**Noname manuscript No.**
(will be inserted by the editor)

# Censored mixed-effects models for irregularly observed repeated measures with applications to HIV viral loads

**Larissa A. Matos · Luis M. Castro · Víctor H. Lachos**

Received: date / Accepted: date

**Abstract** In some AIDS clinical trials, the HIV-1 RNA measurements are collected irregularly over time and are often subject to some upper and lower detection limits, depending on the quantification assays. Linear and nonlinear mixed-effects models, with modifications to accommodate censored observations, are routinely used to analyze this type of data (Vaida and Liu, 2009; Matos et al, 2013a). The paper presents a framework for fitting LMEC/NLMEC with response variables recorded at irregular intervals. To address the serial correlation among the within-subject errors, a damped exponential correlation structure is considered in the random error and an EM-type algorithm is developed for computing the maximum likelihood estimates, obtaining as a byproduct the standard errors of the fixed effects and the likelihood value. The proposed methods are illustrated with simulations and the analysis of two real AIDS case studies.

**Keywords** Censored data, EM Algorithm, HIV viral load, Irregularly observed data, Linear/ nonlinear mixed models

## 1 Introduction

Nowadays, the study of acquired immunodeficiency syndrome (AIDS) and understanding of the dynamics of the human immunodeficiency virus (HIV) have become the focus of biomedical and biostatistical research. As mentioned by many researchers, HIV is an extremely dynamic and variable virus having new subtypes and recombinant forms, about which the scientific community knows little or nothing. HIV/AIDS clinical trails aim to find new ways to prevent, detect and/or treat AIDS by determining whether a new anti-retroviral (ARV) agent/therapy is safe and effective in people. Most of these clinical trials assess the quantitative rates/changes of HIV-1 ribonucleic acid (RNA) levels in plasma (or simply HIV-1 viral load), since is an important surrogate marker to assess the risk of disease progression and to monitor response to ARV therapy in routine medical care of infected patients.

However, modeling HIV-1 viral load presents many challenges from the statistical point of view. Three are of particular importance. First, the viral load measurements are often left or right censored (undetected) due to a lower and/or upper detection limit of quantification. This is because

Larissa A. Matos
Departamento de Estatística, Universidade Estadual de Campinas,
Cidade Universitaria "Zeferino Vaz", Campinas, Sao Paulo, Brazil
E-mail: larissa.amatos@gmail.com

Luis M. Castro
Department of Statistics and CI$^2$MA, Universidad de Concepción, Concepción, Chile
E-mail: luiscastroc@udec.cl

Víctor H. Lachos
Departamento de Estatística, Universidade Estadual de Campinas,
Cidade Universitaria "Zeferino Vaz", Campinas, Sao Paulo, Brazil
E-mail: hlachos@ime.unicamp.br

some quantification assays cannot accurately quantify HIV-1 RNA above/below a specific level. **Particularly, lower detection limits ranging from 400 to 500 RNA copies/mL are considered for standard assays while the range is 50 to 100 RNA copies/mL for ultra-sensitive assays. For example, the Amplicor HIV-1 monitor test 1.5 and Nuclisens HIV-1 QT assay consider a lower detection limit of 400 copies/mL (Antunes et al, 2003), while the Roche Cobas Amplicor HIV-1 Monitor test (versions 1 and 1.5) considers a detection limit of 50 HIV-1 RNA copies/ml and the TaqMan assay, version 1 and 2, considers a lower limit of quantification of 40 and 20 copies/ml respectively (see Swenson et al, 2014).**

Second, as a result of unscheduled follow-up visits of patients and/or missed visits, the viral loads are usually recorded at irregular intervals. **As an example of this situation, Ciesielski and Metler (1997) studied the duration of time between exposure and seroconversion in healthcare workers with occupationally acquired infection with HIV. In this study, the authors mentioned that "because many of the healthcare workers had follow-up testing at irregular intervals, with long periods between tests, it was not possible to define precisely when seroconversion occurred". Another example of this situation was reported by Lopes de Azevedo et al (2010), were a patient diagnosed with HIV-infection in 2003 during her first pregnancy made follow-up visits at irregular intervals. In this particular case, the antiretrovirals were given only for prophylaxis of HIV-infection vertical transmission during pregnancy.** Finally, since the viral load is measured longitudinally over time, the between-subject and within-subject variations have to be taken into account.

Recently, some alternatives for modeling the irregular observation responses and correlations induced by longitudinal data have been proposed in the statistical literature. These proposals consider not only the correlation structure induced by the random effects term but also by other types of correlation in the error term. Particularly, Wang (2013) propose a multivariate Student's-$t$ linear mixed model for outcome variables recorded on irregular occasions considering a damping exponential correlation (DEC) structure as proposed by Muñoz et al (1992). This correlation structure takes into account the autocorrelation generated by the within-subject dependence among irregular occasions. **On the other hand, Lin and Wang (2013) consider a multivariate Student's-$t$ distribution for nonlinear mixed models with multiple outcomes in presence of missing data. To capture the serial correlation among the observations, the authors consider a DEC structure of the error vector.** Moreover, Wang and Fan (2011) consider the multivariate Student's-$t$ linear mixed with autoregressive of order $p$ (AR($p$)) dependence structure for the within-subject errors in the case of multiple outcomes.

**In the case of censored responses, there are several alternatives proposed in the literature to deal with them. For example, Arellano-Valle et al (2012) extend the classic Tobit model (Tobin, 1958) by considering a Student's-$t$ distribution for the error term and proposing an EM-type algorithm for the parameter estimation. More recently, Rocha et al (2015) propose an errors-in-variable Student's-$t$ censored model, obtaining the maximum likelihood estimates (MLE) of the model through an EM algorithm, and Müller and Van de Geer (2015) study a censored linear model for high dimensional data. In the context of linear/nonlinear mixed-effects (LME/NLME) models, Hughes (1999) proposes a likelihood-based Monte Carlo EM algorithm (MCEM) for LME with censored responses (LMEC). Wu (2002) proposes a Monte Carlo EM and a linearization procedure to estimate the parameters of a censored NLME model. In turn, Vaida et al (2007); Vaida and Liu (2009) extend the work of Hughes, proposing a more efficient EM algorithm than Hughes's algorithm. An extended review of these proposals can be found in the book by Wu (2010). Recently, Matos et al (2013a), Matos et al (2013b) and Matos et al (2015) have proposed a likelihood-based estimation and influence analysis for LMEC/NLMEC models, respectively.**

Moreover, stochastic versions of EM such as Monte Carlo EM (Levine and Casella, 2001), SAEM (Deylon et al, 1999) and many other approximations have been proposed in the literature to deal with NLME models under censoring. In fact, Samson et al (2006) propose an extension of the SAEM algorithm to left-censored data in NLME model. However, to the best of the knowledge there is no work considering irregular observations, damping exponential correlation and censored longitudinal responses simultaneously in the context of LMEC/NLMEC models using an exact EM algorithm. Consequently, the aim of this paper is to study the impact of censoring and irregularly timed observed responses under Gaussian LMEC and NLMEC models.

For this purpose, we consider the analysis of two AIDS case studies. The first one investigated the effect of a highly active antiretroviral therapy (HAART) in persons with moderately advanced HIV-1 infection. This case study presented 11% of observations below (left-censored) the detection limits. The second case study evaluated the immune responses to HIV during acute infection, presenting about 22% of measurements lying above (right-censored) the limits of assay quantification. Moreover, in both studies, the viral loads were irregularly measured over time.

The rest of the paper is organized as follows. Section 2 describes the AIDS case studies that motivate this paper. Section 3 introduces the model (DEC-LMEC) and the likelihood function. In Section 4, the related likelihood-based inference is presented, including estimation of the random effects and the expected information matrix. The method for predicting future observations is presented in Section 5. Section 6 presents the extension to the nonlinear case (DEC-NLMEC). The application of the proposed method is presented in Sections 7 and 8 through a simulation study and the analysis of two case studies of HIV viral load. Finally, Section 9 concludes with a short discussion of issues raised by this study and some possible directions for future research.

## 2 Case studies

In this section presents the two motivating datasets, which will be analyzed next.

### 2.1 ACTG 315 data

The ACTG 315 protocol considers 46 HIV-1 infected patients treated with a potent antiretroviral drug cocktail based on protease inhibitor ritonavir and reverse transcriptase inhibitor drugs (zidovudine and lamivudine). Before initiating the antiretroviral therapy, all patients discontinued their own antiretroviral regimen for five weeks as a “washout" period. The aim of this antiretroviral regimen is to show that immunity can be partially restored in people with moderately advance HIV disease.

The viral load was quantified on days 0, 2, 7, 10, 14, 21, 28, 56, 84, 168 and 196 after starting treatment. The dataset includes 361 observations. An immunologic marker known as CD4+ cell count was also measured along with viral load and 72 out of 361 (20%) CD4 values were missing due to a mismatch of the CD4 and the viral load measurement schedules. The number of measurements per subject varied from 4 to 10. Viral load measurements below the detectable threshold of 100 copies/mL (40 out of 361, 11%) were considered left-censored, and the censoring process assumed independence of the complete data. The individual profiles are shown in Figure 1 (left panel).

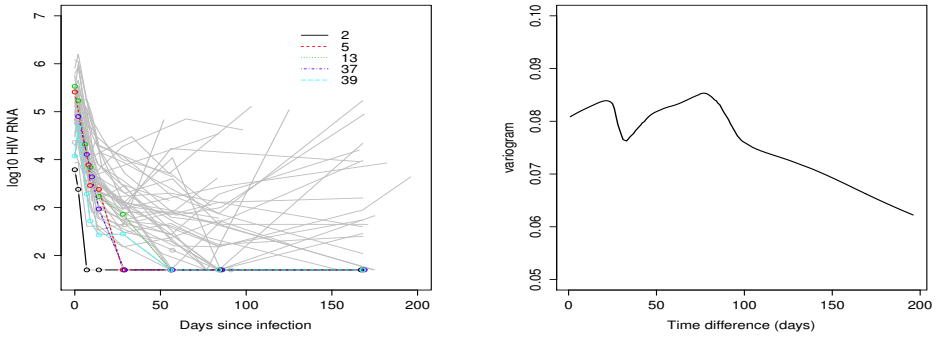


**Fig. 1** ACTG 315 data. (left panel) Individual profiles (in $\log_{10}$ scale) for HIV viral load at different follow-up times. Trajectories for some censored individuals are indicated in different colors. (right panel) Variogram from model residuals using an NLMEC model.

**Figure 1 (right panel) shows the variogram of the residuals under the nonlinear censored mixed model with independent errors. Note that, it indicates long-term autocorrelation, which may be due a serial autocorrelation beyond the random effect model.**

## 2.2 AIEDRP data

The second AIDS case study is from the AIEDRP program. This program, which is a large multicenter observational study of subjects with acute and early HIV infection, covers areas such as the evaluation of immune responses to HIV during acute infection, the assessment of thymic function and T-cell turnover during acute HIV infection and the assessment of transmission and prevalence of HIV resistance among treatment-naive subjects. The aim of this study was to help design future vaccines and to learn the implications of new anti-HIV treatments.

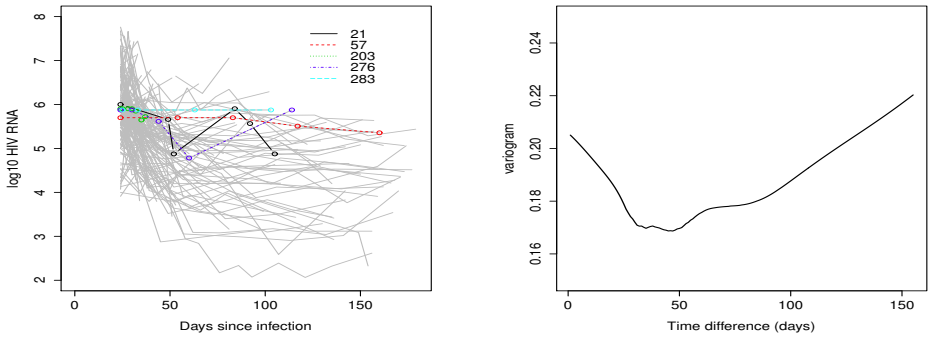


**Fig. 2** AIEDRP data. (left panel) Individual profiles (in $\log_{10}$ scale) for HIV viral load at different follow-up times. Trajectories for some censored individuals are indicated in different colors. (right panel) Variogram from model residuals using an NLMEC model.

We consider 320 untreated individuals with acute HIV infection (See Vaida and Liu (2009) for more details). Of the 830 recorded observations, 185 (22%) were above the limit of assay quantification. The individual profiles are shown in Figure 2 (left panel).

**As in the previous case, Figure 2 (right panel) presents the variogram of the residuals under the nonlinear censored mixed model with independent errors, showing long-term autocorrelation.**

## 3 Model formulation

In the non-censored case, a Gaussian LME model is specified as follows:

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

where $\mathbf{b}_i \overset{iid}{\sim} \mathrm{N}_q(\mathbf{0},\mathbf{D})$ is independent of $\varepsilon_i \overset{ind.}{\sim} \mathrm{N}_{n_i}(\mathbf{0},\Omega_i)$, $i = 1,\ldots,n$; the subscript $i$ is the subject index; $\mathbf{y}_i = (y_{i1},\ldots,y_{in_i})^\top$ is an $n_i \times 1$ vector of observed continuous responses for subject $i$ measured at particular time points $\mathbf{t}_i = (t_{i1},\ldots,t_{n_i})^\top$; $\mathbf{X}_i$ is the $n_i \times p$ design matrix corresponding to the fixed effects, $\beta$, 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 $\mathbf{b}_i$; $\varepsilon_i$ of dimension $(n_i \times 1)$ is the vector of random errors; and the dispersion matrix $\mathbf{D} = \mathbf{D}(\alpha)$ depends on the unknown and reduced parameters $\alpha$. The correlation structure of the error vector is assumed to be $\Omega_i = \sigma^2\mathrm{E}_i$, where the $n_i \times n_i$ matrix $\mathrm{E}_i$ incorporates a time-dependence structure. Consequently, to capture the serial correlation among irregularly observed longitudinal data, such as the ACTG 315 and AIEDRP datasets, it is necessary to consider

a parsimonious parameterization of the matrix $\mathrm{E}_i$. Following Muñoz et al (1992), we adopt a DEC structure for $\mathrm{E}_i$, which is defined as:

$$\mathrm{E}_i = \mathrm{E}_i(\phi, \mathbf{t}_i) = \left[\phi_1^{|t_{ij}-t_{ik}|^{\phi_2}}\right],\ i = 1,\ldots,n,\ \ j,k = 1,\ldots,n_i, \tag{2}$$

where $\phi = (\phi_1, \phi_2)^\top$, the parameter $\phi_1$ describes the autocorrelation between observations separated by the absolute length of two time points, and the parameter $\phi_2$ permits acceleration of the exponential decay of the autocorrelation function, defining a continuous-time autoregressive model.

For practical reasons, the parameter space of $\phi_1$ and $\phi_2$ is confined within $\Phi = \{(\phi_1, \phi_1) : 0 < \phi_1 < 1, \phi_2 > 0\}$. It is important to stress that different values of the damping parameter $\phi_2$ produce a variety of correlation structures for a given value of $\phi_1 > 0$, as follows: (a) if $\phi_2 = 0$, then $\mathrm{E}_i$ generates the compound symmetry correlation structure; (b) when $0 < \phi_2 < 1$, then $\mathrm{E}_i$ presents a decay rate between the compound symmetry structure and the first-order AR (AR (1)) model; (c) if $\phi_2 = 1$, then $\mathrm{E}_i$ generates an AR(1) structure; (d) when $\phi_2 > 1$, $\mathrm{E}_i$ presents a decay rate faster than the AR(1) structure; and (e) if $\phi_2 \to \infty$, then $\mathrm{E}_i$ represents the first-order moving average model, MA(1). A more detailed discussion of the DEC structure presenting more complex scenarios of the parameter space $\Phi$ can be found in Muñoz et al (1992).

As mentioned earlier, the proposed model also considers censored observations, *i.e.*, we assume that the response $Y_{ij}$ is not fully observed for all $i, j$. Let $(\mathbf{V}_i, \mathbf{C}_i)$ be the observed data for the $i$-th subject, where $\mathbf{V}_i$ represents the vector of uncensored readings or censoring level and $\mathbf{C}_i$ is the vector of censoring indicators, such that

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

Note that since the observed response $y_{ij}$ is defined over the real line, extensions to right censored data are straightforward. In fact, the right censored problem can be represented by a left censored problem by simultaneously transforming the response $y_{ij}$ and censoring level $V_{ij}$ to $-y_{ij}$ and $-V_{ij}$. The model defined in (1)-(3), is henceforth called DEC-LMEC.

### 3.1 The log-likelihood function

Following Vaida and Liu (2009), classic inference on the parameter vector $\theta = (\beta^\top, \sigma^2, \alpha^\top, \phi^\top)^\top$ is based on the marginal distribution of $\mathbf{y}_i$. For complete data, the marginal distribution of the vector $\mathbf{y}_i$, for $i = 1,\ldots,n$ is $\mathrm{N}_{n_i}(\mathbf{X}_i\beta, \Sigma_i)$, where $\Sigma_i = \Omega_i + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top$. The strategy followed to compute the likelihood function associated with model (1) and (2) 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 $C_{ij} = 1$ for all elements in $\mathbf{y}_i^c$. After reordering, $\mathbf{y}_i$, $\mathbf{V}_i$, $\mathbf{X}_i$, and $\Sigma_i$ can be partitioned as follows:

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

In this setup, the operator $vec(\cdot)$ denotes the function with stack vectors or matrices of the same number of columns. Consequently, from the marginal-conditional decomposition of the multivariate normal distribution, $\mathbf{y}_i^o \sim \mathrm{N}_{n_i^o}(\mathbf{X}_i^o\beta, \Sigma_i^{oo})$ and $\mathbf{y}_i^c|\mathbf{y}_i^o \sim \mathrm{N}_{n_i^c}(\mu_i, \mathbf{S}_i)$, where $\mu_i = \mathbf{X}_i^c\beta + \Sigma_i^{co}(\Sigma_i^{oo})^{-1}(\mathbf{y}_i^o - \mathbf{X}_i^o\beta)$ and $\mathbf{S}_i = \Sigma_i^{cc} - \Sigma_i^{co}(\Sigma_i^{oo})^{-1}\Sigma_i^{oc}$. 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 $\mathrm{N}_{n_i}(\mathbf{a}, \mathbf{A})$ computed at vector $\mathbf{u}$. From Vaida and Liu (2009) and Matos et al (2013a), the likelihood function for subject $i$ (using conditional probability arguments) is given by:

$$\begin{aligned} L_i(\theta) &= f(\mathbf{y}_i^c \leq \mathbf{V}_i^c|\mathbf{y}_i^o = \mathbf{V}_i^o, \theta) f(\mathbf{y}_i^o = \mathbf{V}_i^o|\theta), \\ &= f(\mathbf{y}_i^c \leq \mathbf{V}_i^c|\mathbf{y}_i^o, \theta) f(\mathbf{y}_i^o|\theta) \\ &= \Phi_{n_i^c}(\mathbf{V}_i^c; \mu_i, \mathbf{S}_i)\phi_{n_i^o}(\mathbf{y}_i^o; \mathbf{X}_i^o\beta, \Sigma_i^{oo}), \end{aligned} \tag{4}$$

which can be easily evaluated computationally.

The log-likelihood function for the observed data, given by

$$\ell(\theta|\mathbf{y}) = \sum_{i=1}^{n}\{\log L_i(\theta)\},$$

is used to compute different model selection criteria, such as:

$$AIC = 2m - 2\ell_{max} \text{ and } BIC = m\,log\,N - 2\ell_{max},$$

where $m$ is the number of model parameters, $N = \sum_{i=1}^{n} n_i$ and $\ell_{max}$ is the maximized log-likelihood value.

## 4 The EM algorithm

This section describes in detail how the proposed model specified in (1)-(3) can be fitted by using the ECM algorithm (Meng and Rubin, 1993). The EM algorithm (proposed originally by Dempster et al (1977)) has several appealing features, such as stability of monotone convergence with each iteration, increasing the likelihood and simplicity of implementation. Due to the computational difficulty at the M-step, we use the ECM algorithm (an extension of the EM algorithm), which shares the appealing features of the EM and presents faster convergence than the original algorithm.

Let $\mathbf{y} = (\mathbf{y}_1^\top,\ldots,\mathbf{y}_n^\top)^\top$, $\mathbf{b} = (\mathbf{b}_1^\top,\ldots,\mathbf{b}_n^\top)^\top$, $\mathbf{V} = vec(\mathbf{V}_1,\ldots,\mathbf{V}_n)$ and $\mathbf{C} = vec(\mathbf{C}_1,\ldots,\mathbf{C}_n)$. Considering $\mathbf{b}$ as the hypothetical missing data, the complete data are denoted by

$$\mathbf{y}_c = (\mathbf{C}^\top,\mathbf{V}^\top,\mathbf{y}^\top,\mathbf{b}^\top)^\top.$$

Hence, the ECM algorithm is applied to the complete data log-likelihood function:

$$\begin{aligned}\ell_i(\theta|\mathbf{y}_c) = -\frac{1}{2}\Big[&n_i\log\sigma^2 + \log(|\mathrm{E}_i|) + \frac{1}{\sigma^2}(\mathbf{y}_i - \mathbf{X}_i\beta - \mathbf{Z}_i\mathbf{b}_i)^\top \mathrm{E}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\beta - \mathbf{Z}_i\mathbf{b}_i)\\ &+ \log|\mathbf{D}| + \mathbf{b}_i^\top\mathbf{D}^{-1}\mathbf{b}_i\Big] + K, \end{aligned} \tag{5}$$

with $K$ being a constant that does not depend on the parameter vector $\theta$. Given the current estimate $\theta = \widehat{\theta}^{(k)}$, the E-step calculates the conditional expectation of the complete data log-likelihood function, given by:

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

where

$$\begin{aligned} Q_{1i}\left(\beta,\sigma^2,\phi|\widehat{\theta}^{(k)}\right) = -\frac{n_i}{2}\log\widehat{\sigma^2}^{(k)} - \frac{1}{2}\log(|\widehat{\mathrm{E}}_i^{(k)}|) - \frac{1}{2\widehat{\sigma^2}^{(k)}}\Big[&\widehat{a}_i^{(k)}(\widehat{\phi}^{(k)})\\ &-2\widehat{\beta}^{(k)\top}\mathbf{X}_i^\top\widehat{\mathrm{E}}_i^{-1(k)}\left(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{\mathbf{b}}_i^{(k)}\right) + \widehat{\beta}^{(k)\top}\mathbf{X}_i^\top\widehat{\mathrm{E}}_i^{-1(k)}\mathbf{X}_i\widehat{\beta}^{(k)}\Big], \end{aligned} \tag{6}$$

$$Q_{2i}\left(\alpha|\widehat{\theta}^{(k)}\right) = -\frac{1}{2}\log|\widehat{\mathbf{D}}^{(k)}| - \frac{1}{2}tr\left(\widehat{\mathbf{b}_i\mathbf{b}_i^\top}^{(k)}\widehat{\mathbf{D}}^{-1(k)}\right), \tag{7}$$

and $\widehat{a}_i^{(k)}(\phi) = tr\left(\widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)}\mathrm{E}_i^{-1} - 2\widehat{\mathbf{y}_i\mathbf{b}_i^\top}^{(k)}\mathbf{Z}_i^\top\mathrm{E}_i^{-1} + \widehat{\mathbf{b}_i\mathbf{b}_i^\top}^{(k)}\mathbf{Z}_i^\top\mathrm{E}_i^{-1}\mathbf{Z}_i\right)$,

$$\widehat{\mathbf{b}}_i^{(k)} = E\left\{\mathbf{b}_i|\mathbf{V}_i,\mathbf{C}_i,\widehat{\theta}^{(k)}\right\} = \widehat{\varphi}_i^{(k)}(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{X}_i\widehat{\beta}^{(k)}),$$

$$\begin{aligned}\widehat{\mathbf{b}_i\mathbf{b}_i^\top}^{(k)} &= E\left\{\mathbf{b}_i\mathbf{b}_i^\top|\mathbf{V}_i,\mathbf{C}_i,\ \widehat{\theta}^{(k)}\right\}\\ &= \widehat{\Lambda}_i^{(k)} + \widehat{\varphi}_i^{(k)}(\widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)} - \widehat{\mathbf{y}}_i^{(k)}\widehat{\beta}^{(k)\top}\mathbf{X}_i^\top - \mathbf{X}_i\widehat{\beta}^{(k)}\widehat{\mathbf{y}}_i^{(k)\top} + \mathbf{X}_i\widehat{\beta}^{(k)}\widehat{\beta}^{(k)\top}\mathbf{X}_i^\top)\widehat{\varphi}_i^{(k)\top},\\ \widehat{\mathbf{y}_i\mathbf{b}_i^\top}^{(k)} &= E\{\mathbf{y}_i\mathbf{b}_i^\top|\mathbf{V}_i,\mathbf{C}_i,\widehat{\theta}^{(k)}\} = (\widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)} - \widehat{\mathbf{y}}_i^{(k)}\widehat{\beta}^{(k)\top}\mathbf{X}_i^\top)\widehat{\varphi}_i^{(k)\top}, \end{aligned}$$

with $\widehat{\Lambda}_i^{(k)} = (\widehat{\mathbf{D}}^{-1(k)} + \mathbf{Z}_i^\top \widehat{\mathrm{E}}_i^{-1(k)} \mathbf{Z}_i / \widehat{\sigma^2}^{(k)})^{-1}$ and $\widehat{\varphi}_i^{(k)} = \widehat{\Lambda}_i^{(k)} \mathbf{Z}_i^\top \widehat{\mathrm{E}}_i^{-1(k)} / \widehat{\sigma^2}^{(k)}$.

It is easy to see from (6) and (7) that the E-step reduces only to the computation of

$$\widehat{\mathbf{y}_i \mathbf{y}_i^\top}^{(k)} = E\{\mathbf{y}_i \mathbf{y}_i^\top | \mathbf{V}_i, \mathbf{C}_i, \widehat{\theta}^{(k)}\} \text{ and } \widehat{\mathbf{y}}_i^{(k)} = E\{\mathbf{y}_i | \mathbf{V}_i, \mathbf{C}_i, \widehat{\theta}^{(k)}\}.$$

These conditional expectations rely on the first and second moments of a multivariate truncated normal distribution and can be determined in closed-form (for more details on the computation of these moments see Vaida and Liu (2009)).

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

$$\widehat{\beta}^{(k+1)} = \left( \sum_{i=1}^{n} \mathbf{X}_i^\top \widehat{\mathrm{E}}_i^{-1(k)} \mathbf{X}_i \right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i^\top \widehat{\mathrm{E}}_i^{-1(k)} \left( \widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i \widehat{\mathbf{b}}_i^{(k)} \right), \tag{8}$$

$$\begin{aligned} \widehat{\sigma^2}^{(k+1)} = \frac{1}{N} \sum_{i=1}^{n} \Bigg[ & \widehat{a}_i^{(k)} - 2\widehat{\beta}^{(k+1)\top} \mathbf{X}_i^\top \widehat{\mathrm{E}}_i^{-1(k)} \left( \widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i \widehat{\mathbf{b}}_i^{(k)} \right) \\ & + \widehat{\beta}^{(k+1)\top} \mathbf{X}_i^\top \widehat{\mathrm{E}}_i^{-1(k)} \mathbf{X}_i \widehat{\beta}^{(k+1)} \Bigg], \end{aligned} \tag{9}$$

$$\widehat{\mathbf{D}}^{(k+1)} = \frac{1}{n} \sum_{i=1}^{n} \widehat{\mathbf{b}_i \mathbf{b}_i^\top}^{(k)}, \tag{10}$$

$$\begin{aligned} \phi^{(k+1)} = \underset{\phi \in (0,1) \times \mathbb{R}^+}{\operatorname{argmax}} \Bigg( & -\frac{1}{2} \log(|\mathrm{E}_i|) - \frac{1}{2\widehat{\sigma^2}^{(k+1)}} \Big[ \widehat{a}_i^{(k)}(\phi) \\ & - 2\widehat{\beta}^{(k+1)\top} \mathbf{X}_i^\top \mathrm{E}_i^{-1} \left( \widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i \widehat{\mathbf{b}}_i^{(k)} \right) + \widehat{\beta}^{(k+1)\top} \mathbf{X}_i^\top \mathrm{E}_i^{-1} \mathbf{X}_i \widehat{\beta}^{(k+1)} \Big] \Bigg). \end{aligned} \tag{11}$$

### 4.1 Estimation of random effects and standard errors

To estimate the random effects, we consider the conditional mean of $\mathbf{b}_i$ given the observed data $\mathbf{V}_i$ and $\mathbf{C}_i$, that is, $E\{\mathbf{b}_i|\mathbf{V}_i, \mathbf{C}_i\}$. Thus, for a given value of $\theta = (\beta^\top, \sigma^2, \alpha^\top, \phi^\top)^\top$, the conditional mean of $\mathbf{b}_i$ given $\mathbf{V}_i$ and $\mathbf{C}_i$ is:

$$\widehat{\mathbf{b}}_i(\theta) = E\{\mathbf{b}_i|\mathbf{V}_i, \mathbf{C}_i\} = \varphi_i(\widehat{\mathbf{y}}_i - \mathbf{X}_i\beta), \tag{12}$$

where $\varphi_i = \Lambda_i \mathbf{Z}_i^\top \mathrm{E}_i^{-1} / \sigma^2$ and $\Lambda_i = (\mathbf{D}^{-1} + \mathbf{Z}_i^\top \mathrm{E}_i^{-1} \mathbf{Z}_i / \sigma^2)^{-1}$. Note that $\widehat{\mathbf{y}}_i = E\{\mathbf{y}_i|\mathbf{Q}_i, \mathbf{C}_i\}$ is given by the first moment of a multivariate truncated normal distribution. In practice, the estimator of $\mathbf{b}_i$, $\widehat{\mathbf{b}}_i$, can be obtained by substituting the ML estimate $\widehat{\theta}$ into (12), leading to $\widehat{\mathbf{b}}_i = \widehat{\mathbf{b}}_i(\widehat{\theta})$. On the other hand, the conditional covariance matrix of $\mathbf{b}_i$ given $\mathbf{V}_i$ and $\mathbf{C}_i$ is:

$$Var\{\mathbf{b}_i|\mathbf{Q}_i, \mathbf{C}_i\} = E\{\mathbf{b}_i \mathbf{b}_i^\top | \mathbf{Q}_i, \mathbf{C}_i\} - \widehat{\mathbf{b}}_i(\theta) \widehat{\mathbf{b}}_i(\theta)^\top = \Lambda_i + \varphi_i Var(\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i) \varphi_i^\top.$$

Note that $Var(\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i)$ can be easily obtained as a byproduct of the proposed ECM algorithm developed in Section 4.

*The empirical information matrix*

Following Lin (2010), we compute the asymptotic covariance of the ML estimates through the empirical information matrix, which is computed as (Meilijson, 1989):

$$\mathbf{I}_e(\theta|\mathbf{y}) = \sum_{i=1}^{n} \mathbf{s}(\mathbf{y}_i \mid \theta) \mathbf{s}^\top(\mathbf{y}_i \mid \theta) - \frac{1}{n} \mathbf{S}(\mathbf{y}_i \mid \theta) \mathbf{S}^\top(\mathbf{y}_i \mid \theta), \tag{13}$$

where $\mathbf{S}(\mathbf{y}_i \mid \theta) = \sum_{i=1}^{n} \mathbf{s}(\mathbf{y}_i \mid \theta)$ and $\mathbf{s}(\mathbf{y}_i \mid \theta)$ is the empirical score function for the subject $i$. According Louis (1982), it is possible to relate the score function of the incomplete data log-likelihood with the conditional expectation of the complete data log-likelihood function. Therefore, the individual score can be determined as

$$\mathbf{s}(\mathbf{y}_i \mid \theta) = \frac{\partial \log f(\mathbf{y}_i \mid \theta)}{\partial \theta} = E\left( \frac{\partial \ell_i(\theta \mid \mathbf{y}_c)}{\partial \theta} \mid \mathbf{V}_i, \mathbf{C}_i, \theta \right), \tag{14}$$

where $\ell_i(\theta \mid \mathbf{y}_c)$ is the complete data log-likelihood formed from the observation $i$. Using the ML estimates $\widehat{\theta}$, that is, $\mathbf{S}(\mathbf{y}_i \mid \widehat{\theta}) = 0$, it follows that (13) can be approximated by:

$$\mathbf{I}_e(\widehat{\theta} \mid \mathbf{y}) = \sum_{i=1}^{n} \widehat{\mathbf{s}}_i \widehat{\mathbf{s}}_i^{\top}, \tag{15}$$

where $\widehat{\mathbf{s}}_i = \mathbf{s}(\mathbf{y}_i \mid \widehat{\theta}) = \left( \widehat{\mathbf{s}}_{i,\beta}^{\top}, \widehat{\mathbf{s}}_{i,\sigma^2}, \widehat{\mathbf{s}}_{i,\alpha}^{\top}, \widehat{\mathbf{s}}_{i,\phi}^{\top} \right)^{\top}$ has elements given by

$$\begin{aligned}
\widehat{\mathbf{s}}_{i,\beta} &= (\widehat{\mathbf{s}}_{i,\beta_1}, \ldots, \widehat{\mathbf{s}}_{i,\beta_p})^{\top} = \frac{1}{\widehat{\sigma^2}} \left[ \mathbf{X}_i^{\top} \widehat{\mathrm{E}}_i^{-1} \left( \widehat{\mathbf{y}}_i - \mathbf{Z}_i \widehat{\mathbf{b}}_i \right) - \mathbf{X}_i^{\top} \widehat{\mathrm{E}}_i^{-1} \mathbf{X}_i \widehat{\beta} \right], \\
\widehat{\mathbf{s}}_{i,\sigma^2} &= -\frac{n_i}{2\widehat{\sigma^2}} + \frac{1}{2\widehat{\sigma^4}} \left[ \widehat{a}_i - 2\widehat{\beta}^{\top} \mathbf{X}_i^{\top} \widehat{\mathrm{E}}_i^{-1} \left( \widehat{\mathbf{y}}_i - \mathbf{Z}_i \widehat{\mathbf{b}}_i \right) + \widehat{\beta}^{\top} \mathbf{X}_i^{\top} \widehat{\mathrm{E}}_i^{-1} \mathbf{X}_i \widehat{\beta} \right], \\
\widehat{\mathbf{s}}_{i,\alpha} &= (\widehat{\mathbf{s}}_{i,\alpha_1}, \ldots, \widehat{\mathbf{s}}_{i,\alpha_r})^{\top}, \\
\widehat{\mathbf{s}}_{i,\phi} &= (\widehat{\mathbf{s}}_{i,\phi_1}, \widehat{\mathbf{s}}_{i,\phi_2})^{\top},
\end{aligned}$$

with $\widehat{a}_i = tr\left( \widehat{\mathbf{y}_i \mathbf{y}_i^{\top}} \widehat{\mathrm{E}}_i^{-1} - 2\widehat{\mathbf{y}_i \mathbf{b}_i^{\top}} \mathbf{Z}_i^{\top} \widehat{\mathrm{E}}_i^{-1} + \widehat{\mathbf{b}_i \mathbf{b}_i^{\top}} \mathbf{Z}_i^{\top} \widehat{\mathrm{E}}_i^{-1} \mathbf{Z}_i \right)$, $\widehat{\mathbf{s}}_{i,\alpha_r} = -\frac{1}{2} tr\left( \widehat{\mathbf{D}}^{-1} \dot{\mathbf{D}}(r) \widehat{\mathbf{D}}^{-1} (\widehat{\mathbf{D}} - \widehat{\mathbf{b}_i \mathbf{b}_i^{\top}}) \right)$,

$$\begin{aligned}
\widehat{\mathbf{s}}_{i,\phi_s} = \frac{1}{2\widehat{\sigma^2}} \Big[ & tr\left( \widehat{\mathbf{y}_i \mathbf{y}_i^{\top}} \widehat{\mathrm{E}}_i^{-1} \dot{\mathrm{E}}_i(s) \widehat{\mathrm{E}}_i^{-1} - 2\widehat{\mathbf{y}_i \mathbf{b}_i^{\top}} \mathbf{Z}_i^{\top} \widehat{\mathrm{E}}_i^{-1} \dot{\mathrm{E}}_i(s) \widehat{\mathrm{E}}_i^{-1} + \widehat{\mathbf{b}_i \mathbf{b}_i^{\top}} \mathbf{Z}_i^{\top} \widehat{\mathrm{E}}_i^{-1} \dot{\mathrm{E}}_i(s) \widehat{\mathrm{E}}_i^{-1} \mathbf{Z}_i \right) \\
& -2\widehat{\beta}^{\top} \mathbf{X}_i^{\top} \widehat{\mathrm{E}}_i^{-1} \dot{\mathrm{E}}_i(s) \widehat{\mathrm{E}}_i^{-1} \left( \widehat{\mathbf{y}}_i - \mathbf{Z}_i \widehat{\mathbf{b}}_i \right) + \widehat{\beta}^{\top} \mathbf{X}_i^{\top} \widehat{\mathrm{E}}_i^{-1} \dot{\mathrm{E}}_i(s) \widehat{\mathrm{E}}_i^{-1} X_i \widehat{\beta} \Big] - \frac{1}{2} tr\left( \widehat{\mathrm{E}}_i^{-1} \dot{\mathrm{E}}_i(s) \right),
\end{aligned}$$

where $\dot{\mathbf{D}}(r) = \frac{\partial \mathbf{D}}{\partial \alpha_r}|_{\alpha = \widehat{\alpha}}$, $r = 1, \ldots, dim(\alpha)$; and $\dot{\mathrm{E}}_i(s) = \frac{\partial \mathrm{E}_i}{\partial \phi_s}|_{\phi = \widehat{\phi}}$, $s = 1, 2$. For the DEC structure we have that

$$\begin{aligned}
\frac{\partial \mathrm{E}_i}{\partial \phi_1} &= |t_{ij} - t_{ik}|^{\phi_2} \phi_1^{|t_{ij} - t_{ik}|^{\phi_2} - 1}, \\
\frac{\partial \mathrm{E}_i}{\partial \phi_2} &= |t_{ij} - t_{ik}|^{\phi_2} \log(|t_{ij} - t_{ik}|) \log(\phi_1) \phi_1^{|t_{ij} - t_{ik}|^{\phi_2}}.
\end{aligned}$$

## 5 Prediction of future observations

The problem related to the prediction of future values has a great impact in many practical applications. Rao (1987) pointed out that the predictive accuracy of future observations can be taken as an alternative measure of "goodness-of-fit". In order to propose a strategy to generate predicted values from the DEC-LMEC model, we use the approach proposed by Wang (2013). Thus, let $\mathbf{y}_{i,obs}$ be an observed response vector of dimension $n_{i,obs} \times 1$ for a new subject $i$ over the first portion of time and $\mathbf{y}_{i,pred}$ be the corresponding $n_{i,pred} \times 1$ response vector over the future portion of time. Let $\bar{\mathbf{X}}_i = (\mathbf{X}_{i,obs}, \mathbf{X}_{i,pred})$ be the $(n_{i,obs} + n_{i,pred}) \times p$ design matrix corresponding to $\bar{\mathbf{y}}_i = (\mathbf{y}_{i,obs}^{\top}, \mathbf{y}_{i,pred}^{\top})$.

To deal with the censored values existing in $\mathbf{y}_{i,obs}$, we use the imputation procedure, by replacing the censored values by $\widehat{\mathbf{y}}_i = E\{\mathbf{y}_i | \mathbf{V}_i, \mathbf{C}_i, \widehat{\theta}\}$ obtained from the EM algorithm. Therefore, when the censored values are imputed, a complete data set, denoted by $\mathbf{y}_{i,obs^*}$, is obtained. The reason to use the imputation procedure is that it avoids computing truncated conditional expectations of the multivariate normal distribution originated by the censoring scheme. Hence, we have that

$$\bar{\mathbf{y}}_i^* = \left( \mathbf{y}_{i,obs^*}^{\top}, \mathbf{y}_{i,pred}^{\top} \right)^{\top} \sim N_{n_{i,obs} + n_{i,pred}} \left( \mathbf{X}_i \beta, \Sigma_i \right),$$

where the matrix $\Sigma_i$, can be represented by $\Sigma_i = \begin{pmatrix} \Sigma_i^{obs^*,obs^*} & \Sigma_i^{obs^*,pred} \\ \Sigma_i^{pred,obs^*} & \Sigma_i^{pred,pred} \end{pmatrix}$. As mentioned in Wang (2013), the best linear predictor of $\mathbf{y}_{i,pred}$ with respect to the minimum mean squared error (MSE) criterion is the conditional expectation of $\mathbf{y}_{i,pred}$ given $\mathbf{y}_{i,obs^*}$, which is given by:

$$\widehat{\mathbf{y}}_{i,pred}(\theta) = \mathbf{X}_{i,pred}\beta + \Sigma_i^{pred,obs^*}\Sigma_i^{obs^*,obs^*-1}\left(\mathbf{y}_{i,obs^*} - \mathbf{X}_{i,obs^*}\beta\right). \tag{16}$$

Therefore, $\mathbf{y}_{i,pred}$ can be estimated directly by substituting $\widehat{\theta}$ into (16), leading to $\widehat{\widehat{\mathbf{y}}_{i,pred}} = \widehat{\mathbf{y}}_{i,pred}(\widehat{\theta})$.

## 6 The nonlinear case

As mentioned in the Introduction, some approximations based on the EM algorithm have been proposed in the statistical literature to deal with NLME models. In this context, we use an approximation of the nonlinear functions mentioned by Vaida and Liu (2009). It is important to stress that this approximation (18) was considered by Matos et al (2013a) in the context of censored nonlinear mixed effects models. In that paper, simulation studies revealed that the approximation can efficiently estimate the model parameters. Recently, Wang (2013) used this approximation to implement an ECM algorithm to carry out ML estimation in Student'-$t$ nonlinear mixed-effects models for multi-outcome longitudinal data with missing values. Consequently, we conclude that this approximation is robust, stable, and does not anticipate any severe consequences in inference when applied to other types of (censored) nonlinear models.

The NLME (without censoring) of Pinheiro and Bates (2000) is defined as:

$$\mathbf{y}_i = \eta(\psi_i, \mathbf{X}_i) + \varepsilon_i,\ \ \psi_i = \mathbf{A}_i\beta + B_i\mathbf{b}_i,\ \ i = 1,\ldots,n, \tag{17}$$

where $\mathbf{b}_i \overset{iid}{\sim} \mathrm{N}_q(0,\mathbf{D})$ and $\varepsilon_i \overset{ind}{\sim} \mathrm{N}_{n_i}(0,\sigma^2\mathbf{E}_i)$ are independent; $\mathbf{y}_i$ is an $(n_i \times 1)$ vector of observed responses for subject $i$; $\eta$ is a nonlinear function of the individual random parameter $\psi_i$; $\mathbf{A}_i$ and $\mathbf{B}_i$ are known design matrices of dimensions $r \times p$ and $r \times q$, respectively, possibly depending on some covariate values; $\beta$ 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 $\theta = (\beta^\top, \sigma^2, \alpha^\top, \phi^\top)^\top$ involves taking the first-order Taylor expansion of $\eta_i$ around the current parameter estimate $\widetilde{\beta}$ and the random effect estimates $\widetilde{\mathbf{b}}_i$ (empirical predictors). This procedure is equivalent to iteratively solving the following LME model (L-step):

$$\widetilde{\mathbf{Y}}_i = \widetilde{\mathbf{W}}_i\beta + \widetilde{\mathbf{H}}_i\mathbf{b}_i + \varepsilon_i,\ \ i = 1,\ldots,n, \tag{18}$$

where $\mathbf{b}_i \overset{iid}{\sim} \mathrm{N}_q(0,\mathbf{D})$ and $\varepsilon_i \overset{ind}{\sim} \mathrm{N}_{n_i}(\mathbf{0},\sigma^2\mathrm{E}_i)$; and $\widetilde{\mathbf{Y}}_i = \mathbf{Y}_i - \eta(\psi(\widetilde{\beta},\widetilde{\mathbf{b}}_i),\mathbf{X}_i)$, with

$$\widetilde{\mathbf{H}}_i = \frac{\partial\eta(\mathbf{A}_i\beta + \mathbf{B}_i\mathbf{b}_i,\mathbf{X}_i)}{\partial\mathbf{b}_i^\top}\Big|_{\mathbf{b}_i=\widetilde{\mathbf{b}}_i} \quad \text{and} \quad \widetilde{\mathbf{W}}_i = \frac{\partial\eta(\mathbf{A}_i\beta + \mathbf{B}_i\mathbf{b}_i,\mathbf{X}_i)}{\partial\beta_i^\top}\Big|_{\beta_i=\widetilde{\beta}_i}.$$

Thus, in the censored case, the model in (18) is an LME with censored data that can be fitted using the strategy explained in Section 4. The model matrices in (18) 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.

## 7 Analysis of case studies

This section illustrates the performance of the proposed methods with the analysis of two HIV datasets, previously analyzed by Wu (2002) and Vaida and Liu (2009), respectively.

### 7.1 ACTG 315 data

Here we reanalyze the HIV viral load data from clinical trial ACTG 315 Wu (2002), considering four different correlation structures, namely the uncorrelated structure (UNC), damped exponential correlation (DEC), continuous-time autoregressive of order 1 (AR(1)) and compound symmetric structure (SYM). As mentioned in Section 2, the dataset consists of 46 HIV-1 infected patients treated with a potent ARV therapy. The viral load was repeatedly quantified on days 0, 2, 7, 10, 14, 21, 28, 56, 84, 168, and 196 after start of treatment, with a total of 361 observations. The viral load detectable limit is 100 copies/mL, and 40 out of 361 (11%) of all viral load measurements are below the detection limit. Wu and Ding (1999) proposed the use of a biphasic model:

$$V(t) = e^{\varphi_1 - \varphi_2 t} + e^{\varphi_3 - \varphi_4 t}, \tag{19}$$

where $V(t)$ is the viral load at time $t$. The parameters $\varphi_2$ and $\varphi_4$ are called the first and second phase viral decay rates, which can represent the minimum turnover rate of productively infected cells and that of latently or long-lived infected cells, respectively. The parameters $\varphi_1$ and $\varphi_3$ are macro-parameters and $e^{\varphi_1} + e^{\varphi_3}$ is the baseline viral load at time $t = 0$.

As noted by Wu and Ding (1999), the inter-subject variation of observed viral loads motivates the use of a NLME model. The viral load trajectories initially exhibit rapid decay (known as first-phase decay), followed by a phase of slow decay for some (the second-phase) with the others rebounding back to the original levels Liu and Wu (2012). Therefore, following Wu (2002) we consider the following NLME model to reflect the dynamics of the HIV viral load:

$$y_{ij} = \log_{10}(e^{\varphi_{1i} - \varphi_{2i} t_{ij}} + e^{\varphi_{3i} - \varphi_{4ij} t_{ij}}) + \varepsilon_{ij}, \tag{20}$$

$$\beta_{1ij} = \varphi_{1i} = \beta_1 + b_{1i},\ \beta_{3ij} = \varphi_{3i} = \beta_3 + b_{3i}, \tag{21}$$

$$\beta_{2ij} = \varphi_{2i} = \beta_2 + b_{2i},\ \beta_{4_{ij}} = \varphi_{4ij} = \beta_4 + \beta_5 \mathrm{CD4}_{ij} + b_{4i}, \tag{22}$$

where $y_{ij}$ is the $\log_{10}$-transformation of the viral load for the $i$th subject at time $t_{ij}$ ($i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n_i$) and $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{in_i})^\top$ represents the vector of within-individual random errors; $\mathrm{CD4}_{ij}$ indicates the observed CD4 values up to time $t_{ij}$; $\beta_{ij} = (\beta_{1ij}, \beta_{2ij}, \beta_{3ij}, \beta_{4ij})^\top$ and $\beta = (\beta_1, \ldots, \beta_5)^\top$ are individual parameters for the $i$-th subject at time $t_{ij}$ and population parameters, respectively and $\mathbf{b}_i = (b_{1i}, \ldots, b_{4i})^\top$ is the random effects vector for subject $i$.

**Table 1** ACTG 315 data. Model selection criteria for the NLMEC model under different correlation structures.

| | NLMEC | | | |
|---|---|---|---|---|
| Criteria | UNC | DEC | AR(1) | SYM |
| $\ell_{max}$ | -281.31 | **-255.83** | -264.99 | -279.33 |
| AIC | 594.61 | **547.66** | 563.97 | 592.66 |
| AIC corr | 596.19 | **549.66** | 565.76 | 594.45 |
| BIC | 656.83 | **617.66** | 630.08 | 658.77 |

The values of $\ell_{max}$, AIC and BIC for the four considered models are presented in Table 1. Note that, based on these criteria, the model presenting the best fit is the model with a damped exponential correlation structure (DEC). **Further, the likelihood ratio test (LRT) for the hypothesis $H_0 : \phi_2 = 1$ and $H_1 : \phi_2 \neq 1$ is performed. The resulting LRT statistic is 18.32 with p-value 0.00002, which is significant compared to $\chi^2_{1,0.05}$, suggesting that the DEC structure is more appropriate than the AR(1) for modeling the dependence among the within-subject errors. Figure 3 shows some individual profiles (in log10 scale) for HIV viral load at different follow-up times and the smooth means of residuals from model fits.**

ML estimates corresponding to the best model are presented in Table 2. Using these estimates, one can quantify the population decay rates and viral load parameters. The first- and second-phase decay rates can be approximated as $\widehat{\varphi}_2 = 31.549$ and $\widehat{\varphi}_4(t) = -0.994 + 0.612\,\mathrm{CD4}$. The population viral load process can be represented as $\widehat{V}(t) = \exp\{11.552 - \widehat{\varphi}_2(t)t\} + \exp\{6.861 - \widehat{\varphi}_4(t)t\}$. **The SE for the parameters estimates are obtained using the empirical information**

**matrix (Section 4.1). Finally, using a bootstrap procedure, one can conclude that all the fixed-effects considered in the model are statistically significant at $\alpha = 0.05$.**

**Table 2** ACTG 315 data. ML estimates with standard errors for the NLMEC model under DEC structure.

| Fixed effects | | | Between-subject variances | | | Within-subject variances | | |
|---|---|---|---|---|---|---|---|---|
| Parameter | Estimative | SE | Parameter | Estimative | SE | Parameter | Estimative | SE |
| $\beta_1$ | 11.552 | 0.266 | $\alpha_{11}$ | 0.155 | 0.045 | $\sigma^2$ | 0.407 | 0.094 |
| $\beta_2$ | 31.549 | 0.040 | $\alpha_{12}$ | -0.808 | 0.127 | $\phi_1$ | 0.188 | 0.152 |
| $\beta_3$ | 6.861 | 0.325 | $\alpha_{22}$ | 5.753 | 0.045 | $\phi_2$ | 0.647 | 0.084 |
| $\beta_4$ | -0.994 | 0.810 | $\alpha_{13}$ | 0.020 | 0.099 | | | |
| $\beta_5$ | 0.612 | 0.195 | $\alpha_{23}$ | 0.110 | 0.114 | | | |
| | | | $\alpha_{33}$ | 0.258 | 0.215 | | | |
| | | | $\alpha_{14}$ | -0.714 | 0.11 | | | |
| | | | $\alpha_{24}$ | 4.625 | 0.069 | | | |
| | | | $\alpha_{34}$ | 0.598 | 0.121 | | | |
| | | | $\alpha_{44}$ | 5.654 | 0.03 | | | |

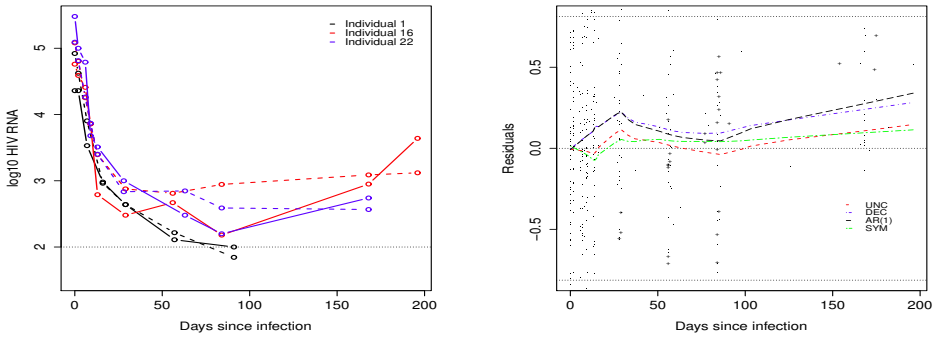


**Fig. 3** ACT315 data. (left panel) Individual profiles (in log10 scale) for HIV viral load at different follow-up times for some subjects, the dashed line are the respective fitted profile. (right panel) Smooth means of residuals from model fits. The residuals from the model with autoregressive of order 1 correlation appear as points.

## 7.2 AIEDRP data

The second case study is taken from the AIEDRP program, a large multicenter observational study of subjects with acute and early HIV infection, which consist of 320 untreated individuals with acute HIV infection. Of the 830 recorded observations, 185 (22%) were above the limit of assay quantification. Therefore, in the spirit of Vaida and Liu (2009), we consider a right-censored five-parameter NLME model (inverted S-shaped curve) as follows:

$$y_{ij} = \lambda_{1i} + \frac{\lambda_2}{1+\exp((t_{ij}-\lambda_3)/\lambda_4)} + \lambda_{5i}(t_{ij}-50) + \varepsilon_{ij}, \tag{23}$$

where $y_{ij}$ is the $\log_{10}$ of the viral load for subject $i$ at time $t_{ij}$. The parameters $\lambda_{1i}$ and $\lambda_2$ represent the subject-specific set-point value and decrease from the maximum HIV-1 RNA. The location parameter $\lambda_3$ indicates the time point at which half of the change in HIV-1 RNA is attained, $\lambda_4$ is a scale parameter modeling the rate of decline and $\lambda_{5i}$ allows increasing the HIV-1 RNA trajectory after day 50. The reparameterization given by $\beta_{1i} = \log(\lambda_{1i}) = \beta_1 + b_{1i}$; $\beta_k = \log(\lambda_k)$, $k = 2,3,4$, and $\lambda_{5i} = \beta_5 + b_{2i}$ is adopted to assure positive values for the model parameters.

As in Section 7.1, the correlation structures UNC, DEC, AR(1) and SYM are considered. Table 3 summarizes the values of $\ell_{max}$, AIC and BIC for all considered models. Note that the values of

**Table 3** AIEDRP data. Model selection criteria for the NLMEC model under different correlation structures.

| | LMEC | | | |
|---|---|---|---|---|
| Criteria | UNC | DEC | AR(1) | SYM |
| $\ell_{max}$ | -783.79 | -769.81 | **-770.10** | -775.62 |
| AIC | 1585.59 | 1561.63 | **1560.19** | 1571.25 |
| AIC corr | 1585.81 | 1561.95 | **1560.46** | 1571.52 |
| BIC | 1628.08 | 1613.56 | **1607.41** | 1618.46 |

$\ell_{max}$ for the DEC and AR(1) models are close. This is explained because the estimated values of $\phi_1$ and $\phi_2$ under the DEC model are 0.83 and 1.15 respectively. Based on this observation and the criteria, the best (parsimonious) fit is obtained using the continuous-time autoregressive of order 1 correlation (AR(1)). **The likelihood ratio test (LRT) for testing the hypothesis** $H_0 : \phi_2 = 1$ **and** $H_1 : \phi_2 \neq 1$ **are also performed. The resulting LRT statistic is 0.58. with p-value 0.446, which is not significant compared to** $\chi^2_{1,0.05}$**, suggesting that the AR(1) structure is more appropriate than the DEC for modeling the dependence among the within-subject errors.** Moreover, the model fit of the AR(1) (and DEC) model is slightly better than the SYM model, with the smooth mean residual curve in Figure 3 (right panel) always being closer to zero.

The ML estimates under this model are presented in Table 4. **As in the previous case, the SE for the parameters estimates are obtained using the empirical information matrix**. One can use the AR(1) model with reasonable confidence for predictions of viral load. For example, at 6 months since infection, the average viral load is 4.537 $\log_{10}$ units. The individual 6-month viral load estimates vary between 1.794 and 6.469, with 5th and 95th quantiles at 3.466 and 5.549. The average slope after day 50 is negative, $\beta_{5i} = -0.004$ $\log_{10}$HIV/day, with 95% CI(-0.006,-0.002). And, for the individual slopes $\alpha_{5i}$ the 5th and 95th quantiles are -0.0061 and -0.0015. **We performed a bootstrap procedure for hypothesis test for the significance of the fixed-effects** ($\alpha = 0.05$)**, concluding that all of them are statistically significant (different from zero).**

**Table 4** AIEDRP data. ML estimates with standard errors for the NLMEC model under AR(1) structure.

| Fixed effects | | | Between-subject variances | | | Within-subject variances | | |
|---|---|---|---|---|---|---|---|---|
| Parameter | Estimative | SE | Parameter | Estimative | SE | Parameter | Estimative | SE |
| $\beta_1$ | 1.614 | 0.011 | $\alpha_{11}$ | 0.01658 | 0.00307 | $\sigma^2$ | 0.308 | 0.024 |
| $\beta_2$ | 0.128 | 0.003 | $\alpha_{12}$ | 0.00020 | 0.00016 | $\phi_1$ | 0.808 | 0.033 |
| $\beta_3$ | 3.516 | 0.025 | $\alpha_{22}$ | 0.00003 | 0.00001 | | | |
| $\beta_4$ | 1.118 | 0.001 | | | | | | |
| $\beta_5$ | -0.004 | 0.001 | | | | | | |

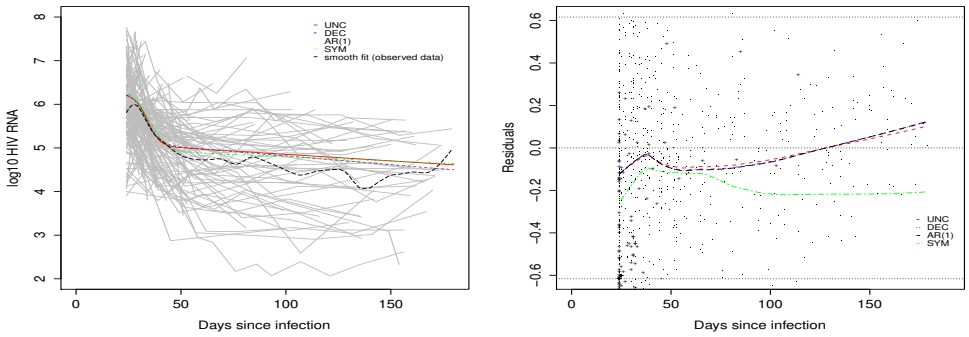


**Fig. 4** AIEDRP data. (Left panel) Individual profiles (in log10 scale) for HIV viral load at different follow-up times with the model fits. (Right panel) Smooth means of residuals from model fits. The residuals from the model with autoregressive of order 1 correlation appear as points.

## 8 Simulation Studies

**In order to examine the performance of the proposed method, here we report three simulation studies to investigate: (a) the consequences for parameter estimation (b) the behavior of the prediction when the correlation structure of the error term is misspecified, and (c) the asymptotic behavior of the parameter estimates.** For this purpose and simplicity reasons, we consider a logistic model similar to that studied in Section 7.2, with random set-points $\lambda_{1i}$ and random decline rates $\lambda_{4i}$, as follows:

$$y_{ij} = \lambda_{1i} + \frac{\lambda_2}{1+\exp((t_{ij}-\lambda_3)/\lambda_{4i})} + \varepsilon_{ij}, \tag{24}$$

with $i = 1,\ldots,100,\ j = 1,\ldots,10$, $\alpha_{1i} = \exp(\beta_1 + b_{1i})$, $\beta_k = \log(\lambda_k)$, $k = 2,3$, $\lambda_{4i} = \exp(\beta_4 + b_{2i})$, $(b_{1i}, b_{2i}) \overset{ind.}{\sim} \mathrm{N}_2(\mathbf{0},\mathbf{D})$, and $\varepsilon_{ij} \overset{ind.}{\sim} \mathrm{N}_{n_i}(\mathbf{0},\Omega_i)$.

The parameters are set at $\beta = (1.6094, 0.6931, 3.8067, 2.3026)^\top$, $\sigma^2 = 0.55$, and $\mathbf{D}$ with elements $\alpha_{11} = 0.05$ and $\alpha_{22} = 0.1$.

For the first study, we simulated several datasets considering different values of the parameter $\phi_1$ under the correlation structure AR(1), with the aim to discover the effect of the correlation level on the estimation. For each value of $\phi_1$, we simulated 100 datasets. In addition, we considered 5% and 20% of censored observations for each value of $\phi_1$. Once the simulated datasets were generated, we fitted the proposed model assuming the uncorrelated (UNC) and AR(1) structures. The model selection criteria (AIC and BIC) as well as the estimates of the model parameters were stored for each simulation. **Summary statistics such as the mean estimate (MC mean), the mean of the approximate standard error obtained through the information-based method described in Section 4.1 (IM SE), the empirical standard error (MC Sd) and the coverage probability at 95% (MC CP) are presented in Tables 5 and 6.**

From the results shown in Tables 5 and 6, one can observe that when the AR(1) is chosen as the true model, the MC CP values are higher than those obtained under the uncorrelated model, even when the correlation parameter $\phi_1$ is small (0.3). Moreover, the biases of fixed effects estimates under the AR(1) structure are lower than those obtained under the uncorrelated structure (see Figures 5 and 6) for different values of the $\phi_1$ parameter. The model selection criteria chose the true model (AR(1)) for moderate values of the $\phi_1$ parameter (greater than 0.5) for the two levels of censoring considered.

The second simulation study analyzes the performance of the prediction of future values described in Section 5. For this purpose, we compared the prediction of the NLMEC model in (24) under the UNC and AR(1) structures. As in the first study, we generated 100 datasets of size $n = 100$ under AR(1) structure with parameter $\phi_1 = 0.9$, considering two different settings of censoring proportions, say 5% and 20%. For the prediction, we excluded the last two measurements of each simulated individual in the datasets. To compare the performance of the prediction, we considered two empirical discrepancy measures, namely the MAE (mean absolute error) and MSE (mean square error). These measures are given by:

$$\mathrm{MAE} = \frac{1}{200}\sum_{i,j}|y_{ij} - y^\star_{ij}| \quad \text{and} \quad \mathrm{MSE} = \frac{1}{200}\sum_{i,j}(y_{ij} - y^\star_{ij})^2, \tag{25}$$

where $y_{ij}$ is the original value and $y^\star_{ij}$ is the predicted value, for $i = 1,\ldots,100$ and $j = 1,\ldots,2$. Table 7 shows the comparison between the predicted values and real ones under the NLMEC model considering the UNC and AR(1) structures. One can see from these results that the model with AR(1) structure generates predictive values close to the real ones.

**Finally, we analyzed the absolute bias (Bias) and mean square error (MSE) of the fixed-effects and variance components estimates obtained from the DEC-LMEC model with different sample sizes. The idea of this simulation is to provide empirical evidence about the consistency of the ML estimates. The bias and MSE measures are defined as:**

$$\textbf{Bias} = \frac{1}{J}\sum_{j=1}^{J}|\widehat{\theta}_i^{(j)} - \theta_i| \ \textbf{ and } \ \textbf{MSE} = \frac{1}{J}\sum_{j=1}^{J}\left(\widehat{\theta}_i^{(j)} - \theta_i\right)^2, \tag{26}$$

**where $\widehat{\theta}_i^{(j)}$ is the ML estimate of the parameter $\theta_i$ for the $j$-th sample, $j = 1,\ldots,J$.**

**Table 5** 5% censored. Summary statistics based on 100 simulated AR(1) samples.

| | | | Parameter estimates | | | | | Criteria | |
|---|---|---|---|---|---|---|---|---|---|
| $\phi_1$ | Corr. Structure | | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\sigma^2$ | MC AIC | MC BIC |
| 0.3 | UNC | MC Mean | 1.67 | 0.53 | 3.73 | 2.11 | 0.55 | 3020 | 3056 |
| | | IM SE | 0.04 | 0.11 | 0.05 | 0.20 | | | |
| | | MC Sd | 0.02 | 0.07 | 0.03 | 0.15 | | | |
| | | MC CP | 79% | 78% | 66% | 89% | | | |
| | AR(1) | MC Mean | 1.61 | 0.71 | 3.83 | 2.27 | 0.55 | 3024 | 3065 |
| | | IM SE | 0.12 | 0.22 | 0.10 | 0.26 | | | |
| | | MC Sd | 0.10 | 0.22 | 0.10 | 0.27 | | | |
| | | MC CP | 84% | 88% | 94% | 91% | | | |
| 0.5 | UNC | MC Mean | 1.66 | 0.54 | 3.74 | 2.12 | 0.55 | 3015 | 3050 |
| | | IM SE | 0.04 | 0.11 | 0.05 | 0.20 | | | |
| | | MC Sd | 0.02 | 0.07 | 0.03 | 0.16 | | | |
| | | MC CP | 82% | 80% | 67% | 91% | | | |
| | AR(1) | MC Mean | 1.60 | 0.71 | 3.83 | 2.27 | 0.55 | 3018 | 3058 |
| | | IM SE | 0.12 | 0.23 | 0.10 | 0.26 | | | |
| | | MC Sd | 0.11 | 0.23 | 0.10 | 0.27 | | | |
| | | MC CP | 84% | 88% | 93% | 91% | | | |
| 0.6 | UNC | MC Mean | 1.66 | 0.56 | 3.74 | 2.16 | 0.54 | 3004 | 3039 |
| | | IM SE | 0.04 | 0.11 | 0.05 | 0.20 | | | |
| | | MC Sd | 0.02 | 0.08 | 0.03 | 0.17 | | | |
| | | MC CP | 82% | 86% | 69% | 92% | | | |
| | AR(1) | MC Mean | 1.60 | 0.71 | 3.83 | 2.27 | 0.55 | 3003 | 3044 |
| | | IM SE | 0.13 | 0.23 | 0.10 | 0.26 | | | |
| | | MC Sd | 0.12 | 0.23 | 0.10 | 0.28 | | | |
| | | MC CP | 84% | 88% | 93% | 91% | | | |
| 0.7 | UNC | MC Mean | 1.65 | 0.62 | 3.73 | 2.27 | 0.52 | 2978 | 30134 |
| | | IM SE | 0.04 | 0.11 | 0.05 | 0.19 | | | |
| | | MC Sd | 0.02 | 0.09 | 0.03 | 0.18 | | | |
| | | MC CP | 90% | 91% | 68% | 94% | | | |
| | AR(1) | MC Mean | 1.59 | 0.72 | 3.84 | 2.27 | 0.55 | 2962 | 3002 |
| | | IM SE | 0.20 | 0.25 | 0.11 | 0.28 | | | |
| | | MC Sd | 0.17 | 0.25 | 0.11 | 0.29 | | | |
| | | MC CP | 84% | 88% | 91% | 92% | | | |
| 0.8 | UNC | MC Mean | 1.62 | 0.75 | 3.74 | 2.51 | 0.47 | 2912 | 2948 |
| | | IM SE | 0.04 | 0.10 | 0.04 | 0.17 | | | |
| | | MC Sd | 0.03 | 0.08 | 0.05 | 0.17 | | | |
| | | MC CP | 99% | 96% | 50% | 76% | | | |
| | AR(1) | MC Mean | 1.60 | 0.72 | 3.84 | 2.27 | 0.55 | 2840 | 2881 |
| | | IM SE | 0.17 | 0.25 | 0.11 | 0.29 | | | |
| | | MC Sd | 0.14 | 0.25 | 0.11 | 0.30 | | | |
| | | MC CP | 84% | 88% | 93% | 91% | | | |
| 0.9 | UNC | MC Mean | 1.60 | 0.90 | 3.73 | 2.77 | 0.36 | 2673 | 2708 |
| | | IM SE | 0.03 | 0.07 | 0.03 | 0.14 | | | |
| | | MC Sd | 0.04 | 0.09 | 0.06 | 0.16 | | | |
| | | MC CP | 95% | 17% | 13% | 12% | | | |
| | AR(1) | MC Mean | 1.61 | 0.70 | 3.83 | 2.26 | 0.53 | 2453 | 2493 |
| | | IM SE | 0.12 | 0.21 | 0.09 | 0.26 | | | |
| | | MC Sd | 0.11 | 0.22 | 0.10 | 0.28 | | | |
| | | MC CP | 83% | 88% | 94% | 91% | | | |

**The censoring proportion was fixed at** $10\%$ **and different sample sizes were considered, say,** $n = 50, 100, 200, 400$ **and** $600$**. Also we considered** $J = 100$**,** *i.e.* **we simulated** $100$ **samples of size** $n$**. For this simulation, an AR (1) structure with parameter** $\phi_1 = 0.8$ **was considered.**

**Figures 7 and 8 show that the MSE of the parameter estimates of** $\beta$**,** $\sigma^2$ **and** $\alpha$ **tends to zero as the sample size increase. Note that, similar results are obtained after the analysis of the absolute bias. In conclusion, the results provide empirical evidence about the consistency of the ML estimates of the DEC-LMEC model even considering the linearization procedure described in (18).**

## 9 Conclusions

The paper proposes a mixed effects model with censored observations based on the multivariate normal distribution. A DEC structure as proposed by Muñoz et al (1992) to model the autocorrelation existing among irregularly observed measures was adopted. **This structure is flexible,**

**Table 6** 20% censored. Summary statistics based on 100 simulated AR(1) samples.

| | | | Parameter estimates | | | | | Criteria | |
|---|---|---|---|---|---|---|---|---|---|
| $\phi_1$ | Corr. Structure | | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\sigma^2$ | MC AIC | MC BIC |
| 0.3 | UNC | MC Mean | 1.67 | 0.50 | 3.72 | 2.08 | 0.55 | 2796 | 2832 |
| | | IM SE | 0.04 | 0.12 | 0.05 | 0.21 | | | |
| | | MC Sd | 0.02 | 0.07 | 0.03 | 0.16 | | | |
| | | MC CP | 68% | 69% | 63% | 87% | | | |
| | AR(1) | MC Mean | 1.59 | 0.72 | 3.84 | 2.28 | 0.55 | 2800 | 2841 |
| | | IM SE | 0.23 | 0.26 | 0.11 | 0.27 | | | |
| | | MC Sd | 0.19 | 0.26 | 0.12 | 0.28 | | | |
| | | MC CP | 87% | 90% | 92% | 91% | | | |
| 0.5 | UNC | MC Mean | 1.67 | 0.51 | 3.72 | 2.09 | 0.54 | 2791 | 2827 |
| | | IM SE | 0.04 | 0.12 | 0.05 | 0.21 | | | |
| | | MC Sd | 0.02 | 0.07 | 0.03 | 0.16 | | | |
| | | MC CP | 76% | 73% | 63% | 87% | | | |
| | AR(1) | MC Mean | 1.60 | 0.71 | 3.83 | 2.27 | 0.55 | 2794 | 2835 |
| | | IM SE | 0.17 | 0.25 | 0.11 | 0.27 | | | |
| | | MC Sd | 0.15 | 0.25 | 0.11 | 0.27 | | | |
| | | MC CP | 87% | 90% | 92% | 91% | | | |
| 0.6 | UNC | MC Mean | 1.67 | 0.52 | 3.72 | 2.12 | 0.54 | 2781 | 2816 |
| | | IM SE | 0.04 | 0.12 | 0.05 | 0.21 | | | |
| | | MC Sd | 0.02 | 0.08 | 0.03 | 0.18 | | | |
| | | MC CP | 78% | 77% | 64% | 88% | | | |
| | AR(1) | MC Mean | 1.60 | 0.71 | 3.83 | 2.27 | 0.55 | 2780 | 2821 |
| | | IM SE | 0.19 | 0.26 | 0.11 | 0.28 | | | |
| | | MC Sd | 0.16 | 0.25 | 0.11 | 0.28 | | | |
| | | MC CP | 87% | 90% | 91% | 92% | | | |
| 0.7 | UNC | MC Mean | 1.66 | 0.58 | 3.72 | 2.22 | 0.52 | 2757 | 2792 |
| | | IM SE | 0.04 | 0.12 | 0.05 | 0.20 | | | |
| | | MC Sd | 0.02 | 0.09 | 0.04 | 0.20 | | | |
| | | MC CP | 85% | 87% | 67% | 92% | | | |
| | AR(1) | MC Mean | 1.59 | 0.72 | 3.83 | 2.27 | 0.55 | 2742 | 2783 |
| | | IM SE | 0.20 | 0.27 | 0.12 | 0.29 | | | |
| | | MC Sd | 0.17 | 0.26 | 0.11 | 0.28 | | | |
| | | MC CP | 86% | 90% | 91% | 94% | | | |
| 0.8 | UNC | MC Mean | 1.63 | 0.71 | 3.72 | 2.48 | 0.48 | 2703 | 2739 |
| | | IM SE | 0.04 | 0.11 | 0.04 | 0.18 | | | |
| | | MC Sd | 0.03 | 0.09 | 0.06 | 0.21 | | | |
| | | MC CP | 96% | 99% | 45% | 80% | | | |
| | AR(1) | MC Mean | 1.62 | 0.68 | 3.82 | 2.24 | 0.55 | 2637 | 2677 |
| | | IM SE | 0.11 | 0.23 | 0.10 | 0.29 | | | |
| | | MC Sd | 0.10 | 0.21 | 0.09 | 0.27 | | | |
| | | MC CP | 86% | 89% | 93% | 94% | | | |
| 0.9 | UN | MC Mean | 1.61 | 0.85 | 3.72 | 2.73 | 0.36 | 2484 | 2520 |
| | | IM SE | 0.03 | 0.07 | 0.03 | 0.16 | | | |
| | | MC Sd | 0.03 | 0.10 | 0.07 | 0.16 | | | |
| | | MC CP | 98% | 43% | 24% | 26% | | | |
| | AR(1) | MC Mean | 1.62 | 0.67 | 3.81 | 2.21 | 0.53 | 2290 | 2331 |
| | | IM SE | 0.09 | 0.20 | 0.09 | 0.26 | | | |
| | | MC Sd | 0.08 | 0.20 | 0.09 | 0.25 | | | |
| | | MC CP | 86% | 89% | 95% | 94% | | | |

**Table 7** Evaluation of the prediction accuracy for the NLMEC model with different correlation structures.

| | 5% censored | | 20% censored | |
|---|---|---|---|---|
| Corr. Structure | MAE | MSE | MAE | MSE |
| UNC | 0.5507 | 0.4739 | 0.6418 | 0.6746 |
| AR(1) | 0.5169 | 0.4299 | 0.6073 | 0.6165 |

**since the parameter $\phi_1$ describes the autocorrelation between observations separated by the absolute length of two time points, and the parameter $\phi_2$ permits acceleration of the exponential decay of the autocorrelation function, defining a continuous-time autoregressive model.** An ECM algorithm to obtain the ML estimates was developed by using the statistical properties of the multivariate truncated normal distribution. The proposed algorithm has a closed-form expression for the E-step, based on the first two moments of the truncated normal distribution. **In this context, it is important to stress that the DEC structure can be easily implemented using**

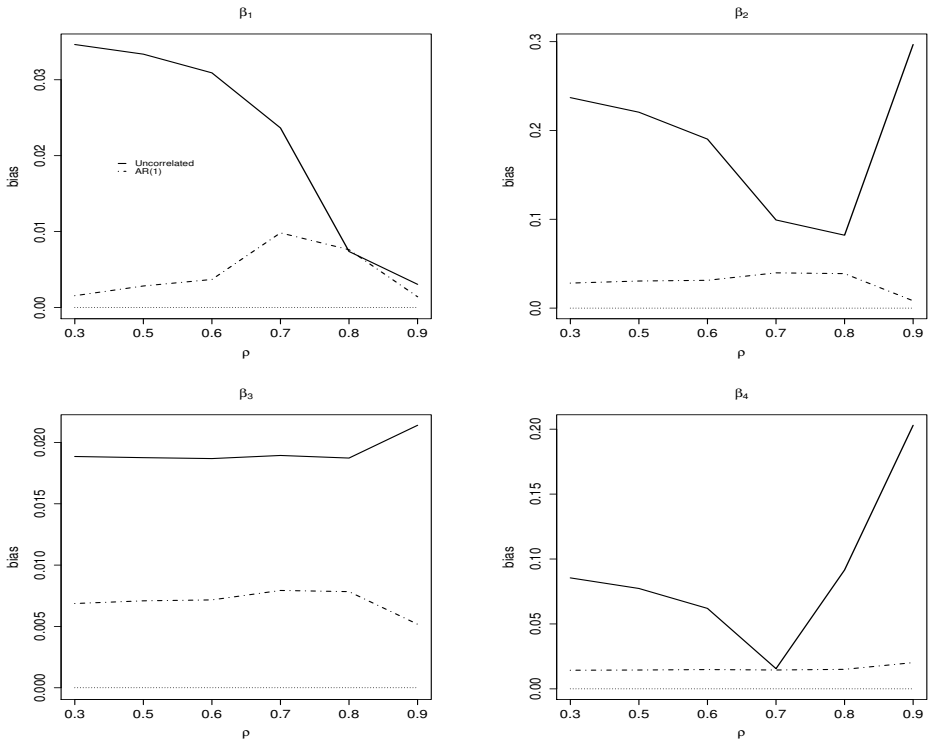


**Fig. 5** Simulation study. 5% censored. Bias of $\beta$ estimates under the uncorrelated and AR(1) models for 6 different values of $\phi_1$.

**the exact EM algorithm, making the proposed approach easy to implement by practitioners. The R codes are available upon request**. The proposed methods were applied to two AIDS case studies and a simulation study was performed, showing the effects of misspecification on the correlation structure over the fixed effects estimates. **Better results were generated than the uncorrelated structure in terms of estimation and prediction.**

Although the LMEC/NLMEC models showed great flexibility to model symmetric data, they can be seriously affected by the presence of outliers. Recently, Garay et al (2014) proposed a remedy to accommodate outliers using a Student'-$t$ regression model with DEC structure. Our methods can be extended by considering the Student's $t$ in the context of LMEC/NLMEC models as in Matos et al (2015), providing satisfactory results at the expense of additional complexity in implementation. Further, it is also of interest to develop an effective Markov chain Monte Carlo algorithm for the DEC-LMEC/NLMEC in a fully Bayesian treatment.

**Acknowledgements** We thank the editor, associate editor and three referees whose constructive comments led to an improved presentation of the paper. L. Matos acknowledges support from FAPESP-Brazil (Grant 2011/22063-9). The research of L.M. Castro was partially supported by CONICYT-Chile through BASAL project CMM, Universidad de Chile and Grant FONDECYT 1130233 from the Chilean government. V.H. acknowledges support from CNPq-Brazil (Grant 305054/2011-2) and FAPESP-Brazil (Grant 2014/02938-9).

## References

Antunes R, Figueiredo S, Bártolo I, Pinheiro M, Rosado L, Soares I, Lourenço H, Taveira N (2003) Evaluation of the clinical sensitivities of three viral load assays with plasma samples from a pediatric population predominantly infected with human inmunodeficiency virus type 1 subtype G and BG recombinants forms. Journal of Clinical Microbiology 41(7):3361–3367

Arellano-Valle RB, Castro L, González-Farías G, Muños Gajardo K (2012) Student-t censored regression model: properties and inference. Statistical Methods and Applications 21(4):453–473

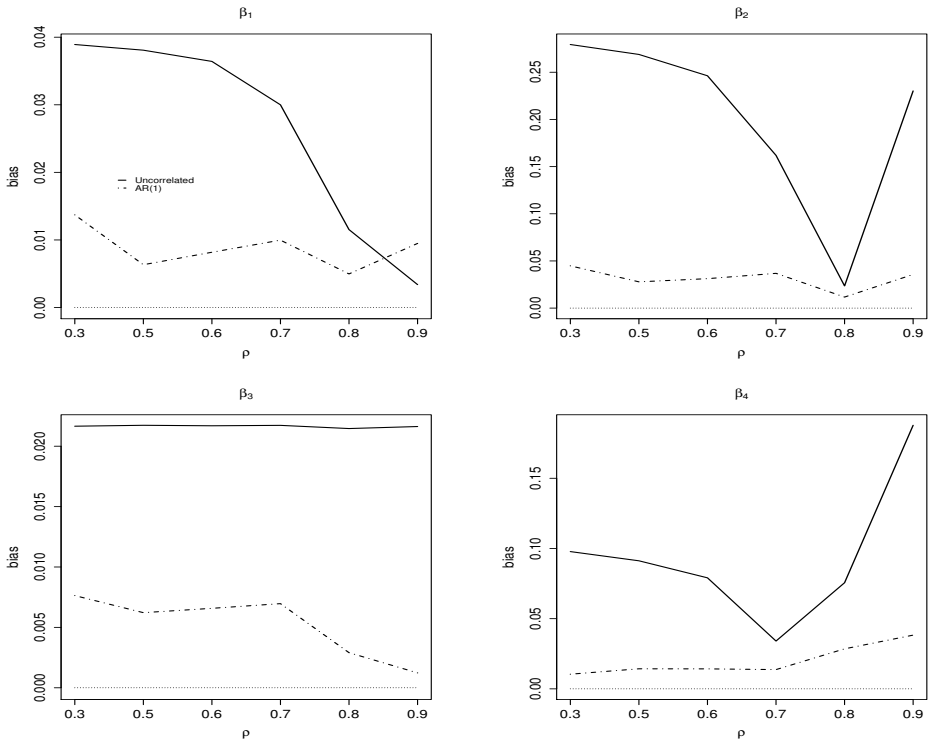


**Fig. 6** Simulation study. 20% censored. Bias of $\beta$ estimates under the uncorrelated and AR(1) models for 6 different values of $\phi_1$.

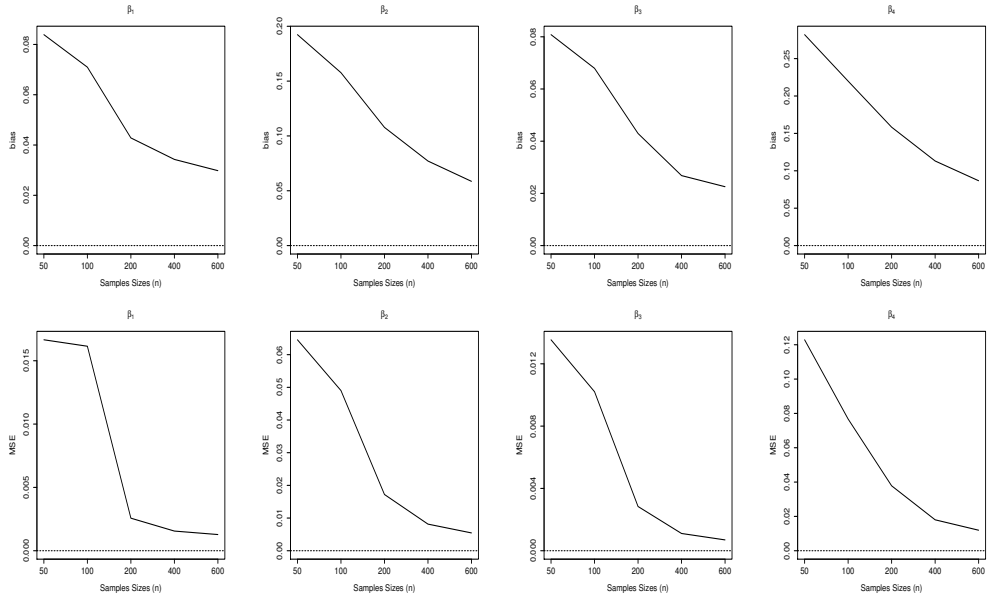


**Fig. 7** Simulation study. 10% censored. Bias and MSE of $\beta$ estimates under the AR(1) model for different sample sizes

Lopes de Azevedo K, Setúbal S, Silami V, Bastos L, Artimos de Oliveira S (2010) Congenital toxoplasmosis transmitted by human immunodeficiency-virus infected women. Brazilian Journal of Infectious Diseases 14(2):186–189

Ciesielski C, Metler R (1997) Duration of time between exposure and seroconversion in healthcare workers with occupationally acquired infection with human immunodeficiency virus. American

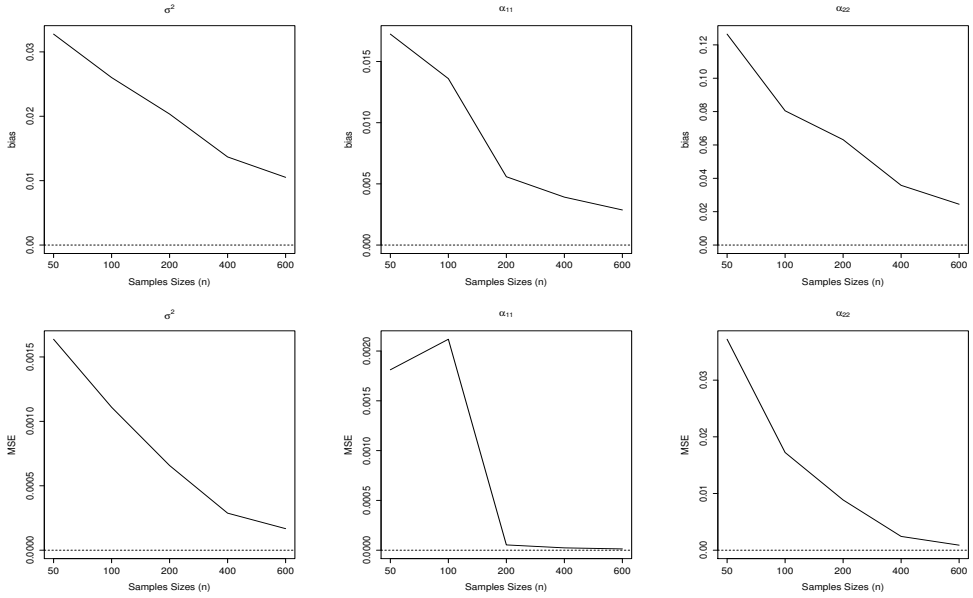


**Fig. 8** Simulation study. 10% censored. Bias and MSE of $\sigma^2$ and $\alpha$ estimates under the AR(1) model for different sample sizes

Journal of Medicine 102(5):115–116

Dempster A, Laird N, Rubin D (1977) Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B, 39:1–38

Deylon B, Lavielle M, Moulines E (1999) Convergence of a stochastic approximation version of the em algorithm. The Annals of Statistics 27(1):94–128

Garay A, Castro L, Leskow J, Lachos VH (2014) Censored linear regression models for irregulary observed longitudinal data using the multivariate-*t* distribution. Statistical Methods in Medical Research DOI: 10.1177/0962280214551191

Hughes J (1999) Mixed effects models with censored data with application to HIV RNA levels. Biometrics 55:625–629

Levine R, Casella G (2001) Implementations of the Monte Carlo EM algorithm. Journal of Computational and Graphical Statistics 10(3):422–439

Lin TI (2010) Robust mixture modeling using multivariate skew t distributions. Statistics and Computing 20(3):343–356

Lin TI, Wang WL (2013) Multivariate skew-normal linear mixed models for multi-outcome longitudinal data. Statistical Modelling 13(3):199–221

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(5):963–978

Louis TA (1982) Finding the observed information matrix when using the em algorithm. Journal of the Royal Statistical Society Series B (Methodological) pp 226–233

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(1):450–464

Matos LA, Prates MO, Chen MH, Lachos VH (2013b) Likelihood-based inference for mixed-effects models with censored response using the multivariate-t distribution. Statistica Sinica 23:1323–1342

Matos LA, Bandyopadhyay D, Castro LM, Lachos VH (2015) Influence assessment in censored mixed-effects models using the multivariate Student′s-*t* distribution. Journal of multivariate analysis 141:104–117

Meilijson I (1989) A fast improvement to the em algorithm on its own terms. Journal of the Royal Statistical Society Series B (Methodological) pp 127–138

Meng XL, Rubin DB (1993) Maximum likelihood estimation via the ECM algorithm: A general framework. Biometrika 80(2):267–278

Muñoz A, Carey V, Schouten JP, Segal M, Rosner B (1992) A parametric family of correlation structures for the analysis of longitudinal data. Biometrics 48:733–742

Müller P, Van de Geer S (2015) Censored linear model in high dimensions. Test DOI: 10.1007/s11749-015-0441-7

Pinheiro JC, Bates DM (2000) Mixed-Effects Models in S and S-PLUS. Springer, New York, NY

Rao CR (1987) Prediction of future observations in growth curve models. Statistical Science 2:434–447

Rocha G, Arellano-Valle RB, Loschi R (2015) Maximum likelihood methods in a robust censored error-in-variables model. Test DOI: 10.1007/s11749-015-0439-1

Samson A, Lavielle M, Mentré F (2006) Extension of the saem algorithm to left-censored data in nonlinear mixed-effects model: application to HIV dynamics model. Computational Statistics & Data Analysis 51(3):1562–1574

Swenson L, Cobb B, Geretti A, Harrigan P, Poljak M, Seguin-Devaux C, Verhofstede C, Wirden M, Amendola A, Boni J, Bourlet T, Huder J, Karasi J, Lepej S, Lunar M, Mukabayire O, Schuurman R, Tomažič J, Van Laethem K, Vandekerckhove L, Wensing A (2014) Comparative performances of HIV-1 RNA load assays at low viral load levels: results of an international collaboration. Journal of Clinical Microbiology 52(2):517–523

Tobin J (1958) Estimation of relationships for limited dependent variables. Econometrica 26:24–36

Vaida F, Liu L (2009) Fast implementation for normal mixed Effects models with censored response. Journal of Computational and Graphical Statistics 18:797–817

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

Wang WL (2013) Multivariate t linear mixed models for irregularly observed multiple repeated measures with missing outcomes. Biometrical Journal 55(4):554–571

Wang WL, Fan TH (2011) Estimation in multivariate t linear mixed models for multivariate longitudinal data. Statistica Sinica 21:1857–1880

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(2):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(460):955–964

Wu L (2010) Mixed Effects Models for Complex Data. Chapman & Hall/CRC, Boca Raton, FL