

Influence diagnostics in linear and nonlinear mixed-effects models with censored data

Larissa A. Matos^a, Victor H. Lachos^{a1}, N. Balakrishnan^b and Filidor V. Labra^a

^a Department of Statistics, Campinas State University, Campinas, São Paulo, Brazil

^b Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada

Abstract

HIV RNA viral load measures are often subjected to some upper and lower detection limits depending on the quantification assays, and consequently the responses are either left or right censored. Linear and nonlinear mixed-effects models with modifications to accommodate censoring (LMEC and NLMEC) are routinely used to analyze this type of data. Recently, Vaida and Liu (2009) proposed an exact EM-type algorithm for LMEC/NLMEC, called SAGE algorithm (Meng and Van Dyk, 1997), that uses closed-form expressions at the E-step, as opposed to Monte Carlo simulations. Motivated by this algorithm, we propose here an exact ECM algorithm (Meng and Rubin, 1993) for LMEC/NLMEC, which enable us to develop local influence analysis for mixed effects models on the basis of the conditional expectation of the complete-data log-likelihood function. This is because the observed data log-likelihood function associated with the proposed model is somewhat complex that makes it difficult to apply directly the approach of Cook (1977, 1986). Some useful perturbation schemes are discussed. Finally, the results obtained from the analyses of two HIV AIDS studies on viral loads are presented to illustrate the newly developed methodology.

Keywords: Censored data, HIV viral load, EM Algorithm, Influential observations, Linear mixed models.

1 Introduction

Studies of HIV viral dynamics, often considered to be a key issue in AIDS research, considers repeated/longitudinal measures over a period of treatment routinely analyzed using linear and non-linear mixed effects models (LME/NLME) to assess rates of changes in HIV-1 RNA level or viral load (Wu, 2005, 2010). Viral load measures

¹*Address for correspondence:* Departamento de Estatística, Universidade Estadual de Campinas, Rua Sérgio Buarque de Holanda, 651, Cidade Universitária Zeferino Vaz, Campinas, São Paulo, Brazil - CEP 13083-859. E-mail: hlachos@ime.unicamp.br

the amount of actively replicating virus and its reduction is frequently used as a primary endpoint in clinical trials of anti-retroviral (ARV) therapy. However, depending on the diagnostic assays used, its measurement may be subjected to some upper and lower detection limits, below or above which they are not quantifiable (resulting in left or right censoring). The proportion of censored data in these studies may not be small (Hughes, 1999) and so the use of crude/adhoc methods, viz., substituting threshold value or some arbitrary point such as mid-point between zero and cut-off for detection (Vaida and Liu, 2009), might lead to biased estimates of fixed effects and variance components (Wu, 2010).

Our motivating datasets in this study are on HIV-1 viral load, (i) after unstructured treatment interruption, or UTI (Saitoh et al., 2008) and (ii) setpoint for acutely infected subjects from the AIEDRP program (Vaida and Liu, 2009). The former has about 7% of observations below (left-censored) the detection-limits, whereas the latter has about 22% lying above (right-censored) the limits of assay quantifications. As an alternative to crude imputation methods, Hughes (1999) proposed a likelihood-based Monte Carlo EM algorithm (MCEM) for LME with censored responses (LMEC). Vaida et al. (2007) proposed a hybrid EM using a more efficient Hughes algorithm, extending it to NLME with censored data (NLMEC). Recently, Vaida and Liu (2009) proposed an exact EM-type algorithm for LMEC/NLMEC, which uses closed-form expressions at the E-step, as opposed to Monte Carlo simulations. Strictly speaking, these algorithms are Space Alternating Generalized EM (SAGE) algorithms (see Vaida et al., 2007). In this paper, in order to perform diagnostics analysis in LMEC/NLMEC models, we first propose a slight modification to the EM-type algorithm proposed by Vaida and Liu (2009), wherein all the parameters are updated (M-step) by considering the random effects and the censoring observations as missing data. Then, the diagnostic measures for assessing the local influence in LMEC/NLMEC are developed and presented.

The study of influence analysis is an important and key step in data analysis subsequent to parameter estimation. This can be carried out by conducting an influence analysis for detecting influential observations. There are two primary approaches for detecting influential observations. The first approach is the case-deletion approach (Cook, 1977) and it is an intuitively appealing method (see also Cook and Weisberg ,1982). Deletion diagnostics such as Cook's distance or the likelihood distance have been applied to many statistical models. The second approach, which is a general statistical technique used to assess the stability of the estimation outputs with respect to the model inputs, is the local influence approach of Cook (1986). Following

the pioneering work of Cook (1986), this method has received considerable attention recently in the statistical literature of mixed effects models (LME/NLME); see, for example, Lesaffre and Verbeke (1998), Zhu and Lee (2001), Lee and Xu (2004), Osorio et al. (2007) and Russo et al. (2009), among others.

Although several diagnostic studies on LME/NLME have appeared in the literature, to the best of our knowledge, no study seems to have been made on influence diagnostics for NLMEC/LMEC and certainly not on the local influence analysis. The main difficulty is due to the fact that the observed log-likelihood function of the NLMEC/LMEC involves intractable integrals (for instance, the pdfs of truncated multinormal distributions), rendering the direct application of Cook's approach (Cook, 1986) to this model to be very difficult if not impossible, since the measures involve the first and second partial derivatives of this function. Zhu and Lee (2001) developed an approach for performing local influence analysis for general statistical models with missing data. And it is based on the Q -displacement function that is closely related to the conditional expectation of the complete-data log-likelihood in the E-step of the EM algorithm. This approach produces results very similar to those obtained from Cook's method. Moreover, the case-deletion can be studied by Q -displacement function following the approach of Zhu et al. (2001). So, we develop here methods to obtain case-deletion measures and local influence measures by using the method of Zhu et al. (2001) (see also Lee and Xu, 2004; Zhu and Lee, 2001) in the context of mixed effects models with censored data. We feel that the results developed here are a necessary supplement to those presented by Vaida and Liu (2009) for the analysis of mixed effects models with censored response and HIV data.

The rest of this paper is organized as follows. In Section 2, the LMEC model is defined, and an EM-type algorithm for obtaining the ML estimates is described. In Section 3, we provide a brief sketch of the local influence approach for models with incomplete data, and also develop a methodology pertinent to the LMEC. Four different perturbation schemes have been considered. In Section 4, the NLMEC model is defined. The methodology has been illustrated in Section 5 with the analysis of two examples involving HIV viral measure and an empirical study. Finally, some concluding remarks are made in Section 6.

2 The linear mixed effects with censored response

Ignoring censoring for the moment, the classical normal LME model is specified as follows (Laird and Ware, 1982):

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

where $\mathbf{b}_i \stackrel{iid}{\sim} N_q(\mathbf{0}, \mathbf{D})$ is independent of $\boldsymbol{\epsilon}_i \stackrel{iid}{\sim} N_{n_i}(\mathbf{0}, \sigma^2\mathbf{I}_{n_i})$, $i = 1, \dots, n$; the subscript i is 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 $n_i \times 1$ vector of observed continuous responses for subject i ; \mathbf{X}_i is the $n_i \times p$ design matrix corresponding to the fixed effects, $\boldsymbol{\beta}$, of dimension $p \times 1$; \mathbf{Z}_i is the $n_i \times q$ design matrix corresponding to the $q \times 1$ vector of random effects \mathbf{b}_i ; $\boldsymbol{\epsilon}_i$ of dimension $(n_i \times 1)$ is the vector of random errors; and the dispersion matrix $\mathbf{D} = \mathbf{D}(\boldsymbol{\alpha})$ depends on unknown and reduced parameters $\boldsymbol{\alpha}$. In the present formulation, we consider the case where the response Y_{ij} is not fully observed for all i, j . Let the observed data for the i -th subject be $(\mathbf{Q}_i, \mathbf{C}_i)$, where \mathbf{Q}_i represents the vector of uncensored reading or censoring level, and \mathbf{C}_i is the vector of censoring indicators such that

$$\begin{aligned} y_{ij} &\leq Q_{ij} && \text{if } C_{ij} = 1, \\ y_{ij} &= Q_{ij} && \text{if } C_{ij} = 0. \end{aligned} \quad (2)$$

For simplicity, we will assume that the data are left-censored with the LMEC being thus defined. The extensions of these results to arbitrary censoring can be easily presented.

2.1 The log-likelihood function

Following Vaida and Liu (2009), classical inference on the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\alpha}^\top)^\top$ is based on the marginal distribution of \mathbf{y}_i . For complete data, we have marginally that $\mathbf{y}_i \stackrel{iid}{\sim} N_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i)$, where $\boldsymbol{\Sigma}_i = \sigma^2\mathbf{I}_{n_i} + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top$. For responses with censoring pattern as in (2), we have $\mathbf{y}_i \sim TN_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i; \mathbb{A})$, where $TN_{n_i}(\cdot; \mathbb{A})$ denotes the truncated normal distribution on the interval \mathbb{A} , where $\mathbb{A}_i = A_{i1} \times \dots \times A_{in_i}$, with A_{ij} being the interval $(-\infty, \infty)$ if $C_{ij} = 0$ and the interval $(-\infty, Q_{ij}]$ if $C_{ij} = 1$. For computing the likelihood function associated with model (1) and (2), 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 follows:

$$\mathbf{y}_i = \text{vec}(\mathbf{y}_i^o, \mathbf{y}_i^c), \mathbf{Q}_i = \text{vec}(\mathbf{Q}_i^o, \mathbf{Q}_i^c), \mathbf{X}_i^\top = (\mathbf{X}_i^o, \mathbf{X}_i^c) \text{ and } \boldsymbol{\Sigma}_i = \begin{pmatrix} \boldsymbol{\Sigma}_i^{oo} & \boldsymbol{\Sigma}_i^{oc} \\ \boldsymbol{\Sigma}_i^{co} & \boldsymbol{\Sigma}_i^{cc} \end{pmatrix},$$

where $\text{vec}(\cdot)$ denote the function which stacks vectors or matrices of the same number of columns. Then, we have $\mathbf{y}_i^o \sim N_{n_i^o}(\mathbf{X}_i^o \boldsymbol{\beta}, \boldsymbol{\Sigma}_i^{oo})$, $\mathbf{y}_i^c | \mathbf{y}_i^o \sim N_{n_i^c}(\boldsymbol{\mu}_i, \mathbf{S}_i)$, where $\boldsymbol{\mu}_i = \mathbf{X}_i^c \boldsymbol{\beta} + \boldsymbol{\Sigma}_i^{co} (\boldsymbol{\Sigma}_i^{oo})^{-1} (\mathbf{y}_i^o - \mathbf{X}_i^o \boldsymbol{\beta})$ and $\mathbf{S}_i = \boldsymbol{\Sigma}_i^{cc} - \boldsymbol{\Sigma}_i^{co} (\boldsymbol{\Sigma}_i^{oo})^{-1} \boldsymbol{\Sigma}_i^{oc}$. Now, let $\Phi_{n_i}(\mathbf{u}; \mathbf{a}, \mathbf{A})$ and $\phi_{n_i}(\mathbf{u}; \mathbf{a}, \mathbf{A})$ be the cdf (left tail) and pdf, respectively, of $N_{n_i}(\mathbf{a}, \mathbf{A})$ computed at vector \mathbf{u} . From Vaida and Liu (2009) and Jacqmin-Gadda et al. (2000), the likelihood function for cluster i (using conditional probability arguments) is given by

$$\begin{aligned} L_i(\boldsymbol{\theta}) = f(\mathbf{y}_i | \boldsymbol{\theta}) &= P(\mathbf{Q}_i | \mathbf{C}_i, \boldsymbol{\theta}) = P(\mathbf{y}_i^c \leq \mathbf{Q}_i^c | \mathbf{y}_i^o = \mathbf{Q}_i^o, \boldsymbol{\theta}) P(\mathbf{y}_i^o = \mathbf{Q}_i^o | \boldsymbol{\theta}), \\ &= P(\mathbf{y}_i^c \leq \mathbf{Q}_i^c | \mathbf{y}_i^o, \boldsymbol{\theta}) f(\mathbf{y}_i^o | \boldsymbol{\theta}) \\ &= \phi_{n_i}(\mathbf{y}_i^o; \mathbf{X}_i^o \boldsymbol{\beta}, \boldsymbol{\Sigma}_i^{oo}) \Phi_{n_i^c}(\mathbf{Q}_i^c; \boldsymbol{\mu}_i, \mathbf{S}_i) = \alpha_i, \end{aligned} \quad (3)$$

which can be evaluated without much computational burden through the routine `mvtnorm()` available in R; see Genz et al. (2008); R Development Core Team (2009). The log-likelihood function for the observed data is thus given by $\ell(\boldsymbol{\theta} | \mathbf{y}) = \sum_{i=1}^n \{\log \alpha_i\}$. Thus 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

As the observed log-likelihood function involves complex expressions, it is very difficult to work directly with $\ell(\boldsymbol{\theta} | \mathbf{y})$, either for the ML estimation or to carry out the influence analysis. For LMEC/NLMEC, an EM-type algorithm was developed by Vaida and Liu (2009) for the ML estimation, in which $\boldsymbol{\beta}$ and σ^2 are updated by integrating out \mathbf{b}_i (marginal model), while \mathbf{D} is updated with \mathbf{y}_i and \mathbf{b}_i as missing data. For our purposes, the strategy mentioned earlier makes the implementation of both the case-deletion methods and the local influence analysis, which are based on the complete log-likelihood function (Zhu and Lee, 2001; Lee and Xu, 2004), quite difficult. For this reason, we propose here an expectation conditional maximization (EMC) algorithm by considering \mathbf{y}_i and \mathbf{b}_i as missing data to update (M-step) all the parameters involved in the model.

Let $\mathbf{y} = (\mathbf{y}_1^\top, \dots, \mathbf{y}_n^\top)^\top$, $\mathbf{b} = (\mathbf{b}_1^\top, \dots, \mathbf{b}_n^\top)^\top$, $\mathbf{Q} = \text{vec}(\mathbf{Q}_1, \dots, \mathbf{Q}_n)$ and $\mathbf{C} = \text{vec}(\mathbf{C}_1, \dots, \mathbf{C}_n)$, and that we observe $(\mathbf{Q}_i, \mathbf{C}_i)$ for the i th subject. In their estimation procedure, \mathbf{b} , \mathbf{Q} and \mathbf{C} are treated as hypothetical missing data, and augmented

with the observed data set $\mathbf{y}_c = (\mathbf{C}^\top, \mathbf{Q}^\top, \mathbf{y}^\top, \mathbf{b}^\top)^\top$. Hence, 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

$$\begin{aligned} \ell_i(\boldsymbol{\theta}|\mathbf{y}_c) &= -\frac{1}{2} \left[\log \sigma^2 + \frac{1}{\sigma^2} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i)^\top (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i) \right. \\ &\quad \left. + \log |\mathbf{D}| - \mathbf{b}_i^\top \mathbf{D}^{-1} \mathbf{b}_i \right] + C, \end{aligned} \quad (4)$$

and C is a constant that is independent of the parameter vector $\boldsymbol{\theta}$. Given the current estimate $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$, the E-step calculates the conditional expectation of the complete data log-likelihood function given by

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = E[\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{Q}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}] = \sum_{i=1}^n Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = \sum_{i=1}^n Q_{1i}(\boldsymbol{\beta}, \sigma^2|\widehat{\boldsymbol{\theta}}^{(k)}) + \sum_{i=1}^n Q_{2i}(\boldsymbol{\alpha}|\widehat{\boldsymbol{\theta}}^{(k)}),$$

where

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

and

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

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

It is clear that the E-step reduces only to the computation of $\widehat{\mathbf{y}}_i \widehat{\mathbf{y}}_i^\top = E\{\mathbf{y}_i \mathbf{y}_i^\top | \mathbf{Q}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\}$ and $\widehat{\mathbf{y}}_i = E\{\mathbf{y}_i | \mathbf{Q}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\}$, that is, the first and second moments of a truncated multinormal distribution. These can be determined in closed-form, as a function of multinormal probabilities, using a sequence of simple transformations. For more details on the computation of these moments one may refer to Vaida and Liu (2009).

The conditional maximization (CM) then conditionally maximizes $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ with respect to $\boldsymbol{\theta}$ and obtains a new estimate $\widehat{\boldsymbol{\theta}}^{(k+1)}$, as follows:

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \left(\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{X}_i \right)^{-1} \sum_{i=1}^n \mathbf{X}_i^\top \left(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i \widehat{\mathbf{b}}_i^{(k)} \right), \quad (5)$$

$$\widehat{\sigma}^2^{(k+1)} = \frac{1}{N} \sum_{i=1}^n \left[\widehat{a}_i^{(k)} - 2\widehat{\boldsymbol{\beta}}^{(k+1)\top} \mathbf{X}_i^\top (\widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i \widehat{\mathbf{b}}_i^{(k)}) + \widehat{\boldsymbol{\beta}}^{(k+1)\top} \mathbf{X}_i^\top \mathbf{X}_i \widehat{\boldsymbol{\beta}}^{(k+1)} \right], \quad (6)$$

$$\widehat{\mathbf{D}}^{(k+1)} = \frac{1}{n} \sum_{i=1}^n \widehat{\mathbf{b}}_i \widehat{\mathbf{b}}_i^\top, \quad (7)$$

where $N = \sum_{i=1}^n n_i$. This process is iterated until some distance between two successive evaluations of the actual log-likelihood $\ell(\boldsymbol{\theta}|\mathbf{y})$ in Section 2.1, such as $|\ell(\widehat{\boldsymbol{\theta}}^{(k+1)}) - \ell(\widehat{\boldsymbol{\theta}}^{(k)})|$ or $|\ell(\widehat{\boldsymbol{\theta}}^{(k+1)})/\ell(\widehat{\boldsymbol{\theta}}^{(k)}) - 1|$, is small enough. The variance of the fixed effects in the LMEC is then given by (Hughes, 1999)

$$Var(\widehat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^n \mathbf{X}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i - \mathbf{X}_i^\top \boldsymbol{\Sigma}_i^{-1} Var(\mathbf{y}_i | \mathbf{Q}_i, \mathbf{C}_i) \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \right)^{-1}. \quad (8)$$

3 Diagnostic analysis

Influence diagnostic techniques consist evaluate the sensitivity of the parameter estimates of a particular model when perturbation occurs either in the data set or in the underlying assumptions of the model. There are primarily two approaches for detecting influential observations. The first approach is the case-deletion technique (Cook, 1977), in which is a common approach for analyzing one or more fitted models after excluding some observations and then assessing by some metrics such as the likelihood distance and the Cook's distance. The second method is the local influence approach (Cook, 1986), which evaluates the changes in the results of the analysis by incorporating a minor perturbation to the model. By using the results of Zhu et al. (2001), we will introduce here the case-deletion measures and the local influence measures to the censored data on the basis of the following Q -function $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$. We discuss first consider the case-deletion measures, then the local influence, and finally the perturbation schemes used.

3.1 Case-deletion measures

Case-deletion is a common approach to study the effects of dropping the i th case from the data set. In the following, a quantity with a subscript "[i]" denotes the original quantity with the i th case deleted. The log-likelihood function of $\boldsymbol{\theta}$, based on the data with the i th case deleted, is denoted by $\ell(\boldsymbol{\theta}|\mathbf{Y}_{c[i]})$. Let $\widehat{\boldsymbol{\theta}}_{[i]} = (\widehat{\boldsymbol{\beta}}_{[i]}^\top, \widehat{\sigma}^2_{[i]}, \widehat{\boldsymbol{\alpha}}_{[i]}^\top)^\top$ be the maximizer of the function $Q_{[i]}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = E\{\ell(\boldsymbol{\theta}|\mathbf{Y}_{c[i]})|\mathbf{Q}, \mathbf{C}, \widehat{\boldsymbol{\theta}}\}$, where $\widehat{\boldsymbol{\theta}}$ is the ML estimate of $\boldsymbol{\theta}$. To assess the influence of the i th case on the ML estimate $\widehat{\boldsymbol{\theta}}$, we compare the difference between $\widehat{\boldsymbol{\theta}}_{[i]}$ and $\widehat{\boldsymbol{\theta}}$. If the deletion of a case seriously influences the estimates, more attention need to be paid to that case. Hence, if $\widehat{\boldsymbol{\theta}}_{[i]}$ is far from $\widehat{\boldsymbol{\theta}}$ in some sense, then the i th case is regarded as influential. As $\widehat{\boldsymbol{\theta}}_{[i]}$ is needed for every case, the required computational effort can be quite heavy, especially when the sample is large. Hence, the following one-step pseudo approximation $\widehat{\boldsymbol{\theta}}_{[i]}^1$ is used to reduce the burden (see Cook and Weisberg, 1982; Zhu et al., 2001):

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

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

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

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

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

where $E_{1[i]} = \sum_{j \neq i} \mathbf{X}_j^\top (\hat{\mathbf{y}}_j - \mathbf{Z}_j \hat{\mathbf{b}}_j - \mathbf{X}_j \hat{\boldsymbol{\beta}})$ and $E_{2[i]} = \sum_{j \neq i} (n_j - \frac{A_j}{\hat{\sigma}^2})$, with $A_j = a_j - 2\hat{\boldsymbol{\beta}}^\top \mathbf{X}_j^\top (\hat{\mathbf{y}}_j - \mathbf{Z}_j \hat{\mathbf{b}}_j) + \hat{\boldsymbol{\beta}}^\top \mathbf{X}_j^\top \mathbf{X}_j \hat{\boldsymbol{\beta}}$. Finally, $\dot{Q}_{[i]\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})$ has its elements as

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

The Hessian matrix $\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})$

Following Zhu and Lee (2001), to obtain the diagnostic measures for case-deletion diagnostic and for local influence of a particular perturbation scheme, it is necessary to compute $\ddot{Q}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^n \partial^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top$, where $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\alpha}^\top)^\top$ is the parameter vector. Hence, the Hessian matrix $\partial^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top$ has its elements as follows:

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

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

$$\ddot{Q}_\beta(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) = -\frac{1}{\hat{\sigma}^2} \mathbf{X}^\top \mathbf{X}, \quad \ddot{Q}_{\sigma^2}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) = -\frac{b}{2\hat{\sigma}^4} \quad \text{and} \quad \ddot{Q}_\alpha(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^n \left(\frac{\partial^2 Q_i(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})}{\partial \alpha_s \partial \alpha_r} \right),$$

where $\mathbf{X} = (\mathbf{X}_1^\top, \dots, \mathbf{X}_n^\top)^\top$ and $b = -\sum_{i=1}^n (n_i - 2A_i/\hat{\sigma}^2)$.

Next, we will obtain the one-step approximation of $\hat{\boldsymbol{\theta}}_{[i]} = (\hat{\boldsymbol{\beta}}_{[i]}^\top, \hat{\sigma}_{[i]}^2, \hat{\boldsymbol{\alpha}}_{[i]}^\top)^\top$, $i = 1, \dots, n$, based on (9), viz., the relationships between the parameter estimates for full data and the data with the i th case deleted.

Theorem 3.1. For the LMEC, the relationships between the parameter estimates for full data and the data with the i th case deleted are as follows:

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{[i]}^1 &= \widehat{\boldsymbol{\beta}} + (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{E}_{1[i]}, \\ \widehat{\sigma}_{[i]}^2 &= \widehat{\sigma}^2 - \frac{1}{b} E_{2[i]}, \\ \widehat{\boldsymbol{\alpha}}_{[i]}^1 &= \widehat{\boldsymbol{\alpha}} + \{-\ddot{Q}_\alpha(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}^{-1} \dot{Q}_{[i]\alpha}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}),\end{aligned}$$

where $\mathbf{E}_{1[i]}$, $E_{2[i]}$ and $\dot{Q}_{[i]\alpha}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$, $i = 1, \dots, n$, are as in (10), (11) and (12), respectively.

From Theorem 3.1, case-deletion measures can be developed for assessing influential observations, such as the generalized Cook distance and the likelihood distance (Zhu and Lee, 2001). To assess the influence of the i th case on the ML estimate $\widehat{\boldsymbol{\theta}}$, we need to compare $\widehat{\boldsymbol{\theta}}_{[i]}$ and $\widehat{\boldsymbol{\theta}}$, and if $\widehat{\boldsymbol{\theta}}_{[i]}$ is far from $\widehat{\boldsymbol{\theta}}$ in some sense, then the i th case is regarded as influential. Based on the metric for measuring the distance between $\widehat{\boldsymbol{\theta}}_{[i]}$ and $\widehat{\boldsymbol{\theta}}$, proposed by Zhu and Lee (2001) based on the EM algorithm, we consider here the following *generalized Cook distance*:

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

Upon substituting (9) into (13), we obtain the approximation

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

Since $\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$ is a diagonal matrix, from Xie et al. (2007), GD_i^1 can be decomposed into three parts that corresponds to the generalized Cook distance for the parameter subsets $\boldsymbol{\beta}$, σ^2 and $\boldsymbol{\alpha}$, which are denoted, respectively, by $GD_i^1(\boldsymbol{\beta})$, $GD_i^1(\sigma^2)$ and $GD_i^1(\boldsymbol{\alpha})$, as follows:

$$GD_i^1 = GD_i^1(\boldsymbol{\beta}) + GD_i^1(\sigma^2) + GD_i^1(\boldsymbol{\alpha}), \quad (14)$$

where

$$\begin{aligned}GD_i^1(\boldsymbol{\beta}) &= \dot{Q}_{[i]\beta}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})^\top \{-\ddot{Q}_\beta(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}^{-1} \dot{Q}_{[i]\beta}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \frac{1}{\widehat{\sigma}^2} \mathbf{E}_{1[i]}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{E}_{1[i]}, \\ GD_i^1(\sigma^2) &= \dot{Q}_{[i]\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})^\top \{-\ddot{Q}_{\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}^{-1} \dot{Q}_{[i]\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \frac{1}{2b} \mathbf{E}_{2[i]}, \\ GD_i^1(\boldsymbol{\alpha}) &= \dot{Q}_{[i]\alpha}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})^\top \{-\ddot{Q}_\alpha(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}^{-1} \dot{Q}_{[i]\alpha}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}).\end{aligned}$$

Another measure for the influence of the i th case is the following Q -distance function, similar to the likelihood distance LD_i (Cook and Weisberg, 1982), defined as

$$QD_i = 2\{Q(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) - Q(\widehat{\boldsymbol{\theta}}_{[i]}|\widehat{\boldsymbol{\theta}})\}. \quad (15)$$

We can calculate an approximation of the likelihood displacement QD_i by substituting (9) into (15), resulting in the following approximation QD_i^1 of QD_i :

$$QD_i^1 = 2\{Q(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) - Q(\widehat{\boldsymbol{\theta}}_{[i]}^1|\widehat{\boldsymbol{\theta}})\}. \quad (16)$$

3.2 Local Influence

In this subsection, we derive the normal curvature of the local influence (Cook, 1986) for some common perturbation schemes either in the model or in the data. We will consider the case-weight, scale matrix perturbation schemes, and response perturbation schemes, for this purpose.

Consider a perturbation vector $\boldsymbol{\omega} = (\omega_1, \dots, \omega_g)^\top$ varying in an open region $\boldsymbol{\Omega} \subset \mathbb{R}^g$. Let $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{y}_c)$ be the complete-data log-likelihood to the perturbed model. We assume that there is a $\boldsymbol{\omega}_0$ in $\boldsymbol{\Omega}$ such that $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}_0 | \mathbf{y}_c) = \ell_c(\boldsymbol{\theta} | \mathbf{y}_c)$ for all $\boldsymbol{\theta}$. Let $\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})$ denote the maximum of the function $Q(\boldsymbol{\theta}, \boldsymbol{\omega} | \widehat{\boldsymbol{\theta}}) = \text{E}[\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{y}_c) | \mathbf{Q}, \mathbf{C}, \widehat{\boldsymbol{\theta}}]$. The influence graph is then defined as $\boldsymbol{\alpha}(\boldsymbol{\omega}) = (\boldsymbol{\omega}^\top, f_Q(\boldsymbol{\omega}))^\top$, where $f_Q(\boldsymbol{\omega})$ is the Q -displacement function defined as follows:

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

Following the approach of Cook (1986) and Zhu and Lee (2001), the normal curvature $C_{f_Q, \mathbf{d}}$ of $\boldsymbol{\alpha}(\boldsymbol{\omega})$ at $\boldsymbol{\omega}_0$ in the direction of some unit vector \mathbf{d} can be used to summarize the local behavior of the Q -displacement function. It can be shown that

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

where $\ddot{Q}(\widehat{\boldsymbol{\theta}} | \widehat{\boldsymbol{\theta}}) = \frac{\partial^2 Q(\boldsymbol{\theta} | \widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \Big|_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}}$ and $\boldsymbol{\Delta}_{\boldsymbol{\omega}} = \frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega} | \widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^\top} \Big|_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})}$.

Following the same procedure as in Cook (1986), the quantity $-\ddot{Q}_{\boldsymbol{\omega}_0}$ is quite useful for detecting influential observations. From the spectral decomposition of a symmetric matrix $-2\ddot{Q}_{\boldsymbol{\omega}_0} = \sum_{k=1}^g \zeta_k \boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k^\top$, where $\{(\zeta_k, \boldsymbol{\varepsilon}_k), k = 1, \dots, g\}$ are eigenvalue–eigenvector pairs of $-2\ddot{Q}_{\boldsymbol{\omega}_0}$ with $\zeta_1 \geq \dots \geq \zeta_r > \zeta_{r+1} = \dots = 0$ and orthonormal eigenvectors $\{\boldsymbol{\varepsilon}_k, k = 1, \dots, g\}$, Zhu and Lee (2001) proposed to inspect all eigenvectors corresponding to nonzero eigenvalues for capturing more information. Based on the work of Zhu and Lee (2001), we consider the following aggregated contribution vector of all eigenvectors that correspond to nonzero eigenvalues. Let $\tilde{\zeta}_k = \zeta_k / (\zeta_1 + \dots + \zeta_r)$, $\boldsymbol{\varepsilon}_k^2 = (\boldsymbol{\varepsilon}_{k1}^2, \dots, \boldsymbol{\varepsilon}_{kg}^2)^\top$ and $M(0) = \sum_{k=1}^r \tilde{\zeta}_k \boldsymbol{\varepsilon}_k^2$. The l th component of $M(0)$, $M(0)_l$, is equal to $\sum_{k=1}^r \tilde{\zeta}_k \boldsymbol{\varepsilon}_{kl}^2$. The assessment of influential cases is based on the visual inspection of the $\{M(0)_l, l = 1, \dots, g\}$ plotted against the index l . The l th case may be regarded as influential if $M(0)_l$ is larger than the benchmark.

The inconvenience on the use of the normal curvature is in deciding about the influence of the observations, since $C_{f_Q, \mathbf{d}}(\boldsymbol{\theta})$ may assume any value and it is not invariant under a uniform change of scale. Based on the work of Poon and Poon (1999) in using a conformal normal curvature, Zhu and Lee (2001) considered the

following conformal normal curvature $B_{f_Q, \mathbf{d}}(\boldsymbol{\theta}) = C_{f_Q, \mathbf{d}}(\boldsymbol{\theta}) / \text{tr}[-2\ddot{Q}_{\boldsymbol{\omega}_0}]$, 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 l th entry as 1 and all other entries as 0. Zhu and Lee (2001) showed that for all l , $M(0)_l = B_{f_Q, \mathbf{d}_l}$. We can therefore obtain $M(0)_l$ via B_{f_Q, \mathbf{d}_l} .

So far, there is no a general rule to judge how large of the influence of a specific case in the data is. Let $\overline{M}(0)$ and $SM(0)$ denote, respectively, the mean and the standard error of $\{M(0)_l : l = 1, \dots, g\}$, where $\overline{M}(0) = 1/g$. Poon and Poon (1999) proposed to use $2\overline{M}(0)$ as a benchmark for $M(0)$. However, we may use different functions of $M(0)$. For instance, Zhu and Lee (2001) proposed to use $\overline{M}(0) + 2SM(0)$ as a benchmark to take into account the variance of $\{M(0)_l : l = 1, \dots, g\}$. According to Lee and Xu (2004), the exact choice of the function of $\overline{M}(0)$ as the benchmark is subjective. More recently, Lee and Xu (2004) proposed to use $\overline{M}(0) + c^*SM(0)$, where c^* is a selected constant, and depending on the application, c^* may be taken to be any value. In this paper, and here we will use $c^* = 3, 5$.

3.3 Perturbation schemes

In this section, we will evaluate the Δ matrix under the following perturbation schemes for LMEC models. *Case-weight* made for detecting observations with outstanding contribution on the log-likelihood function and that may exercise high influence on the maximum likelihood estimates. *Scale perturbation* made on the scale matrix $\Sigma_i = \sigma^2 \mathbf{I}_{n_i} + \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^\top$. It also can be made on either σ^2 or \mathbf{D} which may reveal individuals that are most influential, in the sense, of the likelihood displacement on the scale structure. Finally, *perturbation of response variables* made on the response values, which may indicate observations with large influence on the MLE. In our case, the response variables are \mathbf{Q} 's.

For each perturbation scheme, one has the partitioned form

$$\Delta_{\boldsymbol{\omega}_o} = (\Delta_{\beta}^\top, \Delta_{\sigma^2}^\top, \Delta_{\alpha}^\top)^\top,$$

where $\Delta_{\beta} = \frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega} | \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top} |_{\boldsymbol{\omega}_o} \in \mathbb{R}^{p \times g}$, $\Delta_{\sigma^2} = \frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega} | \hat{\boldsymbol{\theta}})}{\partial \sigma^2 \partial \boldsymbol{\omega}^\top} |_{\boldsymbol{\omega}_o} \in \mathbb{R}^{1 \times g}$ and $\Delta_{\alpha} = (\Delta_{\alpha_1}^\top, \dots, \Delta_{\alpha_{p^*}}^\top)^\top$, with $\Delta_{\alpha_r} = \frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega} | \hat{\boldsymbol{\theta}})}{\partial \alpha_r \partial \boldsymbol{\omega}^\top} |_{\boldsymbol{\omega}_o} \in \mathbb{R}^{1 \times g}$, $r = 1, \dots, p^*$ and g being the dimension of the perturbation vector $\boldsymbol{\omega}$.

Case weight perturbation

First, 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, represented by writing

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

Here, $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^\top$ is an $n \times 1$ vector and $\boldsymbol{\omega}_o = (1, \dots, 1)^\top$. In addition, it is possible to show that the local influence for this perturbation scheme is equivalent to the deletion method discussed in preceding section. For this perturbation scheme, we find

$$\begin{aligned} \Delta_\beta &= \frac{1}{\sigma^2} \mathbf{X}^\top D(\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n), \\ \Delta_{\sigma^2} &= -\frac{1}{2\sigma^2} \mathbf{n}^\top + \frac{1}{2\sigma^4} \mathbf{m}^\top, \\ \Delta_{\alpha_r} &= \left[\frac{\partial Q_1(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \alpha_r}, \dots, \frac{\partial Q_n(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \alpha_r} \right], \quad r = 1, \dots, p^*, \end{aligned}$$

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 $\boldsymbol{\epsilon}_i = \widehat{\mathbf{y}}_i - \mathbf{Z}_i \widehat{\mathbf{b}}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}}_i$ and $\frac{\partial Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \alpha_r} = -\frac{1}{2} \text{tr}[\mathbf{D}^{-1} \dot{\mathbf{D}}(r) - \mathbf{D}^{-1} \dot{\mathbf{D}}(r) \mathbf{D}^{-1} \widehat{\mathbf{b}}_i \widehat{\mathbf{b}}_i^\top]$.

Scale matrix perturbation

To study the effects of departures from the assumption regarding the scale matrix $\boldsymbol{\Sigma}_i$ of the random effects, we consider the perturbations $\mathbf{D}(\omega_i) = \omega_i^{-1} \mathbf{D}$ or $\sigma^2(\omega_i) = \omega_i^{-1} \sigma^2$, for $i = 1, \dots, n$. Under this perturbation scheme, the non-perturbed model is obtained when $\boldsymbol{\omega}_o = (1, \dots, 1)^\top$. Moreover, the perturbed Q -function is as in (5), $\mathbf{D}(\omega_i)$ and $\sigma^2(\omega_i)$ in place of \mathbf{D} and σ^2 , respectively. The matrix $\Delta_{\boldsymbol{\omega}_o}$ has its elements as follows:

- Perturbation on \mathbf{D} : $\Delta_\beta = \mathbf{0}$, $\Delta_{\sigma^2} = \mathbf{0}$ and $\Delta_{\alpha_r} = \frac{1}{2}[g_1, \dots, g_n]$, where $g_i = \mathbf{D}^{-1} \dot{\mathbf{D}}(r) \mathbf{D}^{-1} \widehat{\mathbf{b}}_i \widehat{\mathbf{b}}_i^\top$, $r = 1, \dots, p^*$;
- Perturbation on σ^2 : $\Delta_\beta = \frac{1}{\sigma^2} \mathbf{X}^\top D(\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n)$, $\Delta_{\sigma^2} = \frac{1}{2\sigma^4} \mathbf{m}^\top$ and $\Delta_{\boldsymbol{\alpha}} = \mathbf{0}$.

Response perturbation

A perturbation of the response variables Q_{ij} , $i = 1, \dots, n$, $j = 1, \dots, n_i$, can be introduced by replacing Q_{ij} by $Q_{ij}(\omega) = Q_{ij} + \omega_i s_{ij}$, where s_{ij} is a scale factor. Now substituting $Q_{ij}(\omega)$ into (2), we can write perturbed model as

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

where $\mathbf{y}_{ij}(\omega) = \mathbf{y}_{ij} - \omega_i s_{ij}$. Hence, the perturbed Q-function $Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega})$ is as in Section 2.2, with $\widehat{\mathbf{y}}_i$, $\widehat{\mathbf{y}_i \mathbf{y}_i^\top}$ and $\widehat{\mathbf{y}_i \mathbf{b}_i^\top}$ replaced by with $\widehat{\mathbf{y}}_{i\omega} = \widehat{\mathbf{y}}_i - \omega_i \mathbf{s}_i$, $\widehat{\mathbf{y}_{i\omega} \mathbf{y}_{i\omega}^\top} = \widehat{\mathbf{y}_i \mathbf{y}_i^\top} - \omega_i (\widehat{\mathbf{y}}_i \mathbf{s}_i^\top + \mathbf{s}_i \widehat{\mathbf{y}}_i^\top) + \omega_i^2 \mathbf{s}_i \mathbf{s}_i^\top$ and $\widehat{\mathbf{y}_{i\omega} \mathbf{b}_{i\omega}^\top} = \widehat{\mathbf{y}_i \mathbf{b}_i^\top} - \omega_i \mathbf{s}_i \widehat{\mathbf{b}_i^\top}$, respectively, with $\mathbf{s}_i = (s_{i1}, \dots, s_{in_i})^\top$. Under this perturbation scheme, the vector $\boldsymbol{\omega}_0$, representing no perturbation, is given by $\boldsymbol{\omega}_0 = \mathbf{0}$ and $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0}$ has the following elements:

$$\begin{aligned}\boldsymbol{\Delta}_\beta &= -\frac{1}{\sigma^2} \mathbf{X}^\top D(\mathbf{s}_1, \dots, \mathbf{s}_n), \\ \boldsymbol{\Delta}_{\sigma^2} &= -\frac{1}{\sigma^4} (\mathbf{Y} - \mathbf{Z}\mathbf{b} - \mathbf{X}\boldsymbol{\beta})^\top D(\mathbf{s}_1, \dots, \mathbf{s}_n), \\ \boldsymbol{\Delta}_\alpha &= \mathbf{0},\end{aligned}$$

where $\mathbf{Y} = (\widehat{\mathbf{y}}_1^\top, \dots, \widehat{\mathbf{y}}_n^\top)^\top$, $\mathbf{b} = (\widehat{\mathbf{b}}_1, \dots, \widehat{\mathbf{b}}_n)^\top$ and $D(\mathbf{s}_1, \dots, \mathbf{s}_n)$ is a block-diagonal matrix.

4 The nonlinear case

The NLME (Pinheiro and Bates, 2000) is defined as

$$\mathbf{y}_i = \eta(\boldsymbol{\phi}_i, \mathbf{X}_i) + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\phi}_i = \mathbf{A}_i \boldsymbol{\beta} + \mathbf{B}_i \mathbf{b}_i, \quad i = 1, \dots, n, \quad (17)$$

where $\mathbf{b}_i \stackrel{iid}{\sim} N_q(0, \mathbf{D})$ and $\boldsymbol{\epsilon}_i \stackrel{iid}{\sim} N_{n_i}(0, \sigma^2 \mathbf{I}_{n_i})$ are independent; \mathbf{y}_i is a $(n_i \times 1)$ vector of observed continuous responses for subject i ; η is a nonlinear function of the individual random parameter $\boldsymbol{\phi}_i$; \mathbf{A}_i and \mathbf{B}_i are known design matrices of dimensions $r \times p$ and $r \times q$, respectively, possibly depending on some covariate values; $\boldsymbol{\beta}$ is the $(p \times 1)$ vector of fixed effects, and \mathbf{b}_i is the $(q \times 1)$ vector of random effects.

As mentioned by Vaida and Liu (2009), the linearization (L) procedure to obtain the approximate MLE of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\alpha}^\top)^\top$ involves of taking the first-order Taylor expansion of η_i around the current parameter estimate $\widetilde{\boldsymbol{\beta}}$ and the random effect estimates $\widetilde{\mathbf{b}}_i$ (empirical predictors), which is equivalent to iteratively solving the following LME model (L-step):

$$\widetilde{\mathbf{Y}}_i = \widetilde{\mathbf{W}}_i \boldsymbol{\beta} + \widetilde{\mathbf{H}}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, n, \quad (18)$$

where $\widetilde{\mathbf{Y}}_i = \mathbf{Y}_i - \eta(\boldsymbol{\phi}(\widetilde{\boldsymbol{\beta}}, \widetilde{\mathbf{b}}_i), \mathbf{X}_i)$, $\mathbf{b}_i \stackrel{iid}{\sim} N_q(0, \mathbf{D})$ and $\boldsymbol{\epsilon}_i \stackrel{iid}{\sim} N_{n_i}(\mathbf{0}, \sigma_e^2 \mathbf{I}_{n_i})$, $\widetilde{\mathbf{H}}_i = \frac{\partial \eta(\mathbf{A}_i \boldsymbol{\beta} + \mathbf{B}_i \mathbf{b}_i, \mathbf{X}_i)}{\partial \mathbf{b}_i^\top} \Big|_{\mathbf{b}_i = \widetilde{\mathbf{b}}_i}$ and $\widetilde{\mathbf{W}}_i = \frac{\partial \eta(\mathbf{A}_i \boldsymbol{\beta} + \mathbf{B}_i \mathbf{b}_i, \mathbf{X}_i)}{\partial \boldsymbol{\beta}^\top} \Big|_{\boldsymbol{\beta} = \widetilde{\boldsymbol{\beta}}}$. Thus, for censored response, the linearized model (18) is an LME with censored data, with the same structure as in (1), which is then solved as detailed in the preceding section. The model matrices in (18) depend on the current parameter value, and need to be recalculated at each iteration. The algorithm iterates to convergence between the L-, E-, and CM-steps. Moreover, the influence diagnostic procedures discussed earlier

in Section 3 can be incorporated along with the approximation in (18) to obtain approximate influence diagnostics measures for NLMEC. The performance of the developed influence diagnostics measures are illustrated in the next section.

Table 1: Parameter estimates of the LMEC model and p-values for the UTI data. SE denotes the standard error.

Parameter	Estimate	SE	p-value
$\widehat{\beta}_1$	3.6038	0.1253	< 0.01
$\widehat{\beta}_2$	4.1664	0.1285	< 0.01
$\widehat{\beta}_3$	4.2413	0.1304	< 0.01
$\widehat{\beta}_4$	4.3604	0.1307	< 0.01
$\widehat{\beta}_5$	4.5662	0.1398	< 0.01
$\widehat{\beta}_6$	4.5692	0.1485	< 0.01
$\widehat{\beta}_7$	4.6773	0.1646	< 0.01
$\widehat{\beta}_8$	4.7935	0.2018	< 0.01
$\widehat{\sigma}^2$	0.3414		
$\widehat{\alpha}$	0.76535		

5 Numerical illustrations

We illustrate the performance of the proposed methods with the analysis of two HIV datasets, previously analyzed by Vaida and Liu (2009), and the analysis of a simulated example.

5.1 UTI data

The first application is a study of 72 perinatally HIV-infected children (Saitoh et al., 2008; Vaida and Liu, 2009). The data set is available in the R package *lme*. Primarily due to treatment fatigue, unstructured treatment interruptions (UTI) is common in this population. Suboptimal adherence can lead to ARV resistance and diminished treatment options in the future. The subjects in the study had taken ARV therapy for at least 6 months before UTI, and the medication was discontinued for more than 3 months. The HIV viral load from the closest time points at 0, 1, 3, 6, 9, 12, 18, 24 months after UTI were studied. The number of observations from baseline (month 0) to month 24 are 71, 62, 58, 57, 43, 34, 24, and 13, respectively. Out of 362 observations, 26 (7%) observations were below the detection limits (50 or 400 copies/mL) and were left-censored at these values. Following Vaida and Liu (2009), we consider a profile LME model with random intercepts

b_i as $y_{ij} = b_i + \beta_j + \epsilon_{ij}$, where y_{ij} is the \log_{10} HIV RNA for subject i at time t_j , $t_1 = 0, t_2 = 1, t_3 = 3, t_4 = 6, t_5 = 9, t_6 = 12, t_7 = 18, t_8 = 24$. The \log_{10} transformation of HIV viral load is used to stabilize the variance of the viral load and also to make the viral load more close to be normally distributed. A summary of these parameter estimates and their respective p-values are presented in Table 1. These results conform to those presented in Vaida and Liu (2009). From Table 1, we note that all the regression parameters are significant at 5% level.

5.1.1 Global influence

In order to identify outlying observations under the fitted model, the index plot of the Mahalanobis distance $d_i = (\hat{\mathbf{y}}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})^\top \boldsymbol{\Sigma}_i^{-1} (\hat{\mathbf{y}}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})$, $i = 1, \dots, 72$, is displayed in Figure 1(a). We can see from this figure that observation #42 appears as a possible outlier. To evaluate the effect on the ML estimates when some observation is eliminated, we analyze the QD_i^1 and GD_i^1 index plots, which are shown in Figures 1(b) and 2(a), respectively. We note from these figures that two cases (#20, #42) are potentially influential on the parameter estimates. Figures 2(b)-(d) present the index plots of $GD_i^1(\gamma)$, for $\gamma = \boldsymbol{\beta}, \sigma^2, \alpha$, respectively (see Section 3.1). From these figures, we see that observation #42 is influential with regard to the parameters $\boldsymbol{\beta}$ and σ^2 , while observation #20 is influential with regard to the parameter α .

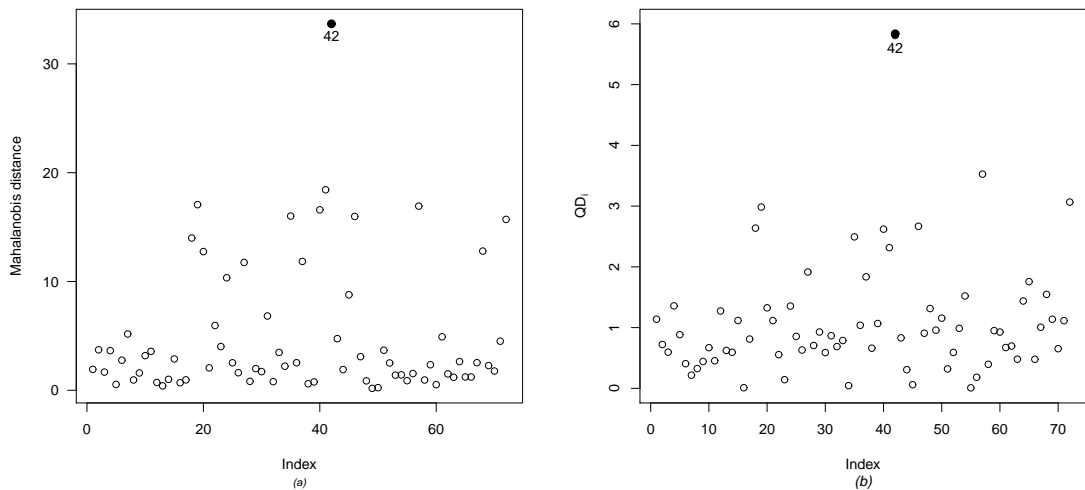


Figure 1: UTI data. (a) Mahalanobis distance and (b) approximate likelihood displacement QD_i^1 . The influential observations are numbered.

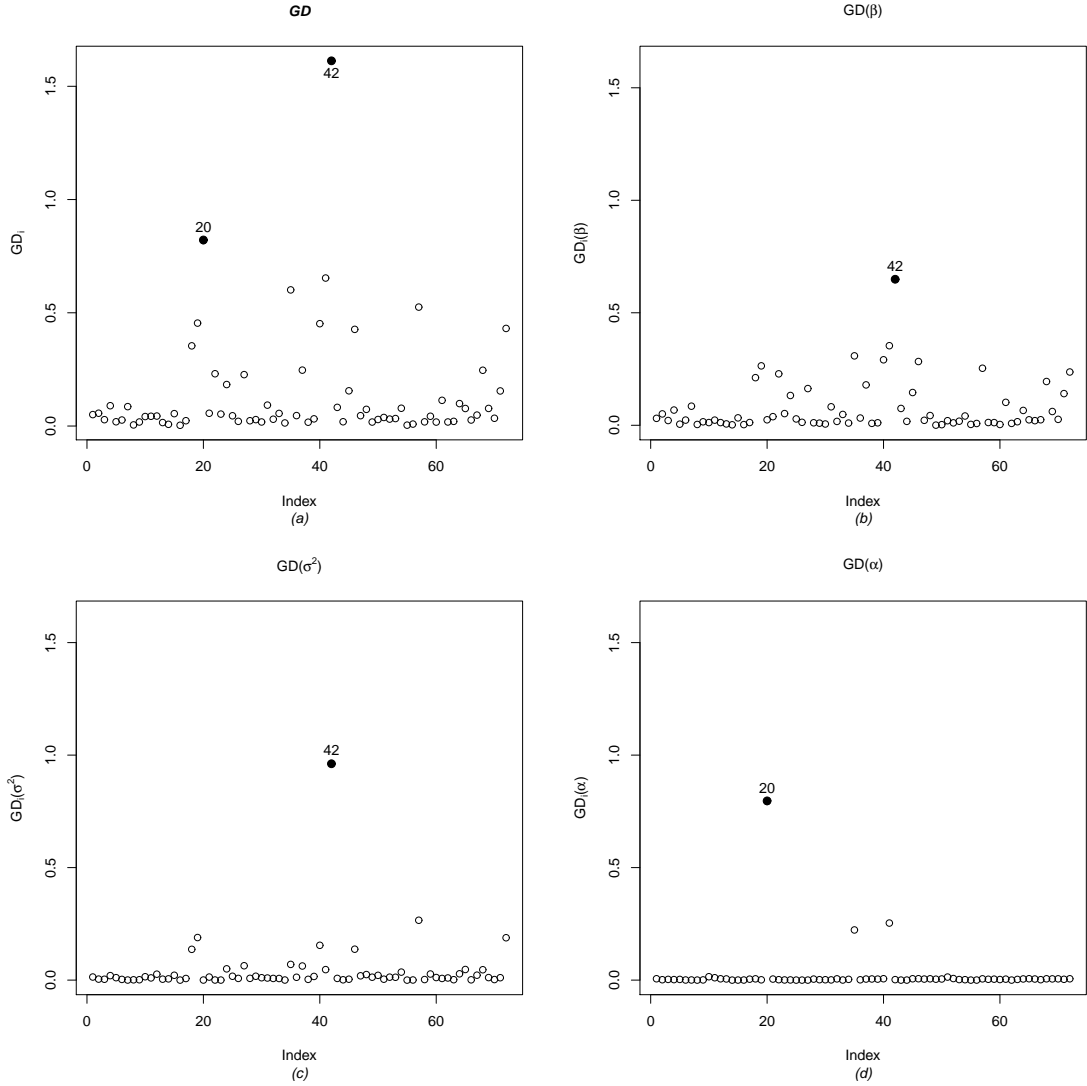


Figure 2: UTI data. (a) Approximate generalized Cook's distance GD_i^1 , (b) GD_i^1 for subset β , (c) GD_i^1 for subset σ^2 and (d) GD_i^1 for subset α . The influential observations are numbered.

5.1.2 Local influence

Next, we conduct a local influence study on the UTI data, based on $M(0)$ with interest focussing on θ . Here we use the criterion $M(0)_i > \overline{M}(0) + 3SM(0)$, $i = 1, \dots, 72$, to discriminate whether an observation is influential or not. Figure 3 presents the index plots of $M(0)$ under the four perturbation schemes described earlier in Section 3.3. From this figure, it is noted that observation #20 appears as influential under the case weight and scale matrix σ^2 perturbation, while observation #42 is more influential under the perturbation on the scale matrix \mathbf{D} . However, no observation appears to be influential under the response variable perturbation.

In order to assess the impact of the two observations that have been highlighted

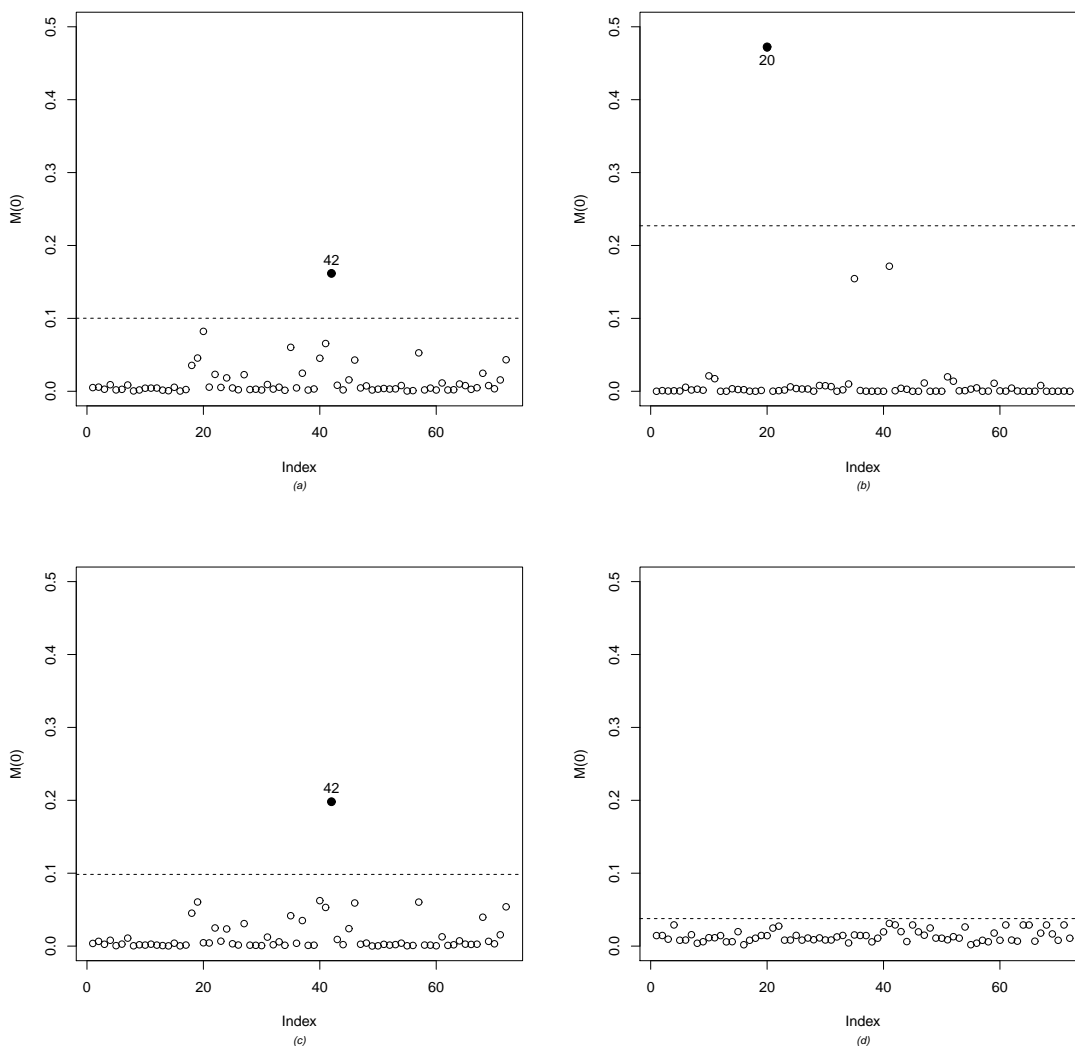


Figure 3: Index plot of $M(0)$ for assessing local influence on θ under (a) case weight perturbation, (b) perturbation on \mathbf{D} , (c) perturbation on σ^2 and (d) perturbation on the response variable for the UTI data. The influential observations are numbered.

as potentially influential on the ML estimates, we refitted the proposed LMEC model by dropping each one of these cases. Let $I_1 = \{20\}$, $I_2 = \{42\}$ and $I_3 = \{20, 42\}$ denotes the sets of observations identified as influential. Table 2 presents the relative changes (RC) in percentage of these estimates defined by

$$RC_{\hat{\gamma}} = \left| \frac{\hat{\gamma} - \hat{\gamma}_{[i]}}{\hat{\gamma}} \right|,$$

where $\gamma = \beta_1, \dots, \beta_8, \sigma^2, \alpha$, and $\hat{\gamma}_{[i]}$ denotes the ML estimate of $\hat{\gamma}$ after the set I_i , ($i = 1, 2, 3$) has been removed. Even though some RC values are large, significant changes in β are not observed. It is of interest to note from Table 2 the coherence with the diagnostic plots in Figure 2 (as we would expect). For instance, the elimination of observation #20 leads to a large change in the RC of α while the

elimination of observation #42 leads to a large change in the RC of σ^2 . Moreover, the elimination of observations #20, #42 leads to a large change in the RC of α and σ^2 .

Table 2: RC (in %) for the UTI data.

Dropped	$RC_{\hat{\beta}_1}$	$RC_{\hat{\beta}_2}$	$RC_{\hat{\beta}_3}$	$RC_{\hat{\beta}_4}$	$RC_{\hat{\beta}_5}$	$RC_{\hat{\beta}_6}$	$RC_{\hat{\beta}_7}$	$RC_{\hat{\beta}_8}$	$RC_{\hat{\sigma}^2}$	$RC_{\hat{\alpha}}$
{#20}	1.28	1.13	1.14	1.14	1.07	0.74	0.71	0.75	0.41	19.07
{#42}	0.49	0.44	0.72	1.10	0.29	0.26	0.59	1.04	10.40	0.93
{#20, #42}	0.93	1.69	1.97	2.36	0.89	0.62	0.24	0.16	10.05	18.48

5.2 AIEDRP study

The second AIDS case study is from the AIEDRP program, a large multicenter observational study of subjects with acute and early HIV infection. We consider 320 untreated individuals with acute HIV infection; for more details on this dataset, one may refer to Vaida and Liu (2009). Of the 830 recorded observations, 185 (22%) were above the limit of assay quantification, and so they were considered to be right-censored. We therefore consider a right-censored version and accommodate it within our NLME. Following Vaida and Liu (2009), we choose a five-parameter NLME model (inverted S-shaped curve) as follows:

$$y_{ij} = \alpha_{1i} + \frac{\alpha_2}{(1 + \exp((t_{ij} - \alpha_3)/\alpha_4))} + \alpha_{5i}(t_{ij} - 50) + \epsilon_{ij}, \quad (19)$$

where y_{ij} is the \log_{10} HIV RNA for subject i at time t_{ij} . The parameters α_{1i} and α_2 represent subject-specific (random) setpoints and decrease from the maximum HIV RNA. In the absence of treatment (following acute infection), the HIV RNA varies around a set-point which may differ among individuals, and so the setpoint is chosen to be subject-specific. The location parameter α_3 indicates the time point at which half of the change in HIV RNA is attained, α_4 is a scale parameter modeling the rate of decline, and α_{5i} allows for increasing HIV RNA trajectory after day 50. To force the parameters to be positive, we re-parameterize then as follows: $\beta_{1i} = \log(\alpha_{1i}) = \beta_1 + b_{1i}$; $\beta_k = \log(\alpha_k)$, $k = 2, 3, 4$, and $\alpha_{5i} = \beta_5 + b_{2i}$. Table 3 presents the ML estimates for the parameters, together with their corresponding standard errors, calculated from (8). From Table 3, we note that all the regression parameters are significant at 5% level, except the parameter β_2 .

Table 3: Parameter estimates of the NLMEC model and p-values for the AIEDRP data. SE denotes the standard error.

Parameter	Estimate	SE	p-value
$\widehat{\beta}_1$	1.6096	0.0137	<0.01
$\widehat{\beta}_2$	0.1422	0.0949	0.1340
$\widehat{\beta}_3$	3.5262	0.0237	<0.01
$\widehat{\beta}_4$	1.0559	0.2677	0.01
$\widehat{\beta}_5$	-0.0035	0.0014	0.01
$\widehat{\sigma}^2$	0.2652		
$\widehat{\alpha}_{11}$	0.0177		
$\widehat{\alpha}_{12}$	0.0002		
$\widehat{\alpha}_{22}$	0.00004		

5.2.1 Global influence

In order to identify outlying observations under the fitted model, the index plot of the Mahalanobis distance is displayed in Figure 4(a). We can see from this figure that observations #9, #166, #230 and #259 all appear as possible outliers. As in the previous application, to evaluate the effect on the ML estimates when some observation is eliminated, we analyze the case deletion measures QD_i^1 and GD_i^1 , which are shown in Figures 4(b) and 5(a), respectively. We note from these figures that cases #9, #166, #195, #230 and #259 are all potentially influential with regard to the full parameter estimate $\boldsymbol{\theta}$. On the other hand, from figures 5(b)-(d), where we present the index plots of $GD_i^1(\gamma)$, for $\gamma = \boldsymbol{\beta}, \sigma^2, \alpha$, respectively, we can see that observations #166, #195 and #230 are influential with regard to the regression parameters $\boldsymbol{\beta}$, while only observation #259 is influential with regard to the parameter α .

5.2.2 Local influence

Next, we conduct a local influence study for the AIEDRP data, based on $M(0)$ with interest focussing on $\boldsymbol{\theta}$. Figure 6 presents the index plots of $M(0)$ under the four perturbation schemes discussed earlier in Section 3.3. From this figure, it is noted that observations #9, #166, #195, #230 and #259 all appear as influential under the case weight and scale- σ^2 perturbations, while only observations #198 and #259 are more influential under the perturbation on the scale matrix \mathbf{D} . It is noted also that different observations #174, #175, #176 and #259 appear out as influential under the response variable perturbation. It is important to emphasize here that, as

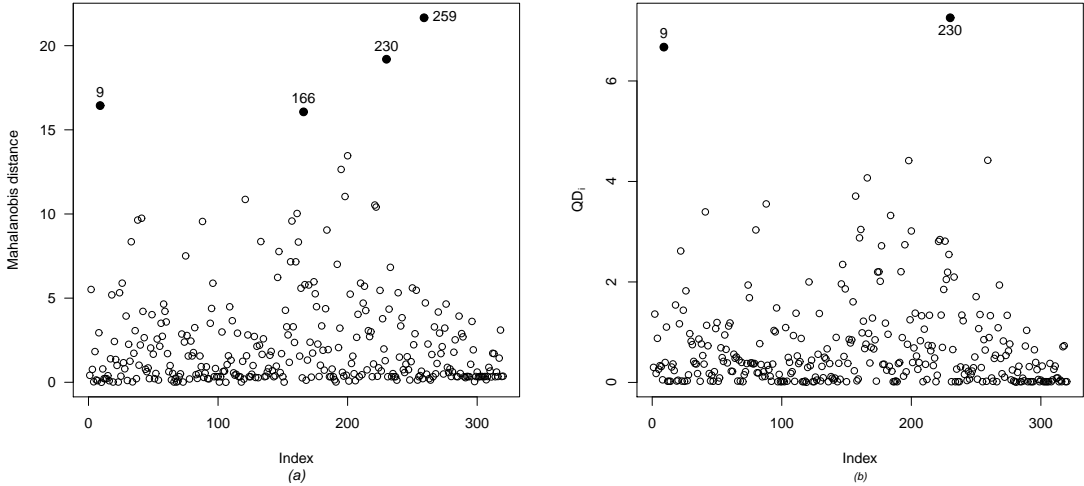


Figure 4: AIEDRP data. (a) Mahalanobis distance and (b) approximate likelihood displacement QD_i^1 . The influential observations are numbered.

in the uncensored case, the influence measure GD_i^1 considered here is closely related to the local influence measure based on the case weight perturbation.

5.3 Simulation Study

Results from analysis of a simulated example are presented here to illustrate the performance of the proposed diagnostic measures. We consider a logistic model similar to the one in (19), with random setpoints α_{1i} and random decline rates: α_{4i} , as follows

$$y_{ij} = \alpha_{1i} + \frac{\alpha_2}{(1 + \exp((t_{ij} - \alpha_3)/\alpha_{4i}))} + \epsilon_{ij},$$

where $i = 1, \dots, 100$, $j = 1, \dots, 10$, $\alpha_{1i} = \exp(\beta_1 + b_{1i})$, $\beta_k = \log(\alpha_k)$, $k = 2, 3$, $\alpha_{4i} = \exp(\beta_4 + b_{2i})$, $(b_{1i}, b_{2i}) \stackrel{ind.}{\sim} N_2(\mathbf{0}, \mathbf{D})$, and $\epsilon_{ij} \stackrel{ind.}{\sim} N_{n_i}(\mathbf{0}, \sigma_e^2 \mathbf{I}_{n_i})$. We set $\boldsymbol{\beta} = (1.6094, 0.6931, 3.8067, 2.3026)^\top$, $\sigma^2 = 0.55$, and \mathbf{D} with elements $D_{11} = 0.0025$, $D_{12} = -0.001$ and $D_{22} = 0.0100$. In addition, twenty percent (20%) of all observations were censored.

After generating the data y_{ij} ($i = 1, \dots, 100$, $j = 1, \dots, 10$), we perturbed the response variable of individual #85 as follows: $y_{i2} \leftarrow y_{i2} + 0.1\sqrt{(\mathbf{y}^{*\top} \mathbf{y}^*)}$, where $\mathbf{y}^* = (y_{12}, \dots, y_{100,2})^\top$. By using the approach described earlier in Section 3, we compute the case deletion measure QD_i^1 and local influence measures based on the response and case weight perturbations. As expected, we observe from Figure 7 the influence of the observation #85. This reveals that the influence measures have

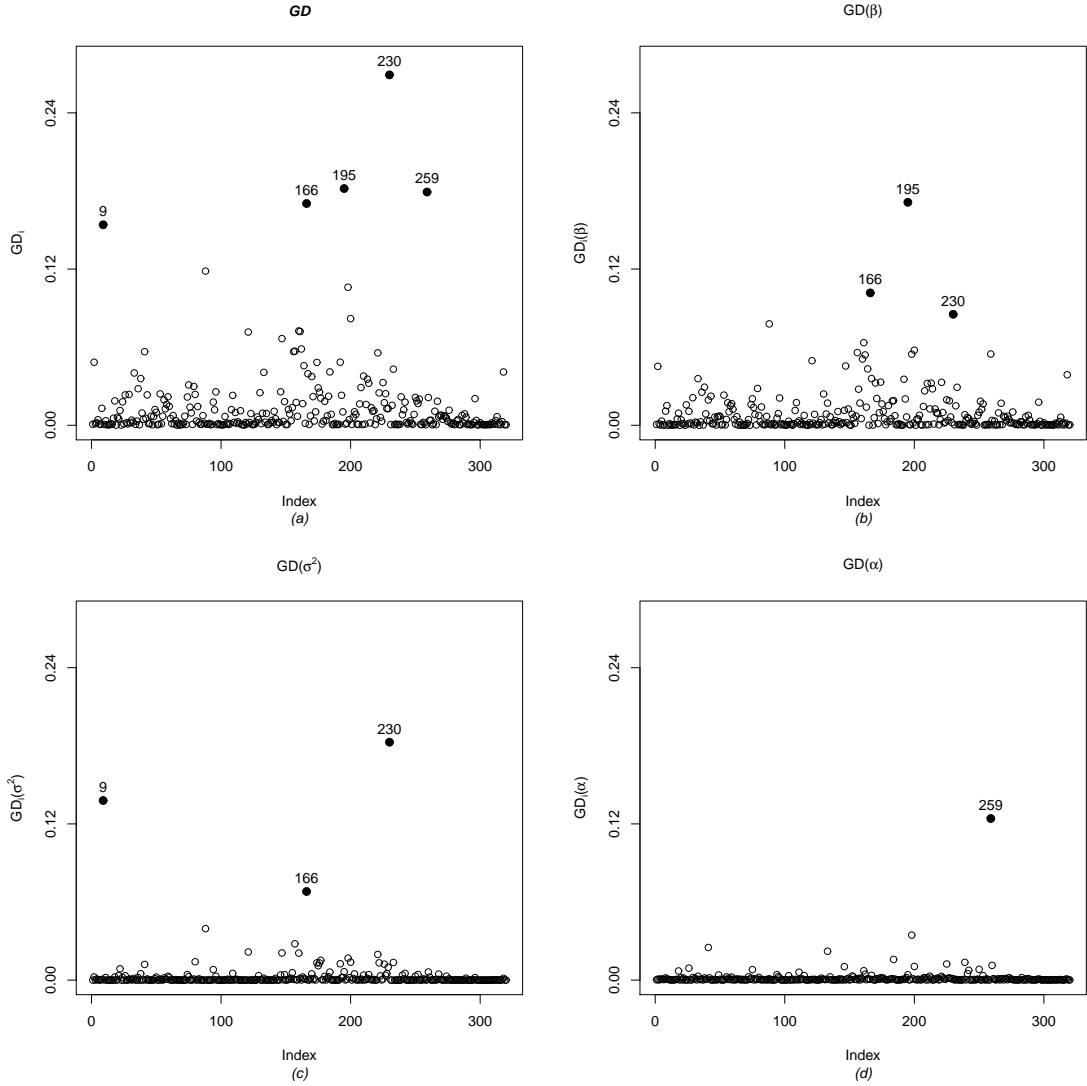


Figure 5: AIEDRP Study data. (a) Approximate generalized Cook's distance GD_i^1 , (b) GD_i^1 for subset β , (c) GD_i^1 for subset σ^2 and (d) GD_i^1 for subset α . The influential observations are numbered.

detected what they are supposed to detect, but at the same time suggest and give no false influential cases.

6 Conclusions

This article provides a new insight into the classical diagnostics methods for censored linear and nonlinear mixed effects models, typically used for analyzing censored HIV viral load outcomes, and also presents an useful expectation conditional maximization (EMC) algorithm, which enable the development of diagnostic influence measures. Explicit expressions are obtained for the Hessian matrix $\ddot{\mathbf{Q}}$ and for the matrix Δ under different perturbation schemes. For NLMEC, the analysis is

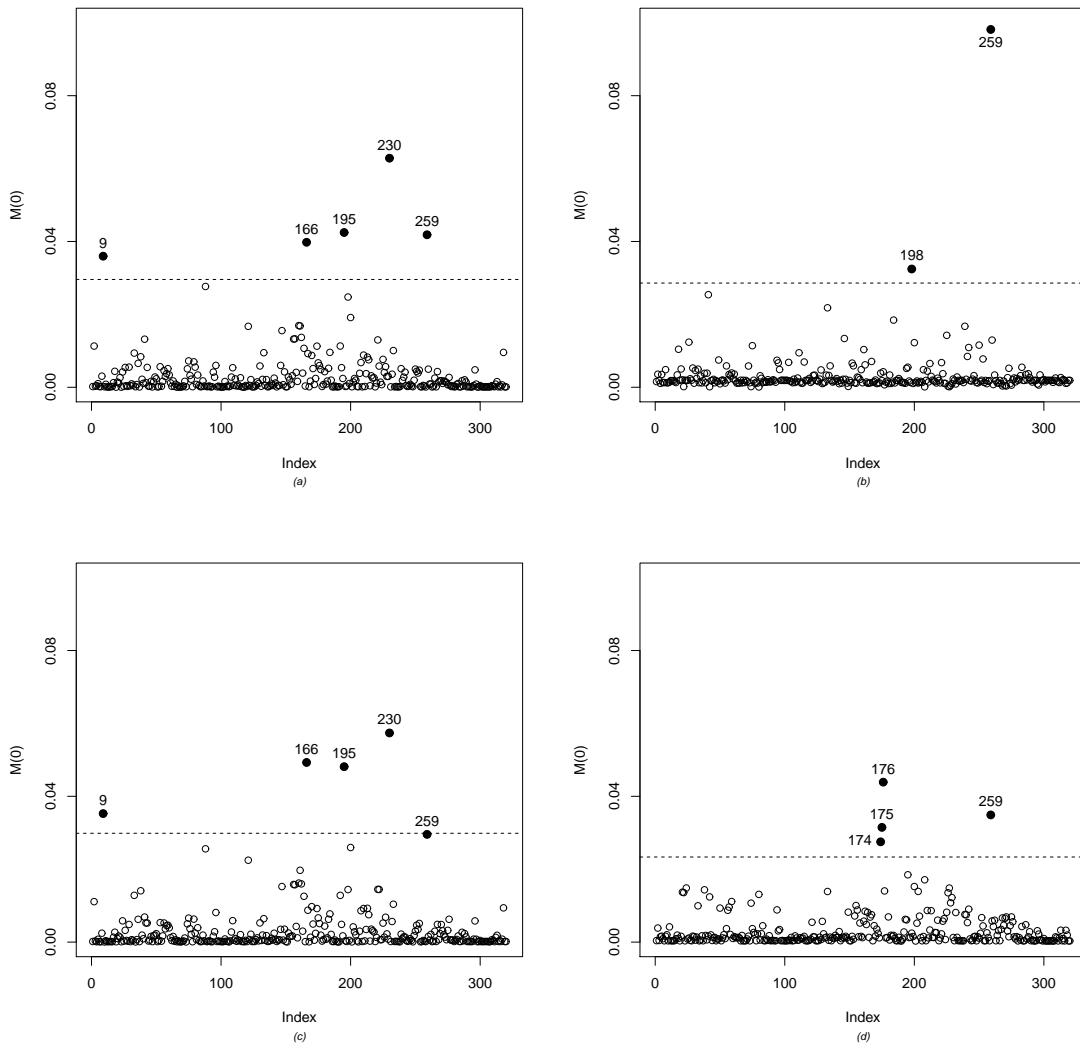


Figure 6: AIEDRP data. Index plot of $M(0)$ for assessing local influence on θ under (a) case weight perturbation, (b) perturbation on \mathbf{D} , (c) perturbation on σ^2 and (d) perturbation on the response variable. The influential observations are numbered.

mathematically (and computationally) feasible through a linearization procedure. The proposed methodology has been applied to two recent (left and right-censored) AIDS studies, which is freely downloadable from R. Our findings about the influential observations for these two datasets agree with those in Lachos et al. (2011) from a Bayesian perspective. The models can be fitted using standard software packages like R, and the code is available from the authors upon request.

The proposed methods can be extended to interval-censored longitudinal data, following the work of Sinha et al. (1999). On the other hand, the models developed here do not consider skewness in the responses because typically in HIV-AIDS studies, the responses (censored viral load) is log-transformed to achieve a ‘close to

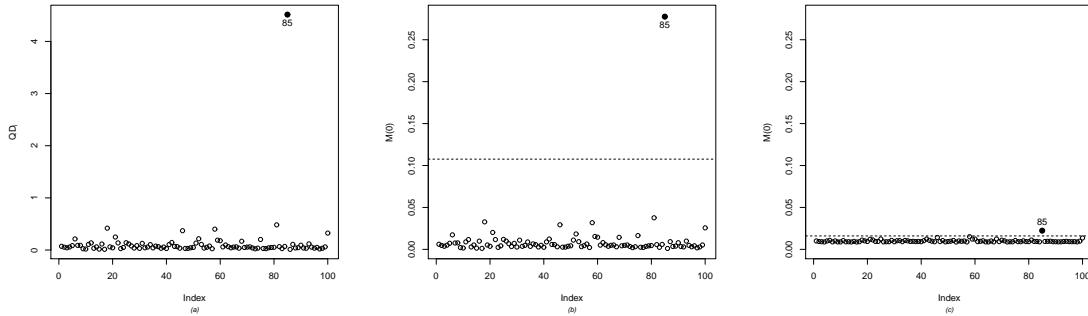


Figure 7: Simulated data set. (a) Approximate generalized Cook's distance GD_i^1 , (b) case weight perturbation and (c) perturbation on the response variable. The influential observations are numbered.

normality' shape. However, features of non-normality, such as skewness and thick-tails, need to be incorporated into the proposed methodology to come up with a more general framework for censored mixed models. This issue is currently under investigation. Incorporating measurement error models (Wu, 2010) within this robust analysis framework for the HIV viral load covariates (viz., CD4 cell counts) will also be of great interest, and we plan to take this on as our future research project.

Acknowledgements

The research of V.H.Lachos was supported by Grant 308109/2008-2 from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-Brazil).

References

- 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.
- Genz, A., Bretz, F., Hothorn, T., Miwa, T., Mi, X., Leisch, F., Scheipl, F., 2008. *mvtnorm: Multivariate normal and t distribution*. R package version 0.9-2, URL <http://CRAN.R-project.org/package=mvtnorm>.

- Hughes, J., 1999. Mixed effects models with censored data with application to HIV RNA levels. *Biometrics* 55, 625–629.
- Jacqmin-Gadda, H., Thiebaut, R., Chene, G., Commenges, D., 2000. Analysis of left-censored longitudinal data with application to viral load in HIV infection. *Biostatistics* 1, 355–368.
- Lachos, V., Bandyopadhyay, D., Dey, D., 2011. Linear and nonlinear mixed-effects models for censored HIV viral loads using normal/independent distributions. *Biometrics*, (to appear).
- Laird, N. M., Ware, J., 1982. Random effects models for longitudinal data. *Biometrics* 38, 963–974.
- Lee, S. Y., Xu, L., 2004. R influence analysis of nonlinear mixed-effects models. *Computational Statistics & Data Analysis* 45, 321–341.
- Lesaffre, E., Verbeke, G., 1998. Local influence in linear mixed models. *Biometrics* 54, 570–582.
- Meng, X., Rubin, D. B., 1993. Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* 81, 633–648.
- Meng, X., Van Dyk, D., 1997. The EM algorithm: An old folk-song sung to a fast new tune. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* 59, 511–567.
- Osorio, F., Paula, G. A., Galea, M., 2007. Assessment of local influence in elliptical linear models with longitudinal structure. *Computational Statistics & Data Analysis* 51, 4354–4368.
- Pinheiro, J. C., Bates, D. M., 2000. *Mixed-Effects Models in S and S-PLUS*. Springer, New York, NY.
- Poon, W., Poon, Y., 1999. Conformal normal curvature and assessment of local influence. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* 61, 51–61.
- R Development Core Team, 2009. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
URL <http://www.R-project.org>

- Russo, C., Paula, G., Aoki, R., 2009. Influence diagnostics in nonlinear mixed-effects elliptical models. *Computational Statistics & Data Analysis* 53, 4143–4156.
- Saitoh, A., Foca, M., Viani, R., Heffernan-Vacca, S., Vaida, F., Lujan-Zilbermann, J., Emmanuel, P., Deville, J., Spector, S., 2008. Clinical outcomes after an unstructured treatment interruption in children and adolescents with perinatally acquired HIV infection. *Pediatrics* 121, 513–521.
- Sinha, D., Chen, M., Ghosh, S., 1999. Bayesian analysis and model selection for interval-censored survival data. *Biometrics* 55, 585–590.
- 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, 171–192.
- Wu, L., 2010. *Mixed Effects Models for Complex Data*. Chapman & Hall/CRC, Boca Raton, FL.
- 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.
- 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.