Published in final edited form as:

Stat Med. 2019 March 15; 38(6): 1074–1102. doi:10.1002/sim.8017.

Flexible longitudinal linear mixed models for multiple censored responses data

Victor H. Lachos¹, Larissa A. Matos², Luis M. Castro³, and Ming-Hui Chen¹

¹Department of Statistics, University of Connecticut, Storrs, Connecticut

²Departamento de Estatística, Universidade Estadual de Campinas, Campinas, Brazil

³Departamento de Estadística, Pontificia Universidad Católica de Chile, Santiago, Chile

Abstract

In biomedical studies and clinical trials, repeated measures are often subject to some upper and/or lower limits of detection. Hence, the responses are either left or right censored. A complication arises when more than one series of responses is repeatedly collected on each subject at irregular intervals over a period of time and the data exhibit tails heavier than the normal distribution. The multivariate censored linear mixed effect (MLMEC) model is a frequently used tool for a joint analysis of more than one series of longitudinal data. In this context, we develop a robust generalization of the MLMEC based on the scale mixtures of normal distributions. To take into account the autocorrelation existing among irregularly observed measures, a damped exponential correlation structure is considered. For this complex longitudinal structure, we propose an exact estimation procedure to obtain the maximum-likelihood estimates of the fixed effects and variance components using a stochastic approximation of the EM algorithm. This approach allows us to estimate the parameters of interest easily and quickly as well as to obtain the standard errors of the fixed effects, the predictions of unobservable values of the responses, and the log-likelihood function as a byproduct. The proposed method is applied to analyze a set of AIDS data and is examined via a simulation study.

Keywords

censored data; HIV viral load; multiple longitudinal responses; SAEM algorithm; outliers

1 | INTRODUCTION

According to Alder,¹ the acquired immune deficiency syndrome (AIDS) is defined as an illness characterized by laboratory evidence of the human immunodeficiency virus (HIV) infection. This type of evidence is commonly determined by using anti-HIV tests based on antibodies to HIV (HIV antigens) and also by detecting HIV genome sequences through the polymerase chain reaction (PCR) technique.

The PCR amplification of HIV RNA is the most accurate way to detect and quantify the virus present in blood. Moreover, PCR amplification also provides rapid access to the HIV genome and can lead to characterization of an HIV isolate to the strain level.² However, this sophisticated molecular technique quantifies the viral load in blood subject to a limit of detection. In other words, assays based on the PCR amplification present threshold values below or above that the measurements of the HIV RNA are not quantifiable. For example, as mentioned in the work of Barletta et al.,³ the Roche Amplicor HIV-1 Monitor Test (versions 1.0 and 1.5; Roche Molecular Systems, Basel, Switzerland) presents a range of detection from 400 to 750 000 HIV-1 RNA copies per milliliter when using the Standard Specimen Processing Procedure (200 μL of sample), and varies from 50 to 100 000 copies per milliliter when using the Ultrasensitive Specimen Processing Procedure (500 μL of sample centrifuged at 23 600g for one hour to concentrate virions).

From the statistical viewpoint, this situation presents the problem of censored statistical models in which the observed response lies in a restricted interval (to the left, right, or both of them). Moreover, since the HIV data are generally obtained from follow-up studies (AIDS clinical trials), censored linear and nonlinear mixed effects models⁴ and regression models with a specific correlation structure of the error terms⁵ are considered to study in detail the effects of specific antiretroviral (ARV) therapies in infected subjects.

In the statistical literature, a Gaussian distribution is often assumed for the random components of the model due to the computational convenience of the parameter estimation.⁶ However, in some cases, specifically in AIDS studies, such an assumption might not be realistic due to the heavy-tailed behavior of the data. In particular, atypical observations and/or the misspecification of the parametric distributions for both random effects and error terms affect the likelihood-based inference. Recently, Matos et al.⁷ have considered the Student- t distribution in the context of censored linear and non-linear mixed effects (λLMC / λNLMEC) models, including influence diagnostics with different perturbation schemes. Even though the λLMC and λNLMEC models have been used for the analysis of longitudinal data, they are restricted to univariate censored repeated measure data.

In longitudinal studies, it is common to observe more than one series of responses repeatedly measured in each subject across time. This type of data is called multivariate longitudinal data and is analyzed (in general) using the multivariate linear mixed effect (MLME) model proposed by Shah et al.⁸ As an extension of this model, Wang⁹ proposed the multivariate t linear mixed model (λMLME), which has been considered to be a robust approach for modeling multioutcome continuous repeated measures in the presence of outliers or heavy-tailed noise. From the censoring viewpoint, Wang et al.¹⁰ extended the λMLME to allow the analysis of multiple longitudinal censored outcomes and heavy-tails, giving rise the censored multivariate t linear mixed (λMLMEC) model. To estimate the parameters of the λMLMEC , an exact EM algorithm for the maximum-likelihood (ML) estimation is developed based on the mean and variance of a truncated multivariate Student- t , as proposed by Ho et al.¹¹

A drawback of those approaches is that they are not appropriate when the distributions of the error terms and random effects are not the same because such models do not consider different levels of heaviness in the tails of the distributions of the random effects and the

error terms, respectively. Our proposed method deals with this situation by considering the use of a class of scale mixtures of normal (SMN) distributions,¹² assuming that the distributions of the random effects and the error terms are different but belong to the SMN class. This approach allows us to introduce more flexibility to the model proposed by Wang⁹ (see also the work of Garay et al¹³), accommodating influential and/or outlying observations generated by the misspecification of the distributions of the error terms and/or random effects.

The motivating dataset considered here is obtained from an AIDS clinical trial – A5055 (see, for example, the works of Wang,⁹ Dagne and Huang,¹⁴ and more recently, Lin and Wang,¹⁵), which involved a total of 44 infected patients with the human immunodeficiency virus type 1 (HIV-1). These patients were treated with one of two potent ARV therapies, namely, IDV 800 mg twice daily (q12h) plus RTV 200 mg q12h (treatment 1) and IDV 400 mg q12h plus RTV 400 mg q12h (treatment 2). In AIDS research, the number of RNA copies (viral load) in blood plasma and its evolutionary trajectories play a prominent role in the diagnosis of HIV-1 disease progression after an ARV treatment regimen.¹⁶ Furthermore, the other two immunologic markers frequently used to monitor disease progression in AIDS studies are the cluster of differentiation 4 (CD4) and the cluster of differentiation 8 (CD8) T cells. The dataset consists of plasma viral load measurements (in copies per milliliter), and CD4 and CD8 cell counts measured roughly at days 0, 7, 14, 28, 56, 84, 112, 140, and 168 of follow-up for each patient. In this study, we focus on investigating the longitudinal trajectories for RNA viral load (in log-base-10 scale), denoted by $\log_{10}(\text{RNA})$, and CD4/CD8 ratio. As discussed earlier, in this case, the lower detection limit for RNA viral load is 50 copies/milliliter, and therefore 33.5% (106 out of 316) of measurements lies below the limits of assay quantification (left-censored). Figure 1 shows the trajectories of the two immunologic responses along time visits.

A typical algorithm for ML estimation in censored models involving the class of SMN distributions is the EM-type algorithm and its variants (see, for example, the work of Matos et al⁷). However, in some cases, EM-type algorithms are not appropriate due to the computational difficulty of the E-step, which involves the computation of expected quantities that cannot be obtained analytically. To deal with this problem, Delyon et al¹⁷ proposed a stochastic approximation version of the EM algorithm called the SAEM algorithm. This algorithm consists of replacing the E-step by a stochastic approximation obtained using simulated data, whereas the M-step remains unchanged. Jank¹⁸ showed that the computational effort of SAEM is much smaller and achieves convergence in just a fraction of the simulation size when compared to Monte Carlo EM (MCEM). In this paper, we develop a full likelihood approach for a censored linear mixed model based on the SMN distribution, including the implementation of the SAEM algorithm for ML estimation. As a byproduct, the proposed approach allows us to obtain the prediction of unobservable values of the responses, to compute the asymptotic standard errors of the parameter estimates, and to carry out the decomposition of AIC and BIC to assess the fit of the multiple longitudinal outcomes of the joint model, separately.

The rest of this paper is organized as follows. In Section 2, we describe the family of SMN distributions, including an outline of the SAEM algorithms. The censored linear mixed

effects model using the SMN class of distributions and the ML estimation procedure based on the SAEM algorithm are described in Section 3. In Section 4, the estimation of the likelihood function and the computation of the approximated standard errors are discussed. To examine the performance of our proposed methods, we carry out several simulation studies in Section 5. Section 6 is devoted to the analysis of the AIDS clinical trial A5055. Finally, Section 7 closes this paper with a brief discussion of issues raised by our study and some possible directions for future research.

2 | PRELIMINARIES

2.1 | Scale mixture of normal distributions (SMN)

A symmetrical class of scale mixtures of multivariate normal distributions (see the work of Andrews and Mallows¹²) is defined as the distribution of the p -variate random vector

$$\mathbf{y} = \boldsymbol{\mu} + \xi(U)^{1/2} \mathbf{Z}, \quad (1)$$

where $\boldsymbol{\mu}$ is a location vector, $\mathbf{Z}$ is a normal random vector with mean vector $\mathbf{0}$ and variance–covariance matrix Σ , U is a positive random variable with the cumulative distribution function (cdf) $H(u; \boldsymbol{\nu})$, and the probability density function (pdf) $h(u; \boldsymbol{\nu})$, $\boldsymbol{\nu}$ is a scalar or parameter vector indexing the distribution of U , $\xi(U)$ is the weight function, and U and $\mathbf{Z}$ are independent. Given $U = u$, $\mathbf{y}$ follows a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\kappa(u)\Sigma$. Hence, the pdf of $\mathbf{y}$ is given by

$f(\mathbf{y}|\boldsymbol{\mu}, \Sigma, \boldsymbol{\nu}) = \int_0^\infty \phi_p(\mathbf{y}; \boldsymbol{\mu}, \xi(u)\Sigma) dH(u; \boldsymbol{\nu})$, where $\phi_p(\cdot; \boldsymbol{\mu}, \Sigma)$ stands for the pdf of the p -variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariate matrix Σ . By convention, we shall write $\mathbf{y} \sim \text{SMN}_p(\boldsymbol{\mu}, \Sigma; H)$. In this paper, we consider $\xi(u) = 1/u$ because it leads to good mathematical properties; and three SMN distributions are commonly used for robust estimation, namely, the *multivariate Student-t distribution* (T), that includes the normal distribution when $\boldsymbol{\nu} \rightarrow \infty$; the *multivariate slash distribution* (SL), including the normal case when $\boldsymbol{\nu} \rightarrow \infty$; and the *multivariate contaminated normal distribution* (CN), which includes the normal distribution when $\boldsymbol{\nu}_2 \rightarrow 1$. Appendix A (Table A1) describes in detail each of these members.

2.2 | The SAEM algorithm

The EM algorithm proposed by Dempster et al¹⁹ has become one of the most useful frequentist approaches for estimating parameters via ML when the data under analysis present missing observations or latent variables. The main features of the EM algorithm are the ease of implementation and the stability of monotone convergence, with the likelihood increasing at each iteration. Let $\boldsymbol{\theta}$ be the parameter vector and let $\mathbf{y}_c = (\mathbf{y}^\top, \mathbf{q}^\top)$ be the vector of complete data, ie, the observed data $\mathbf{y}^\top$ and the missing data (or the latent variables, depending on the situation) $\mathbf{q}^\top$. The EM algorithm consists basically of two steps, ie, the expectation (E-step) and the maximization (M-step). These steps are operated iteratively in the complete data likelihood function ($\ell(\boldsymbol{\theta}|\mathbf{y}_c)$) until convergence. Each iteration is performed as follows.

E-Step: Compute the conditional expectation $Q(\theta|\hat{\theta}^{(k)}) = E[\ell_c(\theta|y_c)|y, \hat{\theta}^{(k)}]$, where $\hat{\theta}^{(k)}$ is the estimate of θ at the k th iteration.

M-Step: Update $\hat{\theta}^{(k)}$ according to $\hat{\theta}^{(k+1)} = \underset{\theta}{\operatorname{argmax}} Q(\theta|\hat{\theta}^{(k)})$.

Although the EM algorithm is a powerful tool when the conditional expectation in the E-step has a closed form, it is problematic when the analytical expression is not available. To overcome this problem, the MCEM was proposed by Wei and Tanner,²⁰ where the E-step is replaced by a Monte Carlo approximation based on a large number of independent simulations of the latent variables. However, for some complex models, the large number of simulations required by the MCEM algorithm can be computationally expensive.

Delyon et al¹⁷ proposed a stochastic approximation of the EM algorithm, called the SAEM algorithm, as an alternative to overcome the computational complexity of the MCEM algorithm. According to Meza et al,²¹ the SAEM algorithm at iteration k includes the following steps.

E-Step: 1. Simulation-step: Draw $\mathbf{q}^{(k,l)} (l = 1, \dots, m)$ from the conditional distribution $f(\mathbf{q}|\mathbf{y}, \hat{\theta}^{(k-1)})$. When the missing data/latent variables cannot be simulated exactly under this conditional distribution, an MCMC procedure such as the Metropolis-Hastings algorithm can be considered. In this case, $\mathbf{q}^{(k,l)} (l = 1, \dots, m)$ can be obtained from the transition probability

$$\prod_{\hat{\theta}^{(k-1)}}(\mathbf{q}^{(k-1)}, \mathbf{q}^k) = r(\mathbf{q}^{(k-1)}, \mathbf{q}^k) \min \left\{ \frac{f(\mathbf{q}^k|\mathbf{y}, \hat{\theta}^{(k-1)})r(\mathbf{q}^k, \mathbf{q}^{(k-1)})}{f(\mathbf{q}^{(k-1)}|\mathbf{y}, \hat{\theta}^{(k-1)})r(\mathbf{q}^{(k-1)}, \mathbf{q}^k)}, 1 \right\},$$

for $\mathbf{q}^k \neq \mathbf{q}^{(k-1)}$, where $r(\mathbf{q}^{(k-1)}, \mathbf{q}^k)$ is any aperiodic recurrent transition density or specifically the proposal distribution in the Metropolis-Hastings algorithm context (see Kuhn and Lavielle²²)

2. Stochastic approximation: update $Q(\theta|\hat{\theta}^{(k)})$ according to

$$Q(\theta|\hat{\theta}^{(k)}) = Q(\theta|\hat{\theta}^{(k-1)}) + \delta_k \left(\frac{1}{m} \sum_{l=1}^m \ell_c(\theta|\mathbf{q}^{(k,l)}, \mathbf{y}) - Q(\theta|\hat{\theta}^{(k-1)}) \right), \quad (2)$$

where $\ell_c(\theta|\mathbf{y}_c) = \sum_{i=1}^n \ell_i(\theta|\mathbf{y}_c)$ is the complete data log-likelihood function and δ_k is a smoothness parameter, ie, a decreasing sequence of positive numbers such that $\sum_{k=1}^{\infty} \delta_k = \infty$ and $\sum_{k=1}^{\infty} \delta_k^2 < \infty$.

M-Step: Maximization: update $\hat{\theta}^{(k)}$ according to $\hat{\theta}^{(k+1)} = \underset{\theta}{\operatorname{argmax}} Q(\theta|\hat{\theta}^{(k)})$.

In the case of considering the MCMC step within the E-Step of the SAEM algorithm, the algorithm is called MCMC-SAEM.²² Following Galarza et al.,²³ we consider the smoothing parameter in (2) as $\delta_k = 1$, if $1 \leq k \leq cW$ and $\delta = 1/(k - cW)$ if $cW + 1 \leq k \leq W$, where W is the maximum number of iterations and c is a cutoff point ($0 \leq c \leq 1$), which determines the percentage of the initial iterations. On the one hand, if $\delta_k = 1$ for all k , the SAEM algorithm has “no memory” and coincides with the MCEM algorithm and converges quickly (convergence in distribution) to a solution neighborhood. On the other hand, when the algorithm has memory, the convergence is slow (almost sure convergence) to the ML solution. Finally, a small number of simulations m (suggested to be $m = 20$) is sufficient to achieve convergence of the SAEM algorithm.

3 | LINEAR MIXED MODELS FOR MULTIPLE CENSORED RESPONSES DATA

3.1 | The statistical model

In this section, we introduce the SMN multivariate linear mixed model for censored responses (the SMN-MLMEC model). Following Wang et al.,¹⁰ suppose that r outcome variables of interest together with several covariates are repeatedly measured for each of n subjects at irregular intervals over a period of time. For the i th subject ($i = 1, \dots, n$), let $\mathbf{Y}_i = [\mathbf{y}_{i1} : \dots : \mathbf{y}_{ir}]$ be an $n_i \times r$ outcome matrix, where each column vector $\mathbf{y}_{ij} = (y_{ij1}, \dots, y_{ijn_i})^\top$ is an $n_i \times 1$ vector of the j th outcome ($j = 1, \dots, r$) over occasion k ($k = 1, \dots, n_i$). Let $\mathbf{X}_{ij}$ be an $n_i \times p_j$ design matrix for fixed effects corresponding to the j th outcome of the i th subject. Furthermore, let $\mathbf{Z}_{ij}$, which is generally a subset of $\mathbf{X}_{ij}$, be an $n_i \times q_j$ design matrix for random effects. In a block-diagonal matrix form, we write $\mathbf{X}_i = \text{diag}\{\mathbf{X}_{i1}, \dots, \mathbf{X}_{ir}\}$, $\mathbf{Z}_i = \text{diag}\{\mathbf{Z}_{i1}, \dots, \mathbf{Z}_{ir}\}$, and $\mathbf{E}_i = [\boldsymbol{\epsilon}_{i1} : \dots : \boldsymbol{\epsilon}_{ir}]$, which is an $n_i \times r$ matrix of within-subject errors, where the j th column $\boldsymbol{\epsilon}_{ij}$ corresponds to the error for the j th outcome $\mathbf{y}_{ij}$.

In writing the model, we ignore the censoring scheme for the moment, using the $\text{vec}(\cdot)$ operator to stack all columns of a matrix vertically. Then, $\mathbf{y}_i = \text{vec}(\mathbf{Y}_i) = (\mathbf{y}_{i1}^\top, \dots, \mathbf{y}_{ir}^\top)^\top$, and $\boldsymbol{\epsilon}_i = \text{vec}(\mathbf{E}_i) = (\boldsymbol{\epsilon}_{i1}^\top, \dots, \boldsymbol{\epsilon}_{ir}^\top)^\top$, which are of dimension $s_i = n_i \times r$. The linear mixed effect model for the i th subject can be written as

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, n, \quad (3)$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \dots, \boldsymbol{\beta}_r^\top)^\top$ is the $p \times 1$ vector of fixed effects associated with the design matrix $\mathbf{X}_i$ and $\mathbf{b}_i = (\mathbf{b}_{i1}^\top, \dots, \mathbf{b}_{ir}^\top)^\top$ is the $q \times 1$ vector of random effects associated with the design matrix $\mathbf{Z}_i$, $p = \sum_{j=1}^r p_j$, and $q = \sum_{j=1}^r q_j$.

Instead of the usual assumption of normality for the errors and random-effects, we replace the multivariate normal distribution with the multivariate SMN distribution. Therefore, the model can be expressed as

$$\mathbf{y}_i | \mathbf{b}_i \stackrel{\text{ind.}}{\sim} \text{SMN}_{s_i}(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i, \mathbf{R}_i; \mathbf{H}_1) \quad \text{and} \quad \mathbf{b}_i \stackrel{\text{ind.}}{\sim} \text{SMN}_q(\mathbf{0}, \mathbf{D}; \mathbf{H}_2), \quad i = 1, \dots, n, \quad (4)$$

where the abbreviation ind. refers to independence. Using the stochastic representation (1), the hierarchical representation (four stages) of the model defined in (3) to (4) is given by

$$\begin{aligned} \mathbf{y}_i | \mathbf{b}_i, \kappa_i &\stackrel{\text{ind.}}{\sim} N_{s_i}(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i, \kappa_i^{-1} \mathbf{R}_i), \\ \mathbf{b}_i | \tau_i &\stackrel{\text{ind.}}{\sim} N_q(\mathbf{0}, \tau_i^{-1} \mathbf{D}), \\ \kappa_i &\stackrel{\text{ind.}}{\sim} H_1(\cdot; \nu), \\ \tau_i &\stackrel{\text{ind.}}{\sim} H_2(\cdot; \eta), \quad i = 1, \dots, n, \end{aligned} \quad (5)$$

where $\mathbf{D} = \mathbf{D}(\boldsymbol{\alpha}) = \mathbf{D}_{jj'}$ is a $q \times q$ dispersion matrix that depends on the unknown and reduced parameters $\boldsymbol{\alpha}$, with $D_{jj'}$ being a partition matrix. In particular, for $j = j'$, D_{jj} is a covariance structure of random effects for the j th outcome, and for $j \neq j'$, $D_{jj'}$ is the covariance for a pair of outcome variables. The variables κ_i and τ_i are assumed to be mutually independent, in which $H_1 = H_1(\cdot; \nu)$ and $H_2 = H_2(\cdot; \eta)$ are the *cdf* generators that determine the specific SMN model that is chosen. We also assume that the within-subject errors for the response at different time points have serial correlations described by an $n_i \times n_i$ autocorrelation matrix $\Omega_i = \Omega_i(\boldsymbol{\phi}; \mathbf{t}_i)$. This matrix presents a parsimonious dependence structure involving only parameter $\boldsymbol{\phi}$ and measurement time $\mathbf{t}_i$ of subject i and that the multiple responses at a particular occasion are correlated with an $r \times r$ variance-covariance matrix $\Sigma = [\sigma_{jj'}]$. Accordingly, $\mathbf{R}_i = \Sigma \otimes \Omega_i$, where $\otimes$ denotes the Kronecker product.

Now, we consider the censoring scheme and focus on the left-censored case. Specifically, we assume that the observations are of the form

$$y_{ijk} \leq V_{ijk}, \quad \text{if } C_{ijk} = 1 \quad \text{and} \quad y_{ijk} = V_{ijk}, \quad \text{if } C_{ijk} = 0, \quad (6)$$

where V_{ijk} represents the uncensored observation or the limit of quantification and C_{ijk} is the censoring indicator, which equals one for a censored observation and zero for an uncensored observation. The observed data for the i th subject are represented by $(\mathbf{V}_i, \mathbf{C}_i)$, where $\mathbf{V}_i = \text{vec}(\mathbf{V}_{i1}, \dots, \mathbf{V}_{in_i})$ is an s_i vector and $\mathbf{C}_i = \text{vec}(\mathbf{C}_{i1}, \dots, \mathbf{C}_{in_i})$ is an $s_i \times 1$ vector, with $\mathbf{V}_{ij} = (V_{ij1}, \dots, V_{ijn_i})^\top$ and $\mathbf{C}_{ij} = (C_{ij1}, \dots, C_{ijn_i})^\top$.

For modeling the correlation structure between the observations of a specific subject, following the work of Muñoz et al.,²⁴ we adopt a damped exponential correlation (DEC) structure for Ω_i , which is defined as

$$\Omega_i = \Omega_i(\boldsymbol{\phi}, \mathbf{t}_i) = \left[\phi_1^{|t_{ij} - t_{ik}|} \phi_2 \right], \quad i = 1, \dots, n, \quad j, k = 1, \dots, n_i, \quad (7)$$

where $\boldsymbol{\phi} = (\phi_1, \phi_2)^\top$, the parameter ϕ_1 describes the autocorrelation between observations separated by the absolute length of two time points, and the parameter ϕ_2 permits acceleration of the exponential decay of the autocorrelation function, defining a continuous-time autoregressive model.

The parameter space of ϕ_1 and ϕ_2 is defined as $\Phi = \{(\phi_1, \phi_2) : 0 < \phi_1 < 1, \phi_2 > 0\}$. Moreover, it can be observed that different values of ϕ_2 produce a variety of correlation structures (for a given value of $\phi_1 > 0$). Particularly, we have (a) $\phi_2 = 0$ generates the compound symmetry correlation structure; (b) $0 < \phi_2 < 1$ presents a decay rate between the compound symmetry structure and the autoregressive of the first-order –AR(1)– model; (c) $\phi_2 = 1$ generates an AR(1) structure; (d) $\phi_2 > 1$ generates a decay rate faster than the AR(1) structure; and (e) $\phi_2 \rightarrow \infty$ represents the first-order moving average model, ie, MA(1).

3.2 | Maximum-likelihood estimation

This section is devoted to the SAEM algorithm for the ML estimation of $\boldsymbol{\theta} = (\boldsymbol{\theta}_\epsilon, \boldsymbol{\theta}_b, \nu, \eta)$, where $\boldsymbol{\theta}_\epsilon = (\boldsymbol{\beta}, \boldsymbol{\sigma}, \boldsymbol{\phi})$ and $\boldsymbol{\theta}_b = \boldsymbol{\alpha}$ in the SMN-MLMEC model defined previously. In this case, $\boldsymbol{\alpha}$ is a vector of the upper triangular parameters of $\mathbf{D}$ and $\boldsymbol{\sigma}$ is a vector of the upper triangular parameters of Σ .

Consider the model defined in (3) to (7). Write $\boldsymbol{\kappa} = (\kappa_1, \dots, \kappa_n)^\top$, $\boldsymbol{\tau} = (\tau_1, \dots, \tau_n)^\top$, $\mathbf{V} = \text{vec}(\mathbf{V}_1, \dots, \mathbf{V}_n)$, and $\mathbf{C} = \text{vec}(\mathbf{C}_1, \dots, \mathbf{C}_n)$ such that we observe $(\mathbf{V}_i, \mathbf{C}_i)$ for the i th subject. Treating $\mathbf{y}$, $\mathbf{b}$, $\boldsymbol{\kappa}$, and $\boldsymbol{\tau}$ as missing data and augmenting with the observed data $\mathbf{V}$, $\mathbf{C}$, we set $\mathbf{y}_c = (\mathbf{V}^\top, \mathbf{C}^\top, \mathbf{y}^\top, \mathbf{b}^\top, \boldsymbol{\kappa}^\top, \boldsymbol{\tau}^\top)^\top$ as the complete data. Therefore, using the representation defined in (5), the complete data log-likelihood function for all subjects can be written as

$$\begin{aligned} \ell_c(\boldsymbol{\theta} | \mathbf{y}_c) &= \sum_{i=1}^n \ell(\boldsymbol{\theta} | \mathbf{y}_c) = \sum_{i=1}^n \left[\log f(\mathbf{y}_i | \mathbf{b}_i, \kappa_i; \boldsymbol{\theta}_\epsilon) + \log f(\mathbf{b}_i | \tau_i; \boldsymbol{\theta}_b) + \log h_1(\kappa_i; \nu) + \log h_2(\tau_i; \eta) \right] \\ &= -\frac{1}{2} \sum_{i=1}^n \log |\mathbf{R}_i| - \frac{1}{2} \sum_{i=1}^n \kappa_i (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i)^\top \mathbf{R}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i) \\ &\quad - \frac{1}{2} \sum_{i=1}^n \log |\mathbf{D}| - \frac{1}{2} \sum_{i=1}^n \tau_i \mathbf{b}_i^\top \mathbf{D}^{-1} \mathbf{b}_i + \sum_{i=1}^n \log h_1(\kappa_i; \nu) + \sum_{i=1}^n \log h_2(\tau_i; \eta) + K, \end{aligned}$$

with K being a constant that does not depend on the parameter vector $\boldsymbol{\theta}$ and $\sum_{i=1}^n s_i = N$

Given the current estimate (at the k th iteration $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(k)}$, and the observed data $\mathbf{V}$, $\mathbf{C}$, the

conditional expectation of the complete data log-likelihood function is given by

$$Q(\theta|\hat{\theta}^{(k)}) = E\left[\ell_c(\theta|y_c)|V, C, \hat{\theta}^{(k)}\right] = \sum_{i=1}^n Q_i(\theta|\hat{\theta}^{(k)}), \text{ where}$$

$$\begin{aligned} Q_i(\theta|\hat{\theta}^{(k)}) &= \widehat{\ell h_{1i}}^{(k)} + \widehat{\ell h_{2i}}^{(k)} - \frac{1}{2} \log |\mathbf{D}| - \frac{1}{2} \text{tr} \left(\widehat{\tau \mathbf{b}_i^2}^{(k)} \mathbf{D}^{-1} \right) - \frac{1}{2} \sum_{i=1}^n \log |\mathbf{R}_i| \\ &\quad - \frac{1}{2} \left[\text{tr} \left(\widehat{\kappa \mathbf{y}_i^2}^{(k)} \mathbf{R}_i^{-1} \right) - 2 \boldsymbol{\beta}^\top \mathbf{X}_i^\top \mathbf{R}_i^{-1} \widehat{\kappa \mathbf{y}_i}^{(k)} + 2 \boldsymbol{\beta}^\top \mathbf{X}_i^\top \mathbf{R}_i^{-1} \mathbf{Z}_i \widehat{\kappa \mathbf{b}_i}^{(k)} \right. \\ &\quad \left. - 2 \text{tr} \left(\mathbf{Z}_i^\top \mathbf{R}_i^{-1} \widehat{\kappa \mathbf{y}_i \mathbf{b}_i}^{(k)} \right) + \text{tr} \left(\mathbf{Z}_i^\top \mathbf{R}_i^{-1} \mathbf{Z}_i \widehat{\kappa \mathbf{b}_i^2}^{(k)} \right) + \widehat{\kappa_i}^{(k)} \boldsymbol{\beta}^\top \mathbf{X}_i^\top \mathbf{R}_i^{-1} \mathbf{X}_i \boldsymbol{\beta} \right] + K, \end{aligned}$$

with

$$\begin{aligned} \widehat{\ell h_{1i}}^{(k)} &= E\left[\log h_1(\kappa_i|\nu)|V_i, C_i, \hat{\theta}^{(k)}\right], \quad \widehat{\ell h_{2i}}^{(k)} = E\left[\log h_2(\tau_i|\eta)|V_i, C_i, \hat{\theta}^{(k)}\right] \\ \widehat{\kappa \mathbf{y}_i^2}^{(k)} &= E\left[\kappa_i \mathbf{y}_i \mathbf{y}_i^\top |V_i, C_i, \hat{\theta}^{(k)}\right], \quad \widehat{\kappa \mathbf{y}_i}^{(k)} = E\left[\kappa_i \mathbf{y}_i |V_i, C_i, \hat{\theta}^{(k)}\right], \\ \widehat{\kappa \mathbf{b}_i^2}^{(k)} &= E\left[\kappa_i \mathbf{b}_i \mathbf{b}_i^\top |V_i, C_i, \hat{\theta}^{(k)}\right], \quad \widehat{\kappa \mathbf{b}_i}^{(k)} = E\left[\kappa_i \mathbf{b}_i |V_i, C_i, \hat{\theta}^{(k)}\right], \\ \widehat{\tau \mathbf{b}_i^2}^{(k)} &= E\left[\tau_i \mathbf{b}_i \mathbf{b}_i^\top |V_i, C_i, \hat{\theta}^{(k)}\right], \quad \widehat{\kappa \mathbf{y}_i \mathbf{b}_i}^{(k)} = E\left[\kappa_i \mathbf{y}_i \mathbf{b}_i^\top |V_i, C_i, \hat{\theta}^{(k)}\right], \\ \widehat{\kappa_i}^{(k)} &= E\left[\kappa_i |V_i, C_i, \hat{\theta}^{(k)}\right]. \end{aligned} \quad (8)$$

Note that the aforementioned conditional expectations do not have a closed form.

Consequently, we need to consider two intermediate steps, namely, the simulation (S) and approximation (A) steps. In this context, for the i th subject, a Gibbs sampler step is considered to generate samples from the full conditional distributions of the latent variables $(\mathbf{y}_i, \mathbf{b}_i, \kappa_i, \tau_i)$.

Step S1: Generate $\mathbf{y}_i^{c(k,l)}$ from $f(\mathbf{y}_i^c | \mathbf{V}_i^c, \mathbf{y}_i^o, \mathbf{b}_i^{(k,l-1)}, \kappa_i^{(k,l-1)}, \tau_i^{(k,l-1)}, \hat{\theta}^{(k)})$, considering a partition of $\mathbf{y}_i$ given by $\mathbf{y}_i = \text{vec}(\mathbf{y}_i^o, \mathbf{y}_i^c)$ with $\dim(\mathbf{y}_i^o) = s_i^o$, $\dim(\mathbf{y}_i^c) = s_i^c$ and $s_i^o + s_i^c = s_i$, where $C_{ijk} = 0$ for all elements in $\mathbf{y}_i^o$, and 1 for all elements in $\mathbf{y}_i^c$. In addition, let $\mathbf{V}_i, \mathbf{X}_i, \mathbf{Z}_i$ and $\mathbf{R}_i$

also be partitioned as $\mathbf{V}_i = \text{vec}(\mathbf{V}_i^o, \mathbf{V}_i^c)$, $\mathbf{X}_i^\top = (\mathbf{X}_i^o, \mathbf{X}_i^c)$, $\mathbf{Z}_i^\top = (\mathbf{Z}_i^o, \mathbf{Z}_i^c)$ and $\mathbf{R}_i = \begin{pmatrix} \mathbf{R}_i^{oo} & \mathbf{R}_i^{oc} \\ \mathbf{R}_i^{co} & \mathbf{R}_i^{cc} \end{pmatrix}$.

Then, we have $\mathbf{y}_i | \mathbf{b}_i, \kappa_i \equiv \mathbf{y}_i | \mathbf{b}_i, \kappa_i, \tau_i$ in which $\mathbf{y}_i | \mathbf{b}_i, \kappa_i \sim N_{s_i}(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i, \kappa_i^{-1} \mathbf{R}_i)$, where

$$\mathbf{y}_i^o | \mathbf{b}_i, \kappa_i \sim N_{s_i^o}(\mathbf{X}_i^o \boldsymbol{\beta} + \mathbf{Z}_i^o \mathbf{b}_i, \kappa_i^{-1} \mathbf{R}_i^{oo}) \quad \text{and} \quad \mathbf{y}_i^c | \mathbf{y}_i^o, \mathbf{b}_i, \kappa_i \sim N_{s_i^c}(\boldsymbol{\mu}_i, \kappa_i^{-1} \mathbf{S}_i), \quad (9)$$

with $\mu_i = (\mathbf{X}_i^c \boldsymbol{\beta} + \mathbf{Z}_i^c \mathbf{b}_i) + \mathbf{R}_i^{co} (\mathbf{R}_i^{oo})^{-1} (\mathbf{y}_i^o - \mathbf{X}_i^o \boldsymbol{\beta} - \mathbf{Z}_i^o \mathbf{b}_i)$ and $\mathbf{S}_i = \mathbf{R}_i^{cc} - \mathbf{R}_i^{co} (\mathbf{R}_i^{oo})^{-1} \mathbf{R}_i^{oc}$.

Conditional on the censoring information, we have $f(\mathbf{y}_i^c | \mathbf{V}_i^c, \mathbf{y}_i^o, \mathbf{b}_i, \kappa_i, \tau_i) = \text{TN}_{s_i^c}(\mu_i, \kappa_i^{-1} \mathbf{S}_i, \mathbb{A}_i)$,

which is a truncated normal distribution with

$\mathbb{A}_i = \{\mathbf{y}_i^c = (y_{i1}^c, \dots, y_{is_i^c}^c)^\top | y_{i1}^c \leq V_{i1}^c, \dots, y_{is_i^c}^c \leq V_{is_i^c}^c\}$. The observation

$\mathbf{y}_i^{(k,l)} = (y_{i1}, \dots, y_{is_i^o}, y_{is_i^o+1}^{c(k,l)}, \dots, y_{is_i^{(k,l)}}^{c(k,l)})$ is a sample generated from a truncated normal

distribution for the s_i^c censored cases and the observed values (uncensored cases).

Step S2: Generate $\mathbf{b}_i^{(k,l)}$ from $f(\mathbf{b}_i | \mathbf{y}_i^{(k,l)}, \kappa_i^{(k,l-1)}, \tau_i^{(k,l-1)}, \hat{\boldsymbol{\theta}}^{(k)})$ using the fact that

$$f(\mathbf{b}_i | \mathbf{y}_i, \kappa_i, \tau_i) = N_q(\Psi_i \mathbf{Z}_i^\top \mathbf{R}_i^{-1} \kappa_i (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}), \Psi_i),$$

where $\Psi_i = (\kappa_i \mathbf{Z}_i^\top \mathbf{R}_i^{-1} \mathbf{Z}_i + \tau_i \mathbf{D}^{-1})^{-1}$.

Step S3: Generate $\kappa_i^{(k,l)}$ from $f(\kappa_i | \mathbf{y}_i^{(k,l)}, \mathbf{b}_i^{(k,l)}, \tau_i^{(k,l-1)}, \hat{\boldsymbol{\theta}}^{(k)})$.

Step S4: Generate $\tau_i^{(k,l)}$ from $f(\tau_i | \mathbf{y}_i^{(k,l)}, \mathbf{b}_i^{(k,l)}, \kappa_i^{(k,l)}, \hat{\boldsymbol{\theta}}^{(k)})$.

Remark 1. Note that, since $\mathbf{y}_i | \mathbf{b}_i$ is independent of κ_i and τ_i are mutually independent, we have $f(\kappa_i | \mathbf{y}_i, \mathbf{b}_i, \tau_i) \propto f(\mathbf{y}_i | \mathbf{b}_i, \kappa_i, \boldsymbol{\theta}_\kappa) h_1(\kappa_i | \nu)$ and $f(\tau_i | \mathbf{y}_i, \mathbf{b}_i, \kappa_i) \propto f(\mathbf{b}_i | \tau_i, \boldsymbol{\theta}_\tau) h_2(\tau_i | \eta)$. Thus, for some particular cases, it is possible to compute the full conditional distributions of $(\kappa_i | \mathbf{y}_i, \mathbf{b}_i, \tau_i)$ and $(\tau_i | \mathbf{y}_i, \mathbf{b}_i, \kappa_i)$ (see Tables A2 and A3 given in Appendix A).

Given the sequence $(\mathbf{y}_i^{(k,l)}, \mathbf{b}_i^{(k,l)}, \kappa_i^{(k,l)}, \tau_i^{(k,l)}), l = 1, \dots, m$ and replacing the conditional expectations in (8) by their corresponding stochastic approximations, at the k th iteration, the stochastic approximation step (Step A) is given as follows:

$$\widehat{\kappa \mathbf{y}_i^2}^{(k)} = \widehat{\kappa \mathbf{y}_i^2}^{(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^m \kappa_i^{(k,l)} \mathbf{y}_i^{(k,l)} \mathbf{y}_i^{(k,l)\top} - \widehat{\kappa \mathbf{y}_i^2}^{(k-1)} \right),$$

$$\widehat{\kappa \mathbf{y}_i}^{(k)} = \widehat{\kappa \mathbf{y}_i}^{(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^m \kappa_i^{(k,l)} \mathbf{y}_i^{(k,l)} - \widehat{\kappa \mathbf{y}_i}^{(k-1)} \right),$$

$$\widehat{\kappa \mathbf{b}_i^2}^{(k)} = \widehat{\kappa \mathbf{b}_i^2}^{(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^m \kappa_i^{(k,l)} \mathbf{b}_i^{(k,l)} \mathbf{b}_i^{(k,l)\top} - \widehat{\kappa \mathbf{b}_i^2}^{(k-1)} \right),$$

$$\widehat{\kappa \mathbf{b}}_i^{(k)} = \widehat{\kappa \mathbf{b}}_i^{(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^m \kappa_i^{(k,l)} \mathbf{b}_i^{(k,l)} - \widehat{\kappa \mathbf{b}}_i^{(k-1)} \right),$$

$$\widehat{\kappa \mathbf{y} \mathbf{b}}_i^{(k)} = \widehat{\kappa \mathbf{y} \mathbf{b}}_i^{(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^m \kappa_i^{(k,l)} \mathbf{y}_i^{(k,l)} \mathbf{b}_i^{(k,l)\top} - \widehat{\kappa \mathbf{y} \mathbf{b}}_i^{(k-1)} \right),$$

$$\widehat{\tau \mathbf{b}}_i^{(k)} = \widehat{\tau \mathbf{b}}_i^{(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^m \tau_i^{(k,l)} \mathbf{b}_i^{(k,l)} \mathbf{b}_i^{(k,l)\top} - \widehat{\tau \mathbf{b}}_i^{(k-1)} \right),$$

$$\widehat{\kappa}_i^{(k)} = \widehat{\kappa}_i^{(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^m \kappa_i^{(k,l)} - \widehat{\kappa}_i^{(k-1)} \right),$$

$$\widehat{\ell h}_{1i}^{(k)} = \widehat{\ell h}_{1i}^{(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^m \log h_1(\kappa_i^{(k,l)} | \widehat{\nu}^{(k-1)}) - \widehat{\ell h}_{1i}^{(k-1)} \right),$$

$$\widehat{\ell h}_{2i}^{(k)} = \widehat{\ell h}_{2i}^{(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^m \log h_2(\tau_i^{(k,l)} | \widehat{\eta}^{(k-1)}) - \widehat{\ell h}_{2i}^{(k-1)} \right).$$

As mentioned earlier, an advantage of the SAEM algorithm is that, even though it performs an MCMC E-step, it requires a small and fixed sample size, making it much faster than the MCEM algorithm. Although in the literature it is recommended that $m = 10$ be large enough, we consider $m = 20$ for being more conservative.

The conditional maximization step leads to updating $\widehat{\boldsymbol{\theta}}^{(k)}$ as follows:

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

$$\widehat{\sigma}_{jl}^{(k+1)} = \begin{cases} \left(\sum_{i=1}^n \widehat{\kappa}_i^{(k)} \right) \sum_{i=1}^n \text{tr} \left(\widehat{\Omega}_i^{-1(k)} \widehat{\kappa \boldsymbol{\epsilon}}_{ijl}^{(k)} \right) & \text{for } j = l, \\ \left(2 \sum_{i=1}^n \widehat{\kappa}_i^{(k)} \right)^{-1} \sum_{i=1}^n \text{tr} \left[\widehat{\Omega}_i^{-1(k)} \left(\widehat{\kappa \boldsymbol{\epsilon}}_{ijl}^{(k)} + \widehat{\kappa \boldsymbol{\epsilon}}_{ilj}^{(k)} \right) \right] & \text{for } j \neq l, \end{cases}$$

$$\hat{\phi}^{(k+1)} = \arg \max_{\phi \in (0, 1) \times \mathbb{R}^+} \left\{ -\frac{r}{2} \sum_{i=1}^n \log |\Omega_i(\phi, \mathbf{t}_i)| - \frac{1}{2} \sum_{i=1}^n \text{tr} \left[\hat{\Sigma}^{(k)} \otimes \Omega_i(\phi, \mathbf{t}_i)^{-1} \hat{\kappa} \mathbf{E}_i \right] \right\},$$

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

$$\hat{\nu}^{(k+1)} = \arg \max_{\nu} \sum_{i=1}^n \widehat{\ell h_{1i}}^{(k)}(\nu), \quad \text{and} \quad \hat{\eta}^{(k+1)} = \arg \max_{\eta} \sum_{i=1}^n \widehat{\ell h_{2i}}^{(k)}(\eta),$$

where

$$\begin{aligned} \widehat{\kappa \mathbf{E}}_i &= \widehat{\kappa \mathbf{y}}_i^{2(k)} - 2\mathbf{X}_i \beta \widehat{\kappa \mathbf{y}}_i^{(k)\top} - 2\mathbf{Z}_i \kappa \widehat{\mathbf{y}}_i^{(k)\top} + \widehat{\kappa}_i^{(k)} \mathbf{X}_i \beta \beta^\top \mathbf{X}_i^\top + 2\mathbf{X}_i \beta \widehat{\kappa \mathbf{b}}_i^{(k)\top} \mathbf{Z}_i^\top + \mathbf{Z}_i \widehat{\kappa \mathbf{b}}_i^{2(k)} \mathbf{Z}_i^\top + \widehat{\kappa}_i^{(k)} \mathbf{X}_i \beta \beta^\top \mathbf{X}_i^\top, \quad \text{and} \quad \widehat{\kappa e_{ijl}}^{(k)} = \\ &\widehat{\kappa \mathbf{y}}_{i[j:l]}^{2(k)} - \widehat{\kappa \mathbf{y}}_{ij}^{(k)} \beta^\top \mathbf{X}_{il}^\top - \widehat{\kappa \mathbf{y}}_{i[j:l]}^{(k)} \mathbf{Z}_{il}^\top - \mathbf{X}_{ij} \beta \widehat{\kappa \mathbf{y}}_{il}^{(k)\top} + \widehat{\kappa}_{ij}^{(k)} \mathbf{X}_{ij} \beta \beta^\top \mathbf{X}_{il}^\top + \mathbf{X}_{ij} \beta \widehat{\kappa \mathbf{b}}_{il}^{(k)\top} \mathbf{Z}_{il}^\top - \mathbf{Z}_{ij} \widehat{\kappa \mathbf{y}}_{i[l:j]}^{(k)\top} + \mathbf{Z}_{ij} \widehat{\kappa \mathbf{b}}_{ij}^{(k)} \beta^\top \mathbf{X}_{il}^\top + \mathbf{Z}_{ij} \widehat{\kappa \mathbf{b}}_{i[j:l]}^{2(k)} \\ &\text{for } j, l = 1, \dots, r, \text{ with } \widehat{\kappa \mathbf{y}}_{ij}^{(k)} \text{ being an } n_i \times 1 \text{ subvector consisting of } ((j-1)n_i + 1)\text{th to } (jn_i)\text{th} \\ &\text{entries of } \widehat{\kappa \mathbf{y}}_i^{(k)}; \widehat{\kappa \mathbf{b}}_{ij}^{(k)} \text{ being a } q_j \times 1 \text{ subvector consisting of } (\sum_{\nu \leq (j-1)q_\nu + 1})\text{th to } \sum_{\nu \leq j} q_\nu \text{th} \\ &\text{entries of } \widehat{\kappa \mathbf{b}}_i^{(k)}; \widehat{\kappa \mathbf{y}}_{i[j:l]}^{2(k)} \text{ being an } n_i \times n_l \text{ submatrix consisting of the } ((j-1)n_i + 1) \text{th to } (jn_i)\text{th} \\ &\text{rows and the } ((l-1)s_l + 1)\text{th to } (ls_l)\text{th columns of } \widehat{\kappa \mathbf{y}}_i^{2(k)}; \widehat{\kappa \mathbf{y}}_{i[j:l]}^{(k)} \text{ is a } s_i \times q_l \text{ submatrix} \\ &\text{consisting of the } ((j-1)s_i + 1)\text{th to } (js_i)\text{th rows and the } (\sum_{\nu \leq (l-1)q_\nu + 1})\text{th to } \sum_{\nu \leq l} q_\nu \text{th} \\ &\text{columns of } \widehat{\kappa \mathbf{y}}_i^{(k)} \text{ and } \widehat{\kappa \mathbf{b}}_{i[j:l]}^{2(k)} \text{ is a } q_j \times q_l \text{ submatrix consisting of } (\sum_{\nu \leq (j-1)q_\nu + 1})\text{th to } \\ &\sum_{\nu \leq j} q_\nu \text{th to rows and } (\sum_{\nu \leq (l-1)q_\nu + 1})\text{th to } \sum_{\nu \leq l} q_\nu \text{th columns of } \widehat{\tau \mathbf{b}}_i^{2(k)}. \end{aligned}$$

Depending on the distribution of the conditional response vector and the random effects, we have closed-form expressions for the parameter estimates $\hat{\nu}$ and $\hat{\eta}$. In Appendix B, we present these expressions.

Remark 2. If parameters ν and η associated with the mixture variables κ_i and τ_i ,

respectively, are known, we have that $\widehat{\kappa \mathbf{y}}_i^{2(k)}$, $\widehat{\kappa \mathbf{y}}_i^{(k)}$, and $\widehat{\kappa}_i^{(k)}$ depend on the minimal sufficient statistics because the complete data likelihood function belongs to the exponential family. Therefore, the parameter estimates of the SAEM algorithm converge, as mentioned in the works of Kuhn and Lavielle²⁵ and Samson et al.²⁶ However, in this paper, we consider these parameters to be unknown. Then, $\widehat{\kappa \mathbf{y}}_i^{2(k)}$, $\widehat{\kappa \mathbf{y}}_i^{(k)}$, and $\widehat{\kappa}_i^{(k)}$ do not depend on minimal sufficient statistics and we cannot guarantee convergence of the parameter estimates. To deal with this problem, we conduct a simulation study to examine the empirical performance of the SAEM algorithm.

3.3 | Imputation of censored components

Let $\mathbf{y}_i^c$ be the true unobserved response vector for the censored components. Now, as a byproduct of the SAEM algorithm, the prediction of the censored components is given by the following expression:

$$\tilde{\mathbf{y}}_i^{c(k)} = \tilde{\mathbf{y}}_i^{c(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^m \mathbf{y}_i^{c(k,l)} - \tilde{\mathbf{y}}_i^{c(k)} \right).$$

Note that $\mathbf{y}_i^{c(k,l)}$ is obtained from **Step S1** of the proposed SAEM algorithm.

4 | ESTIMATION OF THE LIKELIHOOD AND STANDARD ERRORS

4.1 | Likelihood estimation

The likelihood function for the observed data can be computed as

$$L_o(\boldsymbol{\theta}; \mathbf{y}^{\text{obs}}) = \prod_{i=1}^n \int_{\mathbb{R}^q} \left[\int_0^\infty f(\mathbf{y}_i | \mathbf{b}_i, \kappa_i; \boldsymbol{\theta}_\epsilon) h_1(\kappa_i; \nu) d\kappa_i \right] f(\mathbf{b}_i; \boldsymbol{\theta}_b, \eta) d\mathbf{b}_i,$$

where $f(\mathbf{b}_i; \boldsymbol{\theta}_b, \eta) = \int_0^\infty f(\mathbf{b}_i | \tau_i; \boldsymbol{\theta}_b) h_2(\tau_i; \eta) d\tau_i$, for $i = 1, \dots, n$.

Partitioning $\mathbf{y}_i$ as in (9), we obtain

$$\begin{aligned} L_o(\boldsymbol{\theta}; \mathbf{y}^{\text{obs}}) &= \prod_{i=1}^n \int_{\mathbb{R}^q} \left[\int_0^\infty \phi_{s_i^o}(\mathbf{y}_i^o; \mathbf{X}_i^c \boldsymbol{\beta} - \mathbf{Z}_i^c \mathbf{b}_i, \kappa_i^{-1} \mathbf{R}_i^{oo}) \Phi_{s_i^c}(\mathbf{V}_i^c; \boldsymbol{\mu}_i, \kappa_i^{-1} \mathbf{S}_i) h_1(\kappa_i; \nu) d\kappa_i \right] f(\mathbf{b}_i; \boldsymbol{\theta}_b, \eta) d\mathbf{b}_i \\ &= \prod_{i=1}^n \int_{\mathbb{R}^q} g(\mathbf{y}_i | \mathbf{b}_i, \kappa_i; \boldsymbol{\theta}_\epsilon) f(\mathbf{b}_i; \boldsymbol{\theta}_b, \eta) d\mathbf{b}_i, \end{aligned} \quad (10)$$

where $g(\mathbf{y}_i | \mathbf{b}_i, \kappa_i; \boldsymbol{\theta}_\epsilon) = \int_0^\infty \phi_{s_i^o}(\mathbf{y}_i^o; \mathbf{X}_i^c \boldsymbol{\beta} - \mathbf{Z}_i^c \mathbf{b}_i, \kappa_i^{-1} \mathbf{R}_i^{oo}) \Phi_{s_i^c}(\mathbf{V}_i^c; \boldsymbol{\mu}_i, \kappa_i^{-1} \mathbf{S}_i) h_1(\kappa_i; \nu) d\kappa_i$.

The integral involved in (10) can be computed using an importance sampling strategy for any continuous distribution f^* . In fact, we have

$$L_o(\boldsymbol{\theta}; \mathbf{y}^{\text{obs}}) = \prod_{i=1}^n \int_{\mathbb{R}^q} g(\mathbf{y}_i | \mathbf{b}_i, \kappa_i; \boldsymbol{\theta}_\epsilon) \frac{f(\mathbf{b}_i; \boldsymbol{\theta}_b, \eta)}{f^*(\mathbf{b}_i; \boldsymbol{\theta}_b, \eta)} d\mathbf{b}_i,$$

where the distribution f^* is the importance distribution. Consequently, $L_o(\theta; \mathbf{y}_i^{\text{obs}})$ is estimated through the following approximation:

$$L_o(\theta; \mathbf{y}^{\text{obs}}) = \prod_{i=1}^n \left[\frac{1}{M} \sum_{m=1}^M g(\mathbf{y}_i | \mathbf{b}_{im}, k_i; \theta_e) \frac{f(\mathbf{b}_{im}; \theta_b, \eta)}{f^*(\mathbf{b}_{im}; \theta_b, \eta)} \right],$$

with $\mathbf{b}_{i1}, \dots, \mathbf{b}_{im}$ being drawn from $f^*(\mathbf{b}_i; \theta_b, \eta)$, which is chosen to be a normal distribution with empirical mean and empirical variance calculated from draws of $\mathbf{b}_i$ obtained through the SAEM outputs (Step S2).

4.2 | Model selection criteria

Following the idea of Zhang et al,²⁷ our goal is to assess the fit of each longitudinal outcome of the joint model. In this context, the model selection criteria used in this work are AIC = $2m - 2\ell_{\max}$ and BIC = $m \log N - 2\ell_{\max}$, where m is the number of model parameters, $N = \sum_{i=1}^n s_i$, and $\ell_{\max}$ is the maximized log-likelihood value.

To assess the contribution of each longitudinal outcome to the overall fit of the joint model, we need to decompose both model selection criteria into two parts. Let $\mathbf{y}_{i1}^* = (\mathbf{y}_{i1}^T, \dots, \mathbf{y}_{ir}^T)^T$ and $\mathbf{y}_{i2}^* = (\mathbf{y}_{ir+1}^T, \dots, \mathbf{y}_{is}^T)^T$, where $\mathbf{y}_i = (\mathbf{y}_{i1}^T, \mathbf{y}_{i2}^T)^T$ and $r \in \{1, \dots, s\}$. Furthermore, let $f(\mathbf{y}_{i1}^*; \theta_1^*)$ be the marginal density of $\mathbf{y}_{i1}^*$ and $f(\mathbf{y}_{i2}^* | \mathbf{y}_{i1}^*; \theta_2^*, \theta_1^*)$ be the conditional density of the partition $\mathbf{y}_{i2}^*$ given $\mathbf{y}_{i1}^*$, where $\ell(\theta_1^* | \mathbf{y}_i)$ and $\ell(\theta_2^* | \mathbf{y}_i)$ are the log-likelihood functions of $\mathbf{y}_{i2}^* | \mathbf{y}_{i1}^*$ and $\mathbf{y}_{i1}^*$, respectively (see Appendix C for more details about this distribution). Then, the AIC and BIC have the following decompositions:

$$\text{AIC} = \text{AIC}_{\mathbf{y}_1^*} + \text{AIC}_{\mathbf{y}_2^* | \mathbf{y}_1^*} \text{ and } \text{BIC} = \text{BIC}_{\mathbf{y}_1^*} + \text{BIC}_{\mathbf{y}_2^* | \mathbf{y}_1^*},$$

where $\text{AIC}_{\mathbf{y}_1^*} = 2 \dim(\theta_1^*) - 2 \sum_{i=1}^n \ell(\widehat{\theta}_1^* | \mathbf{y}_i)$, $\text{AIC}_{\mathbf{y}_2^* | \mathbf{y}_1^*} = 2 \dim(\theta_2^*) - 2 \sum_{i=1}^n \ell(\widehat{\theta}_2^* | \mathbf{y}_i)$, $\text{BIC}_{\mathbf{y}_1^*} = \dim(\theta_1^*) \log N - 2 \sum_{i=1}^n \ell(\widehat{\theta}_1^* | \mathbf{y}_i)$, $\text{BIC}_{\mathbf{y}_2^* | \mathbf{y}_1^*} = \dim(\theta_2^*) \log N - 2 \sum_{i=1}^n \ell(\widehat{\theta}_2^* | \mathbf{y}_i)$ and $\widehat{\theta}_1^*$ and $\widehat{\theta}_2^*$ are the estimates of θ . Following the work of Zhang et al,²⁷

$$\dim(\theta_2^*) = \dim(\theta) - \dim(\theta_1^*)$$

Finally, we define the model assessment criteria

$$\Delta \text{AIC} = \text{AIC}_{\mathbf{y}_{2,0}^*} - \text{AIC}_{\mathbf{y}_2^* | \mathbf{y}_1^*} \text{ and } \Delta \text{BIC} = \text{BIC}_{\mathbf{y}_{2,0}^*} - \text{BIC}_{\mathbf{y}_2^* | \mathbf{y}_1^*}, \quad (11)$$

where $AIC_{y_{2,0}^\star}$ and $BIC_{y_{2,0}^\star}$ are calculated considering only $y_{2,0}^\star$. The model with a large value of AIC or BIC fits the data better.

4.3 | Empirical information matrix

Following the work of Matos et al,⁵ the individual score can be determined as

$$s(y_i|\theta) = \frac{\partial \log f(y_i|\theta)}{\partial \theta} = E \left(\frac{\partial \ell_{ic}(\theta|y_{ci})}{\partial \theta} | V_i, C_i, \theta \right) \text{ where } \ell_{ic}(\theta|y_{ci}) \text{ is the complete data log-}$$

likelihood formed from the observation y_{ci} (see the work of Louis²⁸). Then, the empirical information matrix is given by

$$I_e(\hat{\theta}|\mathbf{y}) = \sum_{i=1}^n \hat{s}_i \hat{s}_i^\top, \quad (12)$$

where

$$\hat{s}_i = \left(\hat{s}_{i,\beta_1}, \dots, \hat{s}_{i,\beta_p}, \hat{s}_{i,\sigma_1^2}, \dots, \hat{s}_{i,\sigma_r^{*2}}, \hat{s}_{i,\alpha_1}, \dots, \hat{s}_{i,\alpha_{q*}}, \hat{s}_{i,\phi_1}, \hat{s}_{i,\phi_2}, \hat{s}_{i,\nu}, \hat{s}_{i,\eta} \right)^\top. \quad (13)$$

The derivatives of $\hat{s}_i$ are given in Appendix D.

5 | SIMULATION STUDIES

In this section, we present the results of three simulation studies to show the behavior of our proposed method under different situations. In the following, we describe the simulation scenarios.

- **Simulation I:** This study provides an extensive simulation scheme designed to examine the empirical performance of the parameter estimates under different specifications of the parameters ν and η . Recall that those parameters are involved in the distribution of the mixture variables, controlling the heavy-tailed behavior of our resulting model, say, the SMN-MLMEC model.
- **Simulation II:** The second simulation scheme focuses on the study of the finite sample properties of the parameter estimates under different censoring proportions. We generate random samples from a SL/SL-MLMEC model, ie, we consider a SMN-MLMEC model where the distribution of the error term and the random effect follows a slash distribution. In this scheme, we fit several heavy-tailed SMN-MLMEC models and compare them using the AIC and BIC measures.
- **Simulation III:** Finally, the third simulation scheme studies the effect of the misspecification of the distribution of the error and random effect terms on the parameter estimates. As in the previous scenario, we draw random samples from

an SL/SL-MLMEC model, fitting different heavy-tailed SMN-MLMEC models, including the normal one. We compare these models using the AIC measures and log-likelihood values.

It is important to remark that, for all of these studies, we considered the following values of the model parameters: $\beta = (6, -0.2, 4, -0.1)^\top$, $\Omega_i = \Omega_A(0.8, 1)$, t_i , $\Sigma = \begin{pmatrix} 4 & 2 \\ 2 & 4 \end{pmatrix}$, and

$D = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$. For **Simulation I**, we consider $\nu = 3$ and $\eta = 3$, whereas, for **Simulation II**,

parameters ν and η were fixed at values 2 and 3, respectively. Finally, for **Simulation III**, we consider $\nu = 5$ and $\eta = 1.5$. The time points t_{ij} 's at which the longitudinal measures were taken were fixed at 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10. The design matrix of the fixed effects for each outcome X_{ij} contains the intercept and time point t_{ij} and the design matrix of random effects Z_{ij} contains only the intercept component. Therefore, $\mathbf{X}_i = \mathbf{I}_2 \otimes X_{ij}$ and $\mathbf{Z}_i = \mathbf{I}_2 \otimes Z_{ij}$, with $\mathbf{I}_2$ being the identity matrix of dimension 2×2 .

5.1 | Simulation I: effect of the misspecification of ν and η on the parameter estimates

In this scheme, we consider three types of models, namely, M1, M2, and M3, where M1 is the model with the parameters ν and η unknown; M2 is the model with ν and η fixed at the true values; and M3 is the model with ν and η fixed at different values than the true ones. Our goal in this study is to compare how the estimates are affected when the parameters associated with the mixture variables are either estimated or fixed.

We generated 600 simulated datasets from a bivariate longitudinal model with censored responses, considering the slash distribution as the distribution of the error term and the random effect (SL/SL-MLMEC model). Each dataset includes the information of $n = 50$ subjects.

Note that the specification of Ω_i generates an AR(1) structure. Additionally, we considered two different schemes of left censoring proportions, say, 10% and 20% in each dataset. For the SAEM algorithm, the maximum number of iterations was $W = 300$ and a cutoff point $c = 0.25$. It is important to stress that the value of c between 0 and 1 ($0 < c < 1$) will assure an initial convergence in distribution to a solution neighborhood for the first cW iterations and an almost sure convergence for the rest of the iterations (see the work of Galarza et al²³). In addition, and as suggested by Lavielle,²⁹ we consider a graphical approach for monitoring the convergence of the estimates for all the parameters and for determining the value of W and c .

The model selection criteria (AIC and BIC) as well as the estimates of the model parameters were computed for each simulation. Summary statistics such as the Monte Carlo mean estimate (MC mean), the mean of the approximate standard errors obtained through the information-based method described in Section 4.3 (IM SE) and the Monte Carlo empirical standard errors (MC Sd) are presented in Table 1. It can be observed from Table 1 that, in all models, the estimates of the fixed effects are closer to the true values. Moreover, in general, the estimates of the variance components are overestimated in model M3. In fact, these estimates are always greater than those obtained from models M1 and M2. A possible

explanation for this is that model M3 considers the parameters $\nu = 8$ and $\eta = 8$ fixed at different values than the true ones. Finally, we see that the values of MC Sd and IM SE are close.

We also analyzed the absolute bias (Bias) and mean square error (MSE) of the fixed effects and variance component estimates obtained from the models under consideration. In this case, our aim was to compare the Bias and MSE of the ML estimates from all models, showing that they are the same when the parameters ν and η (those parameters associated with the mixture variables) are either estimated or fixed. The Bias and MSE measures are defined as

$$\text{Bias} = \frac{1}{J} \sum_{j=1}^J |\hat{\theta}_i^{(j)} - \theta_i| \quad \text{and} \quad \text{MSE} = \frac{1}{J} \sum_{j=1}^J \left(\hat{\theta}_i^{(j)} - \theta_i \right)^2, \quad (14)$$

where $\hat{\theta}_i^{(j)}$ is the ML estimate of the parameter θ_i for the j th sample, $j = 1, \dots, J$.

Table 1 also shows the Bias and MSE of the parameter estimates. It can be seen that these measures are similar in the case of the fixed effects for all models under study. However, the Bias and MSE are higher in the case of the variance components, specifically in model M3. This fact could reveal the effect of the misspecification related to parameters and. It is important to note that similar results are obtained for the two levels of censoring proportions considered in the study.

Figure 2 shows the boxplots corresponding to the parameter estimates over the 600 simulated datasets. Note that the estimates of the three models are, in general, equivalent. However, the estimates of variance components obtained from models M1 and M2 are generally closer to the true values than those obtained in model M3. Similar results are obtained after the analysis of the data with 10% of censoring (See Appendix E, Figure E1).

Finally, note that the model M2 presents the lowest values of Bias and MSE of the parameter estimates under both censoring proportions. Moreover, the IM SE and MC Sd are always smaller than the ones obtained under the model M1. These results are not surprising since the model M2 considers that the parameters ν and η are fixed at their true values.

5.2 | Simulation II: finite sample properties under different censoring proportions

As mentioned earlier, the goal of this simulation study is to investigate the effects on the parameter estimates under different proportions of censoring and sample sizes. For this purpose, we consider the following SMN-MLMEC models.

1. The N/N-MLMEC model, where the distribution of both the error term ϵ and random effect $\mathbf{b}$ is the normal distribution.
2. The T/T-MLMEC model, where the distribution of both the error term ϵ and random effect $\mathbf{b}$ is the Student- t distribution.

3. The SL/SL-MLMEC model, where the distribution of both the error term ϵ and random effect $\mathbf{b}$ is the slash distribution.

To study the finite sample properties, we compute the MSE (previously defined in Section 5.1) of the parameter estimates obtained after fitting the N/N-, T/T-, and SL/SL-MLMEC models under different sample sizes. The results are presented in Table 2 and Figure 3.

Table 2 shows that the point estimates for the fixed effects and the parameter ϕ_1 are similar for all models considered. However, if we center our attention on the elements to the matrices Σ and $\mathbf{D}$, it can be seen that the SL/SL-MLMEC model (the true model) provides the closest estimates to the true values, even when the sample size is small ($n = 50$) and under both censoring proportions. The T/T-MLMEC model provides better estimates of the matrices Σ and $\mathbf{D}$ when the sample size increases compared to those obtained under normality. This situation is also observed under both censoring proportions considered in this study. Finally, by taking into account the values of the AIC and BIC, we conclude that the heavy-tailed models outperform the N/N MLMEC model when the real dataset is generated from the heavy-tailed SL/SL-MLMEC model.

Figure 3 presents the values of the MSE for all parameter estimates under different sample sizes and censoring proportions. As it can be seen, the N/N-MLMEC model provides the highest values of the MSE in all cases. Finally, we conclude that, based on the empirical evidence, the point estimates of our heavy-tailed SMN-MLMEC models are consistent.

5.3 | Simulation III: effect of the misspecification of the distribution of the error term and random effects

The third simulation scheme studies the effect of the misspecification of the distributional assumption on the error term and random effects. In order to do this, we consider the slash distribution as the distribution of both the error and random effect terms (SL/SL MLMEC model). We generate 100 simulated datasets considering a censoring level of 10%. For each dataset, we consider $n = 100$ subjects.

To study the effect of the misspecification, we consider the following SMN-MLMEC models.

- The N/T-MLMEC model, where the distribution of error term ϵ is normal and the one associated with the random effect $\mathbf{b}$ is the Student- t distribution.
- The N/SL-MLMEC model, where the distribution of error term ϵ is normal and the one associated with the random effect $\mathbf{b}$ is slash.
- The SL/N-MLMEC model, where the distribution of error term ϵ is slash and the one associated with the random effect $\mathbf{b}$ is normal.
- The SL/T-MLMEC model, where the distribution of error term ϵ is slash and the one associated with the random effect $\mathbf{b}$ is the Student- t distribution.

In addition, we consider the N/N- and SL/SL-MLMEC models. Note that the SMN-MLMEC models provide a wide range of heavy-tailed models obtained from different

combinations of the distributions of the error term and random effects. The results are presented in Tables 3 and 4.

Table 3 provides model comparison measures for the SMN-MLMEC models described earlier. According to the AIC value, the best model is the SL/T-MLMEC model. However, the AIC value for this model is very similar to the value obtained under the SL/SL-MLMEC model (the true model). It is important to stress that the N/N-MLMEC provides the worst fit compared to the other heavy-tailed models. Consequently, we conclude that, when the true distribution of the error term and random effects is a heavy-tailed one, the SMN-MLMEC models outperform the normal one.

Parameter estimates as well as standard errors (Monte Carlo and information-based standard errors) are reported in Table 4. Note that the fixed effects estimates are closer to the true values in all models considered in the simulation. However, the standard errors obtained under the N/N-MLMEC model are larger than the ones obtained under SMN-MLMEC models with heavy tails. Other parameter estimates also present larger standard errors under the N/N-MLMEC model.

6 | ANALYSIS OF A5055 CLINICAL TRIAL

We apply the proposed method to carry out an analysis of an HIV dataset, previously analyzed by Wang.⁹ Specifically, we study the HIV viral load data from the clinical trial A5055, considering two different correlation structures, namely, the uncorrelated structure (UNC) and damped exponential correlation (DEC); and different distributions belonging to the SMN class. As discussed in Section 1, the dataset consists of 44 HIV-1 infected patients treated with two potent ARV therapies (treatment 1 and treatment 2). The viral load was quantified on days 0, 7, 14, 28, 56, 84, 112, 140, and 168 of follow-up for each patient, the immunologic markers CD4 and CD8 were also measured. The dataset has 316 observations, where the viral load detection limit is 50 copies/mL, and 106 out of 316 (33%) of all viral load measurements are below the detection limit.

Before fitting the model, we applied a log transformation to the ratio CD4/CD8 to obtain a symmetric distribution. Figure E2 (Appendix E) shows the histogram for the variable CD4/CD8 and log transformation of CD4/CD8.

Let y_{1k} and y_{2k} be $\log_{10}(\text{RNA})$ and $\log(\text{CD4/CD8})$, respectively, for subject i ($i = 1, \dots, 44$) measured at day ik . The model considered for the A5055 dataset is given by

$$y_{1k} = \beta_{10} + \beta_{11}t_{ik} + \beta_{12}\text{treat}_i + \beta_{13}t_{ik}^{0.5} + \beta_{14}\text{treat}_i \times t_{ik} + b_{i10} + b_{i11}t_{ik} + e_{i1k}, \quad (15)$$

$$y_{2k} = \beta_{20} + \beta_{21}t_{ik} + \beta_{22}\text{treat}_i + \beta_{23}\text{treat}_i \times t_{ik} + b_{i20} + b_{i21}t_{ik} + e_{i2k}, \quad (16)$$

where $t_{ik} = \text{day}_{ik}/7$ (week) for $k = 1, \dots, s_i$; treat_i is a treatment indicator ($= 0$ for treatment 1; $= 1$ for treatment 2); $t_{ik}^{0.5}$ is the square root at time t_{ik} ; and b_{ij0} and b_{ij1} are the random intercept and random slope, respectively, for y_{ijk} , $j = 1, 2$. In a preliminary analysis, we fitted several models with different covariate structures and random effects, and the best model was the one given in Equations (15) and (16).

To specify the distribution of the error term and the random effects, we considered many combinations of possible distributions. Table 5 shows the AIC and BIC values obtained for these combinations. From this table, we conclude that the best model is the one with the Student- t distribution for the error and the normal distribution for the random effects.

Figure 4 presents the QQ-plots of the residuals and estimates for the random effects under the N/N-MLMEC model. Note that, for this model, both the residuals and random effects estimates show substantial departure from normality.

Table 6 presents the ML estimates of the parameters of interest under the best model, considering two different correlation structures, namely, the uncorrelated (UNC) and DEC structures. This table also presents the corresponding standard errors of the parameter estimates. It can be observed that, according to the AIC or BIC values, the model with the DEC structure provides a much better fit, outperforming the one with the UNC structure. In fact, the maximum log-likelihood was -357.97 under the UNC and -344.79 under the DEC structure, which gives a likelihood ratio statistic of 26.36 ($p\text{-value} < 0.0001$). Consequently, the model with the DEC structure fits the data substantially better than the model with UNC structure. Moreover, for those models, the estimated ν values of are fairly small, indicating a lack of adequacy of the normal assumption for the error term.

Note that, under the best model (T/N-MLMEC model with the DEC structure), the estimate of β_{11} (0.130) reveals that RNA viral loads change over time. From the estimate of β_{22} (0.252), we conclude that patients receiving different treatments have difference in baseline $\log(\text{CD4}/\text{CD8})$, whereas, from the estimate -0.005 of β_{12} , we do not have significant difference of baseline RNA viral loads. Moreover, we do not find significant evidence related to the treatment and time and interaction ($\hat{\beta}_{14} = -0.007$ with $\text{SE} = 0.025$ and $\hat{\beta}_{23} = -0.003$ with $\text{SE} = 0.007$). For our best model, the convergences of the estimates obtained through the SAEM algorithm are shown in Figures E4, E5, and E6 in Appendix E. These figures show that convergence can be achieved very quickly.

Table 7 shows the AIC and BIC decompositions described in Section 4.2. These measures (AIC and BIC) show that the contribution of considering a joint model ($\log(\text{CD4}/\text{CD8})$ with RNA viral load) is significant.

Finally, to illustrate the performance of the fitted model, Figure E3 (Appendix E) shows the fitted values of repeated measures $\log_{10}(\text{RNA})$ vs $\log(\text{CD4}/\text{CD8})$ for some selected individuals. Here, the fitted results under our best model (red lines) are displayed along with the true observed values (black lines). As expected, the fitted values are close to the observed values. Moreover, Figure 5 shows that the best model accommodates the outliers or influential observations in a better way than the normal model.

7 | CONCLUSIONS

In this paper, we have introduced a robust multivariate linear mixed model for multiple censored responses based on the class of SMN distributions, extending the recent proposal of Wang et al.¹⁰ To model the autocorrelation existing among irregularly observed measures, a damped exponential correlation structure was adopted, as proposed by Muñoz et al.²⁴ The main advantage of the proposed SMN-MLMEC model is that it can reduce the negative impact of distributional misspecification and outliers on the parameter estimation. Moreover, the SMN class allows a convenient framework for implementing the SAEM algorithm, leading to an efficient ML estimation of model parameters. An additional characteristic of our proposed model is that it allows considering different distributions for the error terms, thereby overcoming the aforementioned limitation of the MLMEC model and broadening the scope of censored mixed models. Based on the work of Zhang et al.,²⁷ we also proposed a decomposition of the AIC and BIC, which was shown to be very effective to assess the effect of the joint modeling of censored outcomes. The computing code is available from us on request.

Although the SMN-MLMEC considered here has great flexibility to model symmetric data, its robustness against outliers can be seriously affected by the presence of skewness. Recently, Bandyopadhyay et al.³⁰ proposed a remedy to accommodate skewness and heavy tails simultaneously in the context of LMEC models by using scale mixtures of skew-normal distributions. We conjecture that our methods can be used in MLMEC models and should yield satisfactory results at the expense of additional complexity in the implementation. An in-depth investigation of such extensions is beyond the scope of this paper, but it is an interesting topic for further research.

ACKNOWLEDGEMENTS

We are grateful to two anonymous referees and the associate editor for very useful comments and suggestions, which greatly improved this paper. L. Matos acknowledges support from FAPESP-Brazil (grants 2016/05420–6 and 2018/05013–7). V.H. Lachos acknowledges the support from FAPESP-Brazil (grant 2018/05013–7). Dr. M.-H. Chen's research was partially supported by NIH grants #GM 70335 and #P01 CA142538. L.M. Castro acknowledges support from grant FONDECYT 1170258 from the Chilean government and Iniciativa Científica Milenio - Minecon Núcleo Milenio MiDaS.

Funding information

Fundação de Amparo à Pesquisa do Estado de São Paulo, Grant/Award Number: 2016/05420–6 and 2018/05013–7; National Institutes of Health, Grant/Award Number: GM 70335 and P01 CA142538; Fondo Nacional de Desarrollo Científico y Tecnológico, Grant/Award Number: 1170258; Iniciativa Científica Milenio - Minecon Núcleo Milenio MiDaS

APPENDIX A: SMN CLASS OF DISTRIBUTIONS AND FULL CONDITIONAL DISTRIBUTIONS

TABLE A1

Members of the scale mixtures of normal (SMN) class of distributions

SMN distribution	$\xi(u)$	$h(u \boldsymbol{\nu})$	$f(\mathbf{y} \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu})$
$T_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$	u^{-1}	$\text{Gamma}(\nu/2, \nu/2)$	$\frac{\Gamma(\frac{p+\nu}{2})}{\Gamma(\frac{\nu}{2})\pi^{p/2}} \nu^{-p/2} \boldsymbol{\Sigma} ^{-1/2} \left(1 + \frac{d}{\nu}\right)^{-(p+\nu)/2}$
		$d = (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}), \nu > 0$	
$SL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$	u^{-1}	$\text{Beta}(\nu, 1)$	$\nu \int_0^1 u^{\nu-1} \phi_p(\mathbf{y}; \boldsymbol{\mu}, u^{-1} \boldsymbol{\Sigma}) du, \quad \nu > 0$
$CN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu_1, \nu_2)$	u^{-1}	$\nu_1 \prod_{\{\nu_2\}}(u) + (1 - \nu_1) \prod_{\{1\}}(u)$ $\nu_1, \nu_2 \in (0, 1)$	$\nu_1 \phi_p(\mathbf{y}; \boldsymbol{\mu}, \nu_2^{-1} \boldsymbol{\Sigma}) + (1 - \nu_1) \phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$

TABLE A2

Full conditional distributions of $\boldsymbol{\kappa}_i | \mathbf{y}_i, \mathbf{b}_i, \boldsymbol{\tau}_i$

Distribution of ϵ_i	Distribution of $\boldsymbol{\kappa}_i$	Distribution of $\boldsymbol{\tau}_i \mathbf{y}_i, \mathbf{b}_i, \boldsymbol{\tau}_i$
$T_{s_i}(\mathbf{0}, \mathbf{R}_i, \nu)$	$\text{Gamma}(\nu/2, \nu/2)$	$\text{Gamma}\left((\nu + s_i)/2, (D_{e_i}^2 + \nu)/2\right)$ $D_{e_i}^2 = (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i)^\top \mathbf{R}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i)$
$SL_{s_i}(\mathbf{0}, \mathbf{R}_i, \nu)$	$\text{Beta}(\nu, 1)$	$\text{TGamma}\left(\nu + s_i/2, D_{e_i}^2/2, 1\right)$
$CN_{s_i}(\mathbf{0}, \mathbf{R}_i, \nu_1, \nu_2)$	$\nu_1 \prod_{\{\nu_2\}}(\kappa_i) + (1 - \nu_1) \prod_{\{1\}}(\kappa_i)$	$P(\kappa_i = \nu_2) = 1 - P(\nu_i = 1) = p_1/p_1 + p_2$ $p_1 = \nu_1 \nu_2^{s_i/2} \exp\left\{-\frac{1}{2} D_{e_i}^2 \nu_2\right\}$ $p_2 = (1 - \nu_1) \exp\left\{-\frac{1}{2} D_{e_i}^2\right\}$

TABLE A3

Full conditional distributions of $\boldsymbol{\tau}_i | \mathbf{y}_i, \mathbf{b}_i, \boldsymbol{\kappa}_i$

Distribution of $\mathbf{b}_i$	Distribution of $\boldsymbol{\tau}_i$	Distributions of $\boldsymbol{\tau}_i \mathbf{y}_i, \mathbf{b}_i, \boldsymbol{\kappa}_i$
$T_q(\mathbf{0}, \mathbf{D}, \eta)$	$\text{Gamma}(\eta/2, \eta/2)$	$\text{Gamma}\left((\eta + q)/2, (D_{\mathbf{b}_i}^2 + \eta)/2\right)$ $D_{\mathbf{b}_i}^2 = \mathbf{b}_i^\top \mathbf{D}^{-1} \mathbf{b}_i$
$SL_q(\mathbf{0}, \mathbf{D}, \eta)$	$\text{Beta}(\eta, 1)$	$\text{TGamma}\left(\eta + q/2, D_{\mathbf{b}_i}^2/2, 1\right)$

Distribution of $\mathbf{b}_i$	Distribution of $\boldsymbol{\tau}_i$	Distributions of $\boldsymbol{\tau}_i \mathbf{y}_i, \mathbf{b}_i, \boldsymbol{\kappa}_i$
$\text{CN}_q(\mathbf{0}, \mathbf{D}, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$	$\eta_1 \prod_{\{\eta_2\}} (\boldsymbol{\tau}_i) + (1 - \eta_1) \prod_{\{1\}} (\boldsymbol{\tau}_i)$	$P(\boldsymbol{\tau}_i = \eta_2) = 1 - P(\boldsymbol{\tau}_i = 1) = q_1/q_1 + q_2$ $q_1 = \eta_1 \eta_2^{q/2} \exp\left\{-\frac{1}{2} D_{\mathbf{b}_i}^2 \eta_2\right\}$ $q_2 = (1 - \eta_1) \exp\left\{-\frac{1}{2} D_{\mathbf{b}_i}^2\right\}$

APPENDIX B: MAXIMUM LIKELIHOOD ESTIMATION

In this appendix, we derive the parameters estimates, ie, $\hat{\nu}$ and $\hat{\eta}$, for the some particular distributions. To find the estimates of $\hat{\nu}$ and $\hat{\eta}$, we just need to solve

$$\frac{\partial Q(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}})}{\partial \nu} = \frac{\partial}{\partial \nu} \sum_{i=1}^n \widehat{\ell h_{1i}} = 0 \quad \text{and} \quad \frac{\partial Q(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}})}{\partial \eta} = \frac{\partial}{\partial \eta} \sum_{i=1}^n \widehat{\ell h_{2i}} = 0,$$

where $\widehat{\ell h_{1i}} = E[\log h_1(k_i | \nu) | \mathbf{V}_i, \mathbf{C}_i, \hat{\boldsymbol{\theta}}]$ and $\widehat{\ell h_{2i}} = E[\log h_2(\boldsymbol{\tau}_i | \eta) | \mathbf{V}_i, \mathbf{C}_i, \hat{\boldsymbol{\theta}}]$.

- If the response vector has a Student- t distribution, then $\boldsymbol{\kappa}_i \sim \text{Gamma}(\nu/2, \nu/2)$, and

$$\widehat{\ell h_{1i}} = -\log\left(\Gamma\left(\frac{\nu}{2}\right)\right) + \frac{\nu}{2} \log\left(\frac{\nu}{2}\right) + \left(\frac{\nu}{2} - 1\right) \widehat{\log \kappa_i} - \frac{\nu \hat{\kappa}_i}{2},$$

with $\hat{\kappa}_i = E[k_i | \mathbf{V}_i, \mathbf{C}_i, \hat{\boldsymbol{\theta}}]$ and $\widehat{\log k_i} = E[\log(k_i) | \mathbf{V}_i, \mathbf{C}_i, \hat{\boldsymbol{\theta}}]$. Then, $\hat{\nu}$ can be obtained by using a one-dimensional Newton-Raphson algorithm to solve

$$\log\left(\frac{\nu}{2}\right) - \Psi\left(\frac{\nu}{2}\right) + 1 + \frac{1}{n} \sum_{i=1}^n (\widehat{\log \kappa_i} - \hat{\kappa}_i) = 0,$$

where $\Psi(z) = \frac{d}{dz} \log(\Gamma(z))$ denotes the digamma function.

- If the response vector follows a slash distribution, we have $\boldsymbol{\kappa}_i \sim \text{Beta}(\nu, 1)$ and $\widehat{\ell h_{1i}} = \log(\nu) + (\nu - 1) \widehat{\log \kappa_i}$. Thus, we obtain $\hat{\nu} = -\frac{n}{\sum_{i=1}^n \widehat{\log \kappa_i}}$.

For the random effects, we apply the same strategy.

- If the random effects have a Student- t distribution, we have $\boldsymbol{\tau}_i \sim \text{Gamma}(\eta/2, \eta/2)$ and

$$\widehat{\ell h_{2i}} = -\log\left(\Gamma\left(\frac{\eta}{2}\right)\right) + \frac{\eta}{2} \log\left(\frac{\eta}{2}\right) + \left(\frac{\eta}{2} - 1\right) \widehat{\log \tau_i} - \frac{\eta \hat{\tau}_i}{2},$$

with $\hat{\tau}_i = E[\tau_i | \mathbf{V}_i, \mathbf{C}_i, \hat{\boldsymbol{\theta}}]$ and $\widehat{\log \tau}_i = E[\log(\tau_i) | \mathbf{V}_i, \mathbf{C}_i, \hat{\boldsymbol{\theta}}]$. Then, $\hat{\eta}$ can be obtained by using a one-dimensional Newton-Raphson algorithm to solve

$$\log\left(\frac{\eta}{2}\right) - \Psi\left(\frac{\eta}{2}\right) + 1 + \frac{1}{n} \sum_{i=1}^n (\widehat{\log \tau}_i - \hat{\tau}_i) = 0.$$

- If the random effects have a slash distribution, we have $\tau_i \sim \text{Beta}(\eta, 1)$ and $\widehat{\ell h_{2i}} = \log(\eta) + (\eta - 1)\widehat{\log \tau}_i$. Therefore, $\hat{\eta} = \frac{n}{\sum_{i=1}^n \widehat{\log \tau}_i}$.

We replace the conditional expectations $\hat{\kappa}_i$, $\widehat{\log \kappa}_i$, $\hat{\tau}_i$, and $\widehat{\log \tau}_i$ by the stochastic approximations described in Section 4.

APPENDIX C: AIC AND BIC DECOMPOSITION

Let $\mathbf{y}_{i1}^\star$ be the s_{i1} -vector with $r^\star$ outcomes and $\mathbf{y}_{i2}^\star$ be the s_{i2} -vector with $r^\star - r$ outcomes, where $s_{i1} = n_i \cdot r^\star$ and $s_{i2} = n_i \cdot (r^\star - r)$, with $s_i = s_{i1} + s_{i2}$. After reordering, $\mathbf{y}$, $\mathbf{V}$, $\mathbf{X}$, $\mathbf{Z}$, and $\mathbf{R}_i$ can be partitioned as $\mathbf{y}_i = \text{vec}(\mathbf{y}_{i1}^\star, \mathbf{y}_{i2}^\star)$, $\mathbf{V}_i = \text{vec}(\mathbf{V}_i^1, \mathbf{V}_i^2)$, $\mathbf{X}_i^\top = (\mathbf{X}_i^1, \mathbf{X}_i^2)$, $\mathbf{Z}_i^\top = (\mathbf{Z}_i^1, \mathbf{Z}_i^2)$ and

$$\mathbf{R}_i = \begin{pmatrix} \mathbf{R}_i^{11} & \mathbf{R}_i^{12} \\ \mathbf{R}_i^{21} & \mathbf{R}_i^{22} \end{pmatrix}. \text{ Without considering the censoring for the moment,}$$

$$\mathbf{y}_i | \mathbf{b}_i, \kappa_i \sim N_{s_i}(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i, \kappa_i^{-1} \mathbf{R}_i) \text{ and we have}$$

$$\mathbf{y}_{i1}^\star | \mathbf{b}_i, \kappa_i \sim N_{s_{i1}}(\mathbf{X}_i^1 \boldsymbol{\beta} + \mathbf{Z}_i^1 \mathbf{b}_i, \kappa_i^{-1} \mathbf{R}_i^{11}), \quad (\text{C1})$$

$$\mathbf{y}_{i2}^\star | \mathbf{y}_{i1}^\star, \mathbf{b}_i, \kappa_i \sim N_{s_{i2}}(\boldsymbol{\mu}_i^\star, \kappa_i^{-1} \mathbf{S}_i^\star), \quad (\text{C2})$$

where $\boldsymbol{\mu}_i^\star = (\mathbf{X}_i^2 \boldsymbol{\beta} + \mathbf{Z}_i^2 \mathbf{b}_i) + \mathbf{R}_i^{21} (\mathbf{R}_i^{11})^{-1} (\mathbf{y}_{i1}^\star - \mathbf{X}_i^1 \boldsymbol{\beta} - \mathbf{Z}_i^1 \mathbf{b}_i)$ and $\mathbf{S}_i^\star = \mathbf{R}_i^{22} - \mathbf{R}_i^{21} (\mathbf{R}_i^{11})^{-1} \mathbf{R}_i^{12}$.

Then, using (C2) and (C1), the distributions $f(\mathbf{y}_{i1}^\star; \boldsymbol{\theta}_1^\star)$ and $f(\mathbf{y}_{i2}^\star | \mathbf{y}_{i1}^\star)$ are given by

$$f(\mathbf{y}_{i1}^\star; \boldsymbol{\theta}_1^\star) = \int_{\mathbb{R}^q} \int_0^\infty f(\mathbf{y}_{i1}^\star | \mathbf{b}_i, \kappa_i; \boldsymbol{\theta}_1^\star) f(\mathbf{b}_i; \boldsymbol{\theta}_1^\star) h_1(\kappa_i; \nu) d\kappa_i d\mathbf{b}_i,$$

$$f(\mathbf{y}_{i2}^\star; \boldsymbol{\theta}_2^\star) = \int_{\mathbb{R}^q} \int_0^\infty f(\mathbf{y}_{i2}^\star | \mathbf{y}_{i1}^\star, \mathbf{b}_i, \kappa_i; \boldsymbol{\theta}_2^\star) f(\mathbf{b}_i; \boldsymbol{\theta}_2^\star) h_1(\kappa_i; \nu) d\kappa_i d\mathbf{b}_i,$$

where $f(\mathbf{b}_i; \theta_{\mathbf{b}_\ell}^*, \eta) = \int_0^\infty f(\mathbf{b}_i | \tau_i; \theta_{\mathbf{b}_\ell}^*) h_2(\tau_i; |\eta) d\tau_i$, for $\ell = 1, 2$.

APPENDIX D: EMPIRICAL INFORMATION MATRIX

The elements of $\hat{\mathbf{s}}_i$ in Equation (13) of this paper are given by

$$\hat{\mathbf{s}}_{i, \boldsymbol{\beta}} = \left(\hat{\mathbf{s}}_{i, \beta_1}, \dots, \hat{\mathbf{s}}_{i, \beta_p} \right)^\top = \mathbf{X}_i^\top \widehat{\mathbf{R}}_i^{-1} (\widehat{\boldsymbol{\kappa}} \mathbf{y}_i - \mathbf{Z}_i \widehat{\boldsymbol{\kappa}} \mathbf{b}_i) - \widehat{\boldsymbol{\kappa}}_i \mathbf{X}_i^\top \widehat{\mathbf{R}}_i^{-1} \mathbf{X}_i \hat{\boldsymbol{\beta}},$$

$$\hat{\mathbf{s}}_{i, \alpha_r} = -\frac{1}{2} \text{tr} \left(\widehat{\mathbf{D}}^{-1} \dot{\mathbf{D}}_r \widehat{\mathbf{D}}^{-1} (\widehat{\mathbf{D}} - \tau \mathbf{b}_i^2) \right),$$

$$\hat{\mathbf{s}}_{i, \sigma_j^2} = -\frac{n_i}{2} \log \left| \widehat{\boldsymbol{\Sigma}}_i^{-1} \right| \dot{\boldsymbol{\Sigma}}_j + \frac{1}{2} \text{tr} \left[\left(\widehat{\boldsymbol{\Sigma}}_i^{-1} \dot{\boldsymbol{\Sigma}}_j \widehat{\boldsymbol{\Sigma}}_i^{-1} \otimes \Omega_i^{-1}(\hat{\boldsymbol{\phi}}) \right) \widehat{\boldsymbol{\kappa}} E_i \right],$$

$$\hat{\mathbf{s}}_{i, \boldsymbol{\phi}_s} = -\frac{r}{2} \log \left| \Omega_i^{-1}(\hat{\boldsymbol{\phi}}) \dot{\Omega}_{is} \right| + \frac{1}{2} \text{tr} \left[\left(\widehat{\boldsymbol{\Sigma}}_i^{-1} \otimes \Omega_i^{-1}(\hat{\boldsymbol{\phi}}) \dot{\Omega}_{is} \Omega_i^{-1}(\hat{\boldsymbol{\phi}}) \right) \widehat{\boldsymbol{\kappa}} E_i \right],$$

where $\dot{\mathbf{D}}_r = \frac{\partial \mathbf{D}}{\partial \alpha_r} |_{\boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}}}$, $r = 1, \dots, \dim(\boldsymbol{\alpha})$; $\dot{\boldsymbol{\Sigma}}_j = \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_j^2} |_{\boldsymbol{\sigma} = \hat{\boldsymbol{\sigma}}}$, $j = 1, \dots, \dim(\boldsymbol{\sigma})$; and

$\dot{\Omega}_{is} = \frac{\partial \Omega_i}{\partial \phi_s} |_{\boldsymbol{\phi} = \hat{\boldsymbol{\phi}}}$, $s = 1, 2$. For the DEC structure, we have

$$\frac{\partial \Omega_i}{\partial \phi_1} = \left| t_{ij} - t_{ik} \right|^{\phi_2} \phi_1^{t_{ij} - t_{ik}} - 1,$$

$$\frac{\partial \Omega_i}{\partial \phi_2} = \left| t_{ij} - t_{ik} \right|^{\phi_2} \log \left(\left| t_{ij} - t_{ik} \right| \right) \log(\phi_1) \phi_1^{t_{ij} - t_{ik}}.$$

As shown in Appendix B, to compute $\hat{\mathbf{s}}_{i, \nu}$ and $\hat{\mathbf{s}}_{i, \eta}$, we need to define the distributions of the error term and the random effects. For the Student- t and slash distributions, we have the following.

- If the response vector has a Student- t distribution,

$$\hat{\mathbf{s}}_{i, \nu} = -\frac{1}{2} \left[\Psi(\nu/2) - \log(\nu/2) - 1 - (\widehat{\log \kappa_i} - \hat{\kappa}_i) \right].$$

- If the response vector has a slash distribution,

$$\hat{s}_{i,\nu} = \frac{1}{\nu} + \widehat{\log \kappa_i}.$$

- If the random effects have a Student- t distribution,

$$\hat{s}_{i,\eta} = -\frac{1}{2} \left[\Psi(\eta/2) - \log(\eta/2) - 1 - (\widehat{\log \tau_i} - \tau_i) \right].$$

- If the random effects have a slash distribution,

$$\hat{s}_{i,\eta} = \frac{1}{\eta} + \widehat{\log \tau_i}.$$

APPENDIX E: ADDITIONAL PLOTS

FIGURE E1.

Simulation study: Boxplots of the parameter estimates. Dotted lines indicate the true parameter value. The censoring proportion is 10%

FIGURE E2.

A5055 dataset: Histogram of the CD4/CD8 ratio (left panel) and histogram of the log CD4/CD8 (right panel)

FIGURE E3.

A5055 dataset: Fitted curves of $\log_{10}(\text{RNA})$ and $\log(\text{CD4/CD8})$ for some individuals. Red lines indicate the fitted curves and black lines indicate the trajectories observed [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE E4.**A5055 dataset:** Convergence of the parameter estimates $\hat{\beta}_{10}$, $\hat{\beta}_{11}$, $\hat{\beta}_{12}$, $\hat{\beta}_{13}$, $\hat{\beta}_{14}$, $\hat{\beta}_{20}$, $\hat{\beta}_{21}$, $\hat{\beta}_{22}$ and $\hat{\beta}_{23}$

FIGURE E5.

A5055 dataset: Convergence of the parameter estimates $\widehat{\sigma}_{11}^2$, $\widehat{\sigma}_{12}^2$, $\widehat{\phi}_1$, $\widehat{\phi}_2$, and $\widehat{v}$

FIGURE E6.

A5055 dataset: Convergence of the parameter estimates

$\widehat{\alpha}_{11}$, $\widehat{\alpha}_{12}$, $\widehat{\alpha}_{22}$, $\widehat{\alpha}_{13}$, $\widehat{\alpha}_{23}$, $\widehat{\alpha}_{33}$, $\widehat{\alpha}_{14}$, $\widehat{\alpha}_{24}$, $\widehat{\alpha}_{34}$, and $\widehat{\alpha}_{44}$

REFERENCES

1. Alder M Development of the epidemic. In: Alder M, ed. ABC of AIDS London, UK: BMJ Publishing Group; 2001.
2. Mortimer P, Loveday C. The virus and the tests. In: Alder M, ed. ABC of AIDS London, UK: BMJ Publishing Group; 2001.
3. Barletta J, Edelman D, Constantine N. Lowering the detection limits of HIV-1 viral load using real-time immuno-PCR for HIV-1 p24 antigen. Am J Clin Pathol 2004;112(20–27).
4. Wu L Mixed Effects Models for Complex Data Boca Raton, FL: Chapman & Hall/CRC; 2010.
5. Matos L, Castro L, Lachos V. Censored mixed-effects models for irregularly observed repeated measures with applications to HIV viral loads. TEST 2016:1–27.

6. Vaida F, Liu L. Fast implementation for normal mixed effects models with censored response. *J Comput Graph Stat* 2009;18(4):797–817. [PubMed: 25829836]
7. Matos L, Prates M, Chen MH, Lachos V. Likelihood based inference for linear and nonlinear mixed-effects models with censored response using the multivariate-t distribution. *Statistica Sinica* 2013;23:1323–1345.
8. Shah A, Laird N, Schoenfeld D. A random-effects model for multiple characteristics with possibly missing data. *J Am Stat Assoc* 1997;92(438):775–779.
9. Wang W. Multivariate t linear mixed models for irregularly observed multiple repeated measures with missing outcomes. *Biometrical Journal* 2013;55(4):554–571. [PubMed: 23740830]
10. Wang W, Lin T, Lachos V. Extending multivariate-t linear mixed models for multiple longitudinal data with censored responses and heavy tails. *Stat Methods Med Res* 2018;27(1):48–64. [PubMed: 26668091]
11. Ho HJ, Lin TI, Chen HY, Wang W. Some results on the truncated multivariate t distribution. *J Stat Plan Infer* 2012;142:25–40.
12. Andrews D, Mallows C. Scale mixtures of normal distributions. *J R Stat Soc Ser B Methodol* 1974;99–102.
13. Garay AM, Castro L, Leskow J, Lachos V. Censored linear regression models for irregularly observed longitudinal data using the multivariate-t distribution. *Stat Methods Med Res* 2017;26:542–566. [PubMed: 25296865]
14. Dagne GA, Huang Y. Bayesian semiparametric mixture Tobit models with left censoring, skewness, and covariate measurement errors. *Stat Med* 2013;32(22):3881–3898. [PubMed: 23553914]
15. Lin T, Wang W. Multivariate-t nonlinear mixed models with application to censored multi-outcome AIDS studies. *Biostatistics* 2017;18(4):666–681. [PubMed: 28369172]
16. Paxton W, Coombs R, McElrath M, et al. Longitudinal analysis of quantitative virologic measures in human immunodeficiency virus-infected subjects with 400 CD4 lymphocytes: implications for applying measurements to individual patients. *J Infect Dis* 1997;175(2):247–254. [PubMed: 9203644]
17. Delyon B, Lavielle M, Moulines E. Convergence of a stochastic approximation version of the EM algorithm. *Ann Stat* 1999;27(1):94–128.
18. Jank W. Implementing and diagnosing the stochastic approximation EM algorithm. *J Comput Graph Stat* 2006;15(4):803–829.
19. Dempster A, Laird N, Rubin D. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Ser B Methodol* 1977;39:1–38.
20. Wei G, Tanner M. A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *J Am Stat Assoc* 1990;85(411):699–704.
21. Meza C, Osorio F, De la Cruz R. Estimation in nonlinear mixed-effects models using heavy-tailed distributions. *Stat Comput* 2012;22(1):121–139.
22. Kuhn E, Lavielle M. Coupling a stochastic approximation version of EM with an MCMC procedure. *ESAIM Probab Stat* 2004;8:115–131.
23. Galarza C, Lachos V, Bandyopadhyay D. Quantile regression in linear mixed models: a stochastic approximation EM approach. *Stat Interface* 2017;10(3):471–482. [PubMed: 29104713]
24. Muñoz A, Carey V, Schouten J, Segal M, Rosner B. A parametric family of correlation structures for the analysis of longitudinal data. *Biometrics* 1992;48:733–742. [PubMed: 1420837]
25. Kuhn E, Lavielle M. Maximum likelihood estimation in nonlinear mixed effects models. *Comput Stat Data Anal* 2005;49(4):1020–1038.
26. Samson A, Lavielle M, Mentré F. Extension of the SAEM algorithm to left-censored data in nonlinear mixed-effects model: application to HIV dynamics model. *Comput Stat Data Anal* 2006;51(3):1562–1574.
27. Zhang D, Chen MH, Ibrahim J, Boye M, Wang P, Shen W. Assessing model fit in joint models of longitudinal and survival data with applications to cancer clinical trials. *Statist Med* 2014;33(27):4715–4733.

28. Louis T Finding the observed information matrix when using the EM algorithm. *J R Stat Soc Ser B Methodol* 1982;226–233.
29. Lavielle M *Mixed Effects Models for the Population Approach* Boca Raton, FL: Chapman & Hall/CRC; 2014.
30. Bandyopadhyay D, Lachos V, Castro L, Dey D. Skew-normal/independent linear mixed models for censored responses with applications to HIV viral loads. *Biometrical Journal* 2012;54(3):405–425. [PubMed: 22685005]
31. Lachos VH, Matos LA, Castro LM, Chen M-H. Flexible longitudinal linear mixed models for multiple censored responses data. *Statistics in Medicine* 2018;1–28. 10.1002/sim.8017

FIGURE 1.

Trajectories of $\log_{10}$ RNA (left panel) and CD4/CD8 ratio (right panel) for 44 HIV-1 infected patients who were randomized in two IDV-RTV regimens. Black lines indicate patients under treatment 1 and red lines indicate patients under treatment 2. Dotted line in left panel indicates the censoring level [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 2. Simulation I:
(20% of censoring proportion). Boxplots of the parameter estimates. Dotted lines indicate the true parameter value

FIGURE 3. Simulation II:

Mean square error (MSE) based on 100 simulated samples for the parameter estimates under the N/N-, T/T-, and SL/SL-MLMEC models for different sample sizes and censoring proportions. MLMEC, multivariate censored linear mixed effect [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 4.

QQ-plots of the residuals and estimates for the random effects (upper panel) together with their bivariate normal contour and histogram (lower panel) under the N/N-MLMEC model, where EBE and SN means Empirical Bayer estimates and Standard Normal, respectively. MLMEC, multivariate censored linear mixed effect [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 5.

Random effects estimates together with their bivariate contour and histogram under the T/N-MLMEC model. MLMEC, multivariate censored linear mixed effect [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1

Simulation I. Parameter estimates based on 600 simulated samples. Monte Carlo (MC) mean, MC SD are the respective mean estimates and standard deviations. 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. Bias and mean square error (MSE) of the parameter estimates

	M1		M2		M3	
	MC Mean	IM SE (MC SD)	MC Mean	IM SE (MC SD)	MC Mean	IM SE (MC SD)
Censoring 10%						
β_{10} (6.0)	5.950	0.479 (0.442)	5.952	0.479 (0.442)	5.958	0.489 (0.446)
β_{11} (-0.2)	-0.194	0.054 (0.048)	-0.194	0.054 (0.048)	-0.194	0.056 (0.049)
β_{20} (4.0)	3.997	0.481 (0.434)	3.998	0.482 (0.433)	4.003	0.491 (0.438)
β_{21} (-0.1)	-0.098	0.054 (0.052)	-0.098	0.054 (0.052)	-0.098	0.055 (0.052)
σ_{11} (4.0)	4.061	0.907 (0.782)	3.968	0.894 (0.688)	4.885	1.072 (0.862)
σ_{12} (2.0)	2.027	0.496 (0.426)	1.979	0.488 (0.379)	2.438	0.587 (0.474)
σ_{22} (4.0)	4.066	0.910 (0.774)	3.973	0.895 (0.683)	4.891	1.075 (0.864)
α_{11} (2.0)	2.030	1.229 (0.956)	1.921	1.167 (0.842)	2.468	1.486 (1.116)
α_{12} (1.0)	1.027	0.872 (0.713)	0.978	0.829 (0.660)	1.245	1.054 (0.872)
α_{22} (2.0)	2.047	1.240 (0.939)	1.944	1.178 (0.840)	2.492	1.495 (1.107)
ϕ_1 (0.8)	0.793	0.042 (0.034)	0.793	0.042 (0.034)	0.794	0.043 (0.034)
$\nu(3.0)$	3.728					
$\eta(3.0)$	4.044					
MC AIC	3509.160		3506.556		3515.322	
MC BIC	3572.961		3560.541		3569.308	
Censoring 20%						
β_{10} (6.0)	5.926	0.469 (0.442)	5.929	0.470 (0.442)	5.944	0.479 (0.447)
β_{11} (-0.2)	-0.191	0.054 (0.047)	-0.191	0.054 (0.047)	-0.191	0.055 (0.048)
β_{20} (4.0)	3.988	0.472 (0.432)	3.992	0.472 (0.431)	4.005	0.482 (0.438)
β_{21} (-0.1)	-0.097	0.053 (0.050)	-0.097	0.053 (0.050)	-0.098	0.055 (0.050)
σ_{11} (4.0)	3.952	0.879 (0.810)	3.823	0.860 (0.695)	4.697	1.022 (0.867)
σ_{12} (2.0)	1.957	0.480 (0.430)	1.893	0.469 (0.373)	2.327	0.561 (0.466)
σ_{22} (4.0)	3.952	0.876 (0.778)	3.823	0.856 (0.663)	4.699	1.019 (0.837)
σ_{11} (2.0)	2.021	1.226 (0.949)	1.860	1.140 (0.828)	2.378	1.439 (1.095)
α_{12} (1.0)	1.015	0.862 (0.697)	0.942	0.799 (0.631)	1.193	1.011 (0.839)
α_{22} (2.0)	2.116	1.238 (0.964)	1.950	1.150 (0.847)	2.490	1.451 (1.128)
ϕ_1 (0.8)	0.787	0.043 (0.036)	0.787	0.043 (0.036)	0.787	0.043 (0.036)
$\nu(3.0)$	3.959					
$\eta(3.0)$	4.358					
MC AIC	3381.705		3379.021		3386.550	
MC BIC	3445.506		3433.007		3440.535	

Censoring 10%

Censoring 20%

	Bias			MSE			Bias			MSE		
	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
β_1	0.360	0.360	0.363	0.198	0.197	0.201	0.363	0.363	0.366	0.200	0.200	0.203
β_2	0.039	0.039	0.039	0.002	0.002	0.002	0.039	0.039	0.039	0.002	0.002	0.002
β_3	0.350	0.350	0.352	0.188	0.188	0.192	0.350	0.350	0.353	0.186	0.186	0.191
β_4	0.041	0.041	0.041	0.003	0.003	0.003	0.040	0.040	0.040	0.002	0.002	0.003
σ_{11}	0.631	0.565	0.996	0.614	0.474	1.525	0.649	0.588	0.888	0.658	0.514	1.237
σ_{12}	0.340	0.307	0.518	0.182	0.144	0.416	0.346	0.317	0.450	0.187	0.150	0.324
σ_{22}	0.617	0.554	0.996	0.602	0.466	1.540	0.616	0.559	0.869	0.606	0.470	1.188
α_{11}	0.780	0.695	0.945	0.913	0.714	1.462	0.764	0.686	0.887	0.900	0.704	1.340
α_{12}	0.565	0.526	0.705	0.508	0.436	0.819	0.543	0.500	0.656	0.485	0.400	0.740
α_{22}	0.750	0.676	0.947	0.882	0.708	1.466	0.768	0.676	0.948	0.940	0.719	1.511
ϕ_1	0.028	0.028	0.028	0.001	0.001	0.001	0.030	0.030	0.030	0.001	0.001	0.001

TABLE 2

Simulation II. Parameter estimates and model comparison criteria based on 100 simulated samples under the N/N-, T/T-, and SL/SL-MLMEC models for different sample sizes and censoring proportions. The reported values correspond to the Monte Carlo mean

Distribution ϵ / Distribution b										
	True value	$n = 50$	N/N $n = 100$	$n = 200$	$n = 50$	T/T $n = 100$	$n = 200$	$n = 50$	SL/SL $n = 100$	$n = 200$
Censoring 10%										
β_1	6.0	6.034	6.027	6.016	5.989	6.016	5.989	5.989	6.012	5.985
β_2	-0.2	-0.202	-0.203	-0.194	-0.198	-0.205	-0.197	-0.198	-0.204	-0.196
β_3	4.0	3.967	4.037	4.040	3.976	4.016	4.002	3.974	4.008	4.001
β_4	-0.1	-0.096	-0.100	-0.099	-0.099	-0.101	-0.098	-0.099	-0.101	-0.098
σ_{11}^2	4.0	7.119	7.289	7.287	5.801	5.784	5.669	4.003	3.973	3.855
σ_{12}^2	2.0	3.522	3.715	3.610	2.870	2.914	2.830	1.981	1.998	1.926
σ_{22}^2	4.0	7.002	7.382	7.259	5.696	5.859	5.672	3.927	4.022	3.857
α_{11}	2.0	2.982	2.934	2.959	2.316	2.355	2.429	2.091	2.100	2.122
α_{12}	1.0	1.674	1.442	1.515	1.376	1.198	1.241	1.223	1.073	1.082
α_{22}	2.0	3.228	3.001	2.997	2.541	2.396	2.474	2.266	2.128	2.168
ϕ_1	0.8	0.781	0.792	0.789	0.787	0.797	0.792	0.786	0.796	0.791
ν	2.0	—	—	—	11.314	10.141	9.385	2.224	2.128	2.031
η	3.0	—	—	—	13.329	11.519	11.296	3.984	3.594	3.459
$loglik.$		-1867.3	-3741.7	-7506.5	-1830.4	-3656.9	-7331.1	-1829.2	-3653.9	-7325.1
AIC		3756.6	7505.5	15035.0	3686.8	7339.9	14688.2	3684.3	7333.8	14676.2
BIC		3810.6	7567.1	15104.3	3750.6	7412.7	14770.0	3748.1	7406.6	14758.1
Censoring 20%										
β_1	6.0	5.957	6.040	6.017	5.909	5.965	5.971	5.907	5.969	5.968
β_2	-0.2	-0.185	-0.195	-0.196	-0.188	-0.194	-0.197	-0.188	-0.194	-0.197
β_3	4.0	3.960	4.073	4.037	3.905	4.016	3.981	3.905	4.021	3.981
β_4	-0.1	-0.086	-0.104	-0.096	-0.087	-0.105	-0.096	-0.088	-0.105	-0.096
σ_{11}^2	4.0	7.062	7.031	7.044	5.592	5.623	5.527	3.852	3.877	3.786
σ_{12}^2	2.0	3.559	3.502	3.469	2.797	2.788	2.748	1.928	1.923	1.884
σ_{22}^2	4.0	7.097	7.016	6.967	5.598	5.580	5.500	3.855	3.847	3.768
α_{11}	2.0	3.140	2.915	2.887	2.463	2.316	2.384	2.177	2.039	2.085
α_{12}	1.0	1.514	1.514	1.396	1.196	1.225	1.179	1.053	1.075	1.031
α_{22}	2.0	2.937	3.049	2.942	2.393	2.390	2.392	2.109	2.099	2.106
ϕ_1	0.8	0.780	0.785	0.785	0.785	0.790	0.789	0.784	0.789	0.788
ν	2.0	—	—	—	10.668	10.298	9.603	2.164	2.145	2.072
η	3.0	—	—	—	15.037	13.045	12.412	4.138	3.775	3.688

True value	Distribution ϵ / Distribution b								
	<i>n</i> = 50	N/N <i>n</i> = 100	<i>n</i> = 200	<i>n</i> = 50	T/T <i>n</i> = 100	<i>n</i> = 200	<i>n</i> = 50	SL/SL <i>n</i> = 100	<i>n</i> = 200
<i>loglik.</i>	-1794.4	-3590.1	-7245.5	-1755.2	-3513.8	-7078.5	-1753.6	-3511.4	-7072.3
AIC	3610.8	7202.2	14513.0	3536.4	7053.6	14183.0	3533.3	7048.8	14170.6
BIC	3664.8	7263.8	14582.2	3600.2	7126.4	14264.8	3597.1	7121.6	14252.4

Abbreviations: MLMEC, multivariate censored linear mixed effect.

TABLE 3

Simulation III: AIC and log-likelihood values based on 100 simulated samples under different scale mixtures of normal–multivariate censored linear mixed effect models

	Distribution ϵ / Distribution b					
	N/N	N/T	N/SL	SL/N	SL/T	SL/SL
<i>loglik.</i>	−3609.2	−3602.8	−3602.8	−3451.8	−3447.5	−3413.6
AIC	7240.4	7229.5	7229.6	6927.5	6920.9	6852.9

Table 4

Simulation III: Parameter estimates based on 100 simulated samples. Monte Carlo (MC) mean, MC Sd are the respective mean and standard deviations. IM SE is the average value of the approximate standard error obtained through the information-based method

Parameter		Distribution ϵ / Distribution b					
		N/N	N/T	N/SL	SL/N	SL/T	SL/SL
β_1 (6.0)	MC Mean	6.049	5.974	5.972	5.995	5.967	5.903
	IM SE	0.389	0.358	0.360	0.344	0.330	0.328
	MC Sd	0.394	0.382	0.384	0.355	0.356	0.695
β_2 (-0.2)	MC Mean	-0.201	-0.201	-0.201	-0.200	-0.200	-0.199
	IM SE	0.040	0.040	0.040	0.033	0.033	0.032
	MC Sd	0.039	0.039	0.039	0.033	0.033	0.038
β_3 (4.0)	MC Mean	4.020	3.960	3.960	3.969	3.945	3.907
	IM SE	0.397	0.365	0.367	0.352	0.335	0.333
	MC Sd	0.388	0.374	0.373	0.336	0.337	0.519
β_4 (-0.1)	MC Mean	-0.097	-0.097	-0.097	-0.097	-0.097	-0.096
	IM SE	0.040	0.040	0.040	0.033	0.032	0.032
	MC Sd	0.038	0.038	0.038	0.030	0.030	0.031
σ_{11}^2 (4.0)	MC Mean	5.888	5.881	5.849	2.232	2.166	2.127
	IM SE	0.731	0.719	0.715	0.332	0.332	0.325
	MC Sd	1.036	0.968	0.964	0.463	0.359	0.405
σ_{12}^2 (2.0)	MC Mean	2.920	2.916	2.899	1.105	1.073	1.056
	IM SE	0.402	0.396	0.394	0.179	0.178	0.175
	MC Sd	0.581	0.519	0.518	0.233	0.183	0.207
σ_{22}^2 (4.0)	MC Mean	5.916	5.905	5.873	2.224	2.158	2.122
	IM SE	0.736	0.721	0.718	0.33	0.328	0.322
	MC Sd	1.035	0.933	0.924	0.46	0.354	0.402
α_{11} (2.0)	MC Mean	5.334	2.926	2.254	4.749	0.220	2.496
	IM SE	1.316	1.357	1.415	1.136	1.327	1.441
	MC Sd	1.994	1.152	0.925	1.244	0.855	0.929
α_{12} (1.0)	MC Mean	2.622	1.488	1.146	2.380	0.112	1.277
	IM SE	0.973	0.838	0.800	0.841	0.837	0.817
	MC Sd	1.214	0.776	0.619	0.974	0.452	0.606
α_{22} (2.0)	MC Mean	5.442	2.989	2.300	4.903	0.225	2.546
	IM SE	1.304	1.374	1.421	1.142	1.331	1.428
	MC Sd	1.702	1.180	0.979	1.403	0.879	0.970
ϕ_1 (0.8)	MC Mean	0.782	0.782	0.781	0.792	0.787	0.778
	IM SE	0.026	0.026	0.026	0.026	0.028	0.028
	MC Sd	0.029	0.029	0.028	0.029	0.024	0.082
ν (5.0)	MC Mean	-	-	-	1.294	1.302	1.281
	IM SE	-	-	-	0.226	0.228	0.225
	MC Sd	-	-	-	0.146	0.149	0.192

		Distribution ϵ / Distribution b					
Parameter		N/N	N/T	N/SL	SL/N	SL/T	SL/SL
η (1.5)	MC Mean	-	5.738	1.883	-	7.302	2.166
	IM SE	-	5.108	2.038	-	12.094	2.983
	MC Sd	-	3.364	0.736	-	3.968	0.911

TABLE 5

A5055 dataset: Model comparison criteria for the scale mixtures of normal–multivariate censored linear mixed effect models under the damped exponential correlation structure

	Distribution ϵ / Distribution b								
	N/N	SL/N	T/N	N/SL	N/T	SL/SL	SL/T	T/SL	T/T
AIC	789.85	742.18	739.59	791.98	792.29	744.47	744.54	741.85	741.51
BIC	896.62	853.41	850.81	903.20	903.51	860.14	860.21	857.52	857.19

TABLE 6

A5055 dataset: maximum-likelihood estimates with standard errors under the T/N-MLMEC model

Structure	Fixed effects			Between-subject variances			Within-subject variances		
	Param.	Estimate	SE	Param.	Estimate	SE	Param.	Estimate	SE
DEC	β_{10}	3.743	0.134	d_{11}	0.1446	0.0829	σ_{11}	0.409	0.076
	β_{11}	0.130	0.026	d_{21}	0.0011	0.0133	σ_{21}	-0.039	0.020
	β_{12}	-0.005	0.067	d_{22}	-0.0884	0.1182	σ_{22}	0.050	0.011
	β_{13}	-0.957	0.098	d_{31}	-0.0011	0.0033	ϕ_1	0.704	0.065
	β_{14}	-0.007	0.025	d_{32}	0.0034	0.0027	ϕ_2	0.632	0.131
	β_{20}	-1.284	0.077	d_{33}	-0.0122	0.0116	ν	4.737	0.003
	β_{21}	0.005	0.005	d_{41}	-0.0004	0.0004			
	β_{22}	0.252	0.084	d_{42}	0.2727	0.0861			
	β_{23}	-0.003	0.007	d_{43}	0.0008	0.0015			
				d_{44}	0.0001	0.0001			
	<i>loglik</i>	-344.79		AIC	739.59		BIC	850.81	
	β_{10}	3.718	0.135	d_{11}	0.4089	0.1463	σ_{11}	0.263	0.053
	β_{11}	0.129	0.026	d_{21}	-0.0112	0.0153	σ_{21}	-0.024	0.012
	β_{12}	0.003	0.091	d_{22}	-0.0964	0.1251	σ_{22}	0.028	0.005
UNC	β_{13}	-0.955	0.075	d_{31}	0.0002	0.0030	ν	4.340	0.004
	β_{14}	-0.008	0.027	d_{32}	0.0054	0.0029			
	β_{20}	-1.278	0.076	d_{33}	-0.0132	0.0116			
	β_{21}	0.005	0.004	d_{41}	-0.0006	0.0004			
	β_{22}	0.286	0.081	d_{42}	0.2953	0.0785			
	β_{23}	-0.006	0.006	d_{43}	0.0002	0.0015			
				d_{44}	0.0001	0.0001			
	<i>loglik</i>	-357.97		AIC	761.94		BIC	864.26	

Abbreviations: DEC, damped exponential correlation; MLMEC, multivariate censored linear mixed effect; UNC, uncorrelated.

TABLE 7

A5055 dataset: Decompositions of AIC and BIC under the best scale mixtures of normal–multivariate censored linear mixed effect model

AIC	739.59	BIC	850.81
AIC $y_2^\star y_1^\star$	92.65	BIC $y_2^\star y_1^\star$	158.80
AIC $y_{2,0}^\star$	125.26	BIC $y_{2,0}^\star$	166.58
AIC	32.61	BIC	7.77