Statistica Sinica **23** (2013), 1323-1345
doi:http://dx.doi.org/10.5705/ss.2012.043

LIKELIHOOD-BASED INFERENCE FOR MIXED-EFFECTS MODELS WITH CENSORED RESPONSE USING THE MULTIVARIATE-t DISTRIBUTION

Larissa A. Matos[1], Marcos O. Prates[2], Ming-Hui Chen[3] and Victor H. Lachos[1]

[1] *Universidade Estadual de Campinas,* [2] *Universidade Federal de Minas Gerais and* [3] *University of Connecticut*

Abstract: Mixed-effects models are commonly used to fit longitudinal or repeated measures data. A complication arises when the response is censored, for example, due to limits of quantification of the assay used. Although normal distributions are commonly assumed for random effects and residual errors, such assumptions make inferences vulnerable to outliers. The sensitivity to outliers and the need for heavy tailed distributions for random effects and residual errors motivate us to develop a likelihood-based inference for linear and nonlinear mixed effects models with censored response (NLMEC/LMEC) based on the multivariate Student-t distribution. An ECM algorithm is developed for computing the maximum likelihood estimates for NLMEC/LMEC with the standard errors of the fixed effects and the exact likelihood value as a by-product. The algorithm uses closed-form expressions at the E-step, that rely on formulas for the mean and variance of a truncated multivariate-t distribution. The proposed algorithm is implemented in the R package *tlmec*. It is applied to analyze longitudinal HIV viral load data in two recent AIDS studies. In addition, a simulation study is conducted to examine the performance of the proposed method and to compare it with the approach of Vaida and Liu (2009).

Key words and phrases: Censored data, ECM algorithm, HIV viral load, influential observations, mixed-effects models, outliers.

1. Introduction

Linear and nonlinear mixed-effects models (LME/NLME) are frequently used to analyze grouped data because they are capable of modeling the within-subject correlations often presented in this type of data (Pinheiro and Bates (2000)). Examples of grouped data include longitudinal data, repeated measures, and multilevel data. However, in such longitudinal studies, as those on environmental pollution and infection diseases, measurements of some variables may be subjected to certain threshold values below or above which the measurements are not quantifiable. For instance, viral load measures the amount of actively replicating virus and, depending upon the diagnostic assays used, its measurement may be subjected to detection limits, below or above which they are not quantifiable. The proportion of censored data in these studies may be nontrivial and considering such crude/adhoc methods is substituting values for censored observations (Vaida and Liu (2009)) can lead to biased estimates of fixed effects and variance components (Wu (2010)). As alternatives to crude imputation methods, Hughes (1999) proposed a likelihood-based Monte Carlo expectation-maximization (MCEM) algorithm for LME with censored responses (LMEC). Vaida, Fitzgerald, and DeGruttola (2007) proposed a hybrid EM (HEM) algorithm for linear and nonlinear mixed-effects models with censored responses. Vaida and Liu (2009) proposed an exact EM-type algorithm for LMEC/NLMEC that uses closed-form expressions at the E-step, as opposed to Monte Carlo simulation, leading to an improvement in the speed of computation up to an order of 10. More recently, Matos et al. (2013) provided some additional tools, including influence diagnostics analyses, for LMEC/NLMEC.

In the framework of LMEC/NLMEC, random effects and within-subject errors are routinely assumed to be normal for mathematical convenience. However, this is not always realistic because of atypical observations. To deal with the problem of atypical observations in LME with complete responses, proposals have been made in the literature to replace normality with a more flexible class of distributions. For instance, Pinheiro, Liu, and Wu (2001) proposed a multivariate-t linear mixed model (t-LME) and demonstrated its robustness against outliers through extensive simulations. Lin and Lee (2006) and Lin and Lee (2007) developed some additional tools for t-LME from likelihood-based and Bayesian perspectives. Arellano-Valle et al. (2012) proposed an extension of the normal censored regression model to the case where the error terms follow a univariate-t distribution. Recently, in the context of heavy–tailed LMEC/NLMEC, Lachos, Bandyopadhyay, and Dey (2011) advocated the use of the normal/independent (NI) class of distributions (Liu (1996)) and adopted a Bayesian framework to carry out posterior inference. Although some work with elliptical distributions has recently been published, there are no studies on censored LMEC/NLMEC under the Student-t family from a frequentist perspective. In this paper, we propose a robust parametric modeling of LMEC/NLMEC based on the multivariate-t distribution so that the t-LMEC/t-NLMEC is defined and a fully likelihood-based approach is carried out, including the implementation of an exact ECM algorithm for maximum likelihood (ML) estimation. As in Vaida and Liu (2009), we show that the E-step reduces to computing the first two moments of certain truncated multivariate-t distributions. The general formulas for these moments were derived by Ho et al. (2012). They require the multivariate-t cumulative density function (cdf), for which we use the *mvtnorm* package in R. The likelihood function is easily computed as a by-product of the E-step and is used for monitoring convergence and for model selection using the Akaike information criterion (AIC), the Bayesian information criterion (BIC), or the likelihood ratio test (LRT).

The rest of the paper is organized as follows. In Section 2 we introduce some notation and outline the main results related to the multivariate-t and truncated-t distributions. In Section 3 the t-LMEC and related likelihood-based inference are presented. In Sections 4 the extension to more general t-NLMEC is discussed. The advantage of the proposed methodology is illustrated through the analysis of two case studies of HIV viral load in Section 5. Section 6 presents a simulation study to compare the performance of our methods with other normality-based methods. Section 7 concludes with a short discussion of issues raised by our study and some possible directions for the future research.

2. The Multivariate t and Truncated t-distributions

A random variable $\mathbf{Y}$ having a p-variate t distribution with location vector $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$ and degrees of freedom ν, denoted by $t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, can be written as

$$\mathbf{Y} = \boldsymbol{\mu} + U^{-1/2}\mathbf{Z}, \ \ \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}),\ U \sim \text{Gamma}\Big(\frac{\nu}{2}, \frac{\nu}{2}\Big),$$

where $\mathbf{Z}$ and U are independent and Gamma(a, b) has mean a/b and density $G(.|a, b)$. The density of $\mathbf{Y}$, is

$$t_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \frac{\Gamma((p+\nu)/2)}{\Gamma(\nu/2)\pi^{p/2}}\nu^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\left(1 + \frac{\delta}{\nu}\right)^{-(p+\nu)/2},$$

where $\Gamma(.)$ is the standard gamma function and $\delta = (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})$ is the Mahalanobis distance. The cdf is denoted by $T_p(.|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$. If $\nu > 1$, $\boldsymbol{\mu}$ is the mean of $\mathbf{Y}$, and if $\nu > 2$, $\nu(\nu - 2)^{-1}\boldsymbol{\Sigma}$ is its covariance matrix. As ν tends to infinity, U converges to one with probability one, and so for large ν, $\mathbf{Y}$ is approximately multivariate normal with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

Proposition 1. *Let* $\mathbf{Y} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ *and* $\mathbf{Y}$ *be partitioned as* $\mathbf{Y}^\top = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$, *with* $\dim(\mathbf{Y}_1) = p_1$, $\dim(\mathbf{Y}_2) = p_2$, $p_1 + p_2 = p$, *and where* $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$ *and* $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top$, *are the corresponding partitions of* $\boldsymbol{\Sigma}$ *and* $\boldsymbol{\mu}$. *Then, we have*

(i) $\mathbf{Y}_1 \sim t_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \nu)$; *and*

(ii) *The conditional cdf of* $\mathbf{Y}_2|\mathbf{Y}_1 = \mathbf{y}_1$ *is given by*

$$P(\mathbf{Y}_2 \leq \mathbf{y}_2|\mathbf{Y}_1 = \mathbf{y}_1) = T_{p_2}\left(\mathbf{y}_2|\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}_{22.1}, \nu + p_1\right),$$

where $\widetilde{\boldsymbol{\Sigma}}_{22.1} = [(\nu + \delta_1)/(\nu + p_1)]\boldsymbol{\Sigma}_{22.1}$, $\delta_1 = (\mathbf{y}_1 - \boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)$, $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$, *and* $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)$.

The proof of (i) is straightforward from (1). The proof of (ii) follows from Proposition 4 given in Arellano-Valle and Genton (2010) by setting $\lambda = \tau = 0$.

Let $Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A})$ be the distribution $t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ constrained to lie within the right-truncated hyperplane

$$\mathbb{A} = \{\mathbf{x} = (x_1, \ldots, x_p)^\top | x_1 \leq a_1, \ldots, x_p \leq a_p\}. \tag{2.1}$$

Thus $\mathbf{X} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A})$ if its density is

$$f(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A}) = \frac{t_p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)}{T_p(\mathbf{a}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)} \mathbb{I}_{\mathbb{A}}(\mathbf{x}),$$

where $\mathbf{a} = (a_1, \ldots, a_p)^\top$ and $\mathbb{I}_{\mathbb{A}}(\mathbf{x})$ is the indicator function of $\mathbb{A}$. The proofs of the following propositions are given in the Web Appendix A.

Proposition 2. *If* $\mathbf{X} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A})$ *with* $\mathbb{A}$ *as* (2.1), *then the* k*th moment of* $\mathbf{X}$, $k = 0, 1, 2$, *is*

$$E\left\{ \left(\frac{\nu + p}{\nu + \delta} \right)^r \mathbf{X}^{(k)} \right\} = c_p(\nu, r) \frac{T_p(\mathbf{a}|\boldsymbol{\mu}, \boldsymbol{\Sigma}^*, \nu + 2r)}{T_p(\mathbf{a}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)} E_{\mathbf{W}}\{\mathbf{W}^{(k)}\},$$

$$\mathbf{W} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}^*, \nu + 2r; \mathbb{A}),$$

where

$$c_p(\nu, r) = \left(\frac{\nu + p}{\nu} \right)^r \left(\frac{\Gamma((p+\nu)/2)\Gamma((\nu+2r)/2)}{\Gamma(\nu/2)\Gamma((p+\nu+2r)/2)} \right),$$

$\delta = (\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu})$, $\mathbf{a} = (a_1, \ldots, a_p)^\top$, $\boldsymbol{\Sigma}^* = [\nu/(\nu + 2r)]\boldsymbol{\Sigma}$, $\mathbf{X}^{(0)} = 1$, $\mathbf{X}^{(1)} = \mathbf{X}$, $\mathbf{X}^{(2)} = \mathbf{X}\mathbf{X}^\top$, *and* $\nu + 2r > 0$.

Proposition 3. *Let* $\mathbf{X} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A})$ *with* $\mathbb{A}$ *as* (2.1). *Consider the partition* $\mathbf{X}^\top = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)$ *with* $\dim(\mathbf{X}_1) = p_1$, $\dim(\mathbf{X}_2) = p_2$, $p_1 + p_2 = p$, *and the corresponding partition of the parameters* $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\mathbf{a}$ $(\mathbf{a}^{x_1}, \mathbf{a}^{x_2})$ *and* $\mathbb{A}$ $(\mathbb{A}^{x_1}, \mathbb{A}^{x_2})$. *Then using the notation of Proposition* 1, *the conditional* k*th moment of* $\mathbf{X}_2$ *is*

$$E\Big\{ \Big(\frac{\nu+p}{\nu+\delta} \Big)^r \mathbf{X}_2^{(k)} | \mathbf{X}_1 \Big\} = \frac{d_p(p_1, \nu, r)}{(\nu + \delta_1)^r} \frac{T_{p_2}(\mathbf{a}^{x_2}|\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}^*_{22.1}, \nu + p_1 + 2r)}{T_{p_2}(\mathbf{a}^{x_2}|\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}_{22.1}, \nu + p_1)} E_{\mathbf{W}}\{\mathbf{W}^{(k)}\},$$

where $\mathbf{W} \sim Tt_{p_2}(\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}^*_{22.1}, \nu + p_1 + 2r; \mathbb{A}^{x_2})$, $\delta = (\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu})$,

$$\delta_1 = (\mathbf{X}_1 - \boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}_{11}^{-1} (\mathbf{X}_1 - \boldsymbol{\mu}_1), \quad \mathbf{a}^{x_2} = (a_1, \ldots, a_{p2})^\top,$$

$$\widetilde{\boldsymbol{\Sigma}}^*_{22.1} = \left(\frac{\nu + \delta_1}{\nu + 2r + p_1} \right) \boldsymbol{\Sigma}_{22.1},$$

$$d_p(p_1, \nu, r) = (\nu + p)^r \left(\frac{\Gamma((p+\nu)/2)\Gamma((p_1+\nu+2r)/2)}{\Gamma((p_1+\nu)/2)\Gamma((p+\nu+2r)/2)} \right), \textit{and } \nu + p_1 + 2r > 0.$$

Formulas for $E[\mathbf{W}]$ and $E[\mathbf{W}\mathbf{W}^\top]$, where $\mathbf{W} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A})$, have been recently developed in closed form by Ho et al. (2012); they depend on the multivariate-t cdf. The computation uses existing functions for the cumulative t-distribution, for which the *pmvt*() function of the *mvtnorm* library (Genz et al. (2008)) from R can be used.

3. Linear Mixed-effects with Censored Response

3.1. Model specification

For obtaining robust estimates of the parameters, we proceed as in Pinheiro, Liu, and Wu (2001) by considering a generalization of the classical N-LME as follows:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \text{ with} \tag{3.1}$$

$$(\mathbf{b}_i, \boldsymbol{\epsilon}_i)^\top \sim t_{n_i+q}\{\mathbf{0}, \mathrm{Diag}(\mathbf{D}, \sigma^2\mathbf{I}_{n_i}), \nu\}. \tag{3.2}$$

Here the subscript i is the subject index; $\mathbf{I}_p$ denotes the $p \times p$ identity matrix, Diag($\mathbf{A}$,$\mathbf{B}$) is a block diagonal matrix whose elements are the square matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{y}_i = (Y_{i1}, \ldots, Y_{in_i})^\top$ is a $n_i \times 1$ vector of observed continuous responses for sample unit i, $\mathbf{X}_i$ is the $n_i \times p$ design matrix corresponding to the fixed effects, $\boldsymbol{\beta}$ is a $p \times 1$ vector of population-averaged regression coefficients called fixed effects, $\mathbf{Z}_i$ is the $n_i \times q$ design matrix corresponding to the $q \times 1$ vector of random effects $\mathbf{b}_i$, $\boldsymbol{\epsilon}_i$ is the $n_i \times 1$ vector of random errors, and the dispersion matrix $\mathbf{D} = \mathbf{D}(\boldsymbol{\alpha})$ depends on unknown and reduced parameters $\boldsymbol{\alpha}$.

Note that $\mathbf{b}_i$ and $\boldsymbol{\epsilon}_i$ are uncorrelated, once $\mathrm{Cov}(\mathbf{b}_i, \boldsymbol{\epsilon}_i) = E[\mathbf{b}_i\boldsymbol{\epsilon}_i^\top] = E[E(\mathbf{b}_i\boldsymbol{\epsilon}_i^\top \mid U_i)] = \mathbf{0}$, where U_i is a scalar generated from Gamma($\nu/2, \nu/2$). Classical inference on the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\alpha}^\top, \nu)^\top$ is based on the marginal distribution of $\mathbf{y}_i$, $\mathbf{y}_i \overset{\text{ind.}}{\sim} t_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i, \nu)$, for $i = 1, \ldots, n$, where $\boldsymbol{\Sigma}_i = \sigma^2\mathbf{I}_{n_i} + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top$. The estimates from the multivariate t-LME are more robust against outliers than those based on the standard LME. In a simulation study, Pinheiro, Liu, and Wu (2001) showed that the t-LME substantially outperforms the normal or standard LME when outliers are present in the data. This issue has also been discussed by Wu (2010) in the context of censored mixed-effects models.

Following Vaida and Liu (2009), we consider the case in which the response Y_{ij} is not fully observed for all i, j. Thus, 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 $\mathbf{C}_i$ the vector of censoring indicators:

$$y_{ij} \le Q_{ij} \quad \text{if} \quad C_{ij} = 1, \text{ and } y_{ij} = Q_{ij} \quad \text{if} \quad C_{ij} = 0, \tag{3.3}$$

so that the t-LMEC is defined. For ease of presentation, we assume that the data are left-censored. The extensions to arbitrary censoring are immediate. In the next section, we present the likelihood function. It can be easily computed by using a sequence of simple steps.

3.2 The likelihood function

The first step is to treat separately the observed and censored components of $\mathbf{y}_i$. Partition $\mathbf{y}_i$ into the observed and censored components: $\mathbf{y}_i = vec(\mathbf{y}_i^o, \mathbf{y}_i^c)$, with $C_{ij} = 0$ for all elements in $\mathbf{y}_i^o$, and 1 for all elements in $\mathbf{y}_i^c$; write accordingly $\mathbf{Q}_i = vec(\mathbf{Q}_i^o, \mathbf{Q}_i^c)$, where $vec(.)$ denotes the function which stacks vectors or matrices of the same number of columns, with $\mathbf{\Sigma}_i = \left(\begin{smallmatrix}\mathbf{\Sigma}_i^{oo} & \mathbf{\Sigma}_i^{oc} \\ \mathbf{\Sigma}_i^{co} & \mathbf{\Sigma}_i^{cc}\end{smallmatrix}\right)$. Then, from Proposition 1, we have that $\mathbf{y}_i^o \sim t_{n_i^o}(\mathbf{X}_i^o\boldsymbol{\beta}, \mathbf{\Sigma}_i^{oo}, \nu)$, and $\mathbf{y}_i^c|\mathbf{y}_i^o, \sim t_{n_i^c}(\boldsymbol{\mu}_i^{co}, \mathbf{S}_i^{co}, \nu + n_i^o)$, where

$$\boldsymbol{\mu}_i^{co} = \mathbf{X}_i^c\boldsymbol{\beta} + \mathbf{\Sigma}_i^{co}\mathbf{\Sigma}_i^{oo-1}(\mathbf{y}_i^o - \mathbf{X}_i^o\boldsymbol{\beta}), \;\; \mathbf{S}_i^{co} = \left(\frac{\nu + Q(\mathbf{y}_i^o)}{\nu + n_i^o}\right)\mathbf{\Sigma}_i^{cc.o}, \qquad (3.4)$$

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

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

and the log-likelihood function for the observed data is given by $\ell(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^n \{\log L_i\}$. This can be computed at each step of the EM-type algorithm without additional computational burden, because the L_i's have already been computed at the E-step. In addition, the log-likelihood function can be used to monitor the convergence of the EM algorithm and for the model selection via AIC, BIC, or LRT.

Lucas (1997) carried out an interesting study on the robust aspects of the Student-t M-estimator in the univariate case using influence functions. He showed that the protection against outliers is preserved only if the degrees of freedom parameter are fixed. In this paper, we assume the degrees of freedom and the shape parameters for Student-t to be fixed, and we use a model selection procedure based on AIC or BIC to choose the most appropriate value of ν (see Lange, Little, and Taylor (1989); Meza, Osorio, and De la Cruz (2011)). Thus, hereafter we consider that the parameter vector is $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\alpha}^\top)^\top$.

3.3. The EM algorithm

The EM algorithm originally proposed by Dempster, Laird, and Rubin (1977) has several appealing features such as stability of monotone convergence with each iteration increasing the likelihood and simplicity of implementation. However, ML estimation in model (3.1)–(3.3) is complicated and the EM algorithm is less advisable due to the computational difficulty in the M-step. To cope with this problem, we apply an extension of the EM algorithm, called the ECM algorithm (Meng and Rubin (1993)), that shares the appealing features of the EM and has a typically faster convergence rate than the EM.

Let $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{b} = (\mathbf{b}_1^\top, \ldots, \mathbf{b}_n^\top)^\top$, $\mathbf{u} = (u_1, \ldots, u_n)^\top$, $\mathbf{Q} = vec(\mathbf{Q}_1, \ldots, \mathbf{Q}_n)$, and $\mathbf{C} = vec(\mathbf{C}_1, \ldots, \mathbf{C}_n)$ such that we observe $(\mathbf{Q}_i, \mathbf{C}_i)$ for the i-th subject. Treating $\mathbf{b}$, $\mathbf{u}$, and $\mathbf{y}$ as hypothetical missing data, and augmenting with the observed data $\mathbf{Q}, \mathbf{C}$, we set $\mathbf{y}_c = (\mathbf{C}^\top, \mathbf{Q}^\top, \mathbf{y}^\top, \mathbf{b}^\top, \mathbf{u}^\top)^\top$. Hence, the ECM algorithm is applied to the complete data log-likelihood function $\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta}|\mathbf{y}_c)$, with

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

where C is a constant that is independent of the parameter vector $\boldsymbol{\theta}$ and $h(u_i|\nu)$ is the density of Gamma$(\nu/2, \nu/2)$. Given the current value $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$, the E-step calculates the conditional expectation of the complete data log-likelihood function

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = \sum_{i=1}^n Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = \sum_{i=1}^n Q_{1i}(\boldsymbol{\beta}, \sigma^2|\widehat{\boldsymbol{\theta}}^{(k)}) + \sum_{i=1}^n Q_{2i}(\boldsymbol{\alpha}|\widehat{\boldsymbol{\theta}}^{(k)}), \tag{3.5}$$

where

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

and

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

Here

$$\begin{aligned}
\widehat{a}_i^{(k)} &= \mathrm{tr}\Big(\widehat{u\mathbf{y}_i^2}^{(k)} - 2\widehat{u\mathbf{y}\mathbf{b}_i}^{(k)}\mathbf{Z}_i^\top + \widehat{u\mathbf{b}_i^2}^{(k)}\mathbf{Z}_i^\top\mathbf{Z}_i\Big);\\
\widehat{u\mathbf{b}_i^2}^{(k)} &= E\{u_i\mathbf{b}_i\mathbf{b}_i^\top|\mathbf{Q}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\}\\
&= \widehat{\sigma^2}^{(k)}\widehat{\boldsymbol{\Lambda}}_i^{(k)} + \widehat{\boldsymbol{\varphi}}_i^{(k)}(\widehat{u\mathbf{y}_i^2}^{(k)} - \widehat{u\mathbf{y}}_i^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\widehat{u\mathbf{y}}_i^{(k)\top}\\
&\quad + \widehat{u}_i^{(k)}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top)\widehat{\boldsymbol{\varphi}}_i^\top;\\
\widehat{u\mathbf{b}}_i^{(k)} &= E\{u_i\mathbf{b}_i|\mathbf{Q}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\} = \widehat{\boldsymbol{\varphi}}_i^{(k)}(\widehat{u\mathbf{y}}_i^{(k)} - \widehat{u}_i^{(k)}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)});
\end{aligned}$$

$$\widehat{u\mathbf{y}\mathbf{b}_i}^{(k)} = E\{u_i\mathbf{y}_i\mathbf{b}_i^\top|\mathbf{Q}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\} = (\widehat{u\mathbf{y}_i^2}^{(k)} - \widehat{u\mathbf{y}}_i^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top)\widehat{\boldsymbol{\varphi}}_i^\top,$$

with

$$\widehat{\boldsymbol{\Lambda}}_i^{(k)} = (\widehat{\sigma^2}^{(k)}\widehat{\mathbf{D}}^{-1(k)} + \mathbf{Z}_i^\top\mathbf{Z}_i)^{-1} \quad \text{and} \quad \widehat{\boldsymbol{\varphi}}_i^{(k)} = \widehat{\boldsymbol{\Lambda}}_i^{(k)}\mathbf{Z}_i^\top.$$

Note that in this case we do not consider the computation of $E[h(u_i|\nu)|\mathbf{Q},\mathbf{C},\widehat{\boldsymbol{\theta}}^{(k)}]$ because ν is fixed.

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

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \Big(\sum_{i=1}^n \widehat{u}_i^{(k)}\mathbf{X}_i^\top\mathbf{X}_i\Big)^{-1}\sum_{i=1}^n \mathbf{X}_i^\top\Big(\widehat{u\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{u\mathbf{b}}_i^{(k)}\Big), \tag{3.6}$$

$$\widehat{\sigma^2}^{(k+1)} = \frac{1}{N}\sum_{i=1}^n\Big[\widehat{a}_i^{(k)} - 2\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top(\widehat{u\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{u\mathbf{b}}_i^{(k)}) + \widehat{u}_i^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\Big], \tag{3.7}$$

$$\widehat{\mathbf{D}}^{(k+1)} = \frac{1}{n}\sum_{i=1}^n \widehat{u\mathbf{b}_i^2}^{(k)}, \tag{3.8}$$

where $N = \sum_{i=1}^n n_i$. In our EM algorithm we assume that the scale matrix $\mathbf{D}$ is unstructured and, in this case, $\boldsymbol{\alpha}$ is the upper triangular elements of $\mathbf{D}$. The algorithm is iterated until the distance involving two successive evaluations of the log-likelihood, $|\ell(\widehat{\boldsymbol{\theta}}^{(k+1)})/\ell(\widehat{\boldsymbol{\theta}}^{(k)}) - 1|$, is sufficiently small. From (3.6)−(3.8) it is easy to see that the E-step reduces to the computation of $\widehat{u\mathbf{y}_i^2}$, $\widehat{u\mathbf{y}}_i$, and $\widehat{u}_i$. These expected values can be determined in closed form, using Propositions 1−3, as follows (see the Web Appendix B for details).

1. If individual i has only censored components, from Proposition 2

$$\widehat{u\mathbf{y}_i^2} = E\{u_i\mathbf{y}_i\mathbf{y}_i^\top|\mathbf{Q}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}\} = \frac{T_{n_i}(\mathbf{Q}_i|\widehat{\boldsymbol{\mu}}_i,\widehat{\boldsymbol{\Sigma}}_i^*,\nu+2)}{T_{n_i}(\mathbf{Q}_i|\widehat{\boldsymbol{\mu}}_i,\widehat{\boldsymbol{\Sigma}}_i,\nu)}E\{\mathbf{W}_i\mathbf{W}_i^\top\},$$

$$\widehat{u\mathbf{y}}_i = E\{u_i\mathbf{y}_i|\mathbf{Q}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}\} = \frac{T_{n_i}(\mathbf{Q}_i|\widehat{\boldsymbol{\mu}}_i,\widehat{\boldsymbol{\Sigma}}_i^*,\nu+2)}{T_{n_i}(\mathbf{Q}_i|\widehat{\boldsymbol{\mu}}_i,\widehat{\boldsymbol{\Sigma}}_i,\nu)}E\{\mathbf{W}_i\},$$

$$\widehat{u}_i = E\{u_i|\mathbf{Q}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}\} = \frac{T_{n_i}(\mathbf{Q}_i|\widehat{\boldsymbol{\mu}}_i,\widehat{\boldsymbol{\Sigma}}_i^*,\nu+2)}{T_{n_i}(\mathbf{Q}_i|\widehat{\boldsymbol{\mu}}_i,\widehat{\boldsymbol{\Sigma}}_i,\nu)},$$

   where $\mathbf{W}_i \sim Tt_{n_i}(\widehat{\boldsymbol{\mu}}_i,\widehat{\boldsymbol{\Sigma}}_i^*,\nu+2;\mathbb{A}_i)$, $\widehat{\boldsymbol{\mu}}_i = \mathbf{X}_i\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\Sigma}}_i^* = \dfrac{\nu}{\nu+2}\widehat{\boldsymbol{\Sigma}}_i$, $\widehat{\boldsymbol{\Sigma}}_i = \widehat{\sigma^2}\mathbf{I}_{ni} + \mathbf{Z}_i\widehat{\mathbf{D}}\mathbf{Z}_i^\top$ and $\mathbb{A}_i = \{\mathbf{W}_i = (w_1,\ldots,w_{n_i})^\top|w_1 \le Q_{i1},\ldots,w_{n_i} \le Q_{in_i}\}$.
2. If individual i has only non-censored components, then,

$$\widehat{u\mathbf{y}_i^2} = \frac{\nu+n_i}{\nu+Q(\mathbf{y}_i)}\mathbf{y}_i\mathbf{y}_i^\top,\ \widehat{u\mathbf{y}}_i = \frac{\nu+n_i}{\nu+Q(\mathbf{y}_i)}\mathbf{y}_i,\ \widehat{u}_i = \frac{\nu+n_i}{\nu+Q(\mathbf{y}_i)},$$

where $Q(\mathbf{y}_i) = (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top \boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})$.

3. If individual i has censored and uncensored components, then from Proposition 3 and the fact that $\{\mathbf{y}_i|\mathbf{Q}_i, \mathbf{C}_i\}$, $\{\mathbf{y}_i|\mathbf{Q}_i, \mathbf{C}_i, \mathbf{y}_i^o\}$, and $\{\mathbf{y}_i^c|\mathbf{Q}_i, \mathbf{C}_i, \mathbf{y}_i^o\}$ are equivalent processes, we have

$$\widehat{u\mathbf{y}_i^2} = E\{u_i\mathbf{y}_i\mathbf{y}_i^\top|\mathbf{y}_i^o, \mathbf{Q}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\} = \begin{pmatrix} \mathbf{y}_i^o\mathbf{y}_i^{o\top}\widehat{u}_i & \widehat{u}_i\mathbf{y}_i^o\widehat{\mathbf{w}}_i^{c\top} \\ \widehat{u}_i\widehat{\mathbf{w}}_i^c\mathbf{y}_i^{o\top} & \widehat{u}_i\widehat{\mathbf{w}^{\mathbf{2}}}_i^c \end{pmatrix},$$

$$\widehat{u\mathbf{y}}_i = E\{u_i\mathbf{y}_i|\mathbf{y}_i^o, \mathbf{Q}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\} = vec(y_i^o\widehat{u}_i, \widehat{\mathbf{w}}_i^c),$$

$$\widehat{u}_i = E\{u_i|\mathbf{y}_i^o, \mathbf{Q}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\} = \Big(\frac{n_i^o + \nu}{\nu + Q(\mathbf{y}_i^o)}\Big)\frac{T_p(\mathbf{Q}_i|\boldsymbol{\mu}_i^{co}, \widetilde{\mathbf{S}}^{co}, \nu + n_i^o + 2)}{T_p(\mathbf{Q}_i|\boldsymbol{\mu}_i^{co}, \mathbf{S}^{co}, \nu + n_i^o)},$$

where

$$\widetilde{\mathbf{S}}^{co} = \Big(\frac{\nu + Q(\mathbf{y}_i^o)}{\nu + 2 + n_i^o}\Big)\boldsymbol{\Sigma}_i^{cc.o}, \ \widehat{\mathbf{w}}_i^c = E\{\mathbf{W}_i\},$$

and $\widehat{\mathbf{w}^{\mathbf{2}}}_i^c = E\{\mathbf{W}_i\mathbf{W}_i^\top\}$, with $\mathbf{W}_i \sim Tt_{n_i^c}(\boldsymbol{\mu}_i^{co}, \widetilde{\mathbf{S}}^{co}, \nu + n_i^o + 2; \mathbb{A}_i^c)$ and $\boldsymbol{\Sigma}_i^{cc.o}$, $\boldsymbol{\mu}_i^{co}$, and $\mathbf{S}^{co}$ are as in (3.4).

3.4. Estimation of random effects and the expected information matrix

In this subsection, we consider the conditional approach by using the conditional mean to estimate the random effects (Lin and Lee (2006); Ho et al. (2012)); this is useful for evaluating such subject-specific quantities as individual intercepts and slopes. Thus, if the values of parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\alpha}^\top)^\top$ and ν are known, the conditional mean of $\mathbf{b}_i$ given $\mathbf{C}_i$ and $\mathbf{Q}_i$ is

$$\begin{aligned}\widehat{\mathbf{b}}_i(\boldsymbol{\theta}) &= E\{\mathbf{b}_i|\mathbf{Q}_i, \mathbf{C}_i\} = E\{E\{E\{\mathbf{b}_i|u_i\}|\mathbf{y}_i, u_i\}|\mathbf{Q}_i, \mathbf{C}_i\} \\ &= E\{\boldsymbol{\Lambda}_i\mathbf{Z}_i^\top(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})|\mathbf{Q}_i, \mathbf{C}_i\} = \boldsymbol{\Lambda}_i\mathbf{Z}_i^\top(\widehat{\mathbf{y}}_i - \mathbf{X}_i\boldsymbol{\beta}), \end{aligned} \tag{3.9}$$

where $\boldsymbol{\Lambda}_i$ is defined in Subsection 3.3 and $\widehat{\mathbf{y}}_i = E\{\mathbf{y}_i|\mathbf{Q}_i, \mathbf{C}_i\}$ is the first moment of the truncated multivariate-t distribution ($Tt_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i, \nu; \mathbb{A}_i)$). In practice, the estimators of $\mathbf{b}_i$, $\widehat{\mathbf{b}}_i$, can be obtained by substituting the ML estimate $\widehat{\boldsymbol{\theta}}$ into (3.9), which leads to $\widehat{\mathbf{b}}_i = \widehat{\mathbf{b}}_i(\widehat{\boldsymbol{\theta}})$. The conditional covariance matrix of $\mathbf{b}_i$ given $\mathbf{C}_i$ and $\mathbf{Q}_i$ is

$$\begin{aligned}Var\{\mathbf{b}_i|\mathbf{Q}_i, \mathbf{C}_i\} &= E\{\mathbf{b}_i\mathbf{b}_i^\top|\mathbf{Q}_i, \mathbf{C}_i\} - \widehat{\mathbf{b}}_i(\boldsymbol{\theta})\widehat{\mathbf{b}}_i(\boldsymbol{\theta})^\top \\ &= \frac{\nu + n_i}{\nu + n_i - 2}E\Big\{(\frac{\nu + n_i}{\nu + Q(\mathbf{y}_i)})^{-1}|\mathbf{Q}_i, \mathbf{C}_i\Big\}\boldsymbol{\Lambda}_i\sigma^2 + \boldsymbol{\Lambda}_i\mathbf{Z}_i^\top Var(\mathbf{y}_i|\mathbf{Q}_i, \mathbf{C}_i)\mathbf{Z}_i\boldsymbol{\Lambda}_i.\end{aligned}$$

These expected values can be easily accomplished from Steps 1−3 given in Section 3.3 as a by-product of our proposed ECM algorithm (E-step).

Louis (1982) derived a result that can be used to adjust the variances of the estimated fixed effects for the information lost due to censoring. Using this method, from the results given in Appendix B in Lange, Little, and Taylor (1989), an asymptotic approximation for the variances of the fixed effects is given by, see Web Appendix C,

$$\mathbf{J}_{\boldsymbol{\beta}\boldsymbol{\beta}} = Var(\widehat{\boldsymbol{\beta}}) = \Big(\sum_{i=1}^{n}\frac{\nu+n_i}{\nu+n_i+2}\mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i - \sum_{i=1}^{n}\mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\mathbf{B}_i\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\Big)^{-1}, \quad (3.10)$$

where

$$\mathbf{B}_i = Var\Big\{\frac{\nu+n_i}{\nu+Q(\mathbf{y}_i)}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})|\mathbf{Q}_i, \mathbf{C}_i\Big\}$$

with $\mathbf{y}_i \sim Tt_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i, \nu; \mathbb{A}_i)$. Asymptotic confidence intervals and hypothesis tests for the fixed effects are obtained assuming that the ML estimates $\widehat{\boldsymbol{\beta}}$ has approximately a $N_p(\boldsymbol{\beta}, \mathbf{J}^{-1}_{\boldsymbol{\beta}\boldsymbol{\beta}})$ distribution. In practice, $\mathbf{J}_{\boldsymbol{\beta}\boldsymbol{\beta}}$ is usually unknown and needs to be replaced by its ML estimates $\mathbf{J}_{\widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}}$.

4. The Nonlinear Case

Extending the notation of the previous section and ignoring censoring, we first propose a general mixed-effects model in which the random terms are assumed to follow a multivariate-t distribution (t-NLME). Let $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$ denote the (continuous) response vector for subject i, and $\eta = (\eta(\mathbf{X}_{i1}, \boldsymbol{\phi}_i), \ldots, \eta(\mathbf{X}_{in_i}, \boldsymbol{\phi}_i))^\top$ be a nonlinear vector-valued differentiable function of the random parameter $\boldsymbol{\phi}_i$ and a vector of covariates $\mathbf{X}_i$. The t-NLME can then be expressed as

$$\mathbf{y}_i = \eta(\boldsymbol{\phi}_i, \mathbf{X}_i) + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\phi}_i = \mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \quad (4.1)$$

where the joint distribution of ($\mathbf{b}_i$, $\boldsymbol{\epsilon}_i$) is given as (3.2), $\mathbf{A}_i$ and $\mathbf{B}_i$ are known design matrices of dimensions $r \times p$ and $r \times q$, respectively, possibly depending on some covariable 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. Thus, from the properties of the multivariate-t distribution, we have that marginally, $\boldsymbol{\phi}_i \overset{ind}{\sim} t_r(\mathbf{A}_i\boldsymbol{\beta}, \mathbf{B}_i\mathbf{D}\mathbf{B}_i^\top, \nu)$ and $\boldsymbol{\epsilon}_i \overset{ind.}{\sim} t_{n_i}(\mathbf{0}, \sigma^2\mathbf{I}_{n_i}, \nu)$ and, as in the linear case, they are uncorrelated because $\mathrm{Cov}(\boldsymbol{\phi}_i, \boldsymbol{\epsilon}_i) = \mathbf{0}$. For NI-NLME with non censoring responses, the marginal distribution is

$$f(\mathbf{y}|\boldsymbol{\theta}) = \prod_{i=1}^{n}\int_0^\infty\int_{\mathbb{R}^q} \phi_{n_i}(\mathbf{y}_i; \eta(\boldsymbol{\phi}_i, \mathbf{X}_i), u_i^{-1}\sigma^2\mathbf{I}_{n_i})\phi_q(\boldsymbol{\phi}_i; \mathbf{A}_i\boldsymbol{\beta}, u_i^{-1}\mathbf{B}_i\mathbf{D}\mathbf{B}_i^\top) \times G(u_i|\frac{\nu}{2}, \frac{\nu}{2})d\boldsymbol{\phi}_i du_i,$$

which generally does not have a closed form expression because the model function is not linear in the random effect. In the normal case, various first-order Taylor series expansions of the model function around the conditional mode of

$\mathbf{b}_i$, say $\widetilde{\mathbf{b}}_i$, have been proposed to achieve tractable numerical optimizations (Wu (2010)). Most algorithms for computing the approximate ML estimates $\widehat{\boldsymbol{\theta}}$ and the estimators (predictors) of the random effects $\widehat{\mathbf{b}}_i$ involve the iterative maximization of the approximate log-likelihood functions $\ell(\boldsymbol{\theta}, \widetilde{\mathbf{b}}) = \sum_{i=1}^{n} \log f(\mathbf{y}_i|\boldsymbol{\theta}, \widetilde{\mathbf{b}}_i)$. Following Taylor series expansions, we have a theorem that is useful for the implementation of the EM algorithm; it uses simultaneously a neighborhood of $\mathbf{b}_i$ and $\boldsymbol{\beta}$ as expansions points, with the advantage that the likelihood is completely linearized (in $\mathbf{b}_i$ and $\boldsymbol{\beta}$). This result can be considered as an extension of the results in Lindstrom and Bates (1990) and Pinheiro and Bates (2000) for the Student-t case. The proof is given in Web Appendix A.

Theorem 1. *Let $\widetilde{\mathbf{b}}_i$ and $\widetilde{\boldsymbol{\beta}}$ be expansion points in a neighborhood of $\mathbf{b}_i$ and $\boldsymbol{\beta}$, respectively. Under the t-NLME model in* (4.1) *we have the linearized model*

$$\widetilde{\mathbf{y}}_i = \widetilde{\mathbf{W}}_i\boldsymbol{\beta} + \widetilde{\mathbf{H}}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \;\; i = 1, \ldots, n, \tag{4.2}$$

where $\widetilde{\mathbf{y}}_i = \mathbf{y}_i - \widetilde{\eta}(\mathbf{A}_i\widetilde{\boldsymbol{\beta}} + \mathbf{B}_i\widetilde{\mathbf{b}}_i, \mathbf{X}_i)$, $\mathbf{b}_i \overset{ind}{\sim} t_q(0, \mathbf{D}, \nu)$, $\boldsymbol{\epsilon}_i \overset{ind.}{\sim} t_{n_i}(\mathbf{0}, \sigma^2\mathbf{I}_{n_i}, \nu)$,

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

and $\widetilde{\eta}(\widetilde{\boldsymbol{\beta}}, \widetilde{\mathbf{b}}_i) = \eta(\mathbf{A}_i\widetilde{\boldsymbol{\beta}} + \mathbf{B}_i\widetilde{\mathbf{b}}_i, \mathbf{X}_i) - \widetilde{\mathbf{H}}_i\widetilde{\mathbf{b}}_i - \widetilde{\mathbf{W}}_i\widetilde{\boldsymbol{\beta}}$.

The estimates of the random effects $\widetilde{\mathbf{b}}$, given in (3.9), can be used iteratively in the linearization procedure from Theorem 1. Note that the distribution of $\mathbf{b}_i|\mathbf{y}_i$ is approximately symmetric (Student-t), and thus $\widetilde{\mathbf{b}}_i$ is the mode of the distribution at each step. As in Vaida and Liu (2009), the linearization (L) procedure to obtain the approximate ML estimates of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\alpha}^\top)^\top$ consists of iteratively solving the LME model (L-step) in (4.2). For the censored response, the linearized model (4.2) is an LME with censored data that has the same structure as (3.1)−(3.2), which is then solved as indicated in the previous section. The matrices in (4.2) depend on the current parameter values and need to be recalculated at each iteration. The algorithm iterates to convergence between the L-, E-, and CM-steps. Extension to more general t-LMEC and to t-NLMEC is discussed in the Web Appendix D.

5. Applications to HIV Data

We apply the proposed methods to the two HIV data sets previously analyzed using LMEC models.

5.1. UTI data

The first application is to a study of 72 perinatally HIV-infected children

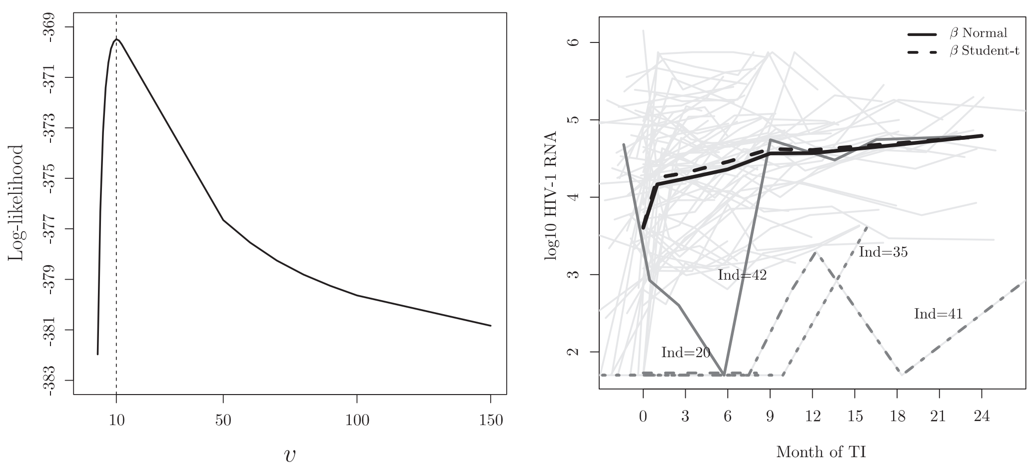


Figure 1. UTI data. (a) Plot of the profile log-likelihood of the degrees of freedom ν. (b) Individual profiles and overall mean (in $\log_{10}$ scale) using the Normal and t distributions for HIV viral load at different follow-up times. The trajectories for the influential individuals are numbered.

(Saitoh et al. (2008)). The data set is available in the R package (R Development Core Team (2009)) *lmec*. Primarily due to treatment fatigue, unstructured treatment interruptions (UTI) are 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. Out of 362 observations, 26 were below the detection limits (50 or 400 copies/mL) and considered left-censored at those values. The individual profiles of viral load at different followup times after UTI are shown in Figure 1 (right panel). 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, with $t_1 = 0$, $t_2 = 1$, $t_3 = 3$, $t_4 = 6$, $t_5 = 9$, $t_6 = 12$, $t_7 = 18$, and $t_8 = 24$. Vaida and Liu (2009) analyzed the same data set by fitting a similar N-LMEC (hereafter LMEC) via the EM algorithm, but from Figure 1 given in Lachos, Bandyopadhyay, and Dey (2011) it is clear that inference based on normality assumptions is questionable (presence of heavy tails). We revisit the UTI data with the aim of carrying out robust inferences, from a frequentist perspective, by using the Student-t distribution. The ML estimates were obtained using the ECM algorithm described in Section 3. Starting values were obtained by using the library *lmec*.

For the Student-t model, we assumed the degree of freedom ν known and, by using the AIC criterion, we found $\nu = 10$ (see left panel in Figure 1), indicating that the normal model is inadequate. Table 1 presents the ML estimates and the

Table 1. ML estimates (MLe) under normal and Student-t models fitted to the UTI data. SE is the corresponding standard error. The Student-t model has 10 degrees of freedom.

Model		Parameter										AIC	BIC
		β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	σ^2	α		
normal	MLe	3.604	4.166	4.241	4.360	4.566	4.569	4.677	4.794	0.341	0.765	844.117	883.034
	SE	0.125	0.129	0.130	0.131	0.140	0.149	0.165	0.202				
Student-t	MLe	3.618	4.253	4.314	4.458	4.623	4.611	4.698	4.787	0.350	0.666	759.015	797.931
	SE	0.125	0.129	0.130	0.131	0.140	0.149	0.165	0.202				

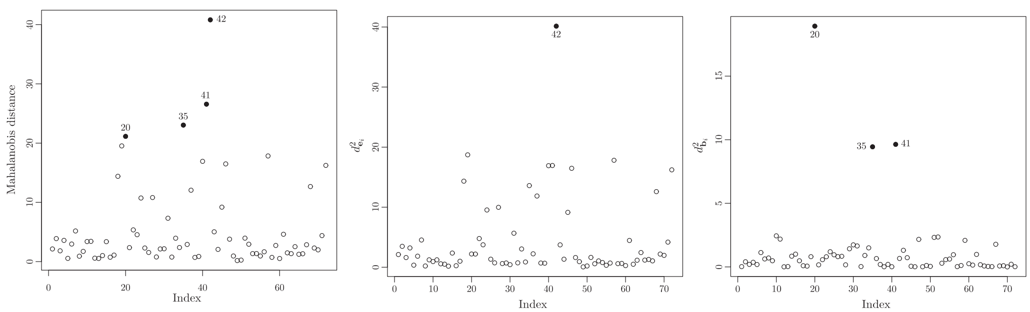


Figure 2. UTI data. (a) Mahalanobis distance, (b) Estimated $d^2_{\mathbf{e}_i}$ (error) and (c) Estimated $d^2_{\mathbf{b}_i}$ (R.E.), for the LMEC model.

corresponding standard errors of the fixed effects $\boldsymbol{\theta}$. Comparing these values we notice a similarity between the estimates under normal and Student-t models. Additionally, the inferences for the variance components are similar for the two models, but are not comparable since they are on different scales. According to the AIC or BIC values given in Table 1, we notice that the t-LMEC model outperforms the LMEC model. For the LRT statistics described in Subsection 3.5, we obtained maximum log-likelihoods of -412.059 for the LMEC model and -369.507 for the t-LMEC model, which gave the corresponding likelihood ratio statistic of $LRT = 42.552$. Here the LRT statistic follows a equally weighted mixture of χ^2_0 and χ^2_1 distributions (see the Web Appendix E). Therefore, the resulting p-value 3.441×10^{-11} suggests the appropriateness of the use of the multivariate-t distribution. With the missing-at-random assumption as in Vaida and Liu (2009), our dropout (censored) model does not bias the inference regarding the mean of β_j. For both models the mean viral load $E(y_{ij}) = \beta_j$ increases gradually throughout 24 months under the two models. For the best model (t-LMEC), it increases from 3.62 at the time of UTI to 4.79 at 24 months. This is in contrast with the mean profiles of the observed data alone, which show a leveling off and a decrease in viral load between 6 and 12 months (see Figure 1 in Vaida and Liu (2009)). The estimates of the between-subject (α) and within-subject (σ^2) scale parameters (in $\log_{10}$ scale) are 0.6662 and 0.3503, respectively.

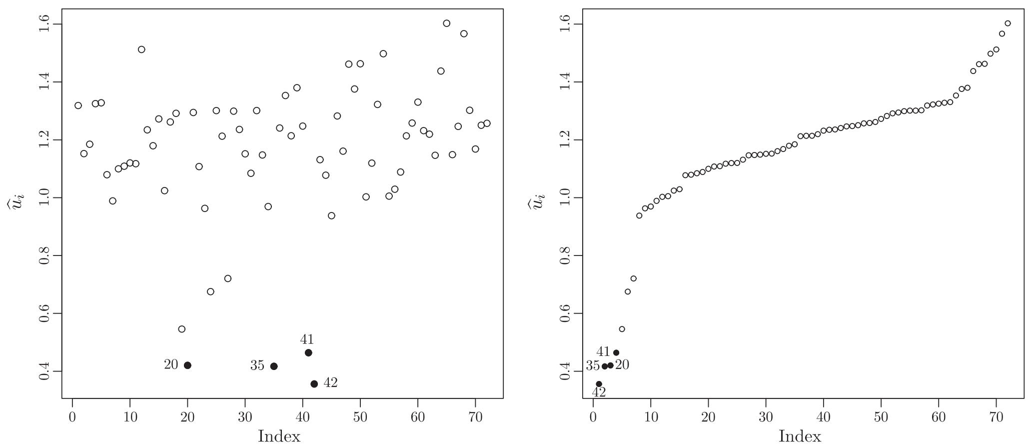


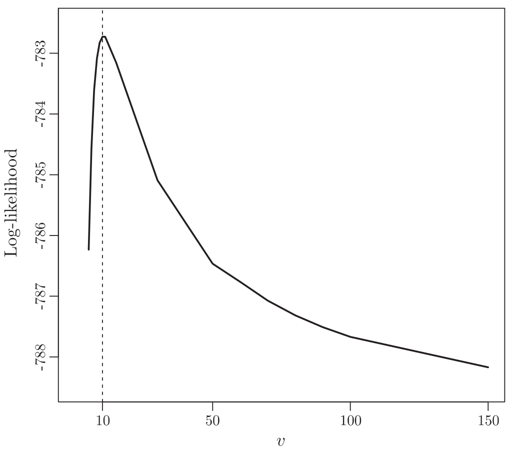
Figure 3. UTI data. (a) Estimated weight $\widehat{u}_i$ for the t-LMEC fit. (b) The influential observations for the LMEC are numbered.

To determine possible influential observations, we used the Mahalanobis distance $d_i^2(\boldsymbol{\theta}) = (\widehat{\mathbf{y}}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}})^\top \boldsymbol{\Sigma}_i^{-1}(\widehat{\mathbf{y}}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}),\ i = 1, \ldots, 72$. As in Pinheiro, Liu, and Wu (2001), replacing $\boldsymbol{\theta}$ and $\mathbf{b}_i$ with their current estimates, we obtain a decomposition for the Mahalanobis distance: $d_i^2(\widehat{\boldsymbol{\theta}}) = -(1/\widehat{\sigma^2})\widehat{\mathbf{e}_i}^\top\widehat{\mathbf{e}}_i + \widehat{\mathbf{b}_i}^\top\widehat{\mathbf{D}}\widehat{\mathbf{b}}_i, = \widehat{d^2_{\mathbf{e}_i}} + \widehat{d^2_{\mathbf{b}_i}}$ where $\widehat{\mathbf{e}_i} = \widehat{\mathbf{y}}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}} - \mathbf{Z}_i\widehat{\mathbf{b}}_i$ and $\widehat{\mathbf{b}}_i$ is as in (3.9). The estimated distances $d^2_{\mathbf{e}_i}$ (Error) and $d^2_{\mathbf{b}_i}$ (Random Effect-R.E.) are useful diagnostic statistics for identifying subjects with outlying observations (see, for example,Meza, Osorio, and De la Cruz (2011)). Figure 2 presents these diagnostic statistics for the LMEC model. Subject #42 has large values of d_i^2 and $d^2_{\mathbf{e}_i}$, suggesting an outlying observation at the within-subject level (**e**-outlier). Moreover, observations #20, #35 and #41 present large values of $d^2_{\mathbf{b}_i}$, suggesting outlying observations at the between-subject level (**b**-outlier). In a Bayesian analysis, these observations were also detected as influential (Lachos, Bandyopadhyay, and Dey (2011)).

It is well known that outlying observations may affect the estimation of the parameters under the normality assumption. However, when we use the Student-t distribution, the EM algorithm allows one to accommodate discrepant observations attributing small weights to them in the estimation procedure. The estimated weights ($\widehat{u}_i,\ i = 1, \ldots, 72$) for the t-LMEC model are presented in Figure 3. We see there that observations #20, #35, #41 and #42, indicated as outliers under the normal model, have smaller values of d_i^2 and $d^2_{\mathbf{e}_i}$, confirming the robust aspects of the ML estimates against outlying observations under the t-LMEC model. This robustness is also observed in Figure 1(b), where the presence of these outliers might lead to the underestimation of the predicted mean curve for the LMEC model as compared to the t-LMEC model. In summary, we see that

Table 2. ML estimates (MLe) under normal and Student-t models fitted to the AIEDRP data. SE are the corresponding standard errors. The Student-t model presented has 10 degrees of freedom.

Model		Parameter									AIC	BIC
		β_1	β_2	β_3	β_4	β_5	σ^2	α_{11}	α_{12}	α_{22}		
normal	MLe	1.610	0.142	3.526	1.056	-0.004	0.265	0.01769	0.00016	0.00004	1610.814	1700.521
	SE	0.015	0.095	0.024	0.268	0.001						
Student-t	MLe	1.611	0.161	3.524	0.987	-0.003	0.207	0.01611	0.00013	0.00004	1581.416	1623.908
	SE	0.013	0.085	0.021	0.246	0.001						



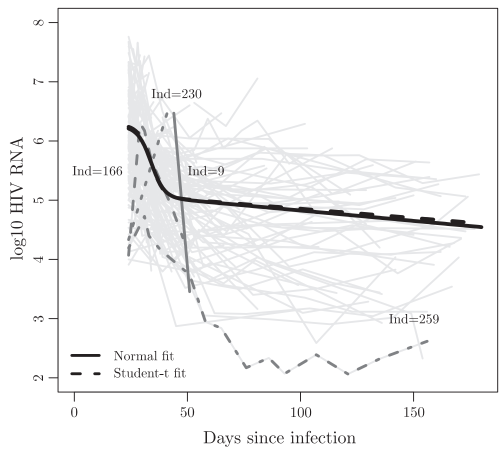


Figure 4. AIEDRP data. (a) plot of the profile log-likelihood of the degrees of freedom ν. (b) Individual profiles and overall mean (in $\log_{10}$ scale) using the Normal and t distributions for HIV viral load at different follow-up times. The trajectories for the influential individuals are numbered.

the robust aspects of the t-LME models (Pinheiro, Liu, and Wu (2001)) against outlying observations can extend to the case in which censoring components are present.

5.2 AIEDRP study

The second AIDS case study is from the AIEDRP program, a large multi-center observational study of subjects with acute and early HIV infection. 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, hence they were considered as right-censored. We considered a right-censored version and accommodate it within our NLME. Following Vaida and Liu (2009), we chose the five-parameter NLME model (inverted S-shaped curve)

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

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) set points 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, hence the set point was 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 reparameterized as follows: $\beta_{1i} = \log(\alpha_{1i}) = \beta_1 + b_{1i}$, $\beta_k = \log(\alpha_k), k = 2, 3, 4$, and $\alpha_{5i} = \beta_5 + b_{2i}$. Within a classical framework, we used the Student-t (t-NLMEC) with the ECM algorithm as described in Section 3. As in the previous application, the estimation of the parameter ν was chosen following the strategy proposed by Lange, Little, and Taylor (1989), which selected $\nu = 10$ (see Figure 4(a)). This parameter acts as a tuning constant in robust estimation methods and in our case we see that is provided an adequate protection against outliers. For the sake of model comparison, we also fit the normal NLMEC (hereafter NLMEC) counterparts, which can be treated as the reduced t-NLMEC as ν tends to infinity.

Table 2 presents the ML estimates of the parameters under the NLMEC model and the t-NLMEC model, together with the corresponding standard errors of the fixed effects and the associated AIC and BIC values. From this table, we observe that the standard errors under the t-NLMEC are smaller, indicating that the Student-t model produces more precise estimates. According to the AIC or BIC values, the t-NLMEC provided a much improved model fit over the NLMEC. In fact, the maximum log-likelihoods were -781.708 for the NLMEC and -775.951 for the t-NLMEC model, which gives the corresponding likelihood ratio statistic of 11.508 (p-value = 0.00035). This further confirms that the t-NLMEC model fits the data substantially better than the NLMEC model.

To identify outlying observations, we computed the Mahalanobis distance $d_i^2(\widehat{\boldsymbol{\theta}})$, $i = 1, \ldots, 320$, and the estimated distances $d^2_{\mathbf{e}_i}$ (Error) and $d^2_{\mathbf{b}_i}$ (Random Effect). Figure 5 presents these diagnostic statistics for the LMEC model. We see there that observations #9, #166, #230 and #259 appear to be outliers. The observations #9, #166 and #230 have large values of $d^2_{\mathbf{e}_i}$, suggesting **e**-outliers; observation #259 presents a large value of $d^2_{\mathbf{b}_i}$, suggesting an **b**-outlier. From Figure 4(b), the fitted viral load curve appears to be underestimated as compared to the t-NLMEC due to the presence of these outliers. This suggests that the t-NLMEC, which downweights the influence of outliers, provides a more appropriate way for achieving robust inference.

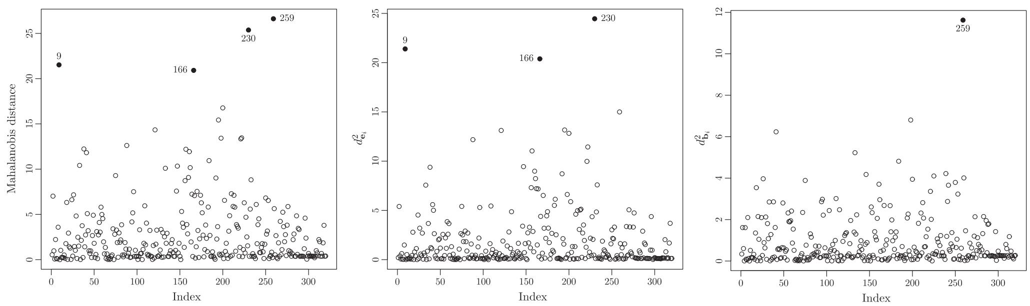


Figure 5. AIEDRP data. (a) Mahalanobis distance, (b) Estimated $d^2_{\mathbf{e}_i}$ (error) and (c) Estimated $d^2_{\mathbf{b}_i}$ (R.E.). The influential observations are numbered.

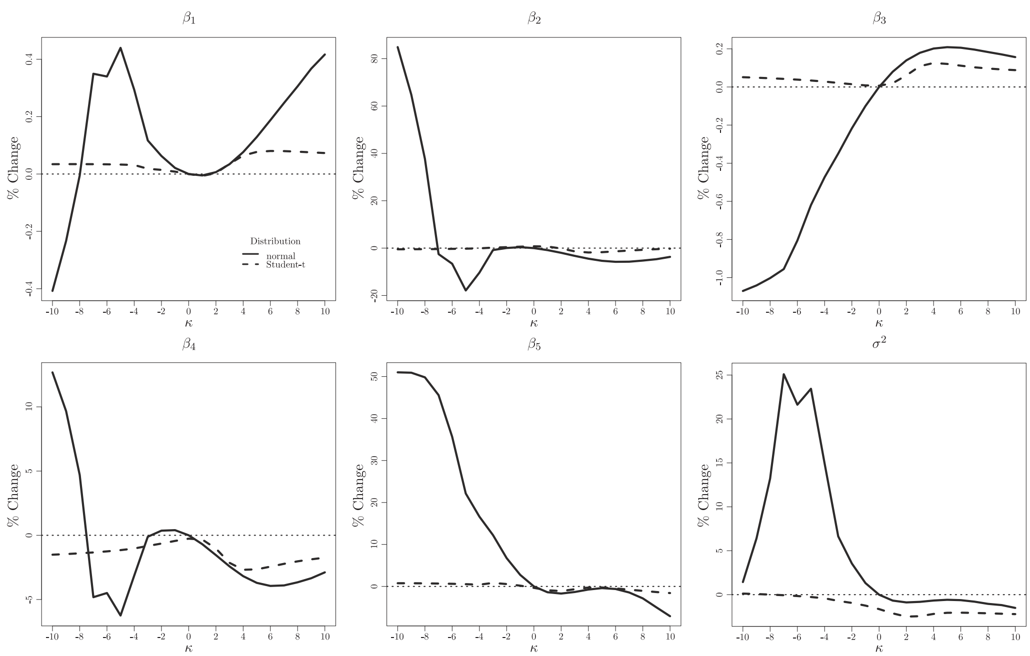


Figure 6. AIEDRP data. Relative changes on the ML estimates of $\boldsymbol{\theta}$ from the normal NLMEC (solid line) and the t-NLMEC (dashed line) for different contaminations κ.

The robustness of the t-LMEC model can also be assessed by considering the influence of a single outlying observation on the ML estimate of $\boldsymbol{\theta}$. In particular, we can assess how much the ML estimate of $\boldsymbol{\theta}$ is influenced by a change of κ units in a single observation y_{ik}. We replace a single observation y_{ik} by $y_{ik}(\delta) = y_{ik}+\kappa$, and record the relative change in the estimates $((\widehat{\theta}(\kappa) - \widehat{\theta})/\widehat{\theta})$, where $\widehat{\theta}$ denotes the original estimate and $\widehat{\theta}(\kappa)$ is the estimate for the contaminated data. In this application we contaminated the first observation on subject 198 and varied δ between -10 and 10. In Figure 6 we present the results of the relative changes of the estimates $\boldsymbol{\beta}$ and σ^2 for different values of κ under the NLMEC and t-NLMEC models. As expected, the estimates from the t-NLMEC were less affected by variations of κ than the NLMEC.

6. Simulation Studies

To examine the performance of our proposed methodology, we conducted a simulation study. The goal was to investigate the consequences on parameter inference when the normality assumption is inappropriate, as well as to investigate whether the model comparison measures, AIC and BIC, determine the best-fitting model to the simulated data. A similar study for the linear case is presented in the Web Appendix F.

Following Vaida and Liu (2009), we considered the nonlinear mixed model

$$y_{ij} = \alpha_{1i} + \frac{\alpha_2}{(1+\exp((t_{ij}-\alpha_3)/\alpha_{4i}))} + \epsilon_{ij}, \quad i = 1,\ldots,100, \quad j = 1,\ldots,10,$$

where $(b_{1i}, b_{2i}) \overset{\text{iid.}}{\sim} t_2(\mathbf{0}, \mathbf{D}, \nu)$ and $\epsilon_{ij} \sim t(0, \sigma^2, \nu)$. We re-parameterized as $\beta_{1i} = \log(\alpha_{1i}) = \beta_1 + b_{1i}$, $\beta_k = \log(\alpha_k)$, $k = 2, 3$, and $\alpha_{4i} = \beta_4 + b_{2i}$. In addition, we set $t_{ij} = (1, 10, 20, 30, 40, 50, 60, 70, 80, 90)$, $\boldsymbol{\beta}^\top = (\beta_1, \beta_2, \beta_3, \beta_4) = (1.6094, 0.6931, 3.8067, 2.3026)$, $\mathbf{D} = \begin{pmatrix} 0.0025 & -0.0010 \\ -0.0010 & 0.0100 \end{pmatrix}$, $\sigma^2 = 0.55$, and $\nu = 4$.

We chose various settings of censoring proportions, 0%, 5%, 10%, 20% and 50%, to study the effect of the level of censoring in the estimation. In this way, we have five settings with 100 simulated data sets under each setting. Once the simulated data were generated, we fit the proposed NLMEC model assuming normal and Student-t distributions to each simulated data set. The model selection criterion AIC and BIC as well as the estimates of the model parameters were recorded for each simulation. For the five censoring patterns, the summary statistics for $\boldsymbol{\beta}$ (the fixed-effects parameters) are presented in Table 3 assuming normal and Student-t distributions.

From Table 3, we observe that for all levels of censoring percentages, the Student-t distribution outperforms the normal distribution and has small standard deviations in the estimates. The arithmetic average (MC AIC and MC BIC) of the model comparison criteria are also strongly in favor of the Student-t model in comparison to the normal model, reinforcing the notion that these measures are capable of detecting departures from normality. In Table 3, $\hat{\sigma}^2$ under the normal distribution model is almost twice the true σ^2, this is because in the normal scenario σ^2 represents the variance and therefore should be compared with $[\nu/(\nu-2)]\sigma^2$, which is 1.10. Notice also that, the Student-t model has a

Table 3. Results based on 100 simulated Student-t samples. MC mean, MC Sd (in parentheses), and MC CP are the respective mean estimates, standard deviations, and coverage probability from fitting LMEC with Student-t and normal assumptions with different settings of censoring proportions. IM SE is the average value of the approximate standard error obtained through the information-based method. MC AIC and MC BIC are the arithmetic averages of the respective model comparison measures.

			Simulated Student-t data						
Censoring	Fit		β_1	β_2	β_3	β_4	σ^2	MC AIC	MC BIC
0%	Normal	MC Mean	1.622	0.661	3.788	2.225	1.071	2986.253	3025.515
		IM SE	0.018	0.070	0.043	0.196			
		MC Sd	(0.016)	(0.068)	(0.048)	(0.156)	(0.187)		
		MC CP	85%	93%	93%	97%			
	Student-t	MC Mean	1.616	0.676	3.798	2.258	0.551	2740.333	2779.595
		IM SE	0.014	0.055	0.032	0.150			
		MC Sd	(0.012)	(0.056)	(0.031)	(0.150)	(0.042)		
		MC CP	93%	90%	93%	98%			
5%	Normal	MC Mean	1.627	0.642	3.796	2.205	0.967	2865.279	2904.541
		IM SE	0.017	0.068	0.041	0.191			
		MC Sd	(0.016)	(0.073)	(0.043)	(0.192)	(0.146)		
		MC CP	81%	87%	96%	95%			
	Student-t	MC Mean	1.615	0.667	3.805	2.230	0.642	2654.928	2694.190
		IM SE	0.015	0.058	0.035	0.161			
		MC Sd	(0.012)	(0.056)	(0.031)	(0.150)	(0.060)		
		MC CP	96%	93%	99%	95%			
10%	Normal	MC Mean	1.623	0.657	3.801	2.235	0.970	2815.475	2854.737
		IM SE	0.018	0.070	0.042	0.191			
		MC Sd	(0.017)	(0.069)	(0.046)	(0.178)	(0.141)		
		MC CP	86%	88%	92%	95%			
	Student-t	MC Mean	1.613	0.676	3.803	2.253	0.629	2608.471	2647.733
		IM SE	0.015	0.059	0.035	0.160			
		MC Sd	(0.014)	(0.057)	(0.036)	(0.150)	(0.057)		
		MC CP	94%	94%	95%	97%			
20%	Normal	MC Mean	1.616	0.683	3.806	2.240	0.975	2705.762	2494.963
		IM SE	0.019	0.070	0.042	0.190			
		MC Sd	(0.016)	(0.069)	(0.042)	(0.183)	(0.145)		
		MC CP	95%	95%	98%	96%			
	Student-t	MC Mean	1.616	0.678	3.797	2.259	0.579	2494.963	2534.225
		IM SE	0.015	0.059	0.035	0.157			
		MC Sd	(0.015)	(0.060)	(0.032)	(0.162)	(0.044)		
		MC CP	89%	92%	99%	95%			
50%	Normal	MC Mean	1.614	0.684	3.781	2.131	0.978	1982.382	2021.644
		IM SE	0.022	0.073	0.043	0.208			
		MC Sd	(0.023)	(0.069)	(0.045)	(0.160)	(0.186)		
		MC CP	94%	95%	90%	93%			
	Student-t	MC Mean	1.624	0.650	3.789	2.226	0.546	1879.266	1918.528
		IM SE	0.022	0.075	0.041	0.187			
		MC Sd	(0.016)	(0.066)	(0.040)	(0.151)	(0.038)		
		MC CP	90%	93%	95%	95%			

smaller confidence interval due to the smaller standard error although its coverage probability is slightly better than that of the normal model. This is a strong evidence of the robustness in estimation of the Student-t method. Table 3 also provides the average values of the approximate standard errors of the EM estimates obtained through the information-based method described in Subsection 3.4 (IM SE) and the Monte Carlo standard deviation (MC Sd) for the parameters. We also see from Table 3 that the estimation method of the standard errors provides relatively close results for the normal and Student-t models, indicating

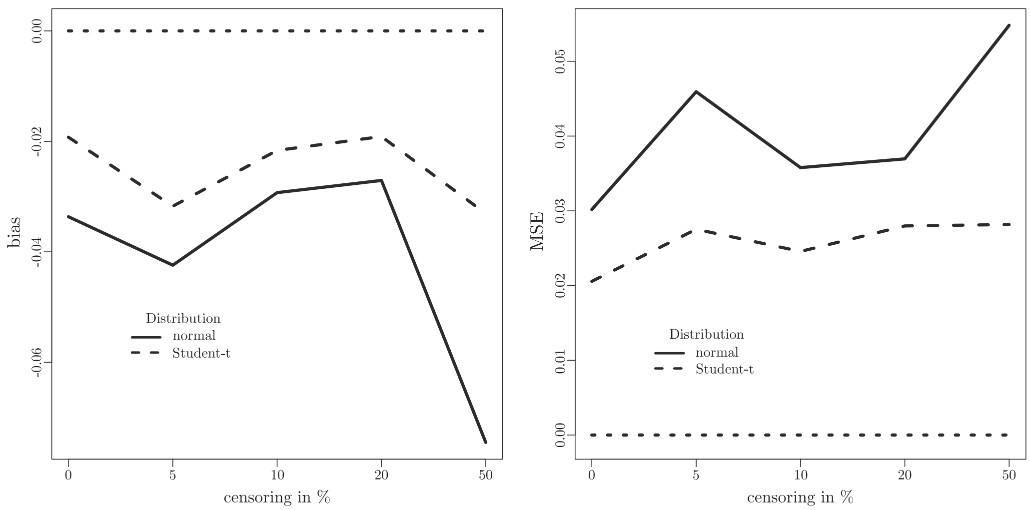


Figure 7. (a) Represents the bias of β_4 in comparison with the true value for the normal and Student-t models for the 5 censoring patterns (0%, 5%, 10%, 20%, 50%) in the NLMEC setup. (b) Presents the Mean Square Error (MSE) for β_4 for the normal and Student-t models.

that the proposed asymptotic approximation for the variances of the fixed effects (Equation (3.10)) is reliable.

In Figure 7 we only present the bias and MSE for the parameter estimate of β_4 under each of the normal and Student-t distributions, a similar pattern was observed for the other parameters. It is clear that the Student-t model is more robust here, providing more accurate estimates when the data has departures from the normality.

7. Conclusions

We have proposed a robust approach to linear and nonlinear mixed effects models with censored observations based on the multivariate-t distribution, called the t-LMEC/t-NLMEC. This offers a great deal of flexibility in dealing with longitudinal data in the presence of outliers. A novel ECM algorithm to obtain approximated ML estimates is developed by exploring the statistical properties of the multivariate truncated Student-t distribution. Our proposed algorithm has a closed-form expression for the E-step, based on formulas for the mean and variance of the truncated Student-t distribution. For NLMEC, the analysis is computationally feasible through approximating the t-NLMEC for a multivariate t distribution with specified parameters. We applied our methodology to two recent AIDS studies (freely downloadable from R) as well as to simulated data to illustrate how the procedures can be used to evaluate model assumptions, identify outliers, and obtain robust parameter estimates. From these results it is encouraging that the use of t-LMEC/t-NLMEC models offer better fitting, better protection against outliers and more precise inferences than their normal counterparts.

An anonymous referee brought up the issue of model identifiability. It is known that the Student-t family tends to the Normal as the degrees of freedom $\nu \to \infty$, and in this case we retrieve the LMEC/NLMEC model proposed by Hughes (1999), Vaida, Fitzgerald, and DeGruttola (2007) and Vaida and Liu (2009). Our model is identifiable since $\mathbf{C}_i$, $\mathbf{Q}_i$, and ν are known, while, the unknown parameters are related to existing well developed mixed-effects models in the statistical literature.

Although the t- LMEC considered here has shown great flexibility to modeling symmetric data, its robustness against outliers can be seriously affected by the presence of skewness. Recently, Lachos, Ghosh, and Arellano-Valle (2010) (see also Ho and Lin (2010)) proposed a remedy to accommodate skewness and heavy-tailedness simultaneously, using scale mixtures of skew-normal (SMSN) distributions. We conjecture that our methodology can be used under LMEC/NLMEC, and should yield satisfactory results at the expense of additional complexity in implementation. An in-depth investigation of such extensions is beyond the scope of the present paper, but it is an interesting topic for further research. Finally, the proposed EM algorithm has been coded and implemented in the R package *tlmec* (Matos, Prates, and Lachos (2012)), which is available for download at CRAN repository.

8. Supplementary Material

The web Appendices referenced in the paper are available under the Paper Information link at the Statistica Sinica website `http://www.stat.sinica.edu.tw/statistica`.

Acknowledgement

The authors wish to thank the Editor, an associate editor, and two referees for their valuable comments and suggestions that led to a substantial improvement of the paper. We would also like to thank Dr. Tsung Lin for supplying his R-code on the truncated Student-t distribution that proved to be very helpful in implementing our procedure. Drs. Lachos and Prates acknowledge the support of the CNPq-Brazil and FAPESP-Brazil. Dr. Prates also acknowledges the support of FAPEMIG-Brazil. Dr. Chen's research was partially supported by the U.S. National Institutes of Health grants #GM 70335 and #CA 74015.

References

Arellano-Valle, R., Castro, L., González-Farías, G. and Muñoz-Gajardo, K. (2012). Student-t censored regression model: properties and inference. *Stat. Methods Appl.* **21**, 453-473.

Arellano-Valle, R. B. and Genton, M. G. (2010). Multivariate extended skew-t distributions and related families. *Metron* LXVIII, 201-234.

Dempster, A., Laird, N. and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39**, 1-38.

Genz, A., Bretz, F., Hothorn, T., Miwa, T., Mi, X., Leisch, F. and Scheipl, F. (2008). mvtnorm: Multivariate Normal and t Distribution. R package version 0.9-2, URL `http://CRAN.R-project.org/package=mvtnorm`.

Ho, H. and Lin, T. (2010). Robust linear mixed models using the skew t distribution with application to schizophrenia data. *Biometrical Journal* **52**, 449-469.

Ho, H. J., Lin, T. I., Chen, H. Y., and Wang, W. L. (2012). Some results on the truncated multivariate t distribution. *J. Statist. Plann. Inference* **142**, 25-40.

Hughes, J. P. (1999). Mixed-effects models with censored data with application to HIV RNA levels. *Biometrics* **55**, 625-629.

Lachos, V. H., Bandyopadhyay, D. and Dey, D. K. (2011). Linear and nonlinear mixed-effects models for censored hiv viral loads using normal/independent distributions. *Biometrics* **67**, 1594-1604.

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

Lange, K. L., Little, R. J. A. and Taylor, J. M. G. (1989). Robust statistical modeling using t distribution. *J. Amer. Statist. Assoc.* **84**, 881-896.

Lin, T. I. and Lee, J. C. (2006). A robust approach to t linear mixed models applied to multiple sclerosis data. *Statist. Medicine* **25**, 1397-1412.

Lin, T. I. and Lee, J. C. (2007). Bayesian analysis of hierarchical linear mixed modeling using the multivariate t distribution. *J. Statist. Plann. Inference* **137**, 484-495.

Lindstrom, M. J. and Bates, D. M. (1990). Nonlinear mixed-effects models for repeated-measures data. *Biometrics* **46**, 673-687.

Liu, C. (1996). Bayesian robust multivariate linear regression with incomplete data. *J. Amer. Statist. Assoc.* **91**, 1219-1227.

Louis, T. A. (1982). Finding the observed information matrix when using the em algorithm. *J. Roy. Statist. Soc. Ser. B* **44**, 226-233.

Lucas, A. (1997). Robustness of the student t based M-estimator. *Comm. Statist. Theory Methods* **26**, 1165-1182.

Matos, L. A., Lachos, V. H., Balakrishnan, N. and Labra, F. V. (2013). Influence diagnostics in linear and nonlinear mixed-effects models with censored data. *Comput. Statist. Data Anal.* **57**, 450-464.

Matos, L. A., Prates, M. O. and Lachos, V. H. (2012). tlmec: Linear Student-t mixed-effects models with censored data. R Package Version 0.0-2 URL `http://CRAN.R-project.org/package=tlmec`.

Meng, X. and Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* **81**, 633-648.

Meza, C., Osorio, F. and De la Cruz, R. (2011). Estimation in nonlinear mixed-effects models using heavy-tailed distributions. *Statist. Comput.* **22**, 1-19.

Pinheiro, J. C. and Bates, Douglas, M. (2000). *Mixed-Effects Models in S and S-PLUS.* Springer, New York.

Pinheiro, J. C., Liu, C. H. and Wu, Y. N. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using a multivariate t-distribution. *J. Comput. Graph. Statist.* **10**, 249-276.

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. `http://www.R-project.org`

Saitoh, A., Foca, M., Viani, R. M., Heffernan-Vacca, S., Vaida, F., Lujan-Zilbermann, J., Emmanuel, P. J., Deville, J. G. and Spector, S. A. (2008). Clinical outcomes after an unstructured treatment interruption in children and adolescents with perinatally acquired HIV infection. *Pediatrics* **121**, e513-e521.

Vaida, F., Fitzgerald, A. P. and DeGruttola, V. (2007). Efficient hybrid EM for linear and nonlinear mixed-effects models with censored response. *Comput. Statist. Data Anal.* **51**, 5718-5730.

Vaida, F. and Liu, L. (2009). Fast implementation for normal mixed effects models with censored response. *J. Comput. Graph. Statist.* **18**, 797-817.

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

Departamento de Estatística, Universidade Estadual de Campinas, Brazil.

E-mail: larissam@ime.unicamp.br

Departamento de Estatística, Universidade Federal de Minas Gerais, Brazil.

E-mail: marcosop@est.ufmg.br

Department of Statistics, University of Connecticut, 215 Glenbrook Road, U-4120, Storrs, Connecticut 06269, U.S.A.

E-mail: ming-hui.chen@uconn.edu

Departamento de Estatística, Universidade Estadual de Campinas, Brazil.

E-mail: hlachos@ime.unicamp.br

(Received February 2012; accepted September 2012)