Journal of Multivariate Analysis 141 (2015) 104–117

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Influence assessment in censored mixed-effects models using the multivariate Student's-t distribution

Larissa A. Matos [a], Dipankar Bandyopadhyay [b,*], Luis M. Castro [c], Victor H. Lachos [a]

[a] Departamento de Estatística, IMECC-UNICAMP, Campinas, São Paulo, Brazil
[b] Division of Biostatistics, University of Minnesota, Minneapolis, MN 55455, United States
[c] Departamento de Estatística, Universidad de Concepción, Chile

ARTICLE INFO

Article history:
Received 4 November 2013
Available online 26 June 2015

AMS subject classifications:
62E99
62E05
62F99
62H99

Keywords:
Censored data
Case-deletion diagnostics
ECM algorithm
Linear mixed-effects model
Multivariate Student's-t distribution
Non-linear mixed-effects model

ABSTRACT

In biomedical studies on HIV RNA dynamics, viral loads generate repeated measures that are often subjected to upper and lower detection limits, and hence these responses are either left- or right-censored. Linear and non-linear mixed-effects censored (LMEC/NLMEC) models are routinely used to analyze these longitudinal data, with normality assumptions for the random effects and residual errors. However, the derived inference may not be robust when these underlying normality assumptions are questionable, especially the presence of outliers and thick-tails. Motivated by this, Matos et al. (2013) recently proposed an exact EM-type algorithm for LMEC/NLMEC models using a multivariate Student's-t distribution, with closed-form expressions at the E-step. In this paper, we develop influence diagnostics for LMEC/NLMEC models using the multivariate Student's-t density, based on the conditional expectation of the complete data log-likelihood. This partially eliminates the complexity associated with the approach of Cook (1977, 1986) for censored mixed-effects models. The new methodology is illustrated via an application to a longitudinal HIV dataset. In addition, a simulation study explores the accuracy of the proposed measures in detecting possible influential observations for heavy-tailed censored data under different perturbation and censoring schemes.

1. Introduction

In AIDS research, the study of the human immunodeficiency virus (HIV) dynamics has received significant attention in the biomedical literature allowing us to understand the pathogenesis of HIV, and assess the effectiveness of the anti-retroviral (ARV) therapy. Most of the clinical trials on ARV therapy assess the rates/changes of viral loads/HIV-1 RNA copies (the amount of actively replicating virus) collected longitudinally over time. The viral load is considered a key primary endpoint because its monitoring is mostly available, a failure in the treatment can be defined virologically, and a new regimen of therapy is recommended as soon as virological rebound occurs [21]. Since the individual viral load trajectories yield large between-subject variations, statistical modeling often focus in formulating the correct linear and nonlinear mixed-effects models (LME/NLME) to estimate these trajectories, and quantify within- and between-subject variations [31,32,25].

* Correspondence to: Division of Biostatistics, University of Minnesota SPH, A460 Mayo MMC 303, 420 Delaware Street SE, Minneapolis, MN 55455, United States.
E-mail address: dbandyop@umn.edu (D. Bandyopadhyay).

http://dx.doi.org/10.1016/j.jmva.2015.06.014
0047-259X/

The statistical modeling of viral load can be challenging. First, depending on the diagnostic assays used, the viral load measures may be subjected to upper or lower detection limits (hence, left or right censored), below and above which they are not quantifiable [30]. Under non-trivial censoring proportions, considering ad-hoc alternatives [12] might lead to bias in fixed effects and variance components estimates. As alternatives to these crude imputation techniques, Vaida and Liu [28] proposed expectation–maximization (EM) schemes for LME/NLME with censored responses (henceforth LMEC/NLMEC). However, all these methods assume normality of the between-subject random effects and within-subject errors. Even though normality is mostly a reasonable model assumption, it may lack robustness in parameter estimation under departures from normality, namely, presence of heavy tails and outliers [24]. Censored HIV viral loads do exhibit heavy-tailed behavior [13]. This is also revealed from the raw histogram and Q–Q plots of viral loads from our motivating example (see Fig. 2, panels a and b in Section 5.1). Although popular data transformations (say, Box–Cox) might render normality, or close to normality with reasonable empirical results, various issues still persist with these transformations [13]. Hence, an appropriate theoretical but 'robust' framework that avoids data transformation is desirable. A variety of proposals (both classical and Bayesian) exist in this direction that uses the univariate or multivariate Student's-t distribution [24,17,18] in the context of LME/NLME models. Some Bayesian propositions in the context of heavy-tailed LMEC/NLMEC models include Lachos et al. [13] who advocated the use of the normal/independent density [14], while Bandyopadhyay et al. [3,2] studied the LMEC model considering both skewness and heavy-tails. Very recently, Matos et al. [20] proposed a full maximum-likelihood (ML) based inference using a computationally convenient exact ECM algorithm for the LMEC/NLMEC models using the multivariate Student's-t distribution (henceforth, the t-LMEC/NLMEC model). Here, the E-step yields closed-form expressions, and all parameters are updated in the M-step by considering the random components and the censored observations as missing data.

A vast majority of model development in the literature for LMEC/NLMEC models focus on estimating the mean function. Hence, developing influence diagnostics is a key in assessing the effect of a single observation on the predicted scores for other observations, and consequently the overall parameter estimates, all based on the mean function. Although diagnostics for the traditional normality based LME and LMEC [19] models exist, those for heavy-tailed LMEC/NLMEC models are not well developed. Influence analysis is generally conducted using two primary approaches. The first one is the case-deletion approach [7] based on the well-known Cook's distance. Under normality assumptions for LME, [4,11,27] focused on case-deletion diagnostics for fixed effects, while Christensen et al. [6] considered a one-step approximation to Cook's distance for the variance components. The other approach is the computationally attractive local influence approach [8], which is a general technique used to assess the stability of the estimation outputs with respect to the model inputs. For elliptical mixed-effects models, this method had been discussed in the literature by [5,16,36,15,22,26], among others.

Developing influence diagnostics for LMEC/NLMEC models in the spirit of [7,8] leads to the underlying observed log-likelihood functions involving intractable integrals. This renders the direct application of Cook's approach to be very difficult if not impossible, since the measures involve first and second derivatives of these functions. In this context, Zhu and Lee [36] and Zhu et al. [37] developed an unified approach for performing local influence and case-deletion diagnostics, respectively, for general missing data models based on the Q-function, *i.e.*, the conditional expectation of the complete-data log likelihood at the E-step in the EM algorithm. This was extended to generalized linear and NLME models by [15,34], respectively. This Q-function approach produces result similar to those obtained using the Cook's approach. Recently, Matos et al. [19] used this Q-function approach for developing influence diagnostics for LMEC/NLMEC models. Stemming from the same difficulty with intractable integrals (for example, the *pdfs* of truncated multivariate Student's-t distributions) in implementing the Cook's diagnostics for the t-LMEC/NLMEC model of [20], we develop case-deletion and influence diagnostics measures using the approach of [37] (see also [15]).

The rest of this paper is organized as follows. Section 2 develops the t-LMEC model specification and an EM-type algorithm for ML estimation. Section 3 presents the global and local influence approaches for the t-LMEC model. For local influence, various perturbation schemes for both subject- and observation-level diagnostics are considered. In Section 4, the t-NLMEC model is defined. The methodology is illustrated in Section 5 using a motivating HIV dataset. Section 6 presents a simulation study evaluating the efficiency of our method in detecting outliers under various degrees of data perturbation and censoring. Finally, Section 7 presents some concluding remarks, with some possible directions for future research.

2. Censored linear mixed effect model

Ignoring censoring for the moment, the t-LME model of [20] is specified as:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \tag{1}$$

where

$$\begin{pmatrix}\mathbf{b}_i\\ \boldsymbol{\epsilon}_i\end{pmatrix} \overset{ind.}{\sim} t_{n_i+q}\left(\begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\mathbf{D} & \mathbf{0}\\ \mathbf{0} & \sigma^2\mathbf{I}_{n_i}\end{pmatrix}, \nu\right), \quad i = 1, \ldots, n,$$

which implies that, marginally,

$$\mathbf{b}_i \overset{iid}{\sim} t_q(\mathbf{0}, \mathbf{D}, \nu) \quad \text{and} \quad \boldsymbol{\epsilon}_i \overset{ind.}{\sim} t_{n_i}(\mathbf{0}, \sigma^2\mathbf{I}_{n_i}, \nu), \quad i = 1, \ldots, n, \tag{2}$$

where $t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ denotes the *pdf* of a multivariate Student's-t distribution with location vector $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$ and degrees of freedom ν. The subscript i refers to the subject index; $\mathbf{I}_p$ denotes the $p \times p$ identity matrix; $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$

is a vector of observed continuous responses for subject i of dimension $n_i \times 1$; $\mathbf{X}_i$ is the $n_i \times p$ design matrix associated with the $p \times 1$ vector of fixed-effects $\boldsymbol{\beta}$; $\mathbf{Z}_i$ is the $n_i \times q$ design matrix corresponding to the $q \times 1$ vector of random effects $\mathbf{b}_i$; $\boldsymbol{\epsilon}_i$ is the $(n_i \times 1)$ vector of random errors and the random effects dispersion matrix $\mathbf{D} = \mathbf{D}(\boldsymbol{\alpha})$ depends on unknown parameters $\boldsymbol{\alpha}$. Following Matos et al. [20], we consider the case where the response Y_{ij} is not fully observed for all i, j. Consequently, the observed data for the ith subject is $(\mathbf{Q}_i, \mathbf{C}_i)$, where $\mathbf{Q}_i$ is the vector of censoring level and $\mathbf{C}_i$ is the vector of censoring indicators such that

$$\begin{aligned} y_{ij} &\le Q_{ij} \quad \text{if } C_{ij} = 1, \\ y_{ij} &= Q_{ij} \quad \text{if } C_{ij} = 0. \end{aligned} \tag{3}$$

For simplicity, we assume that the data are left censored. Extensions to other arbitrary censoring patterns are immediate.

2.1. The likelihood function

The first step is to treat separately the observed and censored components of $\mathbf{y}_i$. Let $\mathbf{y}_i^o$ be the n_i^o-vector of observed outcomes and $\mathbf{y}_i^c$ be the n_i^c-vector of censored observations for subject i with $(n_i = n_i^o + n_i^c)$ such that $C_{ij} = 0$ for all elements in $\mathbf{y}_i^o$, and 1 for all elements in $\mathbf{y}_i^c$. After reordering, $\mathbf{y}_i$, $\mathbf{Q}_i$, $\mathbf{X}_i$, and $\boldsymbol{\Sigma}_i$ can be partitioned as $\mathbf{y}_i = vec(\mathbf{y}_i^o, \mathbf{y}_i^c)$, $\mathbf{Q}_i = vec(\mathbf{Q}_i^o, \mathbf{Q}_i^c)$, $\mathbf{X}_i^\top = (\mathbf{X}_i^o, \mathbf{X}_i^c)$ and $\boldsymbol{\Sigma}_i = \begin{pmatrix} \boldsymbol{\Sigma}_i^{oo} \boldsymbol{\Sigma}_i^{oc} \\ \boldsymbol{\Sigma}_i^{co} \boldsymbol{\Sigma}_i^{cc} \end{pmatrix}$, where $vec(\cdot)$ denotes the function which stacks vectors or matrices of the same number of columns. Using properties of multivariate Student's-t distribution (see [1]), we have $\mathbf{y}_i^o \sim t_{n_i^o}(\mathbf{X}_i^o\boldsymbol{\beta}, \boldsymbol{\Sigma}_i^{oo}, \nu)$, and $\mathbf{y}_i^c|\mathbf{y}_i^o, \sim t_{n_i^c}(\boldsymbol{\mu}_i^{co}, \mathbf{S}_i^{co}, \nu + n_i^o)$, where

$$\boldsymbol{\mu}_i^{co} = \mathbf{X}_i^c\boldsymbol{\beta} + \boldsymbol{\Sigma}_i^{co}\boldsymbol{\Sigma}_i^{oo-1}(\mathbf{y}_i^o - \mathbf{X}_i^o\boldsymbol{\beta}), \qquad \mathbf{S}_i^{co} = \left(\frac{\nu + Q(\mathbf{y}_i^o)}{\nu + n_i^o}\right)\boldsymbol{\Sigma}_i^{cc.o}, \tag{4}$$

with $\boldsymbol{\Sigma}_i^{cc.o} = \boldsymbol{\Sigma}_i^{cc} - \boldsymbol{\Sigma}_i^{co}\boldsymbol{\Sigma}_i^{oo-1}\boldsymbol{\Sigma}_i^{oc}$ and $Q(\mathbf{y}_i^o) = (\mathbf{y}_i^o - \mathbf{X}_i^o\boldsymbol{\beta})^\top\boldsymbol{\Sigma}_i^{oo-1}(\mathbf{y}_i^o - \mathbf{X}_i^o\boldsymbol{\beta})$. Therefore, the likelihood for subject i is

$$\begin{aligned} L_i(\boldsymbol{\theta}|\mathbf{y}) &= f(\mathbf{Q}_i|\mathbf{C}_i, \boldsymbol{\theta}) = f(\mathbf{y}_i^c \le \mathbf{Q}_i^c|\mathbf{y}_i^o = \mathbf{Q}_i^o, \boldsymbol{\theta})f(\mathbf{y}_i^o = \mathbf{Q}_i^o|\boldsymbol{\theta}), \\ &= T_{n_i^c}(\mathbf{Q}_i^c|\boldsymbol{\mu}_i^{co}, \mathbf{S}_i^{co}, \nu + n_i^o)t_{n_i^o}(\mathbf{Q}_i^o|\mathbf{X}_i^o\boldsymbol{\beta}, \boldsymbol{\Sigma}_i^{oo}, \nu) = L_i, \end{aligned}$$

where $T_p(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ denotes the cumulative distribution function (*cdf*) of the multivariate Student's-t distribution with parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and ν. The log-likelihood function for the observed data is given by $\ell(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^n \log L_i$, and the estimates obtained by maximizing the log-likelihood function $\ell(\boldsymbol{\theta}|\mathbf{y})$ are the maximum likelihood estimates (MLEs).

2.2. The EM algorithm

The observed log-likelihood function involves complex expressions, making it very difficult to work directly with $\ell(\boldsymbol{\theta}|\mathbf{y})$, either for the ML estimation, or the corresponding influence analysis. As mentioned above, Matos et al. [20] developed an EM-type algorithm for the t-LMEC/NLMEC models by treating $\mathbf{y} = (\mathbf{y}_1^\top, \dots, \mathbf{y}_n^\top)^\top$, $\mathbf{b} = (\mathbf{b}_1^\top, \dots, \mathbf{b}_n^\top)^\top$, and $\mathbf{u} = (u_1, \dots, u_n)^\top$ as hypothetical missing data, and augmenting those to the observed data vector $(\mathbf{Q}, \mathbf{C})$, where $\mathbf{Q} = vec(\mathbf{Q}_1, \dots, \mathbf{Q}_n)$, and $\mathbf{C} = vec(\mathbf{C}_1, \dots, \mathbf{C}_n)$. Thus, the resulting complete data is $\mathbf{y}_c = (\mathbf{C}^\top, \mathbf{Q}^\top, \mathbf{y}^\top, \mathbf{b}^\top, \mathbf{u}^\top)^\top$, and the EM-type algorithm is applied to the complete data log-likelihood function $\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta}|\mathbf{y}_c)$, where

$$\ell_i(\boldsymbol{\theta}|\mathbf{y}_c) = -\frac{1}{2}\left[n_i \log\sigma^2 + \frac{u_i}{\sigma^2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i) + \log|\mathbf{D}| + u_i\mathbf{b}_i^\top\mathbf{D}^{-1}\mathbf{b}_i\right] + h(u_i|\nu) + C,$$

where C is a constant that does not depend on the vector parameter $\boldsymbol{\theta}$ and $h(u_i|\nu)$ is the *pdf* of a Gamma$(\nu/2, \nu/2)$ distribution. Given a current value $\widehat{\boldsymbol{\theta}}^{(k)}$ of $\boldsymbol{\theta}$, the Q function (the conditional expectation of the complete data log-likelihood function) is given by

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = \sum_{i=1}^n Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = \sum_{i=1}^n Q_{1i}(\boldsymbol{\beta}, \sigma^2|\widehat{\boldsymbol{\theta}}^{(k)}) + \sum_{i=1}^n Q_{2i}(\boldsymbol{\alpha}|\widehat{\boldsymbol{\theta}}^{(k)}), \tag{5}$$

where

$$Q_{1i}(\boldsymbol{\beta}, \sigma^2|\widehat{\boldsymbol{\theta}}^{(k)}) = -\frac{n_i}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\left[\widehat{a}_i^{(k)} - 2\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top(\widehat{u\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{u\mathbf{b}}_i^{(k)}) + \widehat{u}_i^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\right]$$

and $Q_{2i}(\boldsymbol{\alpha}|\widehat{\boldsymbol{\theta}}^{(k)}) = -\frac{1}{2}\log|\mathbf{D}| - \frac{1}{2}\mathrm{tr}\left(\widehat{u\mathbf{b}_i^2}^{(k)}\mathbf{D}^{-1}\right)$. Here, $\widehat{a}_i^{(k)} = \mathrm{tr}\left(\widehat{u\mathbf{y}_i^2}^{(k)} - 2\widehat{u\mathbf{y}\mathbf{b}}_i^{(k)}\mathbf{Z}_i^\top + \widehat{u\mathbf{b}_i^2}^{(k)}\mathbf{Z}_i^\top\mathbf{Z}_i\right)$; $\widehat{u\mathbf{y}_i^2} = E\{u_i\mathbf{y}_i\mathbf{y}_i^\top|\mathbf{Q}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\}$; $\widehat{u\mathbf{b}_i^2}^{(k)} = E\{u_i\mathbf{b}_i\mathbf{b}_i^\top|\mathbf{Q}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\} = \widehat{\sigma^2}^{(k)}\widehat{\boldsymbol{\Lambda}}_i^{(k)} + \widehat{\boldsymbol{\varphi}}_i^{(k)}(\widehat{u\mathbf{y}_i^2}^{(k)} - \widehat{u\mathbf{y}}_i^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\widehat{u\mathbf{y}}_i^{(k)\top} + \widehat{u}_i^{(k)}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top)\widehat{\boldsymbol{\varphi}}_i^\top$; $\widehat{u\mathbf{b}}_i^{(k)} = E\{u_i\mathbf{b}_i|\mathbf{Q}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\} = \widehat{\boldsymbol{\varphi}}_i^{(k)}(\widehat{u\mathbf{y}}_i^{(k)} - \widehat{u}_i^{(k)}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)})$; $\widehat{u\mathbf{y}\mathbf{b}}_i^{(k)} = E\{u_i\mathbf{y}_i\mathbf{b}_i^\top|\mathbf{Q}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\} = (\widehat{u\mathbf{y}_i^2}^{(k)} - \widehat{u\mathbf{y}}_i^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top)\widehat{\boldsymbol{\varphi}}_i^\top$, with $\widehat{\boldsymbol{\Lambda}}_i^{(k)} = (\widehat{\sigma^2}^{(k)}\widehat{\mathbf{D}}^{-1(k)} + \mathbf{Z}_i^\top\mathbf{Z}_i)^{-1}$ and $\widehat{\boldsymbol{\varphi}}_i^{(k)} = \widehat{\boldsymbol{\Lambda}}_i^{(k)}\mathbf{Z}_i^\top$.

It is easy to observe that the E-step reduces to the computation of $\widehat{u\mathbf{y}_i^2} = E\{u_i\mathbf{y}_i\mathbf{y}_i^\top|\mathbf{Q}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\}$, $\widehat{u\mathbf{y}}_i = E\{u_i\mathbf{y}_i|\mathbf{Q}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\}$, and $\widehat{u}_i = E\{u_i|\mathbf{Q}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\}$. These expected values are available in closed form using Propositions available in [20].

Next, the conditional maximization step (CM-step) maximizes $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ conditionally with respect to $\boldsymbol{\theta}$ to obtain new estimates $\widehat{\boldsymbol{\theta}}^{(k+1)}$ as follows:

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \left(\sum_{i=1}^n \widehat{u}_i^{(k)}\mathbf{X}_i^\top\mathbf{X}_i\right)^{-1}\sum_{i=1}^n \mathbf{X}_i^\top\left(\widehat{u\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{u\mathbf{b}}_i^{(k)}\right), \tag{6}$$

$$\widehat{\sigma^2}^{(k+1)} = \frac{1}{N}\sum_{i=1}^n\left[\widehat{a}_i^{(k)} - 2\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top(\widehat{u\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{u\mathbf{b}}_i^{(k)}) + \widehat{u}_i^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\right], \tag{7}$$

$$\widehat{\mathbf{D}}^{(k+1)} = \frac{1}{n}\sum_{i=1}^n \widehat{u\mathbf{b}_i^2}^{(k)}, \tag{8}$$

where $N = \sum_{i=1}^n n_i$, and the scale matrix $\mathbf{D}$ is unstructured with $\boldsymbol{\alpha}$ the upper triangular elements of $\mathbf{D}$. The algorithm is iterated until the distance involving two successive evaluations of the log-likelihood $|\ell(\widehat{\boldsymbol{\theta}}^{(k+1)})/\ell(\widehat{\boldsymbol{\theta}}^{(k)}) - 1|$ is sufficiently small. Here, we do not focus on the ML estimation, and the interested might refer to [20] for further details. In the following section, we derive influence diagnostic measures, given the ML estimate $\widehat{\boldsymbol{\theta}}$.

3. Influence analysis

Influence diagnostics are routinely used in statistical modeling to identify aberrant observations and assess their impact on model fitting and parameter estimation. Recognizing the difficulties following the Cook's [7,8] approach (described in Section 1), we use the Q-function of [37] to develop case-deletion measures, leading to the influence measures for the t-LMEC model.

3.1. Global influence

The case-deletion approach is a commonly used scheme to study the effects of deleting the ith case/observation from the dataset. Henceforth, the subscript '[i]' will denote the original dataset with the ith case deleted. Consequently, the log-likelihood function corresponding to the remaining data is denoted by $\ell(\boldsymbol{\theta}|\mathbf{Y}_{c[i]})$. In order to assess the influence of the ith case on the ML estimate $\widehat{\boldsymbol{\theta}}$, we need to compare the difference between $\widehat{\boldsymbol{\theta}}_{[i]}$ and $\widehat{\boldsymbol{\theta}}$, where $\widehat{\boldsymbol{\theta}}_{[i]} = (\widehat{\boldsymbol{\beta}}_{[i]}^\top, \widehat{\sigma^2}_{[i]}, \widehat{\boldsymbol{\alpha}}_{[i]}^\top)^\top$ is the maximizer of the function $Q_{[i]}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = E\{\ell(\boldsymbol{\theta}|\mathbf{Y}_{c[i]})|\mathbf{Q}, \mathbf{C}, \widehat{\boldsymbol{\theta}}\}$, with $\widehat{\boldsymbol{\theta}}$ being the ML estimate of $\boldsymbol{\theta}$. An observation is regarded as influential if its deletion generates considerable influence on model estimates. In other words, if $\widehat{\boldsymbol{\theta}}_{[i]}$ is fairly far from $\widehat{\boldsymbol{\theta}}$, then the ith observation could considered as influential. Note that, since the estimator $\widehat{\boldsymbol{\theta}}_{[i]}$ is needed for every case, this scheme requires a considerable computational effort, particularly for large sample sizes. For that reason, a one-step approximation (see [9,37]) is used to reduce the burden. This approximation follows:

$$\widehat{\boldsymbol{\theta}}_{[i]}^1 = \widehat{\boldsymbol{\theta}} + \{-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}^{-1}\dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}), \tag{9}$$

where $\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$ represents the Hessian matrix, and $\dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \frac{\partial Q_{[i]}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}}\big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$, $i = 1, \ldots, n$, with its elements given by

$$\dot{Q}_{[i]\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \partial Q_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\beta} = \frac{1}{\widehat{\sigma^2}}E_{1[i]}, \tag{10}$$

$$\dot{Q}_{[i]\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \partial Q_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})/\partial\sigma^2 = -\frac{1}{2\widehat{\sigma^2}}E_{2[i]}, \tag{11}$$

$$\dot{Q}_{[i]\boldsymbol{\alpha}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \partial Q_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\alpha}, \tag{12}$$

where $E_{1[i]} = \sum_{j\neq i}\mathbf{X}_j^\top(\widehat{u\mathbf{y}}_j - \mathbf{Z}_j\widehat{u\mathbf{b}}_j - \widehat{u}_j\mathbf{X}_j\widehat{\boldsymbol{\beta}})$ and $E_{2[i]} = \sum_{j\neq i}(n_j - \frac{A_j}{\widehat{\sigma^2}})$, with $A_j = \mathrm{tr}(\widehat{u\mathbf{y}_j^2} - 2\widehat{u\mathbf{y}\mathbf{b}}_j\mathbf{Z}_j^\top + \widehat{u\mathbf{b}_j^2}\mathbf{Z}_j^\top\mathbf{Z}_j) - 2\widehat{\boldsymbol{\beta}}^\top\mathbf{X}_j^\top(\widehat{u\mathbf{y}}_j - \mathbf{Z}_j\widehat{u\mathbf{b}}_j) + \widehat{u}_j\widehat{\boldsymbol{\beta}}^\top\mathbf{X}_j^\top\mathbf{X}_j\widehat{\boldsymbol{\beta}}$. Finally, the elements of $\dot{Q}_{[i]\boldsymbol{\alpha}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$ are of the form

$$\dot{Q}_{[i]\alpha_r}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = -\frac{1}{2}\sum_{j\neq i}\mathrm{tr}[\mathbf{D}^{-1}\dot{\mathbf{D}}(r) - \mathbf{D}^{-1}\dot{\mathbf{D}}(r)\mathbf{D}^{-1}\widehat{u\mathbf{b}_j^2}].$$

It is necessary to compute the Hessian matrix $\ddot{Q}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ to develop case-deletion, local influence and any particular perturbation schemes, following [36]. The Hessian matrix $\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ has the following elements:

$$\frac{\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} = -\frac{1}{\sigma^2}\mathbf{X}_i^\top \widehat{u}_i \mathbf{X}_i, \qquad \frac{\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\sigma^2} = -\frac{1}{\sigma^4}\mathbf{X}_i^\top(\widehat{u\mathbf{y}}_i - \mathbf{Z}_i\widehat{u\mathbf{b}}_i - \widehat{u}_i\mathbf{X}_i\boldsymbol{\beta}),$$
$$\frac{\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\alpha_r} = \mathbf{0}, \qquad \frac{\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\sigma^2\partial\sigma^2} = \frac{1}{2\sigma^4}[n_i - \frac{2}{\sigma^2}A_i],$$
$$\frac{\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\sigma^2\partial\alpha_r} = 0, \qquad \frac{\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\alpha_s\partial\alpha_r} = \frac{1}{2}\mathrm{tr}(\mathbf{A}(sr)) - \frac{1}{2}\mathrm{tr}(\mathbf{B}(sr)\widehat{u\mathbf{b}_i^2}),$$

where $\mathbf{A}(sr) = \mathbf{D}^{-1}[\dot{\mathbf{D}}(s)\mathbf{D}^{-1}\dot{\mathbf{D}}(r) - \ddot{\mathbf{D}}(s,r)]$ and $\mathbf{B}(sr) = \mathbf{D}^{-1}[\dot{\mathbf{D}}(s)\mathbf{D}^{-1}\dot{\mathbf{D}}(r) + \dot{\mathbf{D}}(r)\mathbf{D}^{-1}\dot{\mathbf{D}}(s) - \ddot{\mathbf{D}}(s,r)]\mathbf{D}^{-1}$, with $\dot{\mathbf{D}}(r) = \partial\mathbf{D}/\partial\alpha_r$, $\ddot{\mathbf{D}}(s,r) = \partial^2\mathbf{D}/\partial\alpha_s\partial\alpha_r$, $r, s = 1, \ldots, p^*$, $p^* = \dim(\boldsymbol{\alpha})$ and $i = 1, \ldots, n$. After some rearrangement and evaluating these derivatives at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, we obtain the Hessian matrix $\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$ (see Appendix A.1) as block-diagonal of the form $\ddot{Q}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \mathrm{diag}(\ddot{Q}_\beta(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}), \ddot{Q}_{\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}), \ddot{Q}_\alpha(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}))$ (the normal case given in [19]), where $\ddot{Q}_\beta(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = -\frac{1}{\widehat{\sigma^2}}\sum_{i=1}^{n}\mathbf{X}_i^\top\widehat{u}_i\mathbf{X}_i$, $\ddot{Q}_{\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = b/2(\widehat{\sigma^2})^2$ and $\ddot{Q}_{\boldsymbol{\alpha}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n}\partial^2 Q_i(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})/\partial\alpha_s\partial\alpha_r$, with $\mathbf{X} = (\mathbf{X}_1^\top, \ldots, \mathbf{X}_n^\top)^\top$ and $b = \sum_{i=1}^{n}(n_i - 2A_i/\widehat{\sigma^2})$. Using (9), the next result proposes the one-step pseudo approximation of $\widehat{\boldsymbol{\theta}}_{[i]} = (\widehat{\boldsymbol{\theta}}_{[i]}^\top, \widehat{\sigma^2}_{[i]}, \widehat{\boldsymbol{\alpha}}_{[i]}^\top)^\top$, $i = 1, \ldots, n$. Its proof is straightforward and is therefore omitted.

Proposition 1. *The one-step pseudo approximation for the parameter estimates of the t-LMEC model with the ith case deleted is given by*

$$\widehat{\boldsymbol{\beta}}_{[i]}^1 = \widehat{\boldsymbol{\beta}} + \left(\sum_{i=1}^{n}\mathbf{X}_i^\top\widehat{u}_i\mathbf{X}_i\right)^{-1} E_{1[i]}$$
$$\widehat{\sigma^2}_{[i]}^1 = \widehat{\sigma^2}\left(1 + \frac{E_{2[i]}}{b}\right)$$
$$\widehat{\boldsymbol{\alpha}}_{[i]}^1 = \widehat{\boldsymbol{\alpha}} + \{-\ddot{Q}_{\boldsymbol{\alpha}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}^{-1}\dot{Q}_{[i]\boldsymbol{\alpha}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$$

where $E_{1[i]}$, $E_{2[i]}$ and $\dot{Q}_{[i]\boldsymbol{\alpha}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$ are as in (10)–(12) respectively, $b = \sum_{i=1}^{n}(n_i - 2A_i/\widehat{\sigma^2})$ and $\ddot{Q}_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n}\partial^2 Q_i(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})/\partial\alpha_s\partial\alpha_r$.

Note that Proposition 1 allows a straightforward influence assessment via the case-deletion approach for the t-LMEC model. One needs to compute the ML estimate $\widehat{\boldsymbol{\theta}}$ for the complete data, the ML estimate $\widehat{\boldsymbol{\theta}}_{[i]}$ with the ith case deleted, and compare both estimates using some metric such as the Cook's or likelihood distance. If the difference between them is fairly large, then the ith case is regarded as influential. The generalized Cook distance [36] is defined as

$$GD_i(\boldsymbol{\theta}) = (\widehat{\boldsymbol{\theta}}_{[i]} - \widehat{\boldsymbol{\theta}})^\top\{-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}(\widehat{\boldsymbol{\theta}}_{[i]} - \widehat{\boldsymbol{\theta}}), \quad i = 1, \ldots, n, \tag{13}$$

Substituting (9) into (13), we have the approximation $GD_i^1(\boldsymbol{\theta}) = \dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}})^\top\{-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}^{-1}\dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}})$, $i = 1, \ldots, n$. Since $\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$ is a diagonal matrix, this approximation can be written as $GD_i^1(\boldsymbol{\theta}) = \sum_{k=1}^{p} GD_i^1(\theta_k)$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^\top$ (for details see [33]). Consequently, for our t-LMEC model we have

$$GD_i^1(\boldsymbol{\theta}) = GD_i^1(\boldsymbol{\beta}) + GD_i^1(\sigma^2) + GD_i^1(\boldsymbol{\alpha}). \tag{14}$$

3.2. Local influence

In this section, we consider local influence analysis [8] focusing on the following perturbation schemes: the case-weight, scale matrix and response perturbation. Here, we consider both subject-level and observation-level diagnostics. The subject-level diagnostics identify if a subject is considered influential or not, and is carried out considering a perturbation function for the ith subject. However, in modeling longitudinal data, we have two level of responses, namely, the subject-level and observation level, and intuitively, an influential subject may/may not contain influential observations [23]. Hence, exploring atypical observations at both levels are warranted. The observation-level diagnostics consider a perturbation in the jth observation of the ith subject.

The theoretical developments in this section proceed in the framework of [8,36]. Let $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_g)^\top$ be a perturbation vector varying in an open region $\boldsymbol{\Omega} \subset \mathbb{R}^g$ and $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{y}_c)$, the complete-data log-likelihood with respect to the perturbed model induced by $\boldsymbol{\omega}$. We assume there exists $\boldsymbol{\omega}_0 \in \boldsymbol{\Omega}$, such that $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}_0|\mathbf{y}_c) = \ell_c(\boldsymbol{\theta}|\mathbf{y}_c)$ for all $\boldsymbol{\theta}$. The Q-displacement function $f_Q(\boldsymbol{\omega})$ is defined as $f_Q(\boldsymbol{\omega}) = 2\left[Q\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right) - Q\left(\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})|\widehat{\boldsymbol{\theta}}\right)\right]$, where $\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})$ is the maximum of the function $Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) =$

$\mathrm{E}[\ell_c(\boldsymbol{\theta},\boldsymbol{\omega}|\mathbf{y}_c)|\mathbf{Q},\mathbf{C},\widehat{\boldsymbol{\theta}}]$. The local behavior of the Q-displacement function can be analyzed by using the normal curvature $C_{f_Q,\mathbf{d}}$ of $\boldsymbol{\alpha}(\boldsymbol{\omega})=(\boldsymbol{\omega}^\top,f_Q(\boldsymbol{\omega}))^\top$ at $\boldsymbol{\omega}_0$ in the direction of some unit vector $\mathbf{d}$. It follows that

$$C_{f_Q,\mathbf{d}}=-2\mathbf{d}^\top\ddot{Q}_{\omega_o}\mathbf{d}\quad\text{and}\quad-\ddot{Q}_{\omega_0}=\boldsymbol{\Delta}_{\omega_0}^\top\left\{-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\right\}^{-1}\boldsymbol{\Delta}_{\omega_0},$$

where $\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})=\frac{\partial^2Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$ and $\boldsymbol{\Delta}_{\boldsymbol{\omega}}=\frac{\partial^2Q(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})}$. For our t-LMEC model, we consider $\boldsymbol{\Delta}_{\omega_0}=(\boldsymbol{\Delta}_\beta^\top,\boldsymbol{\Delta}_{\sigma^2}^\top,\boldsymbol{\Delta}_\alpha^\top)^\top$, where $\boldsymbol{\Delta}_\beta=\frac{\partial^2Q(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}^\top}|_{\omega_o}$, $\boldsymbol{\Delta}_{\sigma^2}=\frac{\partial^2Q(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\sigma^2\partial\boldsymbol{\omega}^\top}|_{\omega_o}$ and $\boldsymbol{\Delta}_\alpha=(\boldsymbol{\Delta}_{\alpha1}^\top,\ldots,\boldsymbol{\Delta}_{\alpha p^*}^\top)^\top$, with $\boldsymbol{\Delta}_{\alpha r}=\frac{\partial^2Q(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\alpha_r\partial\boldsymbol{\omega}^\top}|_{\omega_o}$, $r=1,\ldots,p^*$.

3.2.1. Subject-level diagnostics

Case weight perturbation

We consider an arbitrary attribution of weights for the expected value of the complete-data log-likelihood function (perturbed Q-function), which may capture departures in general directions, by writing

$$Q(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})=\mathrm{E}[\ell_c(\boldsymbol{\theta},\boldsymbol{\omega}|\mathbf{y}_c)|\mathbf{Q},\mathbf{C},\widehat{\boldsymbol{\theta}}]=\sum_{i=1}^n\omega_i\mathrm{E}[\ell_i(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{Q},\mathbf{C},\widehat{\boldsymbol{\theta}}]=\sum_{i=1}^n\omega_iQ_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}).$$

Here, $\boldsymbol{\omega}=(\omega_1,\ldots,\omega_n)^\top$ is an $n\times1$ vector and $\boldsymbol{\omega}_o=(1,\ldots,1)^\top$. Note that the local influence analysis for this perturbation scheme is equivalent to the case-deletion approach discussed in Section 3.1 (see Appendix A.2). Under this perturbation scheme, we have $\boldsymbol{\Delta}_\beta=\frac{1}{\sigma^2}\mathbf{X}^\top D(\boldsymbol{\epsilon}_1,\ldots,\boldsymbol{\epsilon}_n)$, $\boldsymbol{\Delta}_{\sigma^2}=-\frac{1}{2\sigma^2}\mathbf{n}^\top+\frac{1}{2\sigma^4}\mathbf{m}^\top$, $\boldsymbol{\Delta}_{\alpha_r}=[\frac{\partial Q_1(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\alpha_r},\ldots,\frac{\partial Q_n(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\alpha_r}]$ for $r=1,\ldots,p^*$, where $\mathbf{n}=(n_1,\ldots,n_n)^\top$, $\mathbf{m}=(A_1,\ldots,A_n)^\top$, $D(\boldsymbol{\epsilon}_1,\ldots,\boldsymbol{\epsilon}_n)$ is a block-diagonal matrix, with $\boldsymbol{\epsilon}_i=\widehat{u\mathbf{y}}_i-\mathbf{Z}_i\widehat{u\mathbf{b}}_i-\widehat{u}_i\mathbf{X}_i\widehat{\boldsymbol{\beta}}$ and $\frac{\partial Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\alpha_r}=-\frac{1}{2}\mathrm{tr}[\mathbf{D}^{-1}\dot{\mathbf{D}}(r)-\mathbf{D}^{-1}\dot{\mathbf{D}}(r)\mathbf{D}^{-1}\widehat{u\mathbf{b}_i^2}]$.

Scale matrix perturbation

In order to study the effects of perturbation on the scale matrix $\boldsymbol{\Sigma}_i=\sigma^2\mathbf{I}_{n_i}+\mathbf{Z}\mathbf{D}\mathbf{Z}_i^\top$, we consider $\mathbf{D}(\omega_i)=\omega_i^{-1}\mathbf{D}$, or $\sigma^2(\omega_i)=\omega_i^{-1}\sigma^2$, for $i=1,\ldots,n$. The non-perturbed model arises when $\boldsymbol{\omega}_o=(1,\ldots,1)^\top$. The perturbed Q-function follows (5), with $\mathbf{D}(\omega_i)$ and $\sigma^2(\omega_i)$ in place of $\mathbf{D}$ and σ^2, respectively. Considering a perturbation on $\mathbf{D}$ (matrix of random effects), we have $\boldsymbol{\Delta}_\beta=\mathbf{0}$, $\boldsymbol{\Delta}_{\sigma^2}=\mathbf{0}$ and $\boldsymbol{\Delta}_{\alpha_r}=\frac{1}{2}[g_1,\ldots,g_n]$, where $g_i=\mathrm{tr}(\mathbf{D}^{-1}\dot{\mathbf{D}}(r)\mathbf{D}^{-1}\widehat{u\mathbf{b}_i^2})$, $r=1,\ldots,p^*$. Perturbation on σ^2 (the random error variance) yields $\boldsymbol{\Delta}_\beta=\frac{1}{\sigma^2}\mathbf{X}^\top D(\boldsymbol{\epsilon}_1,\ldots,\boldsymbol{\epsilon}_n)$, $\boldsymbol{\Delta}_{\sigma^2}=\frac{1}{2\sigma^4}\mathbf{m}^\top$ and $\boldsymbol{\Delta}_{\boldsymbol{\alpha}}=\mathbf{0}$.

Response perturbation

A general way for perturbing the response variables Q_{ij}, $i=1,\ldots,n$, $j=1,\ldots,n_i$, is introduced by considering $Q_{ij}(\omega)=Q_{ij}+\omega_is_{ij}$, where s_{ij} is a known constant. Hence, for the t-LMEC model, the perturbed response is obtained as $y_{ij}(\omega)\le Q_{ij}$ if $C_{ij}=1$, and $y_{ij}(\omega)=Q_{ij}$ if $C_{ij}=0$, where $\mathbf{y}_{ij}(\omega)=\mathbf{y}_{ij}-\omega_is_{ij}$. Again, the perturbed Q-function follows (5), with $\widehat{u\mathbf{y}}_i$, $\widehat{u\mathbf{y}_i^2}$ and $\widehat{u\mathbf{y}\mathbf{b}}_i$ replaced by $\widehat{u\mathbf{y}_{i\omega}}=\widehat{u\mathbf{y}}_i-\omega_i\mathbf{s}_i\widehat{u}_i$, $\widehat{u\mathbf{y}_{i\omega}^2}=\widehat{u\mathbf{y}_i^2}-\omega_i(\widehat{u\mathbf{y}}_i\mathbf{s}_i^\top+\mathbf{s}_i\widehat{u\mathbf{y}_i^\top})+\omega_i^2\mathbf{s}_i\mathbf{s}_i^\top$ and $\widehat{u\mathbf{y}\mathbf{b}_{i\omega}}=\widehat{u\mathbf{y}\mathbf{b}}_i-\omega_i\mathbf{s}_i\widehat{u\mathbf{b}_i^\top}$, respectively, where $\mathbf{s}_i=(s_{i1},\ldots,s_{in_i})^\top$. The vector $\boldsymbol{\omega}_0=\mathbf{0}$ represents no perturbation. Finally, we have $\boldsymbol{\Delta}_\beta=-\frac{1}{\sigma^2}[\mathbf{X}_1^\top\widehat{u}_1\mathbf{s}_1,\ldots,\mathbf{X}_n^\top\widehat{u}_n\mathbf{s}_n]$, $\boldsymbol{\Delta}_{\sigma^2}=-\frac{1}{\sigma^4}[(\widehat{u\mathbf{y}_1}-\mathbf{Z}_1\widehat{u\mathbf{b}_1}-\widehat{u}_1\mathbf{X}_1\boldsymbol{\beta})^\top\mathbf{s}_1,\ldots,(\widehat{u\mathbf{y}_n}-\mathbf{Z}_n\widehat{u\mathbf{b}_n}-\widehat{u}_n\mathbf{X}_n\boldsymbol{\beta})^\top\mathbf{s}_n]$, and $\boldsymbol{\Delta}_\alpha=\mathbf{0}$.

3.2.2. Observation-level diagnostics

We proceed as above considering a perturbation vector $\boldsymbol{\omega}=(\boldsymbol{\omega}_1,\ldots,\boldsymbol{\omega}_g)^\top$, where $\boldsymbol{\omega}_i=(\omega_{i1},\ldots,\omega_{in_i})^\top$, and noting that all the previous results for the subject-level diagnostics hold for the observation-level cases as well. Also, we denote $\mathbf{u}_i=(u_{i1},\ldots,u_{in_i})^\top$, $\mathbf{v}_i=(v_{i1},\ldots,v_{in_i})^\top$ and $\mathbf{g}_i=(g_{i1},\ldots,g_{in_i})^\top$.

Case weight perturbation

In this case, we have $\boldsymbol{\Delta}_\beta=\frac{1}{\sigma^2}[\mathbf{u}_1,\ldots,\mathbf{u}_n]$, with $u_{ij}=X_{ij}^\top(\widehat{u\mathbf{y}}_{ij}-\mathbf{Z}_{ij}\widehat{u\mathbf{b}}_i-\widehat{u}_i\mathbf{X}_{ij}\widehat{\boldsymbol{\beta}})$; $\boldsymbol{\Delta}_{\sigma^2}=-\frac{1}{2\sigma^2}[\mathbf{v}_1,\ldots,\mathbf{v}_n]$ with $v_{ij}=1-\frac{1}{\sigma^2}A_{ij}$ and $A_{ij}=\mathrm{tr}(\widehat{u\mathbf{y}_{ij}^2}-2\widehat{u\mathbf{y}\mathbf{b}}_{ij}\mathbf{Z}_{ij}^\top+\widehat{u\mathbf{b}_i^2}\mathbf{Z}_{ij}^\top\mathbf{Z}_{ij})-2\widehat{\boldsymbol{\beta}}^\top\mathbf{X}_{ij}^\top(\widehat{u\mathbf{y}}_{ij}-\mathbf{Z}_{ij}\widehat{u\mathbf{b}}_i)+\widehat{u}_i\widehat{\boldsymbol{\beta}}^\top\mathbf{X}_{ij}^\top\mathbf{X}_{ij}\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\Delta}_{\alpha_r}=-\frac{1}{2}[\mathbf{g}_1,\ldots,\mathbf{g}_n]$, with $g_{ij}=\mathrm{tr}(\mathbf{D}^{-1}\dot{\mathbf{D}}(r)\mathbf{D}^{-1}(\mathbf{D}-\widehat{u\mathbf{b}_i^2}))$, $r=1,\ldots,p^*$.

Scale matrix perturbation

Similar to the subject-level, we consider perturbations on $\mathbf{D}$ and σ^2. Consequently, for $\mathbf{D}$ we have that $\boldsymbol{\Delta}_\beta=\mathbf{0}$, $\boldsymbol{\Delta}_{\sigma^2}=\mathbf{0}$ and $\boldsymbol{\Delta}_{\alpha_r}=\frac{1}{2}[\mathbf{g}_1,\ldots,\mathbf{g}_n]$, with $g_{ij}=\mathrm{tr}(\mathbf{D}^{-1}\dot{\mathbf{D}}(r)\mathbf{D}^{-1}\widehat{u\mathbf{b}_i^2})$, $r=1,\ldots,p^*$. In addition, a perturbation on σ^2 generates $\boldsymbol{\Delta}_\beta=\frac{1}{\sigma^2}[\mathbf{u}_1,\ldots,\mathbf{u}_n]$, with $u_{ij}=X_{ij}^\top(\widehat{u\mathbf{y}}_{ij}-\mathbf{Z}_{ij}\widehat{u\mathbf{b}}_i-\widehat{u}_i\mathbf{X}_{ij}\widehat{\boldsymbol{\beta}})$; $\boldsymbol{\Delta}_{\sigma^2}=[\mathbf{v}_1,\ldots,\mathbf{v}_n]$, with $v_{ij}=\frac{1}{2\sigma^4}A_{ij}$ and $A_{ij}=\mathrm{tr}(\widehat{u\mathbf{y}_{ij}^2}-2\widehat{u}_i\mathbf{X}_{ij}\widehat{\boldsymbol{\beta}}+\widehat{u\mathbf{b}_i^2}\mathbf{Z}_{ij}^\top\mathbf{Z}_{ij})-2\widehat{\boldsymbol{\beta}}^\top\mathbf{X}_{ij}^\top(\widehat{u\mathbf{y}}_{ij}-\mathbf{Z}_{ij}\widehat{u\mathbf{b}}_i)+\widehat{u}_i\widehat{\boldsymbol{\beta}}^\top\mathbf{X}_{ij}^\top\mathbf{X}_{ij}\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\Delta}_\alpha=\mathbf{0}$.

Response perturbation

Finally, for the response perturbation case, we have $\boldsymbol{\Delta}_\beta=-\frac{1}{\sigma^2}[\mathbf{u}_1,\ldots,\mathbf{u}_n]$, with $u_{ij}=X_{ij}^\top$; $\boldsymbol{\Delta}_{\sigma^2}=-\frac{1}{\sigma^4}[\mathbf{v}_1,\ldots,\mathbf{v}_n]$, with $v_{ij}=(\widehat{u\mathbf{y}}_{ij}-\mathbf{Z}_{ij}\widehat{u\mathbf{b}}_i-\widehat{u}_i\mathbf{X}_{ij}\widehat{\boldsymbol{\beta}})$ and $\boldsymbol{\Delta}_{\alpha_r}=\mathbf{0}$.

As the reader can note, it is impossible to give details for all perturbation schemes that would be of interest. However, if we can find an appropriate $\boldsymbol{\omega}$ such that the perturbed complete data log-likelihood function $\ell_c(\boldsymbol{\theta},\boldsymbol{\omega}|\mathbf{y}_c)$ is smooth enough and the pertinent derivatives in the diagnostic measures are well-defined, we can conduct the local influence analysis without much difficulty.

In order to quantify the influence of a case in the data, we follow the method based on the function $M(0)_l = \sum_{k=1}^{r} \tilde{\zeta}_k \boldsymbol{\varepsilon}_{kl}^2$, where $\tilde{\zeta}_k = \zeta_k/(\zeta_1 + \cdots + \zeta_r)$ and $\boldsymbol{\varepsilon}_k^2 = (\boldsymbol{\varepsilon}_{k1}^2, \ldots, \boldsymbol{\varepsilon}_{kg}^2)^\top$ with $\{(\zeta_k, \boldsymbol{\varepsilon}_k), k = 1, \ldots, g\}$ the eigenvalue–eigenvector pairs of $-2\ddot{Q}_{\omega_0}$, where $\zeta_1 \geq \cdots \geq \zeta_r > \zeta_{r+1} = \cdots = 0$ and the eigenvectors $\{\boldsymbol{\varepsilon}_k, k = 1, \ldots, g\}$ are orthonormal (for details see [19]). The lth case may be regarded as influential if $M(0)_l$ is larger than the benchmark (cut-off).

Based on the work of [36], we use the following conformal normal curvature $B_{f_Q,\mathbf{d}}(\boldsymbol{\theta}) = C_{f_Q,\mathbf{d}}(\boldsymbol{\theta})/tr[-2\ddot{Q}_{\omega_0}]$, whose computation is quite simple and also has the property that $0 \leq B_{f_Q,\mathbf{d}}(\boldsymbol{\theta}) \leq 1$. Let $\mathbf{d}_l$ be a basic perturbation vector with lth entry as 1 and all other entries as zero. Zhu and Lee [36] showed that for all l, $M(0)_l = B_{f_Q,\mathbf{d}_l}$. Thus, we can obtain $M(0)_l$ via $B_{f_Q,\mathbf{d}_l}$. Following [15], we consider our benchmark as $\overline{M}(0) + c^* SM(0)$, where $\overline{M}(0)$ and $SM(0)$ are the mean and standard error of $\{M(0)_l : l = 1, \ldots, g\}$ respectively; and c^* is a selected constant. The choice of c^* is subjective. In this paper, we will consider $c^* = 4$; following [26,35].

4. Censored nonlinear mixed effects model

In this section, we develop the censored nonlinear mixed effects model under the Student's-t distribution (henceforth, t-NLMEC). Similar to the t-LMEC model, we denote the number of subjects by n, and the number of measurements on the ith subject by n_i. Ignoring censoring for the moment, let us consider x_{ij} the vector incorporating explanatory variables (covariates), the longitudinal time component t_{ij}, $\boldsymbol{\beta}_{ij} = (\beta_{1ij}, \ldots, \beta_{sij})^\top$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top (p > s)$. The Student's-$t$ nonlinear mixed effect model (t-NLME model) can be written as

$$\mathbf{y}_i = \eta_i(t_{ij}, \boldsymbol{\beta}_{ij}) + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\beta}_{ij} = d(x_{ij}, \boldsymbol{\beta}, \mathbf{b}_i), \tag{15}$$

where $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$, with y_{ij} the response for subject i at time t_{ij}, $\eta_i(t_{ij}, \boldsymbol{\beta}_{ij}) = (\eta(t_{i1}, \boldsymbol{\beta}_{i1})^\top, \ldots, \eta(t_{in_i}, \boldsymbol{\beta}_{in_i}))^\top$, with $\eta(\cdot)$ being a nonlinear (known) but differentiable function of vector-valued mixed-effects parameters $\boldsymbol{\beta}_{ij}$, $\boldsymbol{\epsilon} = (\epsilon_{i1}, \ldots, \epsilon_{in_i})^\top$ is the random error vector, $d(\cdot)$ is an s-dimensional linear function, and $\mathbf{b}_i = (b_{1i}, \ldots, b_{qi})^\top$ is the vector of random effects ($q \leq s$). The joint distribution of $(\mathbf{b}_i, \boldsymbol{\epsilon}_i)$ follows (1). From [20], the marginal distribution is given by

$$f(\mathbf{y}|\boldsymbol{\theta}) = \prod_{i=1}^{n} \int_0^\infty \int_{\mathbb{R}^q} \phi_{n_i}(\mathbf{y}_i, \eta_i(t_{ij}, d(x_{ij}, \boldsymbol{\beta}, \mathbf{b}_i)), u_i^{-1}\sigma^2 \mathbf{I}_{n_i})\phi_q(\mathbf{b}_i; 0, u_i^{-1}\mathbf{D}) \times G(u_i|\nu/2, \nu/2) d\mathbf{b}_i du_i,$$

where $G(\cdot|a, b)$ denotes the density of a Gamma(a, b) distribution with mean a/b. The marginal distribution $f(\mathbf{y}|\boldsymbol{\theta})$ does not have a closed form because the model function is not linear in the random effects. However, in order to use all the theory on influence diagnostics developed above for the LMEC model, we use the following approximation proposed by [20] which linearizes the t-NLMEC likelihood in terms of $\mathbf{b}_i$ and $\boldsymbol{\beta}$.

Proposition 2. *Let $\widetilde{\mathbf{b}}_i$ and $\widetilde{\boldsymbol{\beta}}$ be expansion points in the neighborhood of $\mathbf{b}_i$ and $\boldsymbol{\beta}$, respectively. Then, the t-NLME model as defined in (1) and (15) has the following t-LME form*

$$\widetilde{\mathbf{y}}_i = \widetilde{\mathbf{W}}_i \boldsymbol{\beta} + \widetilde{\mathbf{H}}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, n, \tag{16}$$

where $\widetilde{\mathbf{y}}_i = \mathbf{y}_i - \widetilde{\eta}_i(\widetilde{\boldsymbol{\beta}}, \widetilde{\mathbf{b}}_i)$, $\mathbf{b}_i \overset{ind}{\sim} t_q(0, \mathbf{D}, \nu)$, $\boldsymbol{\epsilon}_i \overset{ind}{\sim} t_{n_i}(\mathbf{0}, \sigma^2 \mathbf{I}_{n_i}, \nu)$, $\widetilde{\mathbf{H}}_i = \frac{\partial \eta_i(t_{ij}, d(x_{ij}, \widetilde{\boldsymbol{\beta}}, \mathbf{b}_i))}{\partial \mathbf{b}_i^\top}\big|_{\mathbf{b}_i = \widetilde{\mathbf{b}}_i}$, $\widetilde{\mathbf{W}}_i = \frac{\partial \eta_i(t_{ij}, d(x_{ij}, \boldsymbol{\beta}, \widetilde{\mathbf{b}}_i))}{\partial \boldsymbol{\beta}_i^\top}\big|_{\boldsymbol{\beta}_i = \widetilde{\boldsymbol{\beta}}_i}$ *and* $\widetilde{\eta}(\widetilde{\boldsymbol{\beta}}, \widetilde{\mathbf{b}}_i) = \eta_i(t_{ij}, d(x_{ij}, \widetilde{\boldsymbol{\beta}}, \widetilde{\mathbf{b}}_i)) - \widetilde{\mathbf{H}}_i \widetilde{\mathbf{b}}_i - \widetilde{\mathbf{W}}_i \widetilde{\boldsymbol{\beta}}$.

Proof. See [20].

For the censored case, this model (16) is a t-LMEC model with the same structure as (1)–(3). The model matrices in (16) depend on the current parameter value, and need to be recalculated at each iteration. The algorithm iterates between the L-, E- and CM-steps until convergence. Moreover, the influence diagnostics for t-LMEC discussed earlier in Section 3 can be incorporated along with the approximation in (16) to obtain approximate influence diagnostics for t-NLMEC.

The approximation (16) was initially proposed in [19] in the context of censored nonlinear mixed effects models. In particular, simulation studies in that paper revealed that this approximation can efficiently detect outliers contaminating the generated data. More recently, Wang and Lin [29] used this approximation to implement an efficient ECM algorithm for carrying out ML estimation in Student's-t nonlinear mixed-effects models for multi-outcome longitudinal data with missing values. Consequently, we conclude that this approximation is robust, stable, and we do not anticipate any severe consequences in inference when applied to other types of (censored) non-linear models.

5. Application

5.1. AIEDRP dataset

In this section, we consider an AIDS case study from the AIEDRP program [28]. This program is a multicenter observational study of patients with acute and early HIV infection, covering areas such as the evaluation of immune responses to HIV,

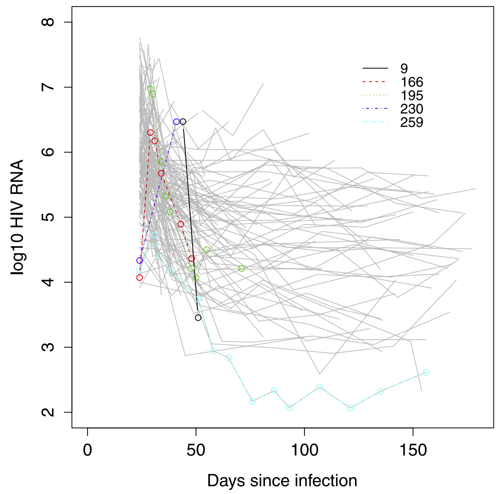

Fig. 1. AIEDRP data. Individual profiles (in $\log_{10}$ scale) for HIV viral load at different follow-up times. Trajectories for some influential individuals are indicated in different colors.

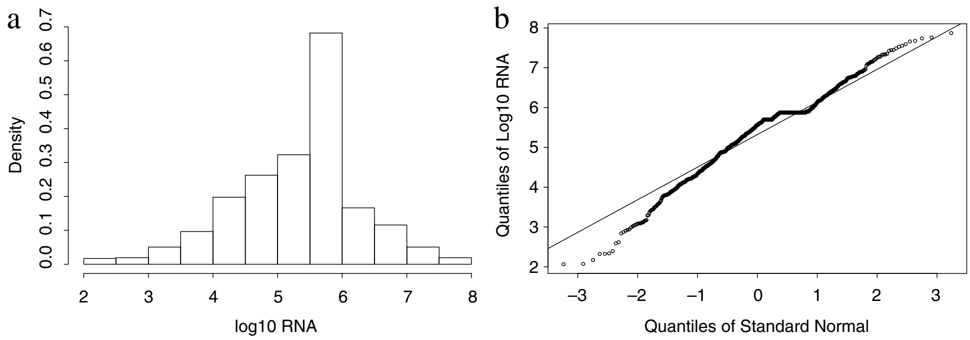

Fig. 2. Plots of raw density histogram (Panel a) and Q–Q plot (Panel b) of viral load.

assessment of thymic function and T-cell turnover during the infection and assessment of transmission and prevalence of HIV resistance. The purpose of the study is to design new vaccines knowing the implications of new antiviral treatments. This dataset has been previously analyzed by some authors in the context of censored non-linear mixed effect models using the Student's-t distribution, see for instance [20] and more recently [10].

In order to illustrate the proposed influence analysis, we consider 320 untreated individuals with HIV infection (see [28] for more details). The dataset consists of 830 observations, with 185 (22%) lying above the limit of assay quantification. The individual profiles are shown in Fig. 1. As was proposed in [28], we consider a right-censored five-parameter NLMEC model as follows:

$$y_{ij} = \lambda_{1i} + \frac{\lambda_2}{1+\exp((t_{ij}-\lambda_3)/\lambda_4)} + \lambda_{5i}(t_{ij}-50) + \epsilon_{ij}, \tag{17}$$

where y_{ij} is the $\log_{10}$ of the viral load for subject i at time t_{ij}. The parameters λ_{1i} and λ_2 represent the subject-specific random setpoints value and decrease from the maximum HIV RNA, respectively. In the absence of treatment (following acute infection), the HIV RNA varies around a setpoint, which may differ among individuals; hence the setpoint is chosen to be subject specific. The location parameter λ_3 indicates the time point at which half of the change in HIV RNA is attained, λ_4 is a scale parameter modeling the rate of decline and λ_{5i} allows for increasing HIV RNA trajectory after day 50. The reparameterization given by $\beta_{1i} = \log(\lambda_{1i}) = \beta_1 + b_{1i}$; $\beta_k = \log(\lambda_k)$, $k = 2, 3, 4$, and $\lambda_{5i} = \beta_5 + b_{2i}$ is adopted to assure positive values for the model parameters. Fig. 2 (Panels a and b) presents raw histogram and Q–Q plot of the log viral load measures, respectively. These plots reveal that viral loads exhibit heavy-tail behavior, and presence of possible outliers. Hence, to accommodate these features, we fit the t-NLMEC model defined in (15) considering the structure given in (17).

5.2. ML estimates using EM algorithm

The model fitting uses the approximated ML method given in Proposition 2 and the ECM algorithm presented in Section 2.2. The degrees of freedom ν is assumed to be known. Using the AIC criterion, we choose $\nu = 10$ which maximizes

Table 1
ML estimates and model comparison criteria for normal and t-NLMEC models. SE are the estimated asymptotic standard errors.

Parameter	N-NLMEC		t-NLMEC	
	MLE	SE	MLE	SE
β_1	1.6093	0.0137	1.6109	0.0133
β_2	0.1449	0.0953	0.1636	0.0854
β_3	3.5256	0.0237	3.5233	0.0207
β_4	1.0599	0.2666	0.9910	0.2450
β_5	−0.0035	0.0015	−0.0031	0.0015
σ^2	0.2621		0.2053	
α_{11}	0.01766		0.01611	
α_{12}	0.00017		0.00014	
α_{22}	0.00005		0.00005	
ν			10	
log-like	−783.8905		−781.8017	
AIC	1585.7812		1581.6034	
BIC	1628.2740		1624.0963	

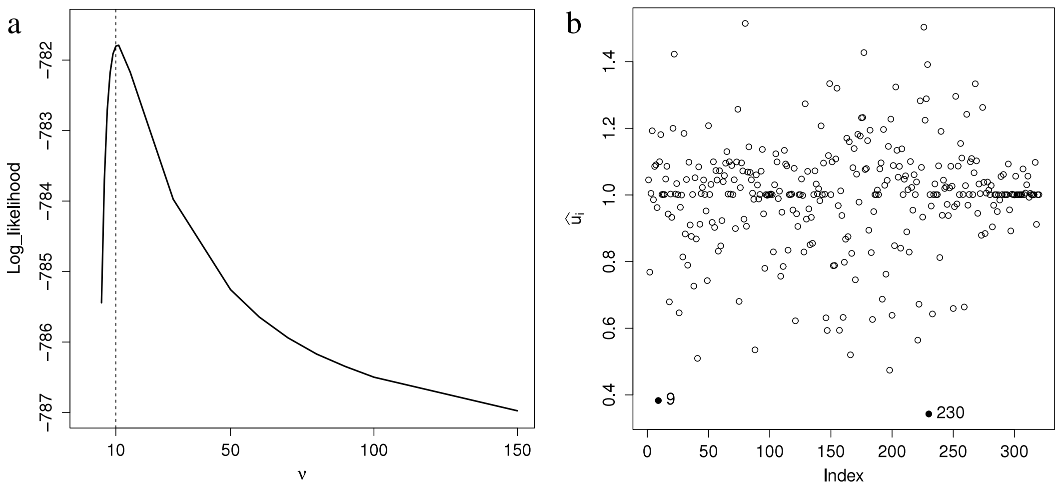

Fig. 3. Plot of the profile log-likelihood versus the degrees of freedom ν (Panel a), and estimated weight $\hat{u}_i$ for the t-NLMEC fit (Panel b), with the influential observations numbered.

the profile log-likelihood (see, Fig. 3, Panel a). This reveals that a fit using a normality-based LMEC model might be inadequate. Further model comparison between the normal and t-NLMEC models using the AIC/BIC criteria presented in Table 1 show that the t-NLMEC model provided a much improved model fit than the normal one.

Because we currently focus on exploring influence diagnostics, details on the estimation and interpretation of the parameter estimates $\boldsymbol{\beta}$ are omitted for brevity. From Fig. 3 (Panel b), we observe that the t-NLMEC model insulates the overall parameter estimation by assigning smaller weights $\hat{u}_i$ to the possible influential observations, which are described later in more details.

5.3. Global influence

In order to evaluate the effect on the ML estimates when some observation is deleted, we analyze the $GD_i^1(\boldsymbol{\theta})$ plot in Fig. 4 (Panel a). The plot reveals that two cases (#195, #259) are potentially influential on the parameter estimates. Fig. 4 (Panels b–d) present plots of $GD_i^1(\boldsymbol{\beta})$, $GD_i^1(\sigma^2)$ and $GD_i^1(\boldsymbol{\alpha})$ respectively, using Proposition 1. From these figures, we infer that subject #195 is influential for $\boldsymbol{\beta}$, #9 and #230 are influential for σ^2, and #259 is influential for α.

5.4. Local influence

Next, we focus on the local influence analysis for the dataset based on $M(0)$, with interest focusing on $\boldsymbol{\theta}$. We study both the subject-level and observation-level diagnostics. It is important to stress that in local influence analysis, there are no general rules so far for selecting the benchmark [15]. Hence, we follow the criterion suggested by [15], *i.e* $M(0)_i > \overline{M}(0) + 3.5SM(0)$, $i = 1, \ldots, 320$, to discriminate whether an observation is influential or not.

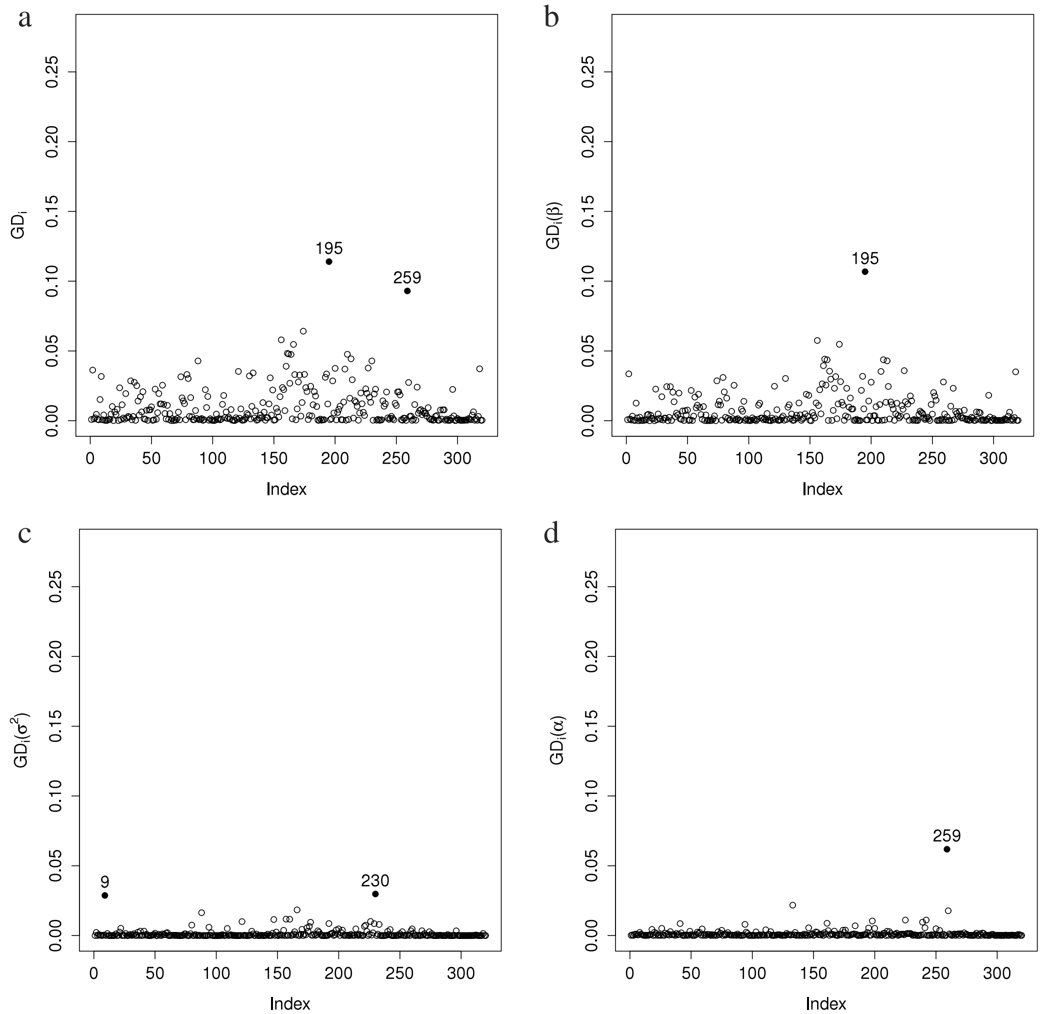

Fig. 4. Global influence. Approximate generalized Cook's distance $GD_i^1(\boldsymbol{\theta})$ (Panel a), GD_i^1 for subset $\boldsymbol{\beta}$ (Panel b), GD_i^1 for subset σ^2 (Panel c), and GD_i^1 for subset $\boldsymbol{\alpha}$ (Panel d). The influential observations are numbered.

5.4.1. Subject-level diagnostics

Fig. 5 presents the index plots of $M(0)$ under the perturbation schemes discussed in Section 3.2.1. We find that subjects #195 and #259 appears influential under case weight perturbation scheme. Moreover, subjects #133 and #159 are potentially influential under perturbation on **D**. For perturbation on σ^2, we find that observations #166, #195 and #259 appear as influential. Finally, for response variable perturbation, observations #174, #175 and #176 are considered as potentially influential. To assess the individual impact of these possible influential observations on the ML estimates, we refitted the t-NLMEC model multiple times by removing one of the following observations: 9, 133, 166, 174, 175, 176, 195, 230 and 259, identified as possibly influential, each time. Table 2 presents the % relative changes (RC) in the parameter estimates presented in Table 1 compared to the parameter estimates obtained after removing the influential observations. Specifically, the RC measure is defined as $\mathrm{RC}_{\widehat{\delta}} = \left|\frac{\widehat{\delta}-\widehat{\delta}_{[i]}}{\widehat{\delta}}\right|$, where $\delta = \beta_1, \ldots, \beta_5, \sigma^2, \boldsymbol{\alpha}$ and $\widehat{\delta_{[i]}}$ denotes the ML estimate of δ with the ith observation removed. From Table 2, we observe that these observations generate greater changes in the RC, in particularly for parameters β_2, α_{12} and α_{22}. These findings are in agreement with the results shown in Fig. 4.

5.4.2. Observation-level diagnostics

Using the perturbation schemes described in Section 3.2.2, Fig. 6 presents the observation-level diagnostics for the dataset. Note that, in the case weight and σ^2 perturbation schemes, the observations #402, #403, #404, #410 (subject #174), #412 (subject #175), #422 (subject #176) and #512, #513, #514 and #515 (subject #203) can be considered influential. For perturbation on **D**, we find that all observations between #680 and #693 can be considered influential. Note, these observations correspond to subject #259, which was considered as possibly influential using the diagnostic

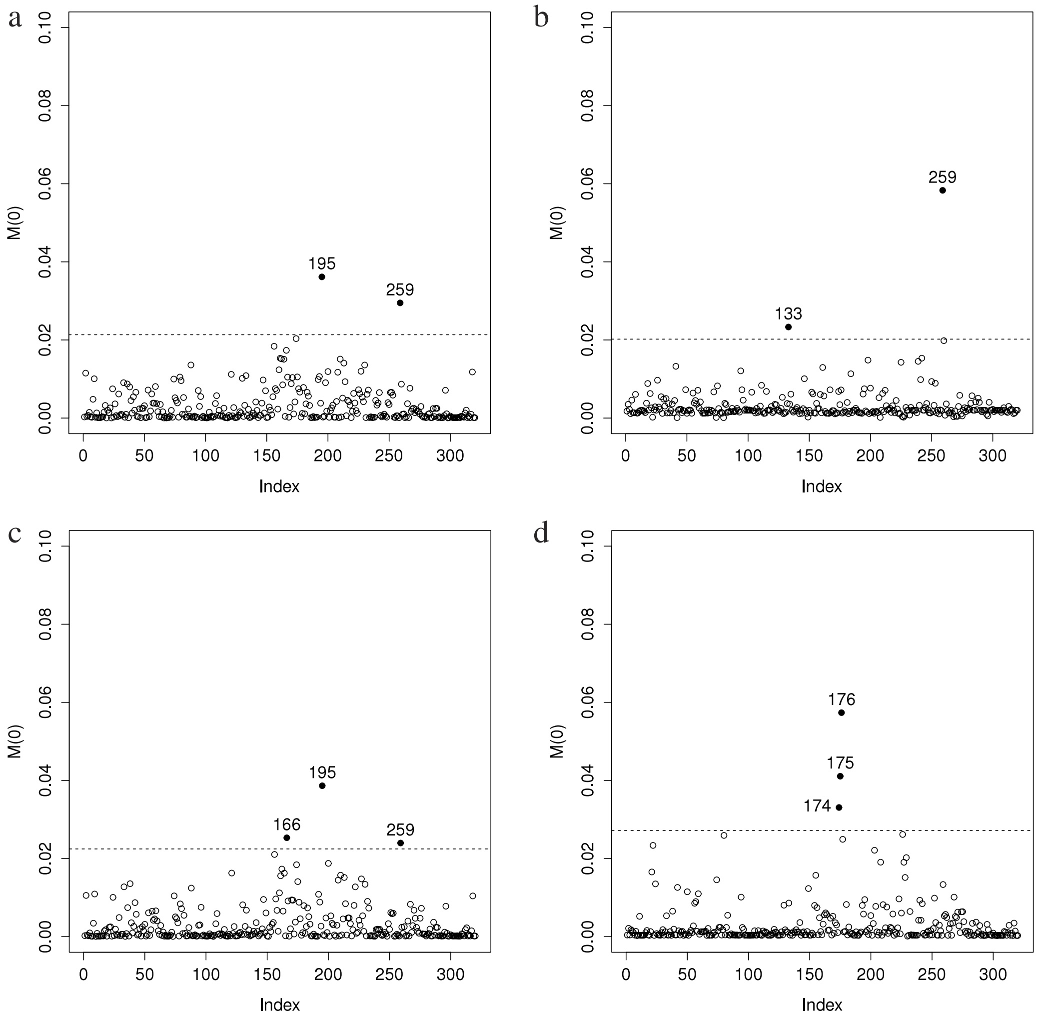

Fig. 5. Index plot of $M(0)$ for assessing local influence on $\boldsymbol{\theta}$ under case weight perturbation (Panel a), perturbation on $\mathbf{D}$ (Panel b), perturbation on σ^2 (Panel c), and perturbation on the response variable (Panel d). The influential observations are numbered.

tools proposed previously (see Sections 5.3 and 5.4.1). Finally, in the case of the perturbation on the response variable, we find that observations #44 (subject #22), #182 and #186 (subject #80), #420 (subject #175), #529 (subject #208), #596 (subject #226), #604 (subject #227) and #616 (subject #229) appear as influential. All these observations with the exception of observation #181 corresponds to the last time observed for the subjects.

6. Simulation studies

In order to assess the finite sample performance of the proposed diagnostic measures for identifying outliers, we conduct a simulation study focusing on subject-level diagnostics. We consider the non-linear mixed-effects model given by

$$y_{ij} = \frac{\beta_1 + b_{i1}}{1 + \exp(-[t_{ij} - (\beta_2 + b_{i2})]/\beta_3)} + \epsilon_{ij}, \quad i = 1, \ldots, 50,\ j = 1, \ldots, 10, \tag{18}$$

where $t_{ij} = 100, 267, 433, 600, 767, 933, 1100, 1267, 1433, 1600$ for all i. The random effects $\mathbf{b}_i = (b_{i1}, b_{i2})^\top$, and the error term $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{i10})^\top$ are non-correlated with

$$\begin{pmatrix} \mathbf{b}_i \\ \boldsymbol{\epsilon}_i \end{pmatrix} \overset{ind.}{\sim} t_{12}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \sigma^2 \mathbf{I}_{10} \end{pmatrix}, 8 \right), \quad i = 1, \ldots, 15.$$

We set the fixed-effects $\boldsymbol{\beta}^\top = (\beta_1, \beta_2, \beta_3) = (200, 700, 350)$, the between-subject covariance matrix $\mathbf{D} = \begin{pmatrix} 4 & -2 \\ -2 & 25 \end{pmatrix}$, and the within-subject variance $\sigma^2 = 25$. Under this model we consider the following perturbation schemes:

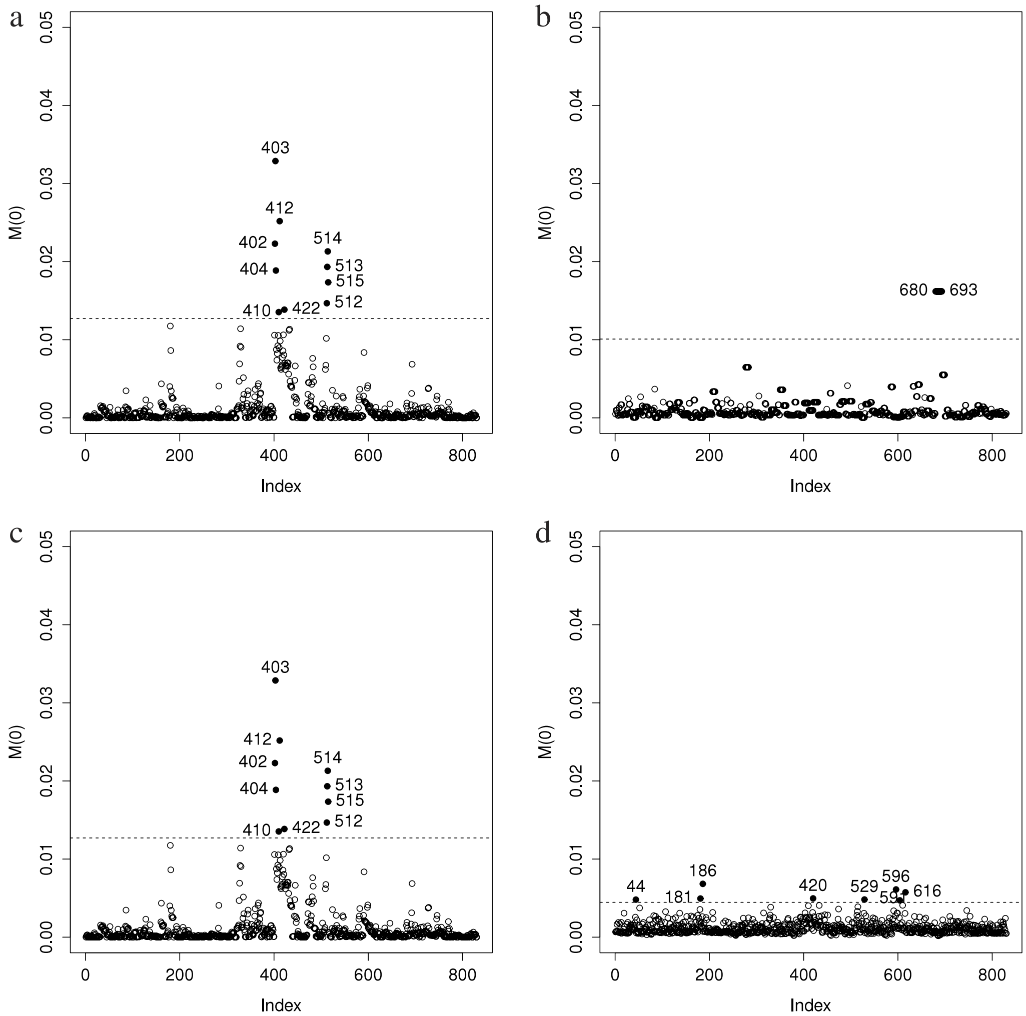

Fig. 6. Index plot of $M(0)$ for assessing local influence on $\boldsymbol{\theta}$ under case weight perturbation (Panel a), perturbation on $\mathbf{D}$ (Panel b), perturbation on σ^2 (Panel c), and perturbation on the response variable (Panel d). The influential observations are numbered.

Table 2
RC (in %) for the AIEDRP study.

Dropped	$RC_{\widehat{\beta}_1}$	$RC_{\widehat{\beta}_2}$	$RC_{\widehat{\beta}_3}$	$RC_{\widehat{\beta}_4}$	$RC_{\widehat{\beta}_5}$	$RC_{\widehat{\sigma^2}}$	$RC_{\widehat{\alpha}_{11}}$	$RC_{\widehat{\alpha}_{12}}$	$RC_{\widehat{\alpha}_{22}}$
9	0.0124	0.9169	0.0170	0.9082	0.0000	2.8738	0.8690	0.0000	20.0000
133	0.1862	8.8020	0.0426	2.6438	3.2258	0.5845	3.5382	0.0000	20.0000
166	0.0062	6.6626	0.1334	3.1887	0.0000	2.4355	1.1173	0.0000	20.0000
174	0.0931	6.9071	0.1845	2.8355	0.0000	1.5100	1.1794	21.4286	0.0000
175	0.0621	1.7726	0.0993	3.0071	3.2258	0.6332	0.8070	7.1429	20.0000
176	0.1800	7.7017	0.0511	2.9162	0.0000	1.0229	1.1794	0.0000	20.0000
195	0.2421	9.4743	0.1760	2.7447	0.0000	0.2923	0.3724	0.0000	0.0000
230	0.0621	4.7066	0.0284	0.7164	0.0000	2.8251	1.0552	0.0000	0.0000
259	0.2111	7.7017	0.1306	4.0464	3.2258	1.1203	6.2073	35.7143	0.0000

(a) Replace the fixed effects $\boldsymbol{\beta}$ by $2\boldsymbol{\beta}$ to generate the responses of the 1st subject $\mathbf{y}_1$,
(b) Replace $\boldsymbol{\beta}$ by $3\boldsymbol{\beta}$ and,
(c) Replace $\boldsymbol{\beta}$ by $4\boldsymbol{\beta}$.

The diagnostic measures were computed for 500 simulated datasets under various censoring proportions, say 0%, 5%, 10%, 20% and 30%. Table 3 reports (in percentage) the number of times the measures correctly identifies $\mathbf{y}_1$ as the most influential.

As expected, the percentage of correctly detecting atypical observations increases for increasing perturbation rates (*i.e.*, for $3\boldsymbol{\beta}$ or $4\boldsymbol{\beta}$) as compared to $2\boldsymbol{\beta}$, and with increased rate of censoring. Interestingly, the % of correct detection when the

Table 3
Simulation study: The values in the table denotes the % of correctly identifying the influential observations using case-deletion, case weight, σ^2 perturbation and matrix **D** perturbation from 500 simulated datasets under the t-NLMEC model specified in (18).

	% of censoring				
	0%	5%	10%	20%	30%
Case-deletion measure (GD_i)					
Pert. $2\boldsymbol{\beta}$	66.8	66.8	74.8	75.8	81.8
Pert. $3\boldsymbol{\beta}$	83.0	83.4	85.8	91.6	94.8
Pert. $4\boldsymbol{\beta}$	93.0	93.2	94.2	97.4	98.4
Case-weight perturbation					
Pert. $2\boldsymbol{\beta}$	66.8	66.8	74.8	75.8	81.8
Pert. $3\boldsymbol{\beta}$	83.0	83.4	85.8	91.6	94.8
Pert. $4\boldsymbol{\beta}$	93.0	93.2	94.2	97.4	98.4
Perturbation on σ^2					
Pert. $2\boldsymbol{\beta}$	13.0	14.4	18.8	19.2	15.2
Pert. $3\boldsymbol{\beta}$	3.60	3.60	4.60	6.00	6.00
Pert. $4\boldsymbol{\beta}$	0.40	0.60	0.80	1.00	0.60
Perturbation on **D**					
Pert. $2\boldsymbol{\beta}$	83.8	83.6	83.2	83.0	84.8
Pert. $3\boldsymbol{\beta}$	95.0	94.6	94.0	94.8	97.4
Pert. $4\boldsymbol{\beta}$	97.2	97.8	97.6	98.8	99.0

influence analysis is focused on σ^2 is not appealing, with a lower percentage of correct detection when the perturbation rate increases. However, higher % of correct detection when the influence analysis is focused on **D** is detected. A possible explanation for this fact is that a perturbation on the fixed-effects of one subject contributes to increasing the between-subject variance, but the within-subject variance remains the same.

7. Conclusions

This article proposes diagnostic tools for detecting outliers and/or influential observations in the context of linear and non-linear mixed-effects censored model where the joint distribution of the random effects and random errors follow the Student's-t distribution. The results presented here supplement the robust likelihood-based inference developed by [20] for LMEC/NLMEC models, appropriate for longitudinal HIV data. Our proposed estimation method relies on the Q-function and the corresponding ECM algorithm. The NLME formulation is mathematically (and computationally) feasible through a linearization. The methodology is implemented using the R software (codes available upon request from the corresponding author), providing practitioners with a convenient tool for further applications in their domain.

For ease of implementation, our current proposal considers an independent within-subject covariance structure, viz. $\sigma^2 I_{n_i}$. Nevertheless, it can be extended to different unstructured covariance matrices (such as AR(1), or ante-dependence) following the work of [23]. This issue is currently under investigation, and we plan to tackle it in a future paper.

Acknowledgments

We thank the Editor, the Associate Editor and two referees whose constructive comments led to an improved presentation. Bandyopadhyay acknowledges support from grants R03DE021762 and R03DE023372 from the US National Institutes of Health. V.H. Lachos acknowledges support from CNPq-Brazil (Grant 305054/2011-2) and from FAPESP-Brazil (Grant 2014/02938-9). L.M. Castro acknowledges funding support by Grant FONDECYT 1130233 from the Chilean government and Grant 2012/19445-0 from FAPESP-Brazil. L.A. Matos acknowledges support from FAPESP-Brazil (Grant 2011/22063-9).

Appendix

A.1. $\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$ is a block-diagonal matrix

From the EM-algorithm, we have $\partial Q(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = 0$. Consequently, for a t-LMEC model:

$$\sum_{i=1}^{n}\mathbf{X}_i^{\top}(\widehat{u\mathbf{y}}_i - \mathbf{Z}_i\widehat{u\mathbf{b}}_i) = \sum_{i=1}^{n}\widehat{u}_i\mathbf{X}_i^{\top}\mathbf{X}_i\widehat{\boldsymbol{\beta}},$$

$$\sum_{i=1}^{n}(n_i\widehat{\sigma^2} - \mathrm{tr}(\widehat{u\mathbf{y}_i^2} - 2\widehat{u\mathbf{y}\mathbf{b}_i}\mathbf{Z}_i^{\top} + \widehat{u\mathbf{b}_i^2}\mathbf{Z}_i^{\top}\mathbf{Z}_i)) = \sum_{i=1}^{n}(2\widehat{\boldsymbol{\beta}}^{\top}\mathbf{X}_i^{\top}(\widehat{u\mathbf{y}}_i - \mathbf{Z}_i\widehat{u\mathbf{b}}_i) - \widehat{u}_i\widehat{\boldsymbol{\beta}}^{\top}\mathbf{X}_i^{\top}\mathbf{X}_i\widehat{\boldsymbol{\beta}}),$$

$$\partial Q(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\alpha} = 0,$$

Finally, from above,

$$\frac{\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial \sigma^2} = -\frac{1}{\sigma^4} \mathbf{X}_i^\top (\widehat{u\mathbf{y}}_i - \mathbf{Z}_j \widehat{u\mathbf{b}}_i - \widehat{u}_i \mathbf{X}_i \boldsymbol{\beta}) = 0$$

and hence, the matrix $\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$ is block-diagonal.

A.2. Equivalence of GD_i^1 and the local influence based on the case weights scheme

For the ith subject, the normal curvature is given by $C_i = 2\boldsymbol{\Delta}_i^\top \left\{-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\right\}^{-1} \boldsymbol{\Delta}_i$, $i = 1, \ldots, n$, where for the case weights perturbation $\boldsymbol{\Delta}_i = \frac{\partial^2 Q_i(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}}|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \frac{\partial Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}$. Since $\dot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = 0$, we can show that

$$\dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = -\dot{Q}_i(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = -\frac{\partial Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}.$$

Then, $\boldsymbol{\Delta}_i = -\dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$ and, as a result, $C_i = 2\dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) \left\{-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\right\}^{-1} \dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$.

Hence, from $GD_i^1(\boldsymbol{\theta}) = \dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}})^\top \{-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}^{-1} \dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}})$, $i = 1, \ldots, n$, we have that $C_i = 2GD_i^1$, and consequently GD_i^1 is equivalent to the local influence based on the case weights scheme.

References

[1] R. Arellano-Valle, H. Bolfarine, On some characterizations of the t-distribution, Statist. Probab. Lett. 25 (1995) 79–85.
[2] D. Bandyopadhyay, L.M. Castro, V.H. Lachos, H.P. Pinheiro, Robust joint non-linear mixed-effects models and diagnostics for censored HIV viral loads with CD4 measurement error, J. Agric. Biol. Environ. Stat. 20 (1) (2015) 121–139.
[3] D. Bandyopadhyay, V. Lachos, L. Castro, D. Dey, Skew normal/independent linear mixed models for censored responses with applications to hiv viral loads, Biometric. J. 54 (3) (2012) 405–425.
[4] M. Banerjee, E.W. Frees, Influence diagnostics for linear longitudinal models, J. Amer. Statist. Assoc. 92 (439) (1997) 999–1005.
[5] R.J. Beckman, C.J. Nachtsheim, R.D. Cook, Diagnostics for mixed–model analysis of variance, Technometrics 29 (4) (1987) 413–426.
[6] R. Christensen, L. Pearson, W. Johson, Case-deletion diagnostics for mixed-models, Technometrics 34 (1992) 38–45.
[7] R. Cook, Detection of influential observation in linear regression, Technometrics (1977) 15–18.
[8] R.D. Cook, Assessment of local influence, J. R. Stat. Soc. Ser. B 48 (1986) 133–169.
[9] R.D. Cook, S. Weisberg, Residuals and Influence in Regression, Chapman & Hall/CRC, Boca Raton, FL, 1982.
[10] A. Garay, L. Castro, J. Leskow, V.H. Lachos, Censored linear regression models for irregulary observed longitudinal data using the multivariate-t distribution, Stat. Methods Med. Res. (2014) http://dx.doi.org/10.1177/0962280214551191. Online early.
[11] J.S. Hodges, Some algebra and geometry for hierarchical models, applied to diagnostics, J. R. Stat. Soc. Ser. B Stat. Methodol. 60 (3) (1998) 497–536.
[12] Y. Huang, G. Dagne, A Bayesian approach to joint mixed-effects models with a skew-normal distribution and measurement errors in covariates, Biometrics 67 (1) (2011) 260–269.
[13] V.H. Lachos, D. Bandyopadhyay, D.K. Dey, Linear and nonlinear mixed-effects models for censored hiv viral loads using normal/independent distributions, Biometrics 67 (2011) 1594–1604.
[14] K.L. Lange, J.S. Sinsheimer, Normal/independent distributions and their applications in robust regression, J. Comput. Graph. Statist. 2 (1993) 175–198.
[15] S.Y. Lee, L. Xu, Influence analysis of nonlinear mixed-effects models, Comput. Statist. Data Anal. 45 (2004) 321–341.
[16] E. Lesaffre, G. Verbeke, Local influence in linear mixed models, Biometrics 54 (1998) 570–582.
[17] T.I. Lin, J.C. Lee, A robust approach to t linear mixed models applied to multiple sclerosis data, Stat. Med. 25 (2006) 1397–1412.
[18] T. Lin, J. Lee, Bayesian analysis of hierarchical linear mixed modeling using the multivariate t distribution, J. Statist. Plann. Inference 137 (2) (2007) 484–495.
[19] L. Matos, V. Lachos, N. Balakrishnan, F. Labra, Influence diagnostics in linear and nonlinear mixed-effects models with censored data, Comput. Statist. Data Anal. 57 (1) (2013) 450–464.
[20] L. Matos, M. Prates, M.-H. Chen, V. Lachos, Likelihood based inference for linear and nonlinear mixed-effects models with censored response using the multivariate-t distribution, Statist. Sinica 23 (2013) 1323–1345.
[21] N. Ndembi, R. Goodall, D. Dunn, A. McCormick, A. Burke, F. Lyagoba, P. Munderi, P. Katundu, C. Kityo, V. Robertson, D. Yirrell, A. Walker, D. Gibb, C. Gilks, P. Kaleebu, D. Pillay, Viral rebound and emergence of drug resistance in the absence of viral load testing: A randomized comparison between Zidovudine–Lamivudine plus Nevirapine and Zidovudine–Lamivudine plus Abacavir, J. Infect. Dis. 201 (1) (2010) 106–113.
[22] F. Osorio, G.A. Paula, M. Galea, Assessment of local influence in elliptical linear models with longitudinal structure, Comput. Statist. Data Anal. 51 (2007) 4354–4368.
[23] J. Pan, Y. Fei, P. Foster, Case-deletion diagnostics for linear mixed models, Technometrics 56 (3) (2014) 269–281.
[24] J.C. Pinheiro, C.H. Liu, Y.N. Wu, Efficient algorithms for robust estimation in linear mixed-effects models using a multivariate t-distribution, J. Comput. Graph. Statist. 10 (2001) 249–276.
[25] W. Qiu, L. Wu, HIV viral dynamic models with censoring and informative dropouts, Stat. Biopharm. Res. 2 (2) (2010) 220–228.
[26] C. Russo, G. Paula, R. Aoki, Influence diagnostics in nonlinear mixed-effects elliptical models, Comput. Statist. Data Anal. 53 (12) (2009) 4143–4156.
[27] F.E. Tan, M.J. Ouwens, M.P. Berger, Detection of influential observations in longitudinal mixed effects regression models, J. R. Stat. Sco. Ser. D. Stat. 50 (3) (2001) 271–284.
[28] F. Vaida, L. Liu, Fast implementation for normal mixed effects models with censored response, J. Comput. Graph. Statist. 18 (4) (2009) 797–817.
[29] W.-L. Wang, T.-I. Lin, Multivariate t nonlinear mixed-effects models for multi-outcome longitudinal data with missing values, Stat. Med. 33 (17) (2014) 3029–3046.
[30] L. Wu, A joint model for nonlinear mixed-effects models with censoring and covariates measured with error, with application to AIDS studies, J. Amer. Statist. Assoc. 97 (460) (2002) 955–964.
[31] H. Wu, Statistical methods for HIV dynamic studies in AIDS clinical trials, Stat. Methods Med. Res. 14 (2) (2005) 171.
[32] L. Wu, W. Liu, X. Hu, Joint inference on HIV viral dynamics and immune suppression in presence of measurement errors, Biometrics (2010) 327–335.
[33] F. Xie, B. Wei, J. Lin, Case-deletion influence measures for the data from multivariate t distributions, J. Appl. Stat. 34 (2007) 907–921.
[34] L. Xu, S. Lee, W. Poon, Deletion measures for generalized linear mixed effects models, Comput. Statist. Data Anal. 51 (2006) 1131–1146.
[35] C.B. Zeller, F.V. Labra, V.H. Lachos, N. Balakrishnan, Influence analyses of skew-normal/independent linear mixed models, Comput. Statist. Data Anal. 54 (5) (2010) 1266–1280.
[36] H. Zhu, S. Lee, Local influence for incomplete-data models, J. R. Stat. Soc. Ser. B Stat. Methodol. 63 (2001) 111–126.
[37] H. Zhu, S. Lee, B. Wei, J. Zhou, Case-deletion measures for models with incomplete data, Biometrika 88 (2001) 727–737.