

Diagnostics for censored mixed-effects models using the multivariate t -distribution

Dipankar Bandyopadhyay^{1*}, Larissa A. Matos², Luis M. Castro³ and Victor H. Lachos²

¹Division of Biostatistics, University of Minnesota, Minneapolis, MN 55455

² Departamento de Estatística, IMECC-UNICAMP, Campinas, São Paulo, Brazil

³ Departamento de Estatística, Universidad de Concepción, Chile

Abstract

In biomedical studies on HIV RNA dynamics, the 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, specially presence of outliers and thick-tails. Motivated by this, Matos et al. (2013b) 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 multivariate Student's- t density, based on the conditional expectation of the complete data log-likelihood which eliminates the complexity associated with the approach of Cook (1977, 1986) for censored mixed-effects models. The new methodology is illustrated through an application to a longitudinal HIV dataset using the NLMEC framework. In addition, a simulation study is presented, which explores the accuracy of the proposed measures in detecting influential observations in heavy-tailed censored data under different perturbation schemes.

Keywords: Censored data; Case-deletion diagnostics; ECM algorithm; linear mixed-effects model; multivariate t ; non-linear mixed-effects model

*Address for correspondence: Division of Biostatistics, University of Minnesota SPH, A460 Mayo MMC 303, 420 Delaware Street SE, Minneapolis, MN 55455. E-mail: dbandyop@umn.edu

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 therapy (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), which are 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 (Ndembu et al. 2010). In addition, the individual viral load trajectories yield large between-subject variations, and covariates like the CD4 counts (the number of CD4+T lymphocytes per microliter of blood) might explain this variation (Satten and Longini Jr 1996). Interest often focus in formulating the correct linear and nonlinear mixed-effects models (LME/NLME) to estimate viral load trajectories, and quantify within- and between-subject variations (Wu 2005; Wu et al. 2010; Qiu and Wu 2010).

However, 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 (Wu 2002). Under non-trivial censoring proportion, considering adhoc alternatives (Huang and Dagne 2011) might lead to bias in fixed effects and variance components estimates. As alternatives to these crude imputation techniques, Vaida et al. (2007); Vaida and Liu (2009) proposed various 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, heavy tails and outliers (Pinheiro et al. 2001). Interestingly, censored HIV viral loads do exhibit heavy-tailed behavior (Lachos et al. 2011). 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 (Lachos et al. 2011). 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- t density (Pinheiro et al. 2001; Lin and Lee 2006, 2007) in the context of LME/NLME models. Some Bayesian propositions in the context of heavy-tailed LMEC/NLMEC models include Lachos et al. (2011) who advocated the use of the normal/independent density (Lange and Sinsheimer 1993), while Bandyopadhyay et al. (2012) studied the LMEC model considering both skewness and heavy-tails. Very recently, Matos et al. (2013b) proposed a full maximum-likelihood (ML) based inference using a computationally convenient exact ECM algorithm for the LMEC/NLMEC models using the multivariate Student- 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, and 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. Although diagnostics for the traditional normality based LME and LMEC (Matos et al. 2013a) 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 (Cook 1977) based on the well-known Cook’s distance. Under normality assumptions for LME, Banerjee and Frees (1997), Hodges (1998), Tan et al. (2001) focused on case-deletion diagnostics for fixed effects, while Christensen et al. (1992) considered a one-step approximation to Cook’s distance for the variance components. The other approach is the computationally attractive local influence approach (Cook 1986), 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 Beckman et al. (1987); Lesaffre and Verbeke (1998); Zhu and Lee (2001); Lee and Xu (2004); Osorio et al. (2007); Russo et al. (2009), among others.

Developing influence diagnostics for LMEC/NLMEC models in the spirit of Cook (1977, 1986) leads to the underlying observed log-likelihood functions involving intractable integrals, rendering 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 (2001) and Zhu et al. (2001) 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 Lee and Xu (2004) and Xu et al. (2006), respectively. This Q -function approach produces result similar to those obtained using the Cook's approach. Recently, Matos et al. (2013a) 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- t distributions) in implementing the Cook's diagnostics for the t -LMEC/NLMEC model of Matos et al. (2013b), we develop case-deletion and influence diagnostics measures using the approach of Zhu et al. (2001) (see also Lee and Xu 2004; Zhu and Lee 2001). The results presented here supplement the robust likelihood-based inference developed by Matos et al. (2013b) for LMEC/NLMEC models, appropriate for longitudinal HIV data.

The rest of this paper is organized as follows. Section 2 presents 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 considering various perturbation schemes for subject-level and observation-level diagnostics. In Section 4, the dynamic t -NLMEC model is defined. The methodology is illustrated in Section 5 using a motivating HIV dataset. Section 6 presents a numerical study comparing the performance of our methods with other normality-based methods. Section 7 concludes, with some possible directions for future research.

2 Censored linear mixed effect model

Ignoring censoring for the moment, the t -LME model of Matos et al. (2013b) is specified as:

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

where

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

which implies that marginally

$$\mathbf{b}_i \stackrel{\text{iid}}{\sim} t_q(\mathbf{0}, \mathbf{D}, \nu) \quad \text{and} \quad \boldsymbol{\epsilon}_i \stackrel{\text{ind.}}{\sim} t_{n_i}(\mathbf{0}, \sigma^2 \mathbf{I}_{n_i}, \nu), \quad i = 1, \dots, n, \quad (3)$$

where $t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ denotes the pdf of a multivariate Student- 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; $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})^\top$ 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 to the $\boldsymbol{\beta}$ ($p \times 1$ vector) fixed effects; \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. (2013b), we consider the case where the response Y_{ij} is not fully observed for all i, j . Consequently, the observed data for the i -th 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

$$\begin{aligned} y_{ij} &\leq Q_{ij} \quad \text{if } C_{ij} = 1, \\ y_{ij} &= Q_{ij} \quad \text{if } C_{ij} = 0. \end{aligned} \quad (4)$$

For simplicity, we assume that the data are left censored, and 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: $\mathbf{y}_i = \text{vec}(\mathbf{y}_i^o, \mathbf{y}_i^c)$, $\mathbf{Q}_i = \text{vec}(\mathbf{Q}_i^o, \mathbf{Q}_i^c)$, $\mathbf{X}_i^\top = (\mathbf{X}_i^o, \mathbf{X}_i^c)$ and $\boldsymbol{\Sigma}_i = \begin{pmatrix} \sum_i^{oo} & \sum_i^{oc} \\ \sum_i^{co} & \sum_i^{cc} \end{pmatrix}$, where $\text{vec}(\cdot)$ denotes the function which stacks vectors or

matrices of the same number of columns. Using properties of multivariate Student- t distribution, we have that $\mathbf{y}_i^o \sim t_{n_i^o}(\mathbf{X}_i^o \boldsymbol{\beta}, \boldsymbol{\Sigma}_i^{oo}, \nu)$, 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

$$\boldsymbol{\mu}_i^{co} = \mathbf{X}_i^c \boldsymbol{\beta} + \boldsymbol{\Sigma}_i^{co} \boldsymbol{\Sigma}_i^{oo-1} (\mathbf{y}_i^o - \mathbf{X}_i^o \boldsymbol{\beta}), \quad \mathbf{S}_i^{co} = \left(\frac{\nu + Q(\mathbf{y}_i^o)}{\nu + n_i^o} \right) \boldsymbol{\Sigma}_i^{cc.o}, \quad (5)$$

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 the subject i is

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

where $T_p(\cdot | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ denoted the cdf of the multivariate Student- 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. (2013b) 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 $\mathbf{Q} = \text{vec}(\mathbf{Q}_1, \dots, \mathbf{Q}_n)$, and $\mathbf{C} = \text{vec}(\mathbf{C}_1, \dots, \mathbf{C}_n)$. 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

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

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

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = \sum_{i=1}^n Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = \sum_{i=1}^n Q_{1i}(\boldsymbol{\beta}, \sigma^2|\widehat{\boldsymbol{\theta}}^{(k)}) + \sum_{i=1}^n Q_{2i}(\boldsymbol{\alpha}|\widehat{\boldsymbol{\theta}}^{(k)}), \quad (6)$$

where

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

$$\text{and } Q_{2i}(\boldsymbol{\alpha}|\widehat{\boldsymbol{\theta}}^{(k)}) = -\frac{1}{2} \log |\mathbf{D}| - \frac{1}{2} \text{tr} \left(\widehat{u}\mathbf{b}_i^{(k)} \mathbf{D}^{-1} \right).$$

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

It is easy to observe that the E-step reduces to the computation of $\widehat{u}\mathbf{y}_i^{(k)} = E\{u_i \mathbf{y}_i \mathbf{y}_i^\top | \mathbf{Q}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\}$, $\widehat{u}\mathbf{y}_i = E\{u_i \mathbf{y}_i | \mathbf{Q}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\}$, 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 Matos et al. (2013b).

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:

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \left(\sum_{i=1}^n \widehat{u}_i^{(k)} \mathbf{X}_i^\top \mathbf{X}_i \right)^{-1} \sum_{i=1}^n \mathbf{X}_i^\top \left(\widehat{u}\mathbf{y}_i^{(k)} - \mathbf{Z}_i \widehat{u}\mathbf{b}_i^{(k)} \right), \quad (7)$$

$$\widehat{\sigma}^{2(k+1)} = \frac{1}{N} \sum_{i=1}^n \left[\widehat{a}_i^{(k)} - 2\widehat{\boldsymbol{\beta}}^{(k)\top} \mathbf{X}_i^\top (\widehat{u}\mathbf{y}_i^{(k)} - \mathbf{Z}_i \widehat{u}\mathbf{b}_i^{(k)}) + \widehat{u}_i^{(k)} \widehat{\boldsymbol{\beta}}^{(k)\top} \mathbf{X}_i^\top \mathbf{X}_i \widehat{\boldsymbol{\beta}}^{(k)} \right], \quad (8)$$

$$\widehat{\mathbf{D}}^{(k+1)} = \frac{1}{n} \sum_{i=1}^n \widehat{u}\mathbf{b}_i^{(k)}, \quad (9)$$

where $N = \sum_{i=1}^n n_i$, and the scale matrix \mathbf{D} unstructured with $\boldsymbol{\alpha}$ being 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 Matos et al. (2013b) for further details. In the following section, we derive influence diagnostic measures, given the ML estimate $\hat{\theta}$.

3 Influence analysis

Influence diagnostic techniques are used in statistical modeling to identify aberrant observations, and assess their impact on model fitting and parameter estimation. Primarily, there are two possible approaches for detecting these influential observations. The most popular one is the case-deletion technique proposed by Cook (1977), where the idea is to analyze a fitted model after deleting cases one at a time, or in small groups, and then to study their impact on the obtained estimates by using some metric, such as the Cook’s or likelihood distance. The other approach is the local influence analysis (Cook 1986), where a minor perturbation of the underlying statistical model is considered, and the stability of the estimation output is assessed. Motivated by the approach of Zhu et al. (2001) that utilizes the Q -function, we develop case-deletion measures, followed by influence measures for the t -LMEC model.

3.1 Case-deletion measures

The case-deletion approach is a commonly used scheme to study the effects of deleting the i th case/observation from the data set. In the rest of the paper, the subscript ‘ $[i]$ ’ will denote the original data set with the i th case deleted. Consequently, the log-likelihood function corresponding to the remaining data is denoted by $\ell(\theta | \mathbf{Y}_{c[i]})$. In order to assess the influence of the i th case on the ML estimate $\hat{\theta}$, we need to compare the difference between $\hat{\theta}_{[i]}$ and $\hat{\theta}$, where $\hat{\theta}_{[i]} = (\hat{\beta}_{[i]}^\top, \hat{\sigma}^2_{[i]}, \hat{\alpha}_{[i]}^\top)^\top$ is the maximizer of the function $Q_{[i]}(\theta | \hat{\theta}) = E\{\ell(\theta | \mathbf{Y}_{c[i]}) | \mathbf{Q}, \mathbf{C}, \hat{\theta}\}$, with $\hat{\theta}$ being the ML estimate of θ . An observation is regarded as influential if its deletion generates considerable influence on model estimates. In other words, if $\hat{\theta}_{[i]}$ is fairly far from $\hat{\theta}$, then the i th observation could be considered as influential. Note that, since the estimator $\hat{\theta}_{[i]}$ is needed for every case, this scheme requires a considerable computational effort, particularly for large sample sizes. For that reason, a one-step pseudo approximation (see Cook and Weisberg 1982; Zhu et al. 2001) is used to reduce

the burden. This approximation follows:

$$\hat{\boldsymbol{\theta}}_{[i]}^1 = \hat{\boldsymbol{\theta}} + \{-\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})\}^{-1}\dot{Q}_{[i]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}), \quad (10)$$

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

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

$$\dot{Q}_{[i]\sigma^2}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) = \partial Q_{[i]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})/\partial \sigma^2 = -\frac{1}{2\sigma^2} E_{2[i]}, \quad (12)$$

$$\dot{Q}_{[i]\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) = \partial Q_{[i]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})/\partial \boldsymbol{\alpha}, \quad (13)$$

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

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

It is necessary to compute the Hessian matrix $\ddot{Q}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^n \partial^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top$, $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\alpha}^\top)^\top$: the parameter vector to develop case-deletion, local influence and any particular perturbation schemes, following Zhu and Lee (2001). The Hessian matrix $\partial^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top$ has the following elements:

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

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 $\dot{\mathbf{D}}(r) = \partial \mathbf{D}/\partial \alpha_r$, $\ddot{\mathbf{D}}(s, r) = \partial^2 \mathbf{D}/\partial \alpha_s \partial \alpha_r$, $r, s = 1, \dots, p^*$, $p^* = \dim(\boldsymbol{\alpha})$ and $i = 1, \dots, n$. After some rearrangement and evaluating these derivatives at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, we obtain the Hessian matrix $\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})$ (see Appendix A) as block-diagonal of the form $\ddot{Q}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) =$

$\text{diag}(\ddot{Q}_\beta(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}), \ddot{Q}_{\sigma^2}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}), \ddot{Q}_\alpha(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}))$ (as in the normal case, see Matos et al. 2013a), where $\ddot{Q}_\beta(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) = -\frac{1}{\sigma^2} \sum_{i=1}^n \mathbf{X}_i^\top \hat{u}_i \mathbf{X}_i$, $\ddot{Q}_{\sigma^2}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) = b/2(\hat{\sigma}^2)^2$ and $\ddot{Q}_\alpha(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^n \partial^2 Q_i(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})/\partial\alpha_s \partial\alpha_r$, with $\mathbf{X} = (\mathbf{X}_1^\top, \dots, \mathbf{X}_n^\top)^\top$ and $b = \sum_{i=1}^n (n_i - 2A_i/\hat{\sigma}^2)$. Using (10), the next result proposes the one-step pseudo approximation of $\hat{\boldsymbol{\theta}}_{[i]} = (\hat{\boldsymbol{\theta}}_{[i]}^\top, \hat{\sigma}_{[i]}^2, \hat{\boldsymbol{\alpha}}_{[i]}^\top)^\top$, $i = 1, \dots, n$. 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 i th case deleted is given by*

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

where $E_{1[i]}$, $E_{2[i]}$ and $\dot{Q}_{[i]}\boldsymbol{\alpha}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})$ are as in (11), (12) and (13) respectively, $b = \sum_{i=1}^n (n_i - 2A_i/\hat{\sigma}^2)$ and $\ddot{Q}_{[i]}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^n \partial^2 Q_i(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})/\partial\alpha_s \partial\alpha_r$.

Note that Proposition 1 allows us to assess influence via the case-deletion approach for the t -LMEC model straightforwardly through computing the ML estimate $\hat{\boldsymbol{\theta}}$ for the complete data, the ML estimate $\hat{\boldsymbol{\theta}}_{[i]}$ with the i th case deleted, and comparing both estimates using some metric, such as the Cook's or likelihood distance. If the difference between them is fairly large, then the i th case is regarded as influential. The generalized Cook distance (Zhu and Lee 2001) is defined as

$$GD_i(\boldsymbol{\theta}) = (\hat{\boldsymbol{\theta}}_{[i]} - \hat{\boldsymbol{\theta}})^\top \{ -\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) \} (\hat{\boldsymbol{\theta}}_{[i]} - \hat{\boldsymbol{\theta}}), i = 1, \dots, n, \quad (14)$$

Substituting (10) into (14), we have the approximation

$$GD_i^1(\boldsymbol{\theta}) = \dot{Q}_{[i]}(\hat{\boldsymbol{\theta}})^\top \{ -\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) \}^{-1} \dot{Q}_{[i]}(\hat{\boldsymbol{\theta}}), i = 1, \dots, n.$$

Since $\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})$ 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 $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^\top$ (for details see Xie et al. 2007). Consequently, for our t -LMEC model

we have

$$GD_i^1(\boldsymbol{\theta}) = GD_i^1(\boldsymbol{\beta}) + GD_i^1(\sigma^2) + GD_i^1(\boldsymbol{\alpha}). \quad (15)$$

3.2 Local Influence

In this section, we consider local influence analysis (Cook 1986) 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 i th 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 (Pan et al. 2013). Hence, exploring atypical observations at both levels are warranted. The observation-level diagnostics consider a perturbation in the j th observation of the i th subject.

The theoretical developments in this section proceed in the framework of Cook (1986) and Zhu and Lee (2001). Let $\boldsymbol{\omega} = (\omega_1, \dots, \omega_g)^\top$ be a perturbation vector varying in an open region $\Omega \subset \mathbb{R}^g$ and $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{y}_c)$, the complete-data log-likelihood respect to the perturbed model induced by $\boldsymbol{\omega}$. We assume there exists $\boldsymbol{\omega}_0 \in \Omega$, 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 follows

$$f_Q(\boldsymbol{\omega}) = 2 \left[Q(\hat{\boldsymbol{\theta}} | \hat{\boldsymbol{\theta}}) - Q(\hat{\boldsymbol{\theta}}(\boldsymbol{\omega}) | \hat{\boldsymbol{\theta}}) \right],$$

where $\hat{\boldsymbol{\theta}}(\boldsymbol{\omega})$ is the maximum of the function $Q(\boldsymbol{\theta}, \boldsymbol{\omega} | \hat{\boldsymbol{\theta}}) = E[\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{y}_c) | \mathbf{Q}, \mathbf{C}, \hat{\boldsymbol{\theta}}]$, which can be written as $\sum_{i=1}^n \omega_i Q_i(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}})$. 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

$$C_{f_Q, \mathbf{d}} = -2\mathbf{d}^\top \ddot{Q}_{\boldsymbol{\omega}_0} \mathbf{d} \quad \text{and} \quad -\ddot{Q}_{\boldsymbol{\omega}_0} = \boldsymbol{\Delta}_{\boldsymbol{\omega}_0}^\top \left\{ -\ddot{Q}(\hat{\boldsymbol{\theta}} | \hat{\boldsymbol{\theta}}) \right\}^{-1} \boldsymbol{\Delta}_{\boldsymbol{\omega}_0},$$

where $\ddot{Q}(\hat{\boldsymbol{\theta}} | \hat{\boldsymbol{\theta}}) = \frac{\partial^2 Q(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \Big|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}}$ and $\boldsymbol{\Delta}_{\boldsymbol{\omega}} = \frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega} | \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^\top} \Big|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}(\boldsymbol{\omega})}$. For our t -LMEC model, we

consider $\Delta_{\omega_o} = (\Delta_{\beta}^{\top}, \Delta_{\sigma^2}^{\top}, \Delta_{\alpha}^{\top})^{\top}$, where $\Delta_{\beta} = \frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega} | \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^{\top}} |_{\boldsymbol{\omega}_o}$, $\Delta_{\sigma^2} = \frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega} | \hat{\boldsymbol{\theta}})}{\partial \sigma^2 \partial \boldsymbol{\omega}^{\top}} |_{\boldsymbol{\omega}_o}$ and $\Delta_{\alpha} = (\Delta_{\alpha_1}^{\top}, \dots, \Delta_{\alpha_{p^*}}^{\top})^{\top}$, with $\Delta_{\alpha_r} = \frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega} | \hat{\boldsymbol{\theta}})}{\partial \alpha_r \partial \boldsymbol{\omega}^{\top}} |_{\boldsymbol{\omega}_o}$, $r = 1, \dots, p^*$.

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

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

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

Scale matrix perturbation

In order to study the effects of perturbation on the scale matrix $\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}_{n_i} + \mathbf{ZDZ}_i^{\top}$, we consider $\mathbf{D}(\omega_i) = \omega_i^{-1} \mathbf{D}$, or $\sigma^2(\omega_i) = \omega_i^{-1} \sigma^2$, for $i = 1, \dots, n$. The non-perturbed model arises when $\boldsymbol{\omega}_o = (1, \dots, 1)^{\top}$. The perturbed Q -function follows (6), 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 $\Delta_{\beta} = \mathbf{0}$, $\Delta_{\sigma^2} = \mathbf{0}$ and $\Delta_{\alpha_r} = \frac{1}{2} [g_1, \dots, g_n]$, where $g_i = \text{tr}(\mathbf{D}^{-1} \dot{\mathbf{D}}(r) \mathbf{D}^{-1} u \mathbf{b}_i^2)$, $r = 1, \dots, p^*$. Perturbation on σ^2 (the random error variance) yields $\Delta_{\beta} = \frac{1}{\sigma^2} \mathbf{X}^{\top} D(\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n)$, $\Delta_{\sigma^2} = \frac{1}{2\sigma^4} \mathbf{m}^{\top}$ and $\Delta_{\alpha} = \mathbf{0}$.

Response perturbation

A general way for perturbing the response variables Q_{ij} , $i = 1, \dots, n$, $j = 1, \dots, n_i$, is

introduced by considering $Q_{ij}(\omega) = Q_{ij} + \omega_i s_{ij}$, where s_{ij} is a known constant. Hence, for the t -LMEC model, the perturbed response is obtained as $y_{ij}(\omega) \leq Q_{ij}$ if $C_{ij} = 1$, and $y_{ij}(\omega) = Q_{ij}$ if $C_{ij} = 0$, where $\mathbf{y}_{ij}(\omega) = \mathbf{y}_{ij} - \omega_i s_{ij}$. Again, the perturbed Q -function follows (6), 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 $\mathbf{s}_i = (s_{i1}, \dots, s_{in_i})^\top$. The vector $\boldsymbol{\omega}_0 = \mathbf{0}$ represents no perturbation. Finally, we have $\boldsymbol{\Delta}_\beta = -\frac{1}{\sigma^2} [\mathbf{X}_1^\top \widehat{u}_1 \mathbf{s}_1, \dots, \mathbf{X}_n^\top \widehat{u}_n \mathbf{s}_n]$, $\boldsymbol{\Delta}_{\sigma^2} = -\frac{1}{\sigma^4} [(\widehat{u\mathbf{y}}_1 - \mathbf{Z}_1 \widehat{u\mathbf{b}}_1 - \widehat{u}_1 \mathbf{X}_1 \boldsymbol{\beta})^\top \mathbf{s}_1, \dots, (\widehat{u\mathbf{y}}_n - \mathbf{Z}_n \widehat{u\mathbf{b}}_n - \widehat{u}_n \mathbf{X}_n \boldsymbol{\beta})^\top \mathbf{s}_n]$, and $\boldsymbol{\Delta}_\alpha = \mathbf{0}$.

3.2.2 Observation-level diagnostics

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

Case weight perturbation

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

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 $\boldsymbol{\Delta}_{\alpha_r} = \frac{1}{2} [\mathbf{g}_1, \dots, \mathbf{g}_n]$, with $g_{ij} = \text{tr}(\mathbf{D}^{-1} \dot{\mathbf{D}}(r) \mathbf{D}^{-1} \widehat{u\mathbf{b}}_i^2)$, $r = 1, \dots, p^*$. In addition, a perturbation on σ^2 generates $\boldsymbol{\Delta}_\beta = \frac{1}{\sigma^2} [\mathbf{u}_1, \dots, \mathbf{u}_n]$, with $u_{ij} = X_{ij}^\top (\widehat{u\mathbf{y}}_{ij} - \mathbf{Z}_{ij} \widehat{u\mathbf{b}}_i - \widehat{u}_i \mathbf{X}_{ij} \boldsymbol{\beta})$; $\boldsymbol{\Delta}_{\sigma^2} = [\mathbf{v}_1, \dots, \mathbf{v}_n]$, with $v_{ij} = \frac{1}{2\sigma^4} A_{ij}$ and $A_{ij} = \text{tr}(\widehat{u\mathbf{y}}_{ij}^2 - 2\widehat{u}_i \mathbf{X}_{ij} \boldsymbol{\beta} + \widehat{u\mathbf{b}}_i^2 \mathbf{Z}_{ij}^\top \mathbf{Z}_{ij}) - 2\widehat{\boldsymbol{\beta}}^\top \mathbf{X}_{ij}^\top (\widehat{u\mathbf{y}}_{ij} - \mathbf{Z}_{ij} \widehat{u\mathbf{b}}_i) + \widehat{u}_i \widehat{\boldsymbol{\beta}}^\top \mathbf{X}_{ij}^\top \mathbf{X}_{ij} \widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\Delta}_\alpha = \mathbf{0}$.

Response perturbation

Finally, for the response perturbation case, we have $\boldsymbol{\Delta}_\beta = -\frac{1}{\sigma^2} [\mathbf{u}_1, \dots, \mathbf{u}_n]$, with $u_{ij} = X_{ij}^\top$;

$$\Delta_{\sigma^2} = -\frac{1}{\sigma^4}[\mathbf{v}_1, \dots, \mathbf{v}_n], \text{ with } v_{ij} = (\widehat{u}\mathbf{y}_{ij} - \mathbf{Z}_{ij}\widehat{u}\mathbf{b}_i - \widehat{u}_i\mathbf{X}_{ij}\widehat{\boldsymbol{\beta}}) \text{ and } \Delta_{\alpha_r} = \mathbf{0}.$$

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 ω such that the perturbed complete data log-likelihood function $\ell_c(\boldsymbol{\theta}, \omega | \mathbf{y}_c)$ 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 $\tilde{\zeta}_k = \zeta_k / (\zeta_1 + \dots + \zeta_r)$ and $\boldsymbol{\varepsilon}_k^2 = (\boldsymbol{\varepsilon}_{k1}^2, \dots, \boldsymbol{\varepsilon}_{kg}^2)^\top$ with $\{(\zeta_k, \boldsymbol{\varepsilon}_k), k = 1, \dots, g\}$ are eigenvalue–eigenvector pairs of $-2\ddot{Q}\omega_0$ with $\zeta_1 \geq \dots \geq \zeta_r > \zeta_{r+1} = \dots = 0$ and orthonormal eigenvectors $\{\boldsymbol{\varepsilon}_k, k = 1, \dots, g\}$ (for details see Matos et al. 2013a). The l th case may be regarded as influential if $M(0)_l$ is larger than the benchmark (cut-off). Following Lee and Xu (2004), we consider our benchmark as $\overline{M}(0) + 3.5SM(0)$, where $\overline{M}(0)$ and $SM(0)$ are the mean and standard error of $\{M(0)_l : l = 1, \dots, g\}$ respectively.

4 Censored nonlinear mixed effect model

In this section, we propose the censored nonlinear mixed effect model under Student- t distribution (t -NLMEC). Similar to the t -LMEC model, we denote the number of subjects by n , and the number of measurements on the i th 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} , $\boldsymbol{\beta}_{ij} = (\beta_{1ij}, \dots, \beta_{sij})^\top$ and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top (p > s)$. The Student- t nonlinear mixed effect model (t -NLME model), can be written as:

$$\mathbf{y}_i = \eta_i(t_{ij}, \boldsymbol{\beta}_{ij}) + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\beta}_{ij} = d(x_{ij}, \boldsymbol{\beta}, \mathbf{b}_i), \quad (16)$$

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

The joint distribution of $(\mathbf{b}_i, \epsilon_i)$ follows (2). From Matos et al. (2013b), 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 result proposed by Matos et al. (2013b) which linearizes the t -NLMEC likelihood in terms of \mathbf{b}_i and $\boldsymbol{\beta}$.

Proposition 2. *Let $\tilde{\mathbf{b}}_i$ and $\tilde{\boldsymbol{\beta}}$ be expansion points in a neighborhood of \mathbf{b}_i and $\boldsymbol{\beta}$, respectively. Then, the t -NLME model as defined in (2) and (16) has the following t -LME form*

$$\tilde{\mathbf{y}}_i = \tilde{\mathbf{W}}_i \boldsymbol{\beta} + \tilde{\mathbf{H}}_i \mathbf{b}_i + \epsilon_i, \quad i = 1, \dots, n, \quad (17)$$

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

Proof: See Matos et al. (2013b).

For the censored case, this model (17) is a t -LMEC model with the same structure as (1)-(4). The model matrices in (17) 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 (17) to obtain approximate influence diagnostics for t -NLMEC. In the next section, we apply these diagnostic methods to the motivating longitudinal HIV data in context of a t -NLMEC model.

5 Application to HIV Data

5.1 ACTG 315 Dataset

In this section, we reanalyze the HIV viral load data from the ACTG 315 clinical trial (Wu 2002), using the t -NLMEC model. In this study, 46 HIV-1 infected patients were treated with a potent ARV therapy. Viral load was recorded on days 0, 2, 7, 10, 14, 21, 28 and weeks 8, 12, 24 and 48 after initiation of treatment, with a total of 361 observations. Measurements below the detectable threshold of 100 copies/mL (40 out of 361, 11%) were considered left-censored, and the censoring process assumed independent of complete data.

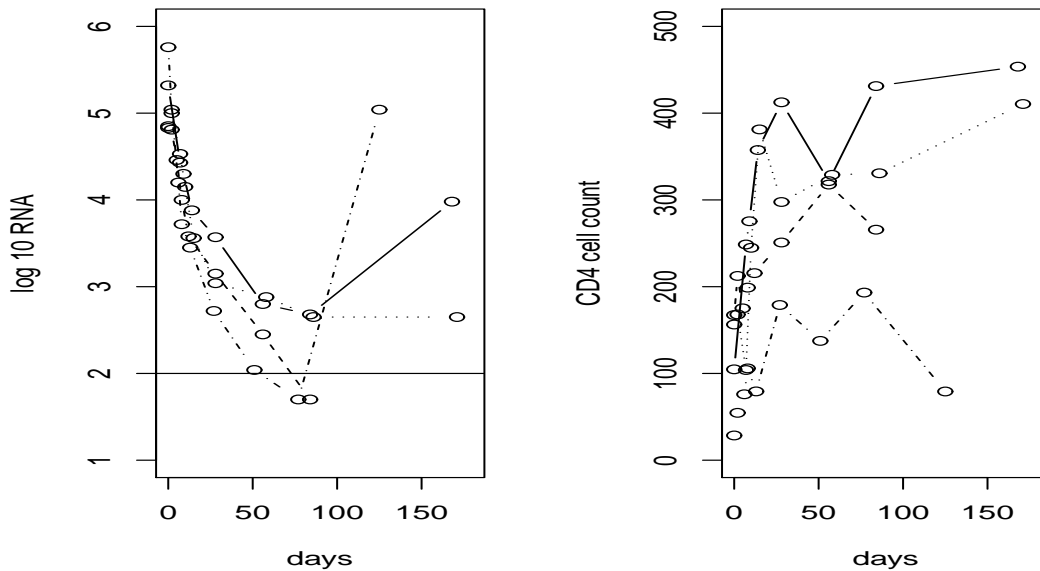


Figure 1: Profiles of viral load in natural \log_{10} scale (Panel a) and CD4 cell count (Panel b) for four randomly selected subjects from the dataset. The horizontal line is below the detectable level of viral load ($2 = \log_{10}(100)$).

Figure 1 plots the viral load (in natural \log_{10} scale) and CD4 longitudinal profiles for four randomly selected patients in the left and right panels, respectively. Clearly, both the viral load and CD4 cell count trajectories exhibit distinct patterns, with the rate change in viral load varying substantially across subjects, possibly reflecting both biological and systematic associations with the subject-level covariate CD4. The viral load trajectories initially exhibit a rapid decay (called

first-phase decay), and after that some have a slower decay (the second-phase), or rebounds back to the original levels (Liu and Wu 2012). In the spirit of Wu (2002) who suggested that a bi-phasic phenomenon might be associated with CD4 counts, we consider the following NLME model:

$$\begin{aligned}
y_{ij} &= \log_{10}(P_{1i}e^{-\lambda_{1ij}t_{ij}} + P_{2i}e^{-\lambda_{2ij}t_{ij}}) + \epsilon_{ij}, \\
\beta_{1ij} &= \log(P_{1i}) = \beta_1 + b_{1i}, \quad \beta_{2ij} = \lambda_{1ij} = \beta_2 + b_{2i}, \\
\beta_{3ij} &= \log(P_{2i}) = \beta_3 + b_{3i}, \quad \beta_{4ij} = \lambda_{2ij} = \beta_4 + \beta_5 CD4_{ij} + b_{4i}
\end{aligned} \tag{18}$$

where y_{ij} is the \log_{10} -transformation of the viral load $V(t_{ij})$ (log-transformation done to stabilize variance) for subject i th at time t_{ij} ($i = 1, 2, \dots, n, j = 1, 2, \dots, n_i$), P_{1i} and P_{2i} are the baseline viral loads, λ_{1ij} and λ_{2ij} are the first- and second-phases of viral decay rates representing the minimum turnover rate of productively infected cells and latently long-lived infected cells, respectively, $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{in_i})^\top$ are within-subject random errors; $\beta_{ij} = (\beta_{1ij}, \beta_{2ij}, \beta_{3ij}, \beta_{4ij})^\top$ and $\beta = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)^\top$ are the subject-level (i th subject at time t_{ij}), and population-level parameters, respectively, $CD4_{ij}$ indicates the observed CD4 counts at time t_{ij} and $b_i = (b_{1i}, \dots, b_{4i})^\top$ are the subject-level random effects. Note that (18) is derived from the bi-phasic exponential decay model $V(t) = P_1e^{-\lambda_1t} + P_2e^{-\lambda_2t}$ of Wu and Ding (1999). Figure 2 (Panel a) presents raw

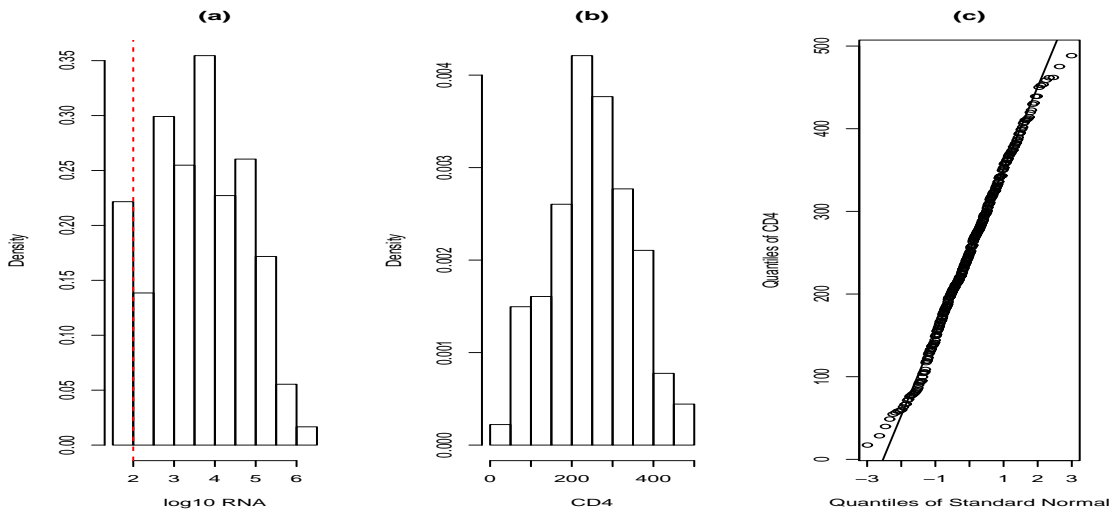


Figure 2: Plots of raw density histogram of viral load (Panel a), density histogram of CD4 cell count (Panel b), and Q-Q plot of CD4 cell count (Panel c) from the HIV dataset. The vertical line in (a) is below the detectable level of viral load ($2 = \log_{10}(100)$).

histogram of the viral load measures, where the viral loads below the threshold are substituted by half the limit, while Figure 2(Panels b and c) display the histogram and the Q-Q plot of the CD4 cell count, respectively. These plots reveal that both viral loads and CD4 counts exhibit heavy-tail behavior, and presence of possible outliers. Hence, to accommodate these features, we fit the t -NLMEC model defined in (16) considering the structure given in (18).

5.2 Model fitting and Diagnostic analysis

The model fitting uses the approximated ML method given in Proposition 2, and the ECM algorithm presented in Section 2.2. To avoid very small/large estimates which might render the estimation method unstable, we standardize the baseline CD4 values, and re-scale the original time t in days between 0 and 1. The degrees of freedom ν is assumed to be known, and using the AIC criterion, $\nu = 8$ which maximizes the profile log-likelihood (See, Figure 3, Panel a). This reveals that a fit using a normality-based LMEC might be inadequate. Further model comparison measures via. the AIC/BIC criteria that compares the normal and t -NLMEC models are presented in Table 1, which show that the t -NLMEC model provided a much improved model fit than the normal one.

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	11.6565	0.1810	11.6457	0.1888
β_2	1.7584	1.6433	31.7590	1.9054
β_3	6.6407	0.3243	6.7695	0.3545
β_4	-0.5095	0.9345	-0.1232	0.8980
β_5	0.3972	0.2547	0.3805	0.2378
σ^2	0.1253		0.1050	
ν			8	
log-likelihood	-281.8423		-258.3164	
AIC	595.6847		548.6329	
BIC	657.9067		610.8549	

Because we currently focus on exploring influence diagnostics, the details on the estimation and interpretation of the parameter estimates of β are omitted for brevity. Figure 3 (Panel b) displays how the t -NLMEC model insulates the overall parameter estimates, and provides a robust estimation scheme by controlling the influence of these influential/outlying observations (numbered in the figure), and attributing smaller weights \hat{u}_i to these observations.

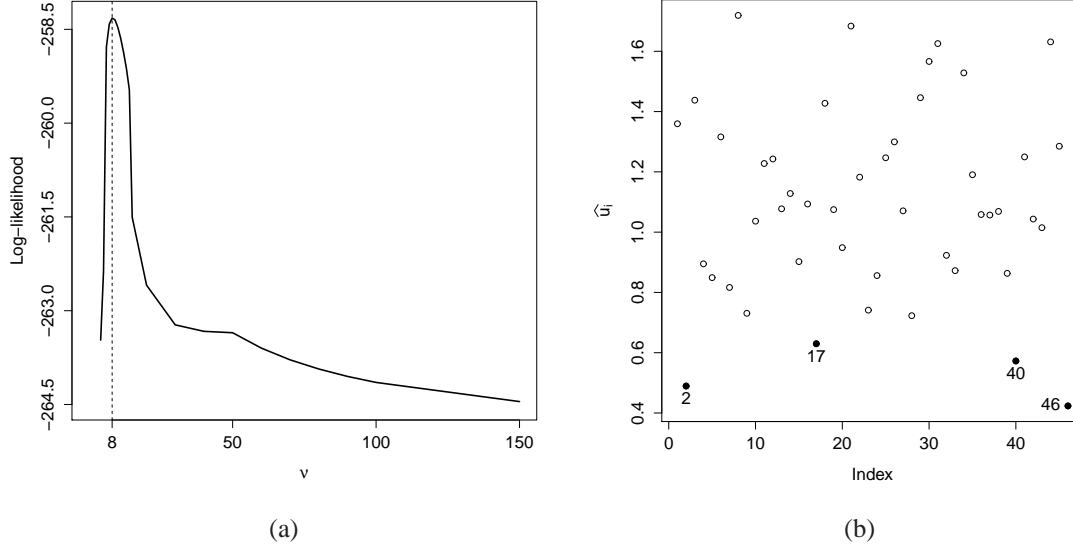


Figure 3: Plot of the profile log-likelihood versus the degrees of freedom ν (Panel a), and estimated weight \hat{u}_i for the t -NLMEC fit (Panel b), with the influential observations numbered.

In order to identify outlying observations, we compute the Mahalanobis distance $d_i^2(\hat{\theta})$, $i = 1, \dots, 46$. Figure 4 (Panels a and b) display the index plot of $d_i^2(\hat{\theta})$ for the normal and t -NLMEC models, respectively, which reveals that subjects #17, #40, #42 and #46 appear to be possible outliers. Moreover, these subjects have large values of $d_{e_i}^2$ in both normal and t -LMEC models, suggesting that they are e-outliers (see Figures 5, Panels a and b). From Figure 5, Panels (c and d), subject #2 with the highest $d_{b_i}^2$ can be considered as b-outlier for both models.

5.2.1 Global influence

In order to evaluate the effect on the ML estimates when some observation is deleted, we analyze the $GD_i^1(\theta)$ plot in Figure 6 (Panel a). The plot reveals that three cases (#7, #24, #40) are potentially influential on the parameter estimates. Figures 6 (Panels b-d) present plots of $GD_i^1(\beta)$, $GD_i^1(\sigma^2)$ and $GD_i^1(\alpha)$ respectively, using Proposition 1. From these figures, we infer that subjects #17, #42 are influential for β , #17, #40, #42 are influential for σ^2 , and observations #7, #24, #40 are influential for α .

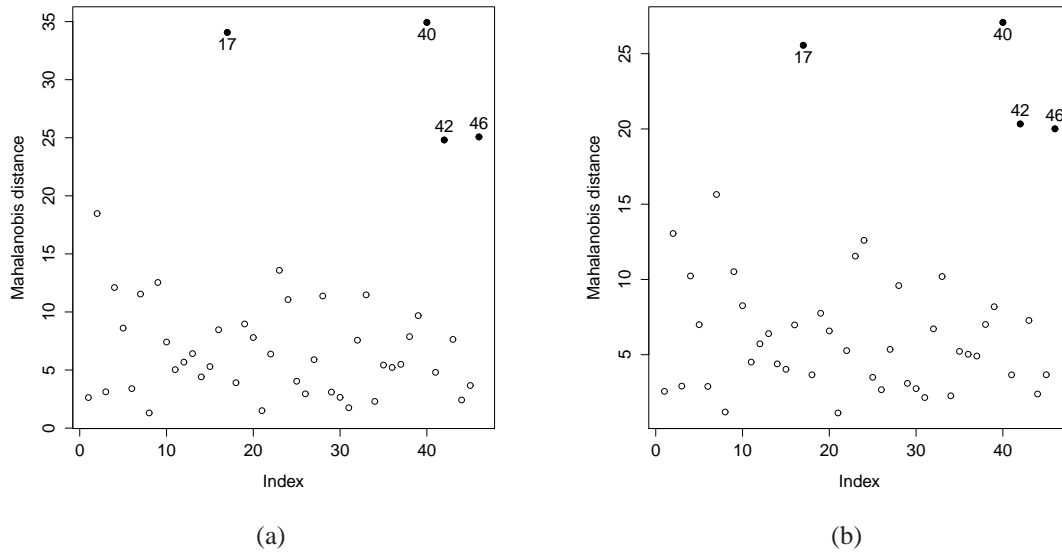


Figure 4: Mahalanobis distance for t -NLMEC (Panel a), Mahalanobis distance for normal NLMEC (Panel b). The influential observations are numbered.

5.2.2 Local influence

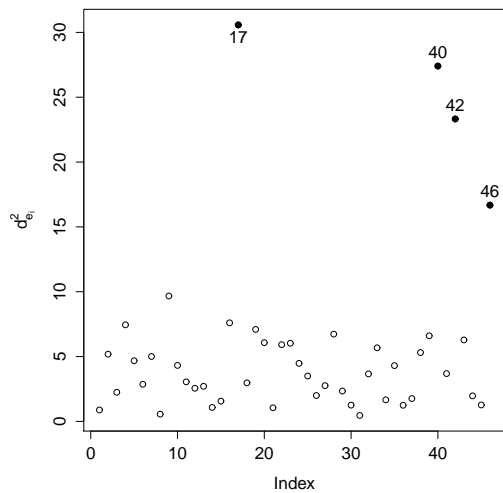
Next, we focus on the local influence analysis for the ACTG 315 data, based on $M(0)$, with interest focussing on θ . We study both the subject-level and observation-level diagnostics. For both of these, we use the criterion $M(0)_i > \overline{M}(0) + 3.5SM(0)$, $i = 1, \dots, 46$, to discriminate whether an observation is influential or not.

Subject-level diagnostics

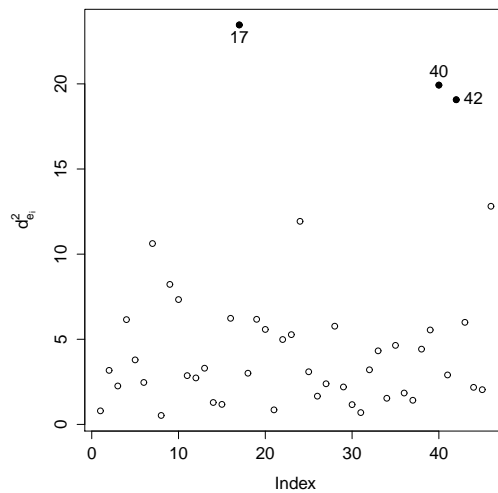
Figure 7 presents the index plots of $M(0)$ under the perturbation schemes discussed in Section 3.2.1. We find that only subject #3 appears influential under response variable perturbation, and no other subjects are influential under the other perturbation schemes. Interestingly, this subject does not have censored responses over time.

Observation-level diagnostics

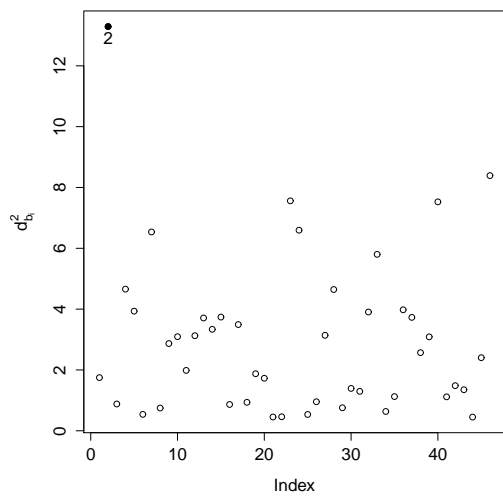
Using the perturbation schemes provided in Section 3.2.2, Figure 8 presents the observation-level diagnostics for the dataset. Note that, in the case weight and σ^2 perturbation schemes, the ob-



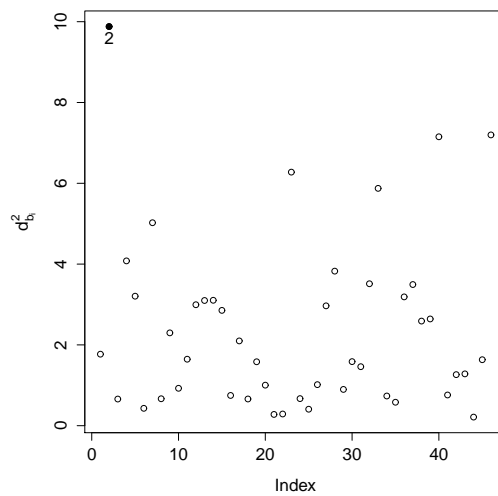
(a)



(b)



(c)



(d)

Figure 5: e-outliers detection for t -NLMEC model (Panel a), e-outliers detection for normal NLMEC model (Panel b), b-outliers detection for t -NLMEC model (Panel c), and b-outliers detection for normal NLMEC model (Panel d). The influential observations are numbered.

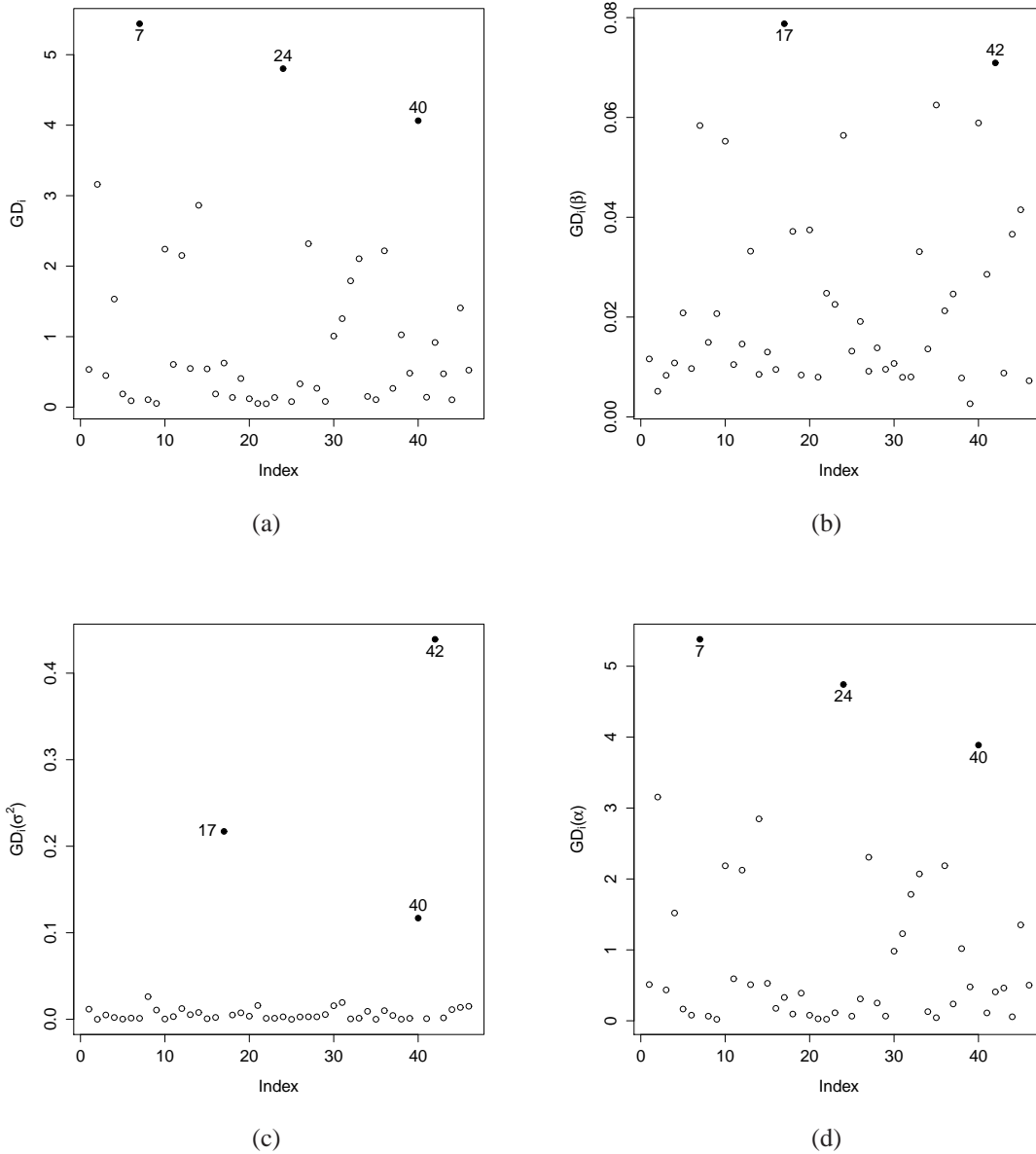


Figure 6: Global influence. Approximate generalized Cook’s distance $GD_i^1(\boldsymbol{\theta})$ (Panel a), GD_i^1 for subset $\boldsymbol{\beta}$ (Panel b), GD_i^1 for subset σ^2 (Panel c), and GD_i^1 for subset $\boldsymbol{\alpha}$ (Panel d). The influential observations are numbered.

servations #53 (subject #7), #76 (subject #10) and #257 (subject #33) could be considered as influential. In the case of the perturbation on the response variable, we find that observations #23 (subject #2), #61 (subject #8), #163 (subject #21), #265 (subject #34) and #348 (subject #44) appear as influential. For perturbation on \mathbf{D} , we do not find influential observations. Interestingly, all these detected observations correspond to the last observed time of the respective subjects.

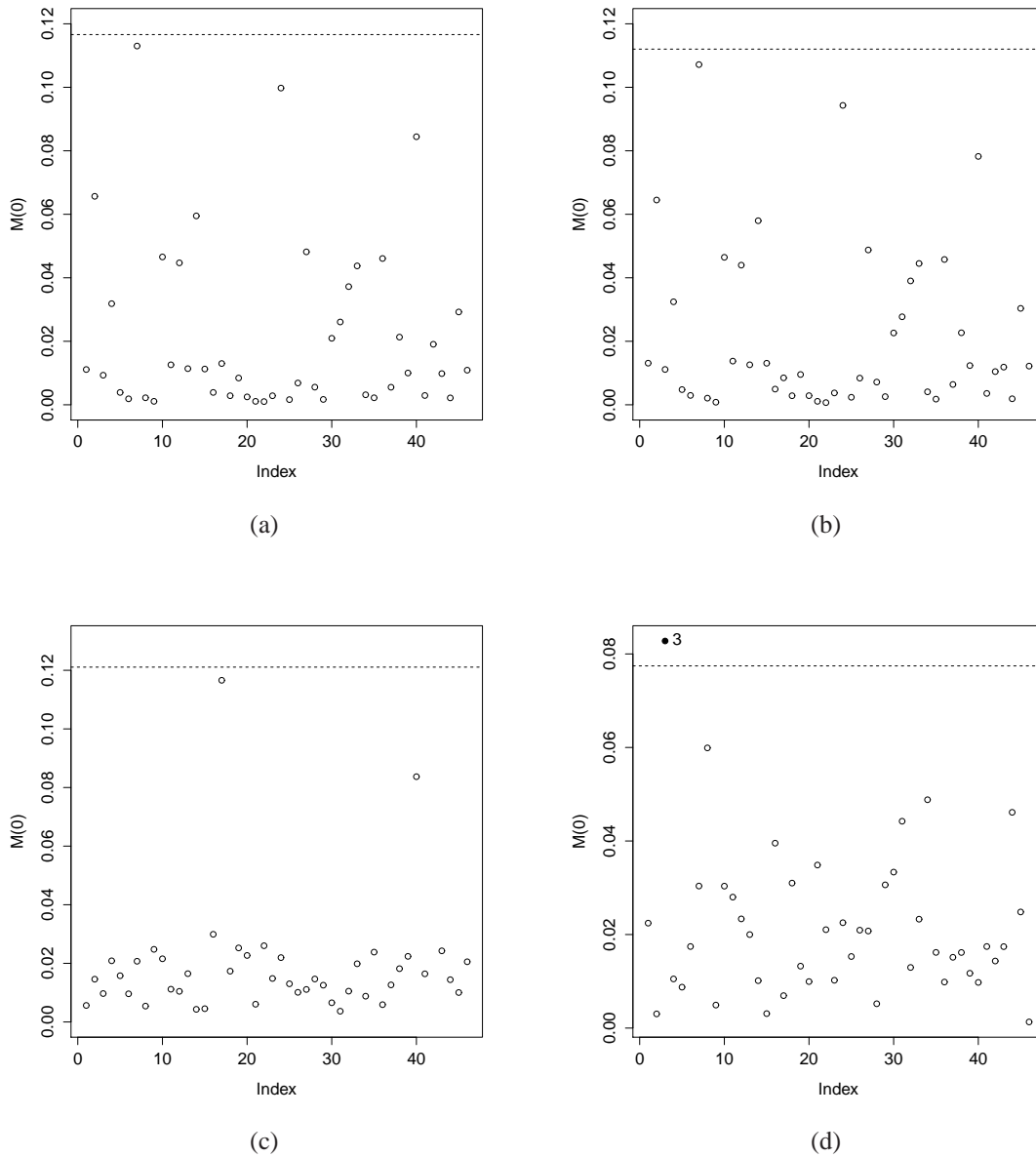


Figure 7: Index plot of $M(0)$ for assessing local influence on θ under case weight perturbation (Panel a), perturbation on \mathbf{D} (Panel b), perturbation on σ^2 (Panel c), and perturbation on the response variable (Panel d). The influential observations are numbered.

6 Simulation studies

To examine the performance of the proposed diagnostic measures for finite samples, we conduct a simulation study focussing on subject-level diagnostics. We consider the non-linear mixed-effects

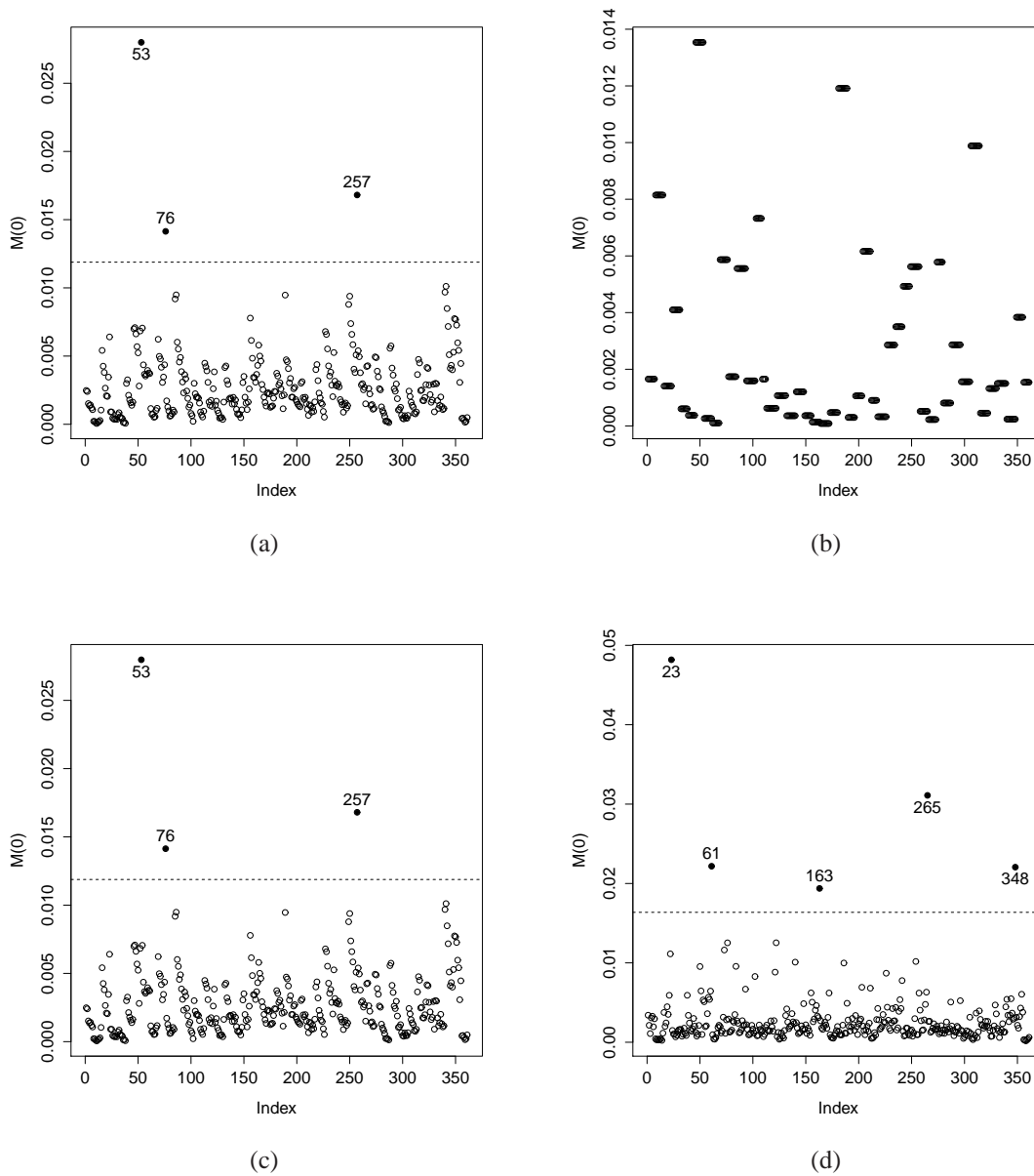


Figure 8: Index plot of $M(0)$ for assessing local influence on θ under case weight perturbation (Panel a), perturbation on \mathbf{D} (Panel b), perturbation on σ^2 (Panel c), and perturbation on the response variable (Panel d). The influential observations are numbered.

model given by

$$y_{ij} = \frac{\beta_1 + b_{i1}}{1 + \exp(-[t_{ij} - (\beta_2 + b_{i2})]/\beta_3)} + \epsilon_{ij}, \quad i = 1, \dots, 15, \quad j = 1, \dots, 10,$$

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 $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{i10})^\top$ are non-correlated with

$$\begin{pmatrix} \mathbf{b}_i \\ \boldsymbol{\epsilon}_i \end{pmatrix} \underset{\text{ind.}}{\sim} t_{12} \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \sigma^2 \mathbf{I}_{10} \end{pmatrix}, 4 \right), i = 1, \dots, 15.$$

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:

- (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 $4\boldsymbol{\beta}$.

Table 2: Simulated data. The values in the table denotes the % of correctly identifying the influential observations using case-deletion, case weight and matrix \mathbf{D} perturbation schemes from 500 simulated datasets under a t -NLMEC model.

Case-deletion measure (GD_i)	% of censoring				
	0%	5%	10%	20%	30%
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	0%	5%	10%	20%	30%
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	0%	5%	10%	20%	30%
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 \mathbf{D}	0%	5%	10%	20%	30%
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

The diagnostic measures were computed for 500 simulated data sets under various settings of censoring proportions, say 0%, 5%, 10%, 20% and 30%. Table 2 reports (in percentage) the

number of times the measures correctly identifies y_1 as the most influential.

To summarize, when the perturbation is high (viz., 3β or 4β), the percentage of correctly detecting atypical observations increases on the overall. This is because the Student- t distribution is a relatively heavy-tailed (as compared to the normal), and hence can accommodate small perturbations (for example 2β) in the data, and has lesser ability to detect influential observations as compared to the normal for smaller perturbations. Interestingly, the % of correct detection increases with the % increase in censored proportions as well, providing more reliability on these measures. Also, the perturbation on the σ^2 do not appear appealing, with a low percentage of correct detection; however the perturbation on \mathbf{D} leads to a substantial increases in detection. 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 constant.

7 Conclusions

This article proposes influence 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 error and random effects follows the Student- t distribution. It supplements the recently published work of Matos et al. (2013b) which considers estimation and inference of those censored models using an EM-type algorithm, and also presents a more robust influence diagnostics framework as compared to the normality-based derivations by Matos et al. (2013a). Our proposed method relies on the Q function, the conditional expectation of the logarithm of the complete-data likelihood, which facilitates the theoretical development of the ECM algorithm in the context of mixed-effects models. The non-linear mixed-effects model formulation is mathematically (and computationally) feasible through a linearization. A simulation study compares the outlier detection accuracy under different censoring and perturbation schemes. For practical demonstration, the methodology is applied to an interesting longitudinal HIV dataset. The methodology is implemented using the R software (codes available upon request from the first author), providing practitioners with a convenient tool for further applications in their domain.

As this is the first work in developing robust outlier detection schemes for censored mixed-

effects models, we decided to keep our estimation strategies simple enough for easy 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 Pan et al. (2013). In addition, some important covariates (such as CD4) in HIV research are often measured with measurement error (Wu 2002), and that might also exert influence on the overall outlier detection strategies. All these issues are currently under investigation, and we plan to tackle these in a future paper.

8 Appendix

8.1 $\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})$ is a block-diagonal matrix

From the EM-algorithm, we know that $\partial Q(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})/\partial \boldsymbol{\theta}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0$. Consequently, we have the following in the t -LMEC model:

$$\begin{aligned} \sum_{i=1}^n \mathbf{X}_i^\top (\widehat{u}\mathbf{y}_i - \mathbf{Z}_i \widehat{u}\mathbf{b}_i) &= \sum_{i=1}^n \widehat{u}_i \mathbf{X}_i^\top \mathbf{X}_i \widehat{\boldsymbol{\beta}}, \\ \sum_{i=1}^n (n_i - \text{tr}(\widehat{u}\mathbf{y}_i^2 - 2\widehat{u}\mathbf{y}_i \mathbf{b}_i \mathbf{Z}_i^\top + \widehat{u}\mathbf{b}_i^2 \mathbf{Z}_i^\top \mathbf{Z}_i)) &= \sum_{i=1}^n (2\widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i^\top (\widehat{u}\mathbf{y}_i - \mathbf{Z}_i \widehat{u}\mathbf{b}_i) - \widehat{u}_i \widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i^\top \mathbf{X}_i \widehat{\boldsymbol{\beta}}), \\ \partial Q(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})/\partial \boldsymbol{\alpha} &= 0, \end{aligned}$$

Finally, note that from the relationships proposed above, the term

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

and therefore the matrix $\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})$ is block-diagonal.

Acknowledgements

Bandyopadhyay acknowledges support from grants R03DE021762 and UL1TR000114 (CTSA Award) from the US National Institutes of Health (NIH), and by NIH grant P30 CA77598 utilizing

the Biostatistics and Bioinformatics Core shared resource of the Masonic Cancer Center, University of Minnesota. V.H. Lachos acknowledges support from CNPq-Brazil (Grant 305054/2011-2) and from FAPESP-Brazil (Grant 2011/17400-6). L.M. Castro acknowledges funding support by Grant FONDECYT 1130233 from the Chilean government and Grant 2012/19445-0 from FAPESP-Brazil.

References

- Bandyopadhyay, D., Lachos, V., Castro, L., Dey, D., 2012. Skewnormal/independent linear mixed models for censored responses with applications to hiv viral loads. *Biometrical Journal* 54, 405–425.
- Banerjee, M., Frees, E.W., 1997. Influence diagnostics for linear longitudinal models. *Journal of the American Statistical Association* 92, 999–1005.
- Beckman, R.J., Nachtsheim, C.J., Cook, R.D., 1987. Diagnostics for mixed-model analysis of variance. *Technometrics* 29, 413–426.
- Christensen, R., Pearson, L., Johson, W., 1992. Case-deletion diagnostics for mixed-models. *Technometrics* 34, 38–45.
- Cook, R., 1977. Detection of influential observation in linear regression. *Technometrics* , 15–18.
- Cook, R.D., 1986. Assessment of local influence. *Journal of the Royal Statistical Society, Series B*, 48, 133–169.
- Cook, R.D., Weisberg, S., 1982. *Residuals and Influence in Regression*. Chapman & Hall/CRC, Boca Raton, FL.
- Hodges, J.S., 1998. Some algebra and geometry for hierarchical models, applied to diagnostics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 60, 497–536.
- Huang, Y., Dagne, G., 2011. A bayesian approach to joint mixed-effects models with a skew-normal distribution and measurement errors in covariates. *Biometrics* 67, 260–269.

- Lachos, V.H., Bandyopadhyay, D., Dey, D.K., 2011. Linear and nonlinear mixed-effects models for censored hiv viral loads using normal/independent distributions. *Biometrics* 67, 1594–1604.
- Lange, K.L., Sinsheimer, J.S., 1993. Normal/independent distributions and their applications in robust regression. *J. Comput. Graph. Stat* 2, 175–198.
- Lee, S.Y., Xu, L., 2004. Influence analysis of nonlinear mixed-effects models. *Computational Statistics and Data Analysis* 45, 321–341.
- Lesaffre, E., Verbeke, G., 1998. Local influence in linear mixed models. *Biometrics* 54, 570–582.
- Lin, T., Lee, J., 2007. Bayesian analysis of hierarchical linear mixed modeling using the multivariate t distribution. *Journal of Statistical Planning and Inference* 137, 484–495.
- Lin, T.I., Lee, J.C., 2006. A robust approach to t linear mixed models applied to multiple sclerosis data. *Statistics in Medicine* 25, 1397–1412.
- Liu, W., Wu, L., 2012. Two-step and likelihood methods for HIV viral dynamic models with covariate measurement errors and missing data. *Journal of Applied Statistics* 39, 963–978.
- Matos, L., Lachos, V., Balakrishnan, N., Labra, F., 2013a. Influence diagnostics in linear and nonlinear mixed-effects models with censored data. *Computational Statistical & Data Analysis* 57, 450–464.
- Matos, L., Prates, M., Chen, M.H., Lachos, V., 2013b. Likelihood based inference for linear and nonlinear mixed-effects models with censored response using the multivariate-t distribution. *Statistica Sinica* 23, 1323–1345.
- Ndembi, N., Goodall, R., Dunn, D., McCormick, A., Burke, A., Lyagoba, F., Munderi, P., Katundu, P., Kityo, C., Robertson, V., Yirrell, D., Walker, A., Gibb, D., Gilks, C., Kaleebu, P., Pillay, D., 2010. 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. *Journal of Infectious Diseases* 201, 106–113.

- Osorio, F., Paula, G.A., Galea, M., 2007. Assessment of local influence in elliptical linear models with longitudinal structure. *Computational Statistics and Data Analysis* 51, 4354–4368.
- Pan, J., Fei, Y., Foster, P., 2013. Case-deletion diagnostics for linear mixed models. Accepted for publication in *Technometrics*, available at <http://amstat.tandfonline.com/doi/abs/10.1080/00401706.2013.810173>.
- Pinheiro, J.C., Liu, C.H., Wu, Y.N., 2001. Efficient algorithms for robust estimation in linear mixed-effects models using a multivariate t-distribution. *Journal of Computational and Graphical Statistics* 10, 249–276.
- Qiu, W., Wu, L., 2010. HIV Viral Dynamic Models With Censoring and Informative Dropouts. *Statistics in Biopharmaceutical Research* 2, 220–228.
- Russo, C., Paula, G., Aoki, R., 2009. Influence diagnostics in nonlinear mixed-effects elliptical models. *Computational Statistics & Data Analysis* 53, 4143–4156.
- Satten, G.A., Longini Jr, I.M., 1996. Markov chains with measurement error: Estimating the true course of a marker of the progression of human immunodeficiency virus disease. *Applied Statistics* , 275–309.
- Tan, F.E., Ouwens, M.J., Berger, M.P., 2001. Detection of influential observations in longitudinal mixed effects regression models. *Journal of the Royal Statistical Society: Series D (The Statistician)* 50, 271–284.
- Vaida, F., Fitzgerald, A., DeGruttola, V., 2007. Efficient hybrid EM for linear and nonlinear mixed effects models with censored response. *Computational statistics & data analysis* 51, 5718–5730.
- Vaida, F., Liu, L., 2009. Fast Implementation for Normal Mixed Effects Models With Censored Response. *Journal of Computational and Graphical Statistics* 18, 797–817.
- Wu, H., 2005. Statistical methods for HIV dynamic studies in AIDS clinical trials. *Statistical Methods in Medical Research* 14, 171.

- Wu, H., Ding, A., 1999. Population HIV-1 dynamics in vivo: applicable models and inferential tools for virological data from AIDS clinical trials. *Biometrics* 55, 410–418.
- Wu, L., 2002. A joint model for nonlinear mixed-effects models with censoring and covariates measured with error, with application to AIDS studies. *Journal of the American Statistical Association* 97, 955–964.
- Wu, L., Liu, W., Hu, X., 2010. Joint Inference on HIV Viral Dynamics and Immune Suppression in Presence of Measurement Errors. *Biometrics* , 327–335.
- Xie, F., Wei, B., Lin, J., 2007. Case-deletion influence measures for the data from multivariate t distributions. *Journal of Applied Statistics* 34, 907–921.
- Xu, L., Lee, S., Poon, W., 2006. Deletion measures for generalized linear mixed effects models. *Computational Statistics & Data Analysis* 51.
- Zhu, H., Lee, S., 2001. Local influence for incomplete-data models. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* 63, 111–126.
- Zhu, H., Lee, S., Wei, B., Zhou, J., 2001. Case-deletion measures for models with incomplete data. *Biometrika* 88, 727–737.