptically Symmetric and Skew-elliptical Linear Mixed Models

Linear mixed models based on elliptically symmetric or skew-elliptical distributions have been proposed as alternatives to the standard Gaussian set-up. Pinheiro *et al.* (2001), Savalli *et al.* (2006), Osorio *et al.* (2007), Arellano-Valle *et al*. (2005, 2007), Bolfarine *et al.* (2007) and Lachos *et al.* (2010), among others, have adopted this approach.

The class of elliptically symmetric distributions includes the Gaussian, multivariate-t, power exponential as well as other distributions with densities of the form

$$f(\mathbf{y}) = |\mathbf{\Sigma}|^{-1/2}g[(\mathbf{y} - \boldsymbol{\mu})^\top\mathbf{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})] \tag{34}$$

where $\mathbf{y}$ is an m-dimensional random vector, $\mathbf{u}$ and $\boldsymbol{\Sigma}$ are, respectively, location and scale parameters and g is a non-negative valued function defined in $I\!R$ such that $\int_0^\infty u^{m/2-1}g(u)du < \infty$. When the second moment exists, $\mathbb{E}(\mathbf{y}) = \boldsymbol{\mu}$, $\mathbb{V}(\mathbf{y}) = \boldsymbol{\Omega} = \alpha\boldsymbol{\Sigma}$ with α denoting a convenient constant ($= \nu/(\nu-2)$ for the multivariate-t distribution with $\nu > 2$ degrees of freedom, for example). We use the notation $ES_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \alpha)$ to represent an elliptically symmetric distribution with parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and α.

Linear mixed models given by (1) with random effects and error terms following (34) may be defined hierarchically as in the standard Gaussian case by

$$\mathbf{y}_i|\mathbf{b}_i \sim ES_{m_i}[\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \mathbf{R}_i(\boldsymbol{\theta}), \gamma_i],\ \mathbf{b}_i \sim ES_q[\mathbf{0}, \mathbf{G}(\boldsymbol{\theta}), \alpha_i],$$

$i = 1, \ldots, n$ with $\mathbf{b}_i$ and $\mathbf{e}_i \sim ES_{m_i}[\mathbf{0}, \mathbf{R}_i(\boldsymbol{\theta}), \gamma_i]$ uncorrelated. Although in some special cases (e.g. the multivariate-t distribution), the joint distribution of $(\mathbf{y}_i^\top, \mathbf{b}_i^\top)^\top$ belongs to the same class; this is not valid in general. For computational convenience and generalisability, many authors consider the following alternative specification for the model

$$\begin{bmatrix} \mathbf{y}_i \\ \mathbf{b}_i \end{bmatrix} \sim ES_{m_i+q}\left\{\begin{bmatrix} \mathbf{X}_i\boldsymbol{\beta} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{Z}_i\mathbf{G}(\boldsymbol{\theta})\mathbf{Z}_i^\top + \mathbf{R}_i(\boldsymbol{\theta}) & \mathbf{Z}_i\mathbf{G}(\boldsymbol{\theta}) \\ \mathbf{G}(\boldsymbol{\theta})\mathbf{Z}_i^\top & \mathbf{G}(\boldsymbol{\theta}) \end{bmatrix}, \alpha_i, \gamma_i\right\}$$

and work with the marginal model $\mathbf{y}_i \sim ES_{m_i}[\mathbf{X}_i\boldsymbol{\beta}, \mathbf{D}_i(\boldsymbol{\theta}), \alpha_i, \gamma_i]$, $\mathbf{D}_i(\boldsymbol{\theta}) = \mathbf{Z}_i\mathbf{G}(\boldsymbol{\theta})\mathbf{Z}_i^\top + \mathbf{R}_i(\boldsymbol{\theta})$, usually assuming that $\mathbf{R}_i(\boldsymbol{\theta}) = \sigma^2\mathbf{I}_{m_i}$. In theory, the parameters α_i and γ_i may be different for different units, but in practice, it is common to set $\alpha_i = \alpha$ and $\gamma_i = \gamma$ or to consider only a few values for such parameters, for example, one for each group of units, as suggested by Pinheiro *et al.* (2001).

Maximum likelihood estimation parallels the Gaussian case and may be obtained by iterating the following expressions:

$$\boldsymbol{\beta}^{(h+1)} = \left\{\sum_{i=1}^n v\left(u_i^{(h)}\right)\mathbf{X}_i^\top[\mathbf{D}_i(\boldsymbol{\theta})^{(h)}]^{-1}\mathbf{X}_i\right\}^{-1}\left\{\sum_{i=1}^n v\left(u_i^{(h)}\right)\mathbf{X}_i^\top[\mathbf{D}_i(\boldsymbol{\theta})^{(h)}]^{-1}\mathbf{y}_i\right\}$$

and

$$\boldsymbol{\theta}^{(h+1)} = \text{argmax}_{\boldsymbol{\theta}}\{l(\boldsymbol{\beta}^{(h)}, \boldsymbol{\theta})\}$$

where $l(\boldsymbol{\beta}, \boldsymbol{\theta})$ denotes the log-likelihood, $u_i^{(h)} = (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}^{(h)})^\top[\mathbf{D}_i(\boldsymbol{\theta})^{(h)}]^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}^{(h)})$ and $v(u) = -2g'(u)/g(u)$, for $h = 0, 1, 2 \ldots$.

For inferential purposes, we must rely on asymptotic results even in special cases, and it is usual [see Pinheiro *et al.* (2001) for example] to consider approximate Gaussian distributions for the MLEs $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$. Savalli *et al.* (2006) suggest that for large n

$$\widehat{\boldsymbol{\beta}} \approx N_p\left(\boldsymbol{\beta}, \mathbf{K}_{\boldsymbol{\beta}}^{-1}\right)$$

where $\mathbf{K}_{\boldsymbol{\beta}} = \sum_{i=1}^n 4d_{gi}/m_i\mathbf{X}_i^\top[\mathbf{D}_i(\boldsymbol{\theta})]^{-1}\mathbf{X}_i$ with $d_{gi} = \mathbb{E}\{[g'(u_i)/g(u_i)]^2u_i\}$ is the corresponding Fisher information matrix. A similar result may be considered for $\widehat{\boldsymbol{\theta}}$. Mitchell (1989) presents a simple expression to compute $\mathbb{E}\{[g'(u_i)/g(u_i)]^2u_i\}$ for a number of

elliptically symmetric distributions. For Student's t- distribution with $\nu > 2$ degrees of freedom, it is given by

$$d_{gi} = \frac{m_i}{4}\left(\frac{\nu + m_i}{\nu + m_i + 2}\right).$$

Assuming that $\boldsymbol{\theta}$ is known, predictors for the random effects may be obtained via a Bayesian argument and are expressed like those obtained in a Gaussian set-up, namely

$$\widehat{\mathbf{b}}_i = \mathbb{E}\left(\mathbf{b}_i|\mathbf{y}_i, \widehat{\boldsymbol{\beta}}, \boldsymbol{\theta}\right) = \mathbf{G}(\boldsymbol{\theta})\mathbf{Z}_i^\top[\mathbf{D}_i(\boldsymbol{\theta})]^{-1}\left(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}\right).$$

The covariance matrix of $\widehat{\mathbf{b}}_i$ is not as easy to obtain as in the Gaussian case because of the weights $v(u_i)$. Assuming fixed weights and letting $\boldsymbol{\Gamma}(\boldsymbol{\theta}) = \mathbf{I}_n \otimes \mathbf{G}(\boldsymbol{\theta})$, $\mathbf{D}(\boldsymbol{\theta}) = \oplus_{i=1}^n \mathbf{D}_i(\boldsymbol{\theta})$, $\mathbf{X} = [\mathbf{X}_1^\top, \ldots, \mathbf{X}_n^\top]^\top$ and $\mathbf{Z} = \oplus_{i=1}^n \mathbf{Z}_i$, it follows that

$$\mathbb{V}(\widehat{\mathbf{b}}) = \boldsymbol{\Gamma}(\boldsymbol{\theta})\mathbf{Z}^\top[\mathbf{D}(\boldsymbol{\theta})]^{-1}\mathbb{V}\left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right)[\mathbf{D}(\boldsymbol{\theta})]^{-1}\mathbf{Z}\boldsymbol{\Gamma}(\boldsymbol{\theta})$$

with

$$\mathbb{V}\left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right) = \mathbb{V}(\widehat{\boldsymbol{\xi}}) = \mathbf{D}(\boldsymbol{\theta})^*\mathbf{Q}(\boldsymbol{\theta})^*\boldsymbol{\Omega}(\boldsymbol{\theta})\mathbf{Q}(\boldsymbol{\theta})^*\mathbf{D}(\boldsymbol{\theta})^*,$$

$\mathbf{D}(\boldsymbol{\theta})^* = \oplus_{i=1}^n[v(u_i)\mathbf{D}_i(\boldsymbol{\theta})]$, $\mathbf{Q}(\boldsymbol{\theta})^* = [\mathbf{D}(\boldsymbol{\theta})^*]^{-1} - [\mathbf{D}(\boldsymbol{\theta})^*]^{-1}\mathbf{X}(\mathbf{X}^\top[\mathbf{D}(\boldsymbol{\theta})^*]^{-1}\mathbf{X})^{-1}\mathbf{X}^\top[\mathbf{D}(\boldsymbol{\theta})^*]^{-1}$. For practical applications, one should replace the unknown terms by their estimates.

Residual and leverage analyses for elliptically symmetric LMM are still not well established. However, the similarity with the Gaussian case suggests some exploratory tools. For example, index plots of the weighted marginal residuals $\widehat{\boldsymbol{\xi}}_i = \sqrt{v(\widehat{u}_i)}[\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}]$ may be used to detect outliers as suggested by Savalli *et al.* (2006). Note that, for the most commonly used non-Gaussian elliptically symmetric distributions such as the multivariate-t or the power exponential distributions, the weights $v(\widehat{u}_i) = (\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}})^\top[\mathbf{D}_i(\widehat{\boldsymbol{\theta}})]^{-1}(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}})$ tend to reduce the influence of units associated to larger values of Mahalanobis's distance and therefore may accommodate outliers possibly detected in a standard Gaussian mixed model set-up. A similar analysis may be carried out for the conditional residuals, $\widehat{\mathbf{e}}_i = \mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}} - \mathbf{Z}_i\widehat{\mathbf{b}}_i$. We believe that the suggestions of Schützenmeister and Piepho (2012) could be useful in this context, but this is a topic that deserves further research.

Local influence, on the other hand, has been considered in detail by Osorio *et al.* (2007). These authors derive the appropriate general expressions for $\mathbf{H}$ and $\ddot{\mathbf{F}}$ in (30) corresponding to perturbations in case weights (w_i), in the scale matrix ($\mathbf{D}_i$) or in the explanatory ($\mathbf{X}_i$) and response ($\mathbf{y}_i$) variables. In practice, one should choose the desired member of the class of elliptically symmetric distributions by specifying g, α and $\mathbf{D}(\boldsymbol{\theta})$ to obtain the corresponding local influence measure (33).

As pointed out by Mudholkar and Hutson (2000), the effects of asymmetry on the appropriateness of normal theory methods are, in general, more serious than those of heavy tails, and for this reason, a considerable research effort has been directed at LMM with alternative asymmetric distributions for the random effects and error terms. Jara *et al.* (2008) adopt a Bayesian approach to deal with skew-elliptical distributions in this context. In particular, they consider a two-stage version of model (1) defined as

$$\mathbf{y}_i \left| \left(\boldsymbol{\beta}, \mathbf{b}_i, \sigma^2, \mathbf{R}_i, \boldsymbol{\Lambda}_{e_i}, g_{\theta_e}^{m_i} \right) \right. \sim SE_{m_i} \left(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i, \sigma^2 \mathbf{R}_i, \boldsymbol{\Lambda}_{e_i}, g_{\theta_e}^{m_i} \right) \tag{35}$$

$$\mathbf{b}_i \left| \left(\mathbf{G}, \boldsymbol{\Lambda}_b, g_{\theta_b}^{q} \right) \right. \sim SE_q \left(\mathbf{0}, \mathbf{G}, \boldsymbol{\Lambda}_b, g_{\theta_b}^{q} \right) \tag{36}$$

where $SE_r(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, g_\theta)$ denotes an r-dimensional skew-elliptical distribution with r-dimensional location vector $\boldsymbol{\mu}$, r-dimensional scale matrix $\boldsymbol{\Sigma}$, r-dimensional asymmetry matrix $\boldsymbol{\Lambda}$ and density generator function g_θ.

Although the models proposed by Jara *et al.* (2008) are quite general, fitting them in practice is quite complicated so that one must consider narrower alternatives, such as the class of skew-normal distributions. Arellano-Valle *et al.* (2007) also adopt a Bayesian approach to analyse skew-normal LMM assuming that

$$\mathbf{y}_i \left| \left[\boldsymbol{\beta}, \mathbf{b}_i, \mathbf{R}_i(\boldsymbol{\theta}), \boldsymbol{\Lambda}_{e_i} \right] \right. \sim SN_{m_i} \left[\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i, \mathbf{R}_i(\boldsymbol{\theta}), \boldsymbol{\Lambda}_{e_i} \right] \tag{37}$$

$$\mathbf{b}_i | [\mathbf{G}(\boldsymbol{\theta}), \boldsymbol{\Lambda}_b] \sim SN_q[\mathbf{0}, \mathbf{G}(\boldsymbol{\theta}), \boldsymbol{\Lambda}_b] \tag{38}$$

where $SN_r(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda})$ denotes a skew-normal distribution with r-dimensional location vector $\boldsymbol{\mu}$, r-dimensional scale matrix $\boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda} = \oplus_{i=1}^{r} \lambda_i$ with $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_r)^\top$ representing the asymmetry parameters. The corresponding density function is

$$f(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}) = 2^r \phi_r(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma} + \boldsymbol{\Lambda}^2) \Phi_r[\boldsymbol{\Lambda}((\boldsymbol{\Sigma} + \boldsymbol{\Lambda}^2)^{-1}(\mathbf{y} - \boldsymbol{\mu})|\mathbf{0}, \mathbf{I}_r + \boldsymbol{\Lambda} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda}]$$

where $\phi_r(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\Phi_r(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, respectively, denote the density and distribution functions of a Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. This implies that

$$\mathbb{E}(\mathbf{y}) = \boldsymbol{\mu} + \sqrt{2/\pi}\, \boldsymbol{\lambda}, \text{ and } \mathbb{V}(\mathbf{y}) = \boldsymbol{\Sigma} + (1 - 2/\pi)\, \boldsymbol{\Lambda}^2.$$

For a special case where $\mathbf{R}_i = \sigma^2 \mathbf{I}_{m_i}$ and $\boldsymbol{\Lambda}_{e_i} = \lambda\, \mathbf{I}_{m_i}$, Arellano-Valle *et al.* (2007) show how Bayesian methodology may be used to fit the model to the marginal distribution of $\mathbf{y}$, mentioning that 'although the associated density functions are quite difficult to handle, we show that the models can be easily fitted using MCMC methods'.

Bolfarine *et al.* (2007) consider the LMM (1) with slightly different formulation for the underlying skew-normal distributions. More specifically, they consider distributions for which the density function is

$$f(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}) = 2\phi_r(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma} + \boldsymbol{\lambda}\boldsymbol{\lambda}^\top) \Phi_1 \left[\frac{\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})}{(1 + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\lambda})^{1/2}} \right]$$

so that

$$\mathbb{E}(\mathbf{y}) = \boldsymbol{\mu} + \sqrt{2/\pi}\, \boldsymbol{\lambda}, \text{ and } \mathbb{V}(\mathbf{y}) = \boldsymbol{\Sigma} + (1 - 2/\pi)\, \boldsymbol{\lambda}\boldsymbol{\lambda}^\top.$$

They restrict their attention to the special case where $\mathbf{R}_i = \sigma^2 \mathbf{I}_{m_i}$ and write the model in a two-stage formulation that facilitates the derivation of the marginal density of $\mathbf{y}_i$, namely

$$f(\mathbf{y}_i|\boldsymbol{\mu}, \mathbf{G}(\boldsymbol{\theta}), \boldsymbol{\lambda}) = 2\phi_{m_i}\{\mathbf{y}_i|\mathbf{X}_i \boldsymbol{\beta}, \mathbf{Z}_i([\mathbf{G}(\boldsymbol{\theta})]^{-1} + \boldsymbol{\lambda}\boldsymbol{\lambda}^\top)\mathbf{Z}_i^\top\} \Phi_1\{\overline{\lambda}_i(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})\}$$

where

$$\overline{\boldsymbol{\lambda}}_i = \frac{\boldsymbol{\lambda}^\top[\mathbf{G}(\boldsymbol{\theta})]^{-1}[\mathbf{D}_i(\boldsymbol{\theta})]^{-1}\mathbf{Z}_i^\top[\mathbf{R}_i(\boldsymbol{\theta})]}{\{1+\boldsymbol{\lambda}^\top[\mathbf{G}(\boldsymbol{\theta})]^{-1}\boldsymbol{\lambda}+\boldsymbol{\lambda}^\top[\mathbf{G}(\boldsymbol{\theta})]^{-1}[\mathbf{D}_i(\boldsymbol{\theta})]^{-1}[\mathbf{G}(\boldsymbol{\theta})]^{-1}\boldsymbol{\lambda}\}^{1/2}}$$

with

$$\mathbf{D}_i(\boldsymbol{\theta}) = \{[\mathbf{G}(\boldsymbol{\theta})+\boldsymbol{\lambda}\boldsymbol{\lambda}^\top]^{-1}+\mathbf{Z}_i[\mathbf{R}_i(\boldsymbol{\theta})]^{-1}\mathbf{Z}_i\}.$$

In this context, they develop an EM algorithm to compute maximum likelihood estimates of the model parameters and propose local influence diagnostic measures based on perturbation of the dispersion matrix for the random effects ($\mathbf{G}(\boldsymbol{\theta})$), of case weights ($w_i$), of the explanatory variables ($\mathbf{X}_i$) and of the response variables ($\mathbf{y}_i$). They also consider a case-deletion diagnostic measure and mention that model adequacy is still an open problem.

As with LMMs based on elliptically symmetric distributions, other diagnostic tools such as residual or leverage analysis are still not available.

5.3 Robust Linear Mixed Models

More recently, Koller (2013) proposed robust methods for analysing LMM. He considered model (3), reparametrising it as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{U}_b(\boldsymbol{\theta})\mathbf{b}^*+\mathbf{U}_e\mathbf{e}^*,$$

with $\mathbf{b}^* \sim N(\mathbf{0},\sigma^2\mathbf{I}_q)$, $\mathbf{e}^* \sim N(\mathbf{0},\sigma^2\mathbf{I}_N)$ so that $\boldsymbol{\Gamma}(\boldsymbol{\theta}) = \mathbf{U}_b(\boldsymbol{\theta})\mathbf{U}_b(\boldsymbol{\theta})^\top$ and $\mathbf{R} = \mathbf{U}_e\mathbf{U}_e^\top$, that is, $\mathbf{U}_b(\boldsymbol{\theta})$ and $\mathbf{U}_e$ represent the lower parts of the Cholesky decomposition of $\boldsymbol{\Gamma}(\boldsymbol{\theta})$ and $\mathbf{R}$, respectively.

We outline the proposal in a special case where the covariance structure is similar for all units and where the robustifying bounded functions ψ are equal for random effects and error terms. First, we let $d_i = \mathbf{b}_i^{*\top}, i = 1,\ldots,n$ and consider the weights

$$w(d) = \begin{cases} \psi(\sqrt{d})/\sqrt{d} & \text{if } d \neq 0 \\ \psi'(0) & \text{if } d = 0. \end{cases}$$

Then, we let $\mathbf{w}(\mathbf{d}) = [w(d_1/\sigma),\ldots,w(d_n/\sigma)] \otimes \mathbf{1}_q^\top$ and $\mathbf{W}_b(\mathbf{d}) = diag[\mathbf{w}(\mathbf{d})]$. The corresponding robust estimating equations are

$$\begin{aligned} \mathbf{X}^\top\mathbf{U}_e{}^{-\top}\boldsymbol{\psi}(\widehat{\mathbf{e}}^*/\sigma) &= \mathbf{0} \\ \mathbf{U}_b{}^\top\mathbf{Z}^\top\mathbf{U}_e{}^{-\top}\boldsymbol{\psi}(\widehat{\mathbf{e}}^*/\sigma)-\mathbf{W}_b(\widehat{\mathbf{d}})\mathbf{b}^*/\sigma &= \mathbf{0} \end{aligned} \tag{39}$$

For given $\boldsymbol{\theta}$ and σ, the estimation of the fixed and random effects can be obtain using iteratively reweighted least squares. Defining $\mathbf{W}_e = diag[\mathbf{w}(\widehat{\mathbf{e}}^*)]$ similarly to $\mathbf{W}_b$, the estimating equations are then given by

$$\begin{bmatrix} \mathbf{X}^\top\mathbf{U}_e{}^{-\top}\mathbf{W}_e\mathbf{U}_e^{-1}\mathbf{X} & \mathbf{X}^\top\mathbf{U}_e{}^{-\top}\mathbf{W}_e\mathbf{U}_e^{-1}\mathbf{Z}\mathbf{U}_b \\ \mathbf{U}_b\mathbf{Z}^\top\mathbf{U}_e{}^{-\top}\mathbf{W}_e\mathbf{U}_e^{-1}\mathbf{X} & \mathbf{U}_b^\top\mathbf{Z}^\top\mathbf{U}_e{}^{-\top}\mathbf{W}_e\mathbf{U}_e^{-1}\mathbf{Z}\mathbf{U}_b+\mathbf{W}_b \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\mathbf{b}}^* \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top\mathbf{U}_e{}^{-\top}\mathbf{W}_e\mathbf{y} \\ \mathbf{U}_b\mathbf{Z}^\top\mathbf{U}_e{}^{-\top}\mathbf{W}_e\mathbf{y} \end{bmatrix}. \tag{40}$$

An efficient algorithm to obtain estimates of the fixed and random effects consists in alternating the computation of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{b}}^*$ with updated weights. The initial values may correspond to all weights equal to a constant and we stop the algorithm when the relative change of the estimates is small.

In practice, the values of $\boldsymbol{\theta}$ and σ are unknown and may be estimated by robustifying the restricted maximum likelihood equations. Details may be obtained in Koller (2013).

6 Computation and Examples

The numerous approaches available to analyse repeated measures via LMM may turn out to be a source of concern when fitting such models to real data. First, it is difficult to decide which class of models and what analysis strategy to adopt. Second, suitable software is not always available or may not accommodate the peculiarities that accompany practical problems. It is possible to adapt existing software to deal with specific details of the problem under investigation; however, this may require more time than the practitioner is allowed to.

The first step in specifying the model, that is, choosing $\mathbf{X}$ and $\mathbf{Z}$, may be approached via the construction of individual and mean *profile plots* originally proposed by Rao and Rao (1966). An examination of the mean profile may suggests a structure for the fixed effects. For example, the degree of a polynomial relating (at least approximately) the mean response to the time metameter (dose, for example) may help in the specification of $\mathbf{X}$. The individual profiles may also suggest the degree of the unit-specific polynomials as well as possible heteroskedasticity and/or the within-unit covariance structure, giving an idea about how $\mathbf{Z}$ and $\mathbf{G}$ may be defined. Guidelines in this direction may be found in Fearn (1977), Weiss and Lazaro (1992) or Rutter and Elashoff (1994).

The modelling strategy should include an iterative procedure according to which after each new model is fitted, appropriate diagnostic tools should be employed to check whether the new proposal is more adequate than the previous one. This procedure may be complemented by comparing the models via criteria like AIC, BIC or likelihood ratios, when appropriate. Obviously, an optimal fit may never be attained, but the increasing acquaintance with the data will certainly improve the analysis.

We also note that the diagnostic tools described in Section 4 depend on the covariance structure of the model which, in general, is unknown. For practical applications, we must replace the corresponding covariance parameters with estimates and start the diagnostic procedure with an examination of plots of the modified Lesaffre–Verbeke index $\mathcal{V}_i^*$ versus unit indices. When there is some indication that the proposed covariance structure is not adequate for some units, we may try to respecify it. In this process, we must face the dilemma of modifying either the between-units or the within-units variance components or both, namely, $\mathbf{G}$ and $\mathbf{R}_i$ in (2). Strategies to accomplish this depend on the specific problems under investigation; we outline some alternatives in the examples. Among the various statistical software packages for fitting linear and non-linear mixed models, including `GenStat`, `SPSS` and `ASREML`, `SAS proc MIXED` and the libraries `lme4` and `nlme` in the free software package `R` are the most popular. We focus our attention on the latter given its flexibility and the possibility of modifying the available functions or programming new ones. The appropriate `R` functions to fit linear mixed models as well as more general mixed models are indicated in Table 4.

Software for diagnostics is generally only available from a few statistical packages or from authors and mainly designed for use in specific cases. A recent review is given by West and Galecki (2011). We have developed R-based functions to produce residual, leverage and case deletion plots for Gaussian LMM. The functions are designed to work

Table 4. *R-based software for fitting mixed models.*

Library	Function	Fits	Random effects distribution	$\mathbf{G}$ or $\mathbf{R}_W$ matrix	Error distribution	$\mathbf{R}_i$ matrix
lme4	lmer	LMM	Gaussian	Unstructured $\mathbf{G}$	Gaussian	$\sigma^2\mathbf{I}_{m_i}$
	nlmer	NLMM	Gaussian	Unstructured $\mathbf{G}$	Gaussian	Structured
	glmer	GLMM	Gaussian	Unstructured $\mathbf{G}$	exponential family	NA
nlme	lme	LMM	Gaussian	Structured $\mathbf{G}$	Gaussian	Structured
	nlme	NLMM	Gaussian	Structured $\mathbf{G}$	Gaussian	Structured
	gls	LM	NA	NA	Gaussian	Structured
gee	gee	GEE-based model	NS	Structured $\mathbf{R}_W$ matrix	Exponential family or NS	NA
geepack	geeglm	GEE-based model	NS	Structured $\mathbf{R}_W$ matrix	Exponential family or NS	NA
heavy	heavyLme	ES-LMM	Elliptically symmetric	Unstructured $\mathbf{G}$	Elliptically symmetric	NA
robustlmm	rlmer	Robust LMM	Symmetric	Diagonal or unstructured $\mathbf{G}$	Symmetric	$\sigma^2\mathbf{I}_{m_i}$

$\mathbf{R}_W$, working covariance marix; NA, not applicable; NS, not specified.

with objects generated via the `lme4` and `nlme` libraries and may be downloaded from www.ime.usp.br/∼jmsinger/lmmdiagnostics.zip.

We discuss three examples to illustrate how the diagnostic tools may be used to modify the Gaussian model proposed initially. In the first example, directed at the prediction of random effects, we show how an heteroskedastic model identified via an analysis of the residual plots may accommodate outliers in the conditional errors and generate better predictors of random effects. Although it is possible to identify these outliers by looking at the data, this example allows a clear view of the features of the diagnostic procedures. In the second example, we show how some ad hoc changes in the covariance structure may cope with an apparent long-tailedness of the underlying distributions. In both cases, we compare the results with those obtained from other elliptically symmetric and robust models. The third example is considered to indicate how diagnostic procedures may be useful to identify and accommodate serial correlation in the conditional errors.

Computation for the three examples were carried out via the function `lme` available in the library `nlme` in `R` using restricted maximum likelihood (REML) and a Cholesky decomposition for the covariance matrices. Although there is some debate with respect to the appropriate criteria to compare mixed models [see Muller *et al.* (2013), for example], we used the default AIC and BIC in `lme`, defined as

$$AIC = -2l_R(\widehat{\theta}) + 2(p+t)$$

$$BIC = -2l_R(\widehat{\theta}) + 2(p+t)\log(N),$$

where p denotes the number of fixed parameters, t denotes the number of covariance parameters in the model, N is the total number of observations used to fit the model and $l_R(\widehat{\theta})$ denotes the REML log-likelihood.

Decisions based on residual, Cook's distance and leverage plots must be considered subjectively and with an exploratory spirit; the idea is to identify units or observations that do not follow the overall pattern. To facilitate the analysis, reference limits corresponding to the third quartile plus 1.5 times the interquartile range (dashed line) are indicated in the plots. In the leverage plots, we also include a dotted line corresponding to $2p/n$, which is commonly suggested in the literature.

6.1 Passive Filter Example

The first example is extracted from a study conducted at the Laboratory for Atmospheric Pollution, University of São Paulo, Brazil, with the objective of evaluating the Ozone concentration (as an indicator of atmospheric pollution) in different periods. Given the high cost of the standard equipment used to measure Ozone concentration directly, the reflectance of passive filters impregnated with a solution of indigo carmine, whose colour fades when exposed to that pollutant, is used as an indirect measure. The fading measured by a reflectometer constitutes the response. The results will be employed by the investigators to estimate a calibration curve. Because of the passive filter design, rain and other factors may interfere in the reflectance measures leading to possible outliers.

Each of nine sets consisting of three passive filters placed next to the standard equipment employed by the state environment protection agency was observed in nine different periods of 7 days spaced by 10 days. After each period of 7 days, the responses of the three filters in the corresponding set were recorded. The observed reflectance of each filter is displayed in Table 5.

A standard LMM for this problem is

$$y_{ij} = \mu + a_i + e_{ij}, \quad i = 1, \ldots, 9, \ j = 1, 2, 3 \tag{41}$$

where y_{ij} denotes the reflectance of filter j in period i, μ is the expected reflectance over all periods, $a_i \sim N(0, \sigma_a^2)$ and $e_{ij} \sim N(0, \sigma^2)$, with independence for all random terms. Fitting this model to the data results in $\widehat{\mu} = 46.4$, $\widehat{\sigma}_a^2 = 100.4$, $\widehat{\sigma}^2 = 104.8$. The predicted reflectances obtained as BLUPs under this model are given by

$$\widehat{y}_{ij} = \overline{y} + k(\overline{y}_i - \overline{y}) \tag{42}$$

where $\overline{y}_i = 1/3 \sum_{j=1}^{3} y_{ij}$, $\overline{y} = 1/9 \sum_{i=1}^{9} \overline{y}_i$ and the shrinkage constant

$$k = \frac{\sigma_a^2}{\sigma_a^2 + \sigma^2/3} \tag{43}$$

is estimated as $\widehat{k} = 0.75$. The residual diagnostic plot for the modified Lesaffre–Verbeke index is presented in Figure 1 and suggests that the proposed covariance structure may not be adequate for units 3 and 5.

Table 5. *Reflectance of passive filters.*

Period	Filter	Reflectance	Period	Filter	Reflectance
1	1	27.0	6	1	47.9
1	2	34.0	6	2	60.4
1	3	17.4	6	3	47.3
2	1	24.8	7	1	50.4
2	2	29.9	7	2	50.7
2	3	32.1	7	3	55.9
3	1	35.4	8	1	54.9
3	2	63.2	8	2	43.2
3	3	27.4	8	3	52.1
4	1	51.2	9	1	38.8
4	2	54.5	9	2	59.9
4	3	52.2	9	3	61.1
5	1	77.7			
5	2	53.9			
5	3	48.2			

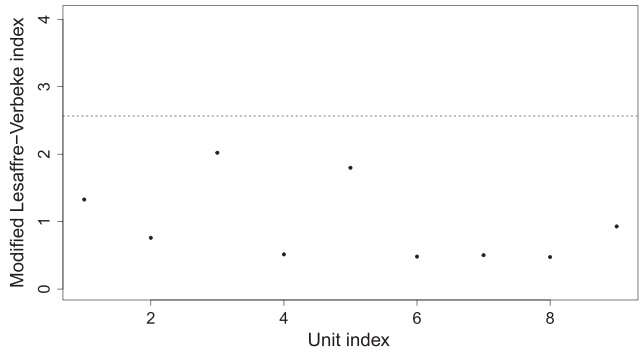

Figure 1. *Passive filter example–modified Lesaffre–Verbeke unit index plot for model (41). Dashed line: third quartile + 1.5 interquartile range.*

Given the experimental design, there is no reason to consider modification of the between-units covariance structure (defined by σ_a^2). Keeping in mind that uncontrolled variables such as wind, humidity and temperature might affect the within-period variability, we propose to modify model (41) by including period-specific variances, for which the shrinkage constant is $k_i = \sigma_a^2/(\sigma_a^2 + \sigma_i^2/3)$. The (BLUP) predicted reflectances under this model are given by

$$\widehat{y}_{ij} = \widehat{\mu} + k_i\,(\overline{y}_i - \widehat{\mu}) \tag{44}$$

where $\widehat{\mu} = \sum_{i=1}^{9}(w_i/\sum_{i=1}^{9} w_i)\overline{y}_i$ is a weighted mean with $w_i = (\sigma_a^2 + \sigma_i^2/3)^{-1}$.

For parsimony reasons, instead of a completely heteroskedastic model, we consider setting model (41) with

$$\sigma_i^2 = \tau^2, i = 3, 5 \text{ and } \sigma_i^2 = \sigma^2, \text{ otherwise.} \tag{45}$$

As a result, we obtain $\widehat{\mu} = 45.9$, $\widehat{\sigma}_a^2 = 114.3$, $\widehat{\sigma}^2 = 49.3$, $\widehat{\tau}^2 = 274.0$, $\widehat{k}_{i\neq 3,5} = 0.87$, $\widehat{k}_{i=3,5} = 0.56$. Here, the expected predicted reflectance for the periods (3 and 5) with largest error variances ($\widehat{y}_3 = 43.8$ and $\widehat{y}_5 = 53.7$) are closer to the mean than in the original model ($\widehat{y}_3 = 43.1$ and $\widehat{y}_5 = 56.4$), accommodating the possible effect of the inappropriate covariance structure in the original model. In the corresponding plot of the modified Lesaffre–Verbeke index (Figure 2), periods 3 and 5 are no longer flagged, but period 9 and possibly period 1 show up, suggesting that further modelling might be considered.

We fitted a final model with different within-period variances for periods (1, 9) and (3, 5), namely, (41) with

$$\sigma_i^2 = \tau^2, i = 3, 5, \sigma_i^2 = \nu^2, i = 1, 9 \text{ and } \sigma_i^2 = \sigma^2, \text{ otherwise,} \tag{46}$$

obtaining the plots presented in the Supporting Information, Figures 1–10. Given the small amount of data, we feel that there is not enough evidence against this model. The reduction in AIC, BIC and log-likelihood from 214.9, 218.6, −104.4 to 208.5, 214.7 and −99.2, respectively, also suggest that the reduction of model (41) to (41 with 46) seems reasonable. As expected, given the small number of observations, a full heteroskedastic model does not lead to a better fit (AIC = 215.9, BIC = 229.7 and log-likelihood = −97.0).

Note that periods 1 and 2 are identified as possible outliers (Figure 3 of the Supporting Information); this, however, is expected because they correspond to periods in which expected Ozone concentration is different from the average; therefore, no changes in the model are required.

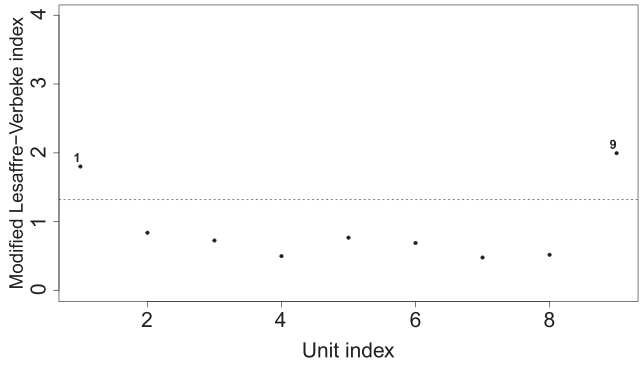

Figure 2. *Passive filter example–modified Lesaffre–Verbeke unit index plot for model (45). Dashed line: third quartile + 1.5 interquartile range.*

Table 6. *Passive filters example: best linear unbiased predictors for the reflectance latent values under the different models.*

Period	Sample mean	Homoskedastic LMM	Heteroskedastic LMM (3 variances)	t (df=21.8)	Robust
1	26.1	31.4	31.9	31.8	29.2
2	28.9	33.4	30.2	33.8	31.6
3	42.0	43.1	44.0	43.3	39.4
4	52.6	51.0	52.2	50.9	51.5
5	59.9	56.4	53.5	56.3	54.4
6	51.9	50.4	51.5	50.4	50.9
7	52.3	50.8	51.9	50.8	51.3
8	50.1	49.1	49.8	49.1	49.4
9	53.3	51.5	51.3	51.4	52.7
Mean	46.4	46.4	46.2	46.6	45.7

LMM, linear mixed models.

For comparison purposes, we fitted the original model (41) adopting elliptically symmetric distributions for both the random effects and error terms and also considered a robust estimation approach (using the default options for both the `heavy` and `rlmer` functions). The predicted latent values are compared in Table 6.

The first observation for period 5 (that corresponds to the largest reflectance value in the data) is flagged in Figures 7–10 of the Supporting Information, suggesting that it is influential with respect to the both the fixed and random effects. The predicted latent value obtained via model (41) with (46), namely, 53.5, is closer to the sample mean obtained with the omission of that observation (51.1) than the predictions obtained with the two competing models (56.3 for the model based on the t-distribution and 54.4 for the robust approach), indicating that the proposed model downweighs that observation with more intensity than the two competitors.

6.2 House Prices Example

Next, we analyse an example originally considered in Harrison and Rubinfeld (1978) with the objective of studying the association between air quality and house prices in the Boston Metropolitan Area. The data, collected in 1970, involve 14 variables observed in 506 standard metropolitan statistical areas (SMSA) arising from 92 towns. More details, along with the

definition of the variables, may be obtained from the Ecdat package in R (Croissant, 2015). Belsley *et al.* (1980) fitted a standard linear model to the data via ordinary least squares (OLS) and constructed a standard raw residual QQ plot (Figure 3), suggesting that the error distribution should have heavier tails than that of the adopted Gaussian model.

Longford (1993), arguing that the observations from the SMSAs in the same town should be considered as clusters, reanalysed the data using an LMM having town as a random effect. He obtained a conditional residual QQ plot (Figure 4), concluding that the violation of the Gaussian assumption was not as evident as with the OLS analysis but could not be completely ignored. Both authors did not take the heteroskedasticity and dependence of the raw residual into account when constructing the QQ plots. We fitted the same model proposed by Longford (1993) and considered the more detailed residual analysis outlined in Section 4. Besides the (slight) violation of the Gaussian assumption (Figure 4) for the conditional error terms, the random effects also do not seem to follow a normal distribution, as depicted in Figure 5. A plot of the modified Lesaffre–Verbeke measure (Figure 6) suggests that the adopted covariance structure was not adequate at least for towns labelled 76 (Back Bay, 5 SMSAs), 77 (Beacon Hill, 2 SMSAs), 80 (East Boston, 10 SMSAs), 81 (South Boston, 12 SMSAs), 82 (South Bay, 7 SMSAs), 90 (Chelsea, 4 SMSAs) and 92 (Winthrop, 4 SMSAs).

To take a possible spatial correlation into account, we should model the within-town covariance structure taking the between-SMSA distances into account. As we do not have that information, we adopted a simpler ad hoc model, introducing different compound symmetry within-unit covariance matrices ($\mathbf{R}_i$) for these towns, refitted the model and generated a new plot of the modified Lesaffre-Verbeke index that suggested a poorly fitted covariance structure for some additional towns, namely, 4 (Salem, 7 SMSAs), 15 (Manchester, 1 SMSA), 28 (Sommerville, 14 SMSAs), 29 (Cambridge, 29 SMSAs), 31 (Belmont, 7 SMSAs), 66 (Hull, 1 SMSA), 75 (Allston-Brighton, 7 SMSAs), 78 (North End, 1 SMSA), 79 (Charleston, 6 SMSAs), 83 (Roxbury, 18 SMSAs), 84 (Savin Hill, 22 SMSAs), 85 (Dorchester, 10 SMSAs), 87 (Forest

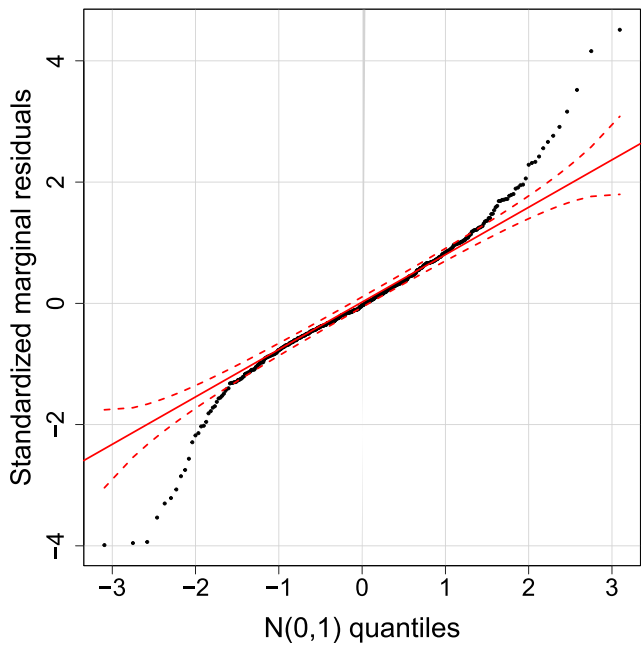

Figure 3. *House prices example–QQ plot for standardised conditional residuals (Belsley model).[Colour figure can be viewed at wileyonlinelibrary.com]*

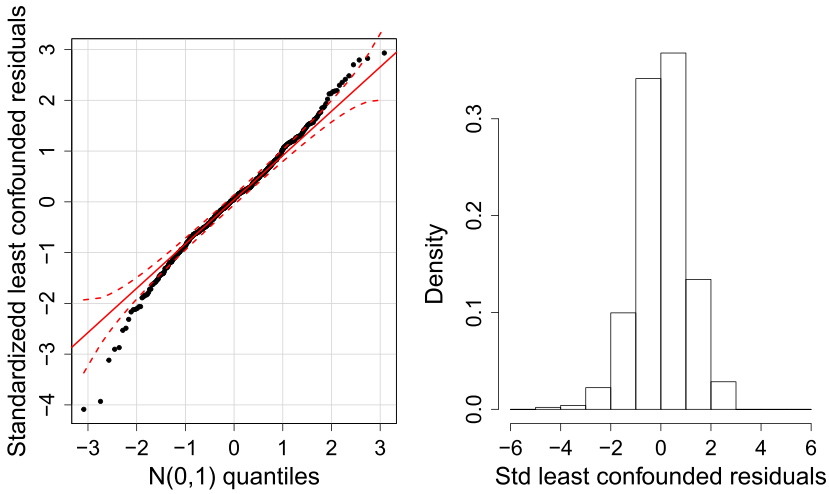

Figure 4. *House prices example–QQ plot and histogram for standardised conditional residuals (Longford model). [Colour figure can be viewed at wileyonlinelibrary.com]*

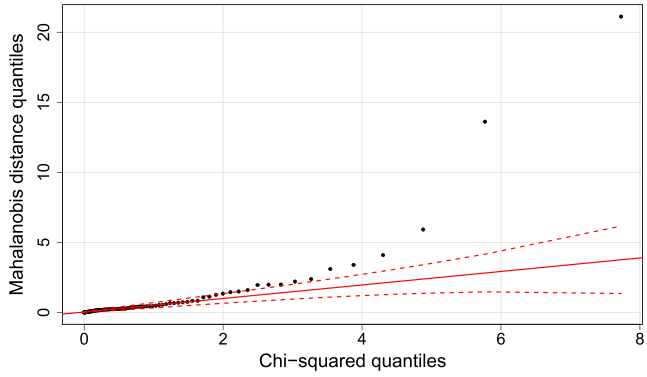

Figure 5. *House prices example–QQ for Mahalanobis distance (Longford model). [Colour figure can be viewed at wileyonlinelibrary.com]*

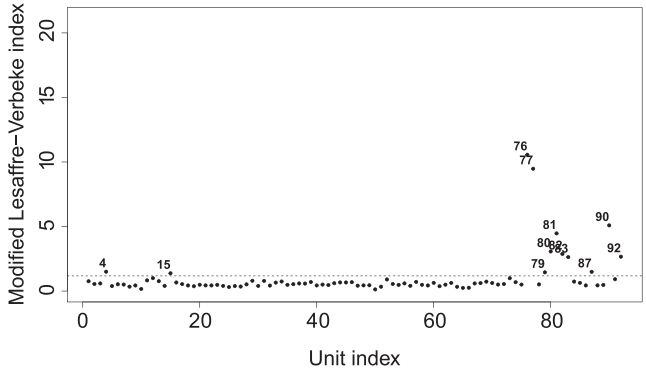

Figure 6. *House prices example–modified Lesaffre–Verbeke unit index plot (Longford model). Dashed line: third quartile + 1.5 interquartile range.*

Hills, 6 SMSAs) and 91 (Revere, 7 SMSAs). We refitted the model with a similar covariance structure for these additional towns; in this case, a plot of the Lesaffre–Verbeke index (Figure 7) does not suggest other units for which the proposed covariance structure seems inadequate; also, QQ plots for the standardised marginal and least confounded conditional residuals (Figures 8 and 9) show no evidence against the adopted Gaussian assumption for both the random effects and error terms. Other diagnostic plots presented in Figures 11–20 in the Supporting Information do not show strong evidence against the adopted assumptions. For this final model, the values for AIC, BIC and log-likelihood were −614.7, −455.2 and 345.4, respectively, in contrast with −314.7, −247.5 and 173.3 for the Longford model, confirming the better fit of the former.

To assess the effect of adopting a heavier tailed distribution for both the random effects and errors, we fitted a similar model using t-distributions with estimated degrees of freedom. We also adopted the robust estimation approach proposed by Koller (2013). Estimates of the model parameters are displayed in Table 7 along with those of the two Gaussian LMM (Longford and Final) and the standard linear model (Belsley).

As already pointed by Pinheiro *et al.* (2001), the major differences relate to a consistent decrease in the magnitude of the standard errors obtained under a t-distribution, which are consistently larger than those of the competitors. Unfortunately, neither the function `heavy` nor the function `rlmer` in `R` allow fitting models with structured covariance matrices for the associated random terms, complicating comparison among the different models.

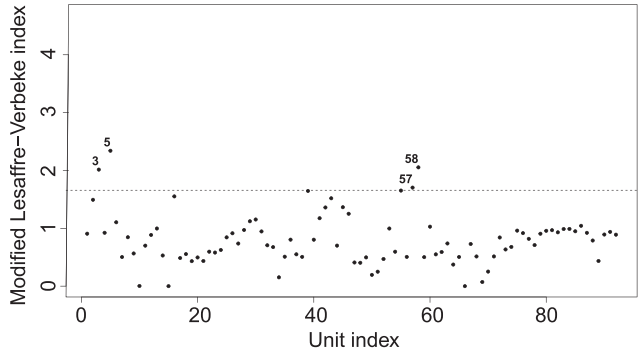

Figure 7. *House prices example–modified Lesaffre–Verbeke unit index plot (final model). Dashed line: third quartile + 1.5 interquartile range.*

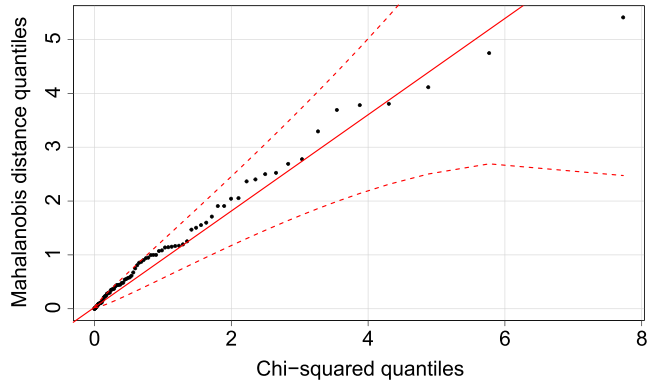

Figure 8. *House prices example–QQ for Mahalanobis distance (final model). [Colour figure can be viewed at wileyonlinelibrary.com]*

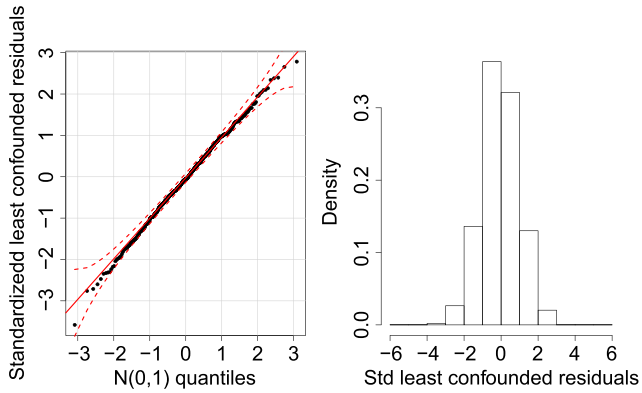

Figure 9. *House prices example–QQ and histogram for conditional least confounded residuals (final model). [Colour figure can be viewed at wileyonlinelibrary.com]*

The estimates are similar across the four LMM. Furthermore, only for the standard linear model, a different conclusion with respect to the significant explanatory variables was obtained.

6.3 Calf Weight Example

Data for the third example were obtained from a study on the growth of calves conducted by the Brazilian Agricultural Research Corporation (EMBRAPA) and are presented in Table 1 of the Supporting Information. Weights (in kg) of 28 calves subjected to an experimental diet were recorded every 2 weeks for 26 weeks after birth. The corresponding profile plot is displayed in Figure 10 and suggest that a second degree polynomial may represent the population-averaged weight growth along the 26 weeks. The 'parallel' aspect of the individual profiles in the plot also suggests that random effects should be included in a model. A longitudinal plot of the sample variances (Figure 11) confirms that the response variances increase with time.

Table 7. *House prices example: parameter estimates and standard errors.*

			Gaussian LMM				*t*-LMM		Robust	
	Belsley		Longford		Final		df=2.19123		LMM	
Variable	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
(Intercept)	9.7560	0.1496	9.6721	0.2140	9.6691	0.1253	9.6820	0.2500	9.6120	0.1577
crim	−0.0119	0.0012	−0.0071	0.0010	−0.0064	0.0015	−0.0061	0.0016	−0.0055	0.0008
zn	0.0001	0.0005	0.0000	0.0007	0.0002	0.0004	0.0005	0.0006	0.0012	0.0005
indus	0.0002	0.0024	0.0024	0.0046	0.0027	0.0026	0.0026	0.0056	0.0036	0.0033
chasyes	0.0914	0.0332	−0.0141	0.0288	0.0173	0.0175	0.0159	0.0415	−0.0043	0.0218
nox	−0.0064	0.0011	−0.0059	0.0012	−0.0049	0.0011	−0.0052	0.0025	−0.0054	0.0009
rm	0.0063	0.0013	0.0092	0.0012	0.0176	0.0008	0.0182	0.0016	0.0142	0.0009
age	0.0001	0.0005	−0.0010	0.0005	−0.0024	0.0003	−0.0023	0.0006	−0.0016	0.0004
dis	−0.1913	0.0334	−0.1248	0.0467	−0.1164	0.0287	−0.1173	0.0577	−0.1087	0.0347
rad	0.0957	0.0191	0.0975	0.0296	0.0617	0.0175	0.0731	0.0268	0.0781	0.0216
tax	−0.0004	0.0001	−0.0004	0.0002	−0.0003	0.0001	−0.0003	0.0002	−0.0004	0.0001
ptratio	−0.0311	0.0050	−0.0299	0.0102	−0.0241	0.0057	−0.0253	0.0107	−0.0263	0.0075
blacks	0.3638	0.1031	0.5822	0.1005	0.6568	0.0881	0.6393	0.1860	0.6592	0.0762
lstat	−0.3710	0.0250	−0.2817	0.0238	−0.1133	0.0157	−0.0985	0.0324	−0.1952	0.0180

SE, standard errors; LMM, linear mixed models.

In view of the descriptive analysis, the data may be modelled by (1) with $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$, where y_{ij} denotes the j-th observation of the weight of the i-th calf, $i = 1, \ldots, 28$, $j = 1, \ldots, m_i$ and m_i represents the corresponding number of observations,

$$\mathbf{X}_i = \mathbf{Z}_i = \begin{bmatrix} 1 & t_1 & t_1^2 \\ \vdots & \vdots & \vdots \\ 1 & t_{m_i} & t_{m_i}^2 \end{bmatrix}, \ \boldsymbol{\beta} = \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}, \ \mathbf{b}_i = \begin{pmatrix} a_i \\ b_i \\ c_i \end{pmatrix}, \text{ and } \mathbf{e}_i = \begin{pmatrix} e_{i1} \\ \vdots \\ e_{im_i} \end{pmatrix}. \tag{47}$$

The terms a_i, b_i and c_i, respectively, represent the random intercept, linear and quadratic effects associated to the i-th calf. Initially, we assume that $\mathbf{G}$ is unstructured and $\mathbf{R}_i = \sigma^2 \mathbf{I}_{m_i}$. This model induces the heteroskedastic nature of the response detected in the descriptive analysis.

Fitting polynomial models to longitudinal data may lead to convergence problems induced by possible ill-conditioning of the model-specification matrices ($\mathbf{X}$ and $\mathbf{Z}$). This may be handled by choosing appropriate algorithms, by using orthogonal polynomials or by rescaling the time variable. To avoid such convergence problems in this example, we used a quasi-Newton algorithm known by the acronym BFGS (Broyden–Fletcher–Goldfarb–Shanno optimisation algorithm). For details, see Lange (2013), for example. We also rescaled the time variable dividing its values

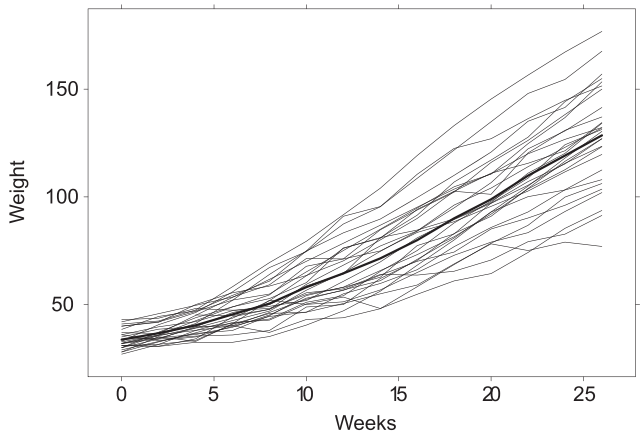

Figure 10. *Calf weight example–profile plot.*

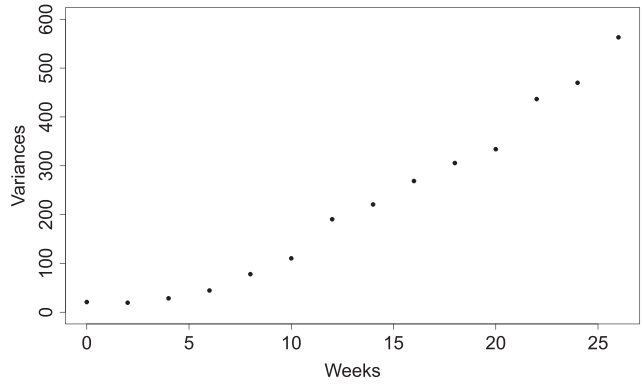

Figure 11. *Calf weight example—variance plot.*

by 4.29 (=30/7) so that the time unit may be expressed in months (30 days) instead of in weeks; this favours convergence of the algorithm without affecting the interpretation of the results.

Estimates of the corresponding parameters are presented in Table 8, and a plot of the modified Lesaffre–Verbeke index is displayed in Figure 12, suggesting that the proposed covariance structure may not be adequate for units 1, 15 and 22. Units 15 and 22 are also highlighted either in the standardised marginal residual or the conditional Cook's distance plots (see Figures 22–24 in the Supporting Information).

A plot of the within-calves correlations versus time lags as suggested by Grady and Helms (1995) is displayed in Figure 13 giving a clear indication that the within-calves observations are serially correlated. This is also evident from an auto-correlation plot (see Figure 25 in the SM).

To incorporate this characteristic and keeping in mind that there are missing observations, we fitted model (47) with the following continuous AR(1) specification to the data.

$$\mathbb{C}ov(e_{ij}, e_{ik}) = \sigma^2\{\exp[-\phi(t_{ij} - t_{ik})]\},\ t_{ij} \geq t_{ik} \tag{48}$$

Estimates for the parameters of model (47)–(48) are also presented in Table 8. An auto-correlation function plot is displayed in Figure 14 and does not show sufficient evidence of further serial correlation in the conditional errors.

The corresponding plot of the modified Lesaffre–Verbeke index is displayed in Figure 15 and still suggests that the covariance structure is not adequate for some animals. Other diagnostic plots for this model are presented in Figures 26–39 in the SM.

Table 8. *Calf weight example: parameter estimates and standard errors for model (47) with different covariance structures.*

	Conditional independence		CAR1 complete		CAR1 (3 groups) complete		CAR1 (3 groups) -obs 22.14 and unit 21	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Fixed effects								
Intercept	32.71	0.91	33.49	0.97	33.01	0.94	32.90	0.98
Linear	7.95	1.19	7.80	1.02	7.84	0.97	7.56	1.04
Quadratic	1.34	0.16	1.31	0.14	1.32	0.13	1.36	0.13
Variance components								
Intercept (I)	18.88		0.66		1.29		0.001	
Linear (L)	36.79		19.37		17.02		19.33	
Quadratic (Q)	0.62		0.28		0.24		0.19	
Corr(I,L)	−0.32		0.92		0.99		0.99	
Corr(I,Q)	0.16		−0.56		−0.52		−0.53	
Corr(L,Q)	−0.74		−0.55		−0.50		−0.52	
Residual								
Single group	8.04		25.34	—	—		—	
Group 1	—		—		37.07		39.26	
Group 2	—		—		18.26		20.02	
Group 3	—		—		28.08		30.79	
AIC	2135.5		2038.6		2019.2		1966.5	
BIC	2174.6		2081.6		2070.0		2017.0	
Loglik	−1057.8		−1008.3		−996.6		−970.2	
ϕ			0.62		0.64		0.67	

CAR1: continuous AR(1).
Group1: units 17, 22 and 23.
Group2: units 4, 5, 11, 12 and 28.
Group3: remaining units.

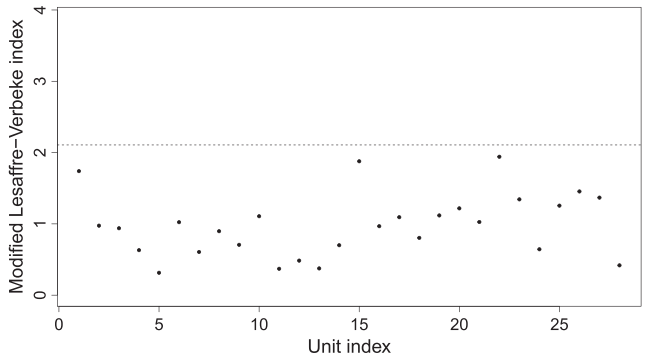

Figure 12. *Calf weight example–modified Lesaffre–Verbeke unit index plot for model (47). Dashed line: third quartile + 1.5 interquartile range.*

A possible justification is that there might be calf-specific variability for some animals, so we considered an ad hoc grouping according to the values of the modified Lesaffre–Verbeke index to avoid over-parametrisation. In particular, we considered a group formed by those animals with the highest values of the modified Lesaffre–Verbeke index (calves 17, 22 and 23), a group with those associated to the lowest values (calves 4, 5, 11, 12 and 28) and a third group with the remaining animals. In fact, these groups correspond to animals with, respectively, the largest and lowest variances in a complete heteroskedastic model. Diagnostic plots for this model are displayed in Figures 40–53 in the SM, and those corresponding to Cook's distances flag the 14th observation of calf 22 as that being more influential. This observation corresponds to the lowest weight among all calves at the 26th week. Also, calf 21 is singled out in Figures 51–53 in the Supporting Information as being of high leverage. In fact, half of the observations are missing for this calf, and this may be the reason for it being highlighted in those plots.

Estimates of the parameters of the same model fitted without the 14th observation of calf 22 and elimination of the data for calf 21 are also presented in Table 8. A plot of the modified Lesaffre–Verbeke index for this model is displayed in Figure 16 and does not flag units for which the proposed covariance structure is inadequate.

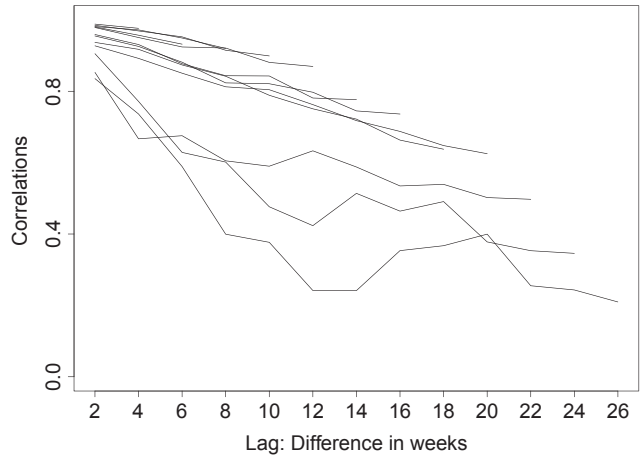

Figure 13. *Calf weight–Grady–Helms correlation plot versus time lags for model (47).*

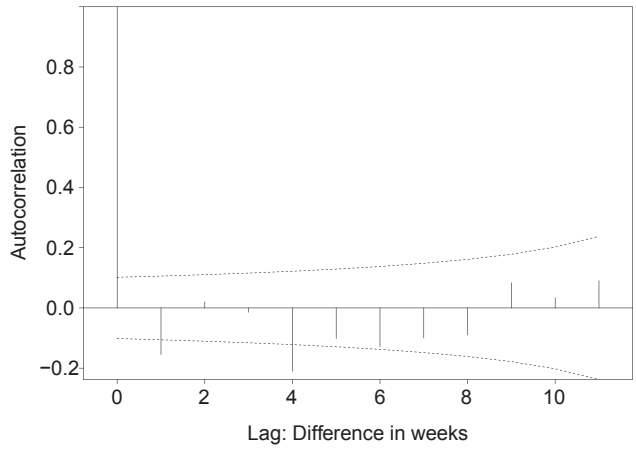

Figure 14. *Calf weight–auto-correlation function plot for model (47)–(48).*

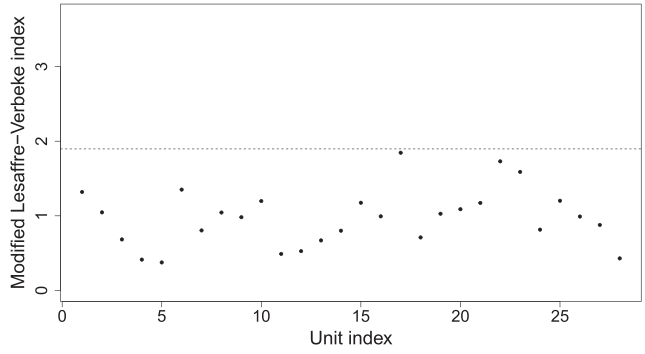

Figure 15. *Calf weight example—modified Lesaffre–Verbeke unit index plot for model (47)–(48). Dashed line: third quartile + 1.5 interquartile range.*

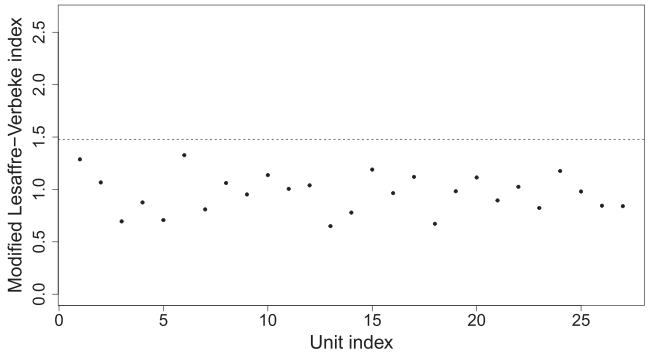

Figure 16. *Calf weight example—modified Lesaffre–Verbeke unit index plot for model (47)–(48) with three groups of unit-specific variances and exclusion of calf 21 and the 14th observation of calf 22. Dashed line: third quartile + 1.5 interquartile range.*

Other diagnostic plots for this model are displayed in Figures 54–67 in the Supporting Information and do not present considerable evidence against its fit.

The AIC and BIC values for this model are 1966.5 and 2017.0, respectively, suggesting a better fit relatively to the previous model for which the corresponding values are 2019.2 and 2070.0, respectively. The impact on the parameter estimates is not considerable as indicated in Table 8.

7 Discussion

Analysing repeated measures or longitudinal data via Gaussian LMM is convenient for various reasons. These models are very flexible (they include the class of linearisable models, as considered in Alencar *et al.* (2012), for example), easily interpretable and may be fitted via a series of very efficient algorithms for which software is widely available. Furthermore, if both the fixed and random components are well specified, the results obtained in practical applications are usually very similar to those generated by other classes of models as evidenced in Pinheiro *et al.* (2001), Savalli *et al.* (2006) or Alencar *et al.* (2012), for example. Such models are also convenient because in addition to the population parameters, they provide insight on the covariance structure as well as on the individual components.

Given this flexibility, selection of appropriate models in this class may be a difficult task. A few criteria proposed for such purposes are well summarised in Müller *et al.* (2013) but none is known to perform better than the others. In fact, they should be used with a complementary spirit, and in this context, the diagnostics tools we consider may be useful additional means to select reasonable linear mixed models. We believe that a careful use of such tools may provide the means to better understand the problem under investigation and to suggest appropriate analyses.

Although some functions designed to generate many of such diagnostic plots are available in the literature, they are still not implemented in the Comprehensive R Archive Network (CRAN)-repository. These functions may be obtained from the authors, but their use in practical applications may not be a straightforward task, because extracting the necessary information from the fitting function output, specially for more complex models, that is, with different random effects and/or different error covariance matrices for different groups, may not be an easy task. To bypass this problem, we constructed an R-function that provides the desired plots.

Obtaining diagnostic plots is more problematic for the other classes (ES and SE LMM) of models described in Section 5, where even the definition of residuals is not well established. Some effort in designing and testing flexible functions for such purposes could help to identify cases where the use of the Gaussian LMM may not be adequate and to disseminate the use of models that can accommodate outlying or influential observations. This opens a promising field for research.

Other alternatives to analyse repeated measures data include GLMM or generalised linear models with working covariance matrices, but here also, corresponding diagnostic tools still require investigation for their widespread practical application. Some efforts in this direction have been carried out in Xiang *et al.* (2002) or Venezuela *et al.* (2007). Generalised additive models for location, scale and shape proposed by Rigby and Stasinopoulos (2005) may also be employed for such purposes, but in general, are useful only for large sets of data.

Notes

[1] We will use $\widehat{\boldsymbol{\theta}}$ to represent either the MLE or the RMLE throughout the paper.

Acknowledgements

This research received financial support from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq: projects 308613/2011-2, 445294/2014-0 and 306770/2014-8) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP: project 2013/10528-2), Brazil. The authors are grateful to the editor and two anonymous referees for their meticulous and constructive comments.

References

Alencar, A.P., Singer, J.M. & Rocha, F.M.M. (2012). Competing regression models for longitudinal data. *Biom. J.*, **54**, 663–671.

Arellano-Valle, R., Bolfarine, H. & Lachos, V.H. (2005). Skew-normal linear mixed models. *J. Data Sci.*, **3**, 415–438.

Arellano-Valle, R., Bolfarine, H. & Lachos, V.H. (2007). Bayesian inference for skew-normal linear mixed models. *J. Appl. Stat.*, **34**, 663–682.

Atkinson, A. & Riani, M. (2000). *Robust Diagnostic Regression Analysis.* New York: Springer.

Banerjee, M. & Frees, E.W. (1997). Influence diagnostics for linear longitudinal models. *J. Amer. Statist. Assoc.*, **92**, 999–1005.

Beckman, R.J., Nachtsheim, C.J. & Cook, R.D. (1987). Diagnostics for mixed-model analysis of variance. *Technometrics*, **29**, 413–426.

Belsley, D.A., Kuh, E. & Welsch, R.E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity.* New York: Wiley.

Bolfarine, H., Montenegro, L.C. & Lachos, V.H. (2007). Influence diagnostics for skew-normal linear mixed models. *Sankhya*, **69**, 648–670.

Chatterjee, S. & Hadi, A.S. (1988). *Sensitivity Analysis in Linear Regression.* New York: Wiley.

Christensen, R., Pearson, L.M & Johnson, W. (1992). Case-deletion diagnostics for mixed models. *Technometrics*, **34**, 38–45.

Cook, R.D. (1977). Detection of influential observation in linear regression. *Technometrics*, **19**, 15–18.

Cook, R.D. (1986). Assessment of local influence (with discussion). *J. R. Stat. Soc. B.*, **48**, 133–169.

Cook, R.D. & Weisberg, S. (1982). *Residuals and Influence Regression.* New York: Chapman & Hall.

Croissant, Y. (2015). Ecdat package. https://cran.r-project.org/web/packages/Ecdat/index.html. Accessed 15 January 2015.

Daniels, M.J. & Hogan, J. W. (2008). *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis.* New York: Chapman & Hall.

Demidenko, E. (2013). *Mixed Models: Theory and Applications with R*, 2nd ed. New York: Wiley.

Demidenko, E. & Stukel, T.A. (2005). Influence analysis for linear mixed-effects models. *Stat. Med.*, **24**, 893–909.

Dey, D. K., Ghosh, S. K. & Mallick, B. K. (2000). *Generalized Linear Models: A Bayesian Perspective.* New York: Marcel Dekker.

Diggle, P., Heagerty, P., Liang, K-Y. & Zeger, S. (2002). *Analysis of Longitudinal Data*, 2nd ed. New York: Oxford University Press.

Fearn, T. (1977). A two stage model for growth curves which leads to Rao's covariance adjusted estimators. *Biometrika*, **64**, 141–143.

Fei, Y. & Pan, J. (2003). Influence assessments for longitudinal data in linear mixed models. In *18th International Workshop on Statistical Modelling*, Eds. G. Verbeke, G. Molenberghs, M. Aerts & S. Fieuws, pp. 143–148. Belgium: Leuven.

Fitzmaurice, G.M., Davidian, M., Verbeke, M. & Molenberghs, G. (2008). *Longitudinal Data Analysis: A Handbook of Modern Statistical Methods.* New York: Chapmann & Hall.

Fitzmaurice, G.M., Laird, N.M. & Ware, J.H. (2011). *Applied Longitudinal Analysis*, 2nd ed. New York: Wiley.

Fung, W.K., Zhu, Z.Y., Wei, B.C. & He, X. (2002). Influence diagnostics and outliers tests for semiparametric mixed models. *J. R. Stat. Soc. B.*, **64**, 565–579.

Grady, J.J. & Helms, R.W. (1995). Model selection techniques for the covariance matrix for incomplete longitudinal data. *Stat. Med.*, **14**, 1397–1416.

Gumedze, F.N., Welham, S.J., Gogel, B.J. & Thompson, R. (2010). A variance shift model for detection of outliers in the linear mixed model. *Comput. Stat. Data Anal.*, **54**, 2128–2144.

Harrison, D. Jr. & Rubinfeld, D. (1978). Hedonic house prices and the demand for clean air. *J. Environ. Econ. Manage.*, **5**, 81–102.

Harville, D. (1976). Extension of the Gauss–Markov Theorem to include the estimation of random effects. *Ann. Stat.*, **4**, 384–395.

Hedecker, D. & Gibbons, R.D. (2006). *Longitudinal Data Analysis.* Hoboken, NJ: Wiley.

Henderson, C.R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, **31**, 423–447.

Hilden-Minton, J.A. (1995). *Multilevel diagnostics for mixed and hierarchical linear models*, Unpublished PhD Thesis, University of California, Los Angeles.

Hoaglin, D.C. & Welsch, R.E. (1978). The hat matrix in regression and ANOVA. *Am. Stat.*, **32**, 17–22.

Jara, A., Quintana, F. & San Martín, E. (2008). Linear mixed models with skew-elliptical distributions: a Bayesian approach. *Comput. Stat. Data Anal.*, **52**, 5033–5045.

Jiang, J. (1996). REML estimation: asymptotic behavior and related topics. *Ann. Stat.*, **24**, 255–286.

Koller, M. (2013). Robust estimation of linear mixed models. Doctoral dissertation ETH. Available from: http://dx.doi.org/10.3929/ethz-a-007632241. Accessed 23 October 2014.

Lachos, V.H., Ghosh, P. & Arellano-Valle, R.B. (2010). Likelihood based inference for skew-normal independent linear mixed models. *Statistica Sinica*, **20**, 303–322.

Laird, N.M. & Ware, J.H. (1982). Random effects models for longitudinal data. *Biometrics*, **38**, 963–974.

Lange, K. (2013). *Optimization*, 2nd Edition. New York: Springer.

Lesaffre, E. & Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics*, **54**, 570–582.

Longford, N.T. (1993). *Random Coefficient Models.* Oxford: Oxford University Press.

Loy, A. & Hofman, H. (2013). Diagnostic tools for hierarchical linear models. *WIREs Comput. Stat.*, **5**, 48–61.

McCulloch, C.E., Searle, S.R. & Neuhaus, J.M. (2008). *Generalized, Linear and Mixed Models*, 2nd ed. New York: Wiley.

Mitchell, A.F.S. (1989). The information matrix, skewness tensor and $\alpha-$connection for the general multivariate elliptic distribution. *Ann. Inst. Stat. Math.*, **41**, 289–304.

Molenberghs, G. & Verbeke, G. (2005). *Models for Discrete Longitudinal Data.* New York: Springer.

Mudholkar, G.S. & Hutson, A.D. (2000). The epsilon-skew-normal distribution for analyzing near-normal data. *J. Stat. Plan. Inference*, **83**, 291–309.

Müller, S., Scealy, J.L. & Welsh, A.H. (2013). Model selection in linear mixed models. *Stat. Sci.*, **28**, 135–167.

Mun, J. & Lindstrom, M.J. (2013). Diagnostics for repeated measurements in linear mixed models. *Stat. Med.*, **32**, 1361–1375.

Nobre, J.S. & Singer, J.M. (2007). Residual analysis for linear mixed models. *Biom. J.*, **49**, 863–875.

Nobre, J.S. & Singer, J.M. (2011). Leverage analysis for linear mixed models. *J. Appl. Stat.*, **38**, 1063–1072.

Osorio, F., Paula, G.A. & Galea, M. (2007). Assessment of local influence in elliptical linear models with longitudinal structure. *Comput. Stat. Data Anal.*, **51**, 4354–4368.

Patterson, H.D. & Thompson, R. (1971). Recovery of interblock information when block sizes are unequal. *Biometrika*, **58**, 545–554.

Pinheiro, J.C., Liu, C. & Wu, Y.N. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t-distribution. *J. Comp. Graph. Stat.*, **10**, 249–276.

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

Pregibon, D. (1981). Logistic regression diagnostics. *Ann. Stat.*, **9**, 705–724.

Rao, M.N. & Rao, C.R. (1966). Linked cross-sectional study for determining norms and growth rates - a pilot survey of Indian school-going boys. *Sankhya B*, **28**, 237–258.

Rigby, R.A. & Stasinopoulos, D.M. (2005). Generalized additive models for location, scale and hape, (with discussion). *Appl. Statist.*, **54**, 507–554.

Robinson, G.K. (1991). That BLUP is a good thing: the estimation of random effects. *Stat. Sci.*, **6**, 15–51.

Rocha, F.M.M. & Singer, J.M. (2016). Selection of terms in random coefficients models. *Submitted*.

Rutter, C.M. & Elashoff, R.M. (1994). Analysis of longitudinal data: random coefficient regression modelling. *Stat. Med.*, **13**, 1211–1231.

Savalli, C., Paula, G.A. & Cysneiros, F.J.A. (2006). Assessment of variance components in elliptical linear mixed models. *Stat. Model.*, **6**, 59–76.

Schützenmeister, A. & Piepho, H.P. (2012). Residual analysis of linear mixed models using a simulation approach. *Comput. Stat. Data Anal.*, **56**, 1405–1416.

Sen, P.K., Singer, J.M. & Pedroso-de-Lima, A.C. (2009). *From Finite Sample to Asymptotic Methods in Statistics.* Cambridge: Cambridge University Press.

Singer, J.M. & Andrade, D.F. (2000). Analysis of longitudinal data. In *Handbook of Statistics, Bio-Environmental and Public Health Statistics*, Vol. 18, Eds. P.K. Sen & C.R. Rao, pp. 115–160. North Holland: Amsterdam.

Singer, J.M. & Cúri, M. (2006). Modelling regression and dispersion parameters in a complex repeated measures experiment. *Environ. Ecol. Stat.*, **13**, 53–68.
Tan, F.E.S., Ouwens, M.J.N. & Berger, M.P.F. (2001). Detection of influential observations in longitudinal mixed effects regression models. *The Statistician*, **50**, 271–284.
Venezuela, M.K., Botter, D.A. & Sandoval, M.C. (2007). Diagnostic techniques in generalized estimating equations. *J. Stat. Comput. Simul.*, **77**, 879–888.
Verbeke, G. & Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data.* New York: Springer.
Vonesh, E. & Chinchilli, V.M. (1997). *Linear and Non-linear Models for the Analysis of Repeated Measures.* New York: Marcel Dekker.
Wei, B.C., Hu, Y.Q. & Fung, W.K. (1998). Generalized leverage and its applications. *Scand. J. Stat.*, **25**, 25–37.
Weiss, R.E & Lazaro, C.G. (1992). Residual plots for repeated measures. *Stat. Med.*, **11**, 115–124.
West, B.T. & Galecki, A.T. (2011). An overview of current software procedures for fitting linear mixed models. *Am. Stat.*, **65**, 274–282.
Wu, H. & Zhang, J.T. (2006). *Nonparametric Regression Methods for Longitudinal Data Analysis: Mixed-Effects Modeling Approaches.* Hoboken, NJ: Wiley.
Xiang, L., Tse, S.-K. & Lee, A.H. (2002). Influence diagnostics for generalized linear mixed models: applications to clustered data. *Comput. Stat. Data Anal.*, **40**, 759–774.
Zellner, A. (1962). An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias. *J. Amer. Statist Assoc.*, **57**, 348–368.
Zewotir, T. (2008). Multiple cases deletion diagnostics for linear mixed models. *Commun. Stat. Theory Methods*, **37**, 1071–1084.
Zewotir, T. & Galpin, J.S. (2005). Influence diagnostics for linear mixed models. *J. Data Sci.*, **3**, 153–177.
Zewotir, T. & Galpin, J.S. (2007). A unified approach on residuals, leverages and outliers in the linear mixed models. *Test*, **16**, 58–75.
Zhu, H.T., Ibrahim, J.G., Lee, S. & Zhang, H. (2007). Perturbation selection and influence measures in local influence analysis. *Ann. Stat.*, **35**, 2565–2588.

[*Received June 2015, accepted May 2016*]

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web site.