Thailand Statistician
July 2025; 23(3): 692-709
http://statassoc.or.th
Contributed paper

Multivariate Skew Normal Independent Nonlinear Mixed Model for Longitudinal Data

Maha E. Qura [a], Ahmed M. Gad*[b] and Zohdi Nofal [a]
[a] Department of Statistics, Mathematics & Insurance, Faculty of Commerce, Banha University, Egypt.
[b] Business Administration Department, Faculty of Business Administration, Economics and Political Science, British University in Egypt (BUE), Egypt.
*Corresponding author; e-mail: ahmed.gad@feps.edu.eg

Received: 28 September 2021
Revised: 23 September 2023
Accepted: 9 October 2023

Abstract

The multivariate nonlinear mixed effects models (MNLMM) have received increasing attention due to their flexibility in analyzing and modeling multivariate longitudinal data. In the framework of MNLMM, the random effects and within-subject errors are assumed to be normally distributed for mathematical tractability and computational simplicity. However, such assumption might not offer robust inference if the data, even after being transformed, exhibit skewness. In this paper, we propose a multivariate skew normal independent nonlinear mixed model (MSNI-NLMM) constructed by assuming a multivariate skew normal independent distribution for the random effects and a multivariate normal independent distribution for the random errors. We develop a new model which can flexibly handle asymmetric, unbalanced, and irregularly observed multivariate longitudinal data. Also, we present two different iterative algorithms for maximum likelihood estimation of the MSNI-NLMM. They are the penalized nonlinear least squares coupled to the multivariate linear mixed effects (PNLS-MLME) procedure and the pseudo-data expectation conditional maximization (ECM) algorithm. The proposed approaches are illustrated through an application to ACTG 315 data and a simulation study.

Keywords: AR(1) correlation, multivariate longitudinal data, nonlinear mixed effects, damped exponential correlation, unstructured correlation.

1. Introduction

Multivariate longitudinal data where more than one response can be measured over time, for each subject. Marshall et al. (2006) propose the multivariate nonlinear mixed effects models (MNLMMs) for multivariate longitudinal data. Lachos et al. (2010) assume that the random effects follow multivariate skew normal/independent distribution, and the random errors follow symmetric normal/independent distribution. Meza et al. (2012) consider heavy-tailed multivariate distributions, such as the t-distribution, the contaminated normal and slash, for both random effects and errors. Pereira and Russo (2019) present a nonlinear mixed effects model with skewed and heavy-tailed distributions, where the nonlinearity is incorporated only in the fixed effects. Schumacher et al. (2021b) provide an extension of the skew-normal/independent linear mixed model, where the error term has a dependence structure, such as damped exponential correlation or autoregressive correlation of order p. Schumacher et al. (2021a) present a class of asymmetric nonlinear mixed effects models if the random effects follow a multivariate scale mixture of skew-normal distribution, and the random errors follow a symmetric scale mixture of normal distribution, providing an appealing robust alternative to the usual normal distribution.

Lin and Wang (2013) propose a multivariate skew-normal linear mixed model, assuming a multivariate skew-normal distribution for the random effects, and a multivariate normal distribution for the random errors. Wang and Lin (2014) consider a joint multivariate t-distribution for the random effects and within subject errors, called the multivariate t nonlinear mixed-effects model (Mt-NLMM). Wang (2015) assume that the random effects and the within subject errors are normally distributed to handle symmetric multivariate longitudinal data. However, such an assumption is not always applicable, especially when data contain outliers or heavy-tailed. Wang and Lin (2017) propose the multivariate t nonlinear mixed model with censored responses (Mt-NLMMC) for multivariate longitudinal data exhibiting nonlinear growth patterns with censorship and heavy-tailed behavior. Multivariate skew normal/independent nonlinear mixed effects models (MSNI-NLMMs) are considerably more complicated and computationally intensive than MSNI-LMMs. The nonlinearity offers no closed-form solutions to the model parameters.

In this article we provide hierarchical forms of multivariate skew normal independent nonlinear mixed effects models. Also, we introduce the skew normal independent (SNI) distribution. The SNI distribution is an attractive class of skew heavy-tailed distributions. Special cases of the SNI distribution are the skew normal, the skew Student's-t, the skew slash, and the skew contaminated normal distributions. We propose a multivariate skew nonlinear mixed effects model (MSNLMM), which is an extension of the multivariate nonlinear mixed effects model. In the proposed MSNLMM model we assume a multivariate-skew normal independent (MSNI) distribution for random effects, and a multivariate normal independent (MNI) distribution for within-subject errors. The proposed MSNLMM model can be used to fit multivariate longitudinal data exhibiting nonlinear growth pattern. We suggest two different iterative estimation algorithms. They are the penalized nonlinear least squares coupled to the multivariate linear mixed-effects (PNLS-MLME) procedure, and the pseudo-data expectation conditional maximization (pseudo-ECM) algorithm.

The rest of the article is organized as follows. Section 2 introduces the model formulation, addresses some relevant properties. In Section 3, we present the hierarchal forms of the proposed model and discuss the computational aspects of PNLS-MLME procedure and pseudo-ECM algorithm. A method of obtaining approximate standard errors of ML estimates is also provided. The proposed techniques are applied to ACTG 315 data in Section 4. A simulation study is also conducted to evaluate the proposed techniques in Section 5. Some concluding remarks and future works are given in Section 6.

2. Statistical Models

2.1. Multivariate skew normal/independent (MSNI) distribution

The skew normal/independent (SNI) distribution can be defined as (Lachos et al. 2010):

$$\boldsymbol{Y} = \boldsymbol{\mu} + U^{-1/2}\boldsymbol{Z}, \tag{1}$$

where $\boldsymbol{\mu}$ is a location vector, U is a positive random variable with cdf of $\mathrm{H}(u;\boldsymbol{\nu})$ and a pdf of $h(u;\boldsymbol{\nu})$, $\mathbf{v}$ is a vector of parameters and $\boldsymbol{Z}$ is a multivariate skew normal random vector (Arellano-

Valle et al., 2005) with location vector 0, scale matrix $\mathbf{\Sigma}$ and skewness parameter vector $\boldsymbol{\lambda}$, i.e. $\boldsymbol{Z} \sim SN_d\left(0, \mathbf{\Sigma}, \boldsymbol{\lambda}\right)$. Given $U = u$, $\boldsymbol{Y}$ follows a multivariate skew normal distribution with location vector 0, scale matrix $u^{-1}\mathbf{\Sigma}$ and skewness parameter vector $\boldsymbol{\lambda}$, i.e., $\boldsymbol{Y} \mid U = u \sim SN_d\left(\boldsymbol{\mu}, u^{-1}\mathbf{\Sigma}, \boldsymbol{\lambda}\right)$. The marginal pdf of $\boldsymbol{Y}$ is

$$f\left(\boldsymbol{y}\right) = 2\int_0^\infty \phi_d\left(\boldsymbol{y}; \boldsymbol{\mu}, u^{-1}\mathbf{\Sigma}\right)\Phi\left(u^{1/2}\boldsymbol{\lambda}^T\mathbf{\Sigma}^{-1/2}\left(\boldsymbol{y}-\boldsymbol{\mu}\right)\right)dH\left(u; \boldsymbol{v}\right),\ \boldsymbol{y} \in \mathbb{R}^d, \tag{2}$$

where $\phi_d\left(.; \boldsymbol{\mu}, \mathbf{\Sigma}\right)$ stands for the pdf of the d-variate normal distribution with mean vector μ and dispersion matrix $\mathbf{\Sigma}$ and $\Phi(\cdot)$ represents the cdf of the standard univariate normal distribution.

2.2. Multivariate skew-normal/independent nonlinear mixed model (MSNI-NLMM)

Suppose that there are m subjects and the subject i has n_i observations on each of the r responses. Let $\boldsymbol{Y}_i = \left[\boldsymbol{y}_{i1} : \ldots : \boldsymbol{y}_{ir}\right] = \left[\boldsymbol{y}_{i,1}^T : \ldots : \boldsymbol{y}_{i,n_i}^T\right]^T$ be an $n_i \times r$ response matrix for the i^{th} subject collected longitudinally $(i=1,\ldots,m)$, in which $\boldsymbol{y}_{ij} = (y_{ij,1},\ldots,y_{ij,n_i})^T$ is a column vector of responses for the j^{th} outcome $(j=1,\ldots,r)$ and $\boldsymbol{y}_{i,k} = (y_{i1,k},\ldots,y_{ir,k})$ is a row vector of responses collected at the k^{th} occasion $(k=1,\ldots,n_i)$. Let $\boldsymbol{E}_i = \left[\boldsymbol{e}_{i1} : \ldots. : \boldsymbol{e}_{ir}\right] = \left[\boldsymbol{e}_{i,1}^T : \ldots. : \boldsymbol{e}_{i,n_i}^T\right]^T$ be the $n_i \times r$ matrix of within-subject errors, in which $\boldsymbol{e}_{ij} = (e_{ij,1},\ldots,e_{ij,n_i})^T$ is a column vector corresponding to $\boldsymbol{y}_{ij}$, and $e_{i,k} = (e_{i1,k},\ldots,e_{ir,k})$ is a row vector corresponding to $\boldsymbol{y}_{i,k}$. Introducing the vec$(\cdot)$ operator, which strings out the columns of a matrix vertically, we can obtain $\boldsymbol{y}_i = \text{vec}(\boldsymbol{Y}_i)$ and $\boldsymbol{e}_i = \text{vec}(\boldsymbol{E}_i)$ denoted by the stacked $n_i r$-dimensional vectors of all responses and within-subject errors, respectively.

The MSNI-NLMM for the i^{th} subject takes the form

$$\boldsymbol{y}_i = \boldsymbol{\mu}_i\left(\phi_i, \boldsymbol{X}_i\right) + \boldsymbol{e}_i, \tag{3}$$

where $\boldsymbol{\mu}_i\left(\phi_i, \boldsymbol{X}_i\right) = \boldsymbol{\mu}_i\left(\boldsymbol{\beta}, \boldsymbol{b}_i\right)$, $\boldsymbol{\mu}_i$ is a nonlinearly differentiable function of the parameters vector ϕ_i and the covariate vector $\boldsymbol{X}_i$, and $\boldsymbol{e}_i$ is the error term in the model. The parameter vector ϕ_i can be incorporated into the model as

$$\phi_i = \boldsymbol{X}_i\,\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i, \tag{4}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T,\ldots,\boldsymbol{\beta}_r^T)^T$ are the regression coefficients with each p_j vector $\boldsymbol{\beta}_j$, used to describe the fixed effects of response j; $\boldsymbol{b}_i = (\boldsymbol{b}_{i1}^T,\ldots,\boldsymbol{b}_{ir}^T)^T$ is a q-variate random effects. Moreover, $\boldsymbol{X}_i =$ diag$\{\boldsymbol{X}_{i1},\ldots,\boldsymbol{X}_{ir}\}$ and $\boldsymbol{Z}_i = \text{diag}\{\boldsymbol{Z}_{i1},\ldots,\boldsymbol{Z}_{ir}\}$, where $\boldsymbol{X}_{ij}$ is a $n_i \times p_j$ covariate matrix of full column rank for fid effects associated with $\boldsymbol{y}_{ij}$, and $\boldsymbol{Z}_{ij}$ is a $n_i \times q_j$ design matrix for random effects. Typically, $\boldsymbol{Z}_{ij}$ is a subset of $\boldsymbol{X}_{ij}$. The block diagonal structures of $\mathbf{X}_{\text{i}}$ and $\boldsymbol{Z}_i$ allow us to analyze multivariate longitudinal data with different numbers of measurements and/ or unequal sets of occasions per subject. Also, to specify distinct design matrices for each response. For easy notation, we let $s_i = n_i r$, $p = \sum_{j=1}^r p_j$ and $q = \sum_{j=1}^r q_j$. Meanwhile, we assume that

$$\begin{bmatrix}\boldsymbol{b}_i \\ \boldsymbol{e}_i\end{bmatrix} \sim SNI_{s_i+q}\left(\begin{bmatrix}0 \\ 0\end{bmatrix}, \begin{bmatrix}\boldsymbol{D} & 0 \\ 0 & \boldsymbol{R}_i\end{bmatrix}, \begin{bmatrix}\boldsymbol{\lambda}_b \\ 0\end{bmatrix}, H\right), \quad i = 1,\ldots,m. \tag{5}$$

where the matrices $\boldsymbol{R}_i$ and $\boldsymbol{D}$, $i = 1,\ldots, m$, are dispersion matrices, corresponding to the within and between subjects, respectively, $\boldsymbol{\lambda}_b$ is the skewness parameter vector of the random effects and $H = H(.;\boldsymbol{v})$ is the cdf of the assumed distribution. The $\boldsymbol{b}_i \sim SNI_q(0, \boldsymbol{D}, \boldsymbol{\lambda}_b; H)$ and $\boldsymbol{e}_i \sim NI_{s_i}(0, \boldsymbol{R}_i; H)$ are indexed by the same scale mixing factor u_i, so, they are not independent in general.

The mixed effects model assumes that $\boldsymbol{e}_i$ are symmetrically distributed, while the distribution of random effects is assumed to be asymmetric. That is, the skewness parameter $\boldsymbol{\lambda}$, a measure of asymmetry, are all zeros, where the MSNI-NMM reduced to the MNI-NMM (Arellano-Valle et al. 2005; Lin and Lee 2008). When $\lambda = 0$ and U is distributed as $\text{Gamma}(\nu/2, \nu/2)$, the MSNI-LMM reduced to the (hierarchical) Mt-NLMM.

When the dimensions of $\boldsymbol{R}_i$ are large, estimation can be burdensome. Thus, a parsimonious structure for $\boldsymbol{R}_i$ is used. Accordingly, we assume that $\boldsymbol{e}_{i,k} \sim NI_r(0, \boldsymbol{W}, H)$, where $\boldsymbol{W} = [\sigma_{jj'}]$ explains the unstructured varinaces/covariances among the r outcome variables. Also, we assume that $\boldsymbol{e}_{ij} \sim NI_{n_i}(0, \sigma_{jj'}\boldsymbol{C}_i, H)$, where $\boldsymbol{C}_i$ is a time varying dependence structure for the autocorrelation among n_i occasions. Hence, we have $\boldsymbol{R}_i = \boldsymbol{W} \otimes \boldsymbol{C}_i(\xi, \psi)$, where $\otimes$ denotes the Kronecker product.

3. Maximum Likelihood Inference

3.1. Hierarchical formulation of MSNI-NLMM

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{D}, \mathbf{W}, \xi, \psi, \boldsymbol{\lambda}, \mathbf{v})$ be the entire model parameters. The likelihood function of $\boldsymbol{\theta}$ is formed by the product of the marginal density of each $\mathbf{y}_i$ for $i = 1,\ldots,m$, obtained by multiplying the marginal density of $\mathbf{b}_i$ by the conditional density of $\mathbf{y}_i$ given $\mathbf{b}_i$ and then integrating out $\mathbf{b}_i$ The exact marginal distribution of $\mathbf{y}_i$ cannot be analytically determined because marginalizing out $\mathbf{b}_i$ involves complex multidimensional integration. For ease of computation and theoretical derivation, we reparametrize $\mathbf{D} = \mathbf{F}'\mathbf{F}$, namely, $\mathbf{F} = \mathbf{D}^{1/2}$. The models in Equation (3) and Equation (5) have four-level hierarchical specification as

$$\mathbf{y}_i \left| (\mathbf{b}_i, \gamma_i, u_i) \sim N_{s_i}\left(\boldsymbol{\mu}_i(\boldsymbol{\beta}, \mathbf{b}_i), u_i^{-1}\mathbf{R}_i\right), \quad \mathbf{b}_i \right| (\gamma_i, u_i) \sim N_q\left(\boldsymbol{\alpha}\gamma_i, u_i^{-1}\boldsymbol{\Lambda}\right),$$
$$\gamma_i | u_i \sim TN\left((0, u_i^{-1}) | (0, \infty)\right), \quad u_i \sim H(u_i | \mathbf{v}), \tag{6}$$

for $i = 1,\ldots,m$, where $\boldsymbol{\alpha} = \mathbf{F}\boldsymbol{\delta}, \boldsymbol{\Lambda} = \mathbf{D} - \boldsymbol{\alpha}\boldsymbol{\alpha}^T$ with $\boldsymbol{\delta} = \boldsymbol{\delta}(\boldsymbol{\lambda}) = \dfrac{\boldsymbol{\lambda}}{\sqrt{1 + \boldsymbol{\lambda}^T\boldsymbol{\lambda}}} \in (-1,1)^q$, and $\mathbf{F}$ being the square root of $\mathbf{D}$ containing $q(q+1)/2$ distinct elements. $TN((\mu, \tau)|(a,b))$ denotes the univariate normal distribution $(N(\mu, \tau))$ truncated on the interval (a,b).

Let $\mathbf{y}_c = (\mathbf{y}_i^T, \mathbf{b}_i^T, \gamma_i^T, u_i^T)^T$, with $\mathbf{y} = (\mathbf{y}_1^T, \ldots, \mathbf{y}_m^T)^T$, $\mathbf{b} = (\mathbf{b}_1^T, \ldots, \mathbf{b}_m^T)^T$, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)^T$, $\mathbf{u} = (u_1, \ldots, u_m)^T$ and let $\hat{\boldsymbol{\theta}}^{(c)} = \left(\hat{\boldsymbol{\beta}}^{(c)T}, \hat{\mathbf{F}}^{(c)T}, \hat{\mathbf{W}}^{(c)T}, \hat{\xi}^{(c)T}, \hat{\psi}^{(c)T}, \hat{\boldsymbol{\lambda}}^{(c)T}, \hat{\mathbf{v}}^{(c)T}\right)^T$, denote the estimate $\boldsymbol{\theta}$

at the c^{th} iteration. The marginal density of $\mathbf{y}_i$, when $\mathbf{y}_i|(\mathbf{b}_i,\gamma_i,u_i) \sim N_{s_i}\left(\boldsymbol{\mu}_i(\boldsymbol{\beta},\mathbf{b}_i),u_i^{-1}\mathbf{R}_i\right)$, $\mathbf{b}_i|(\gamma_i,u_i) \sim N_q(\boldsymbol{\alpha}\gamma_i,u_i^{-1}\boldsymbol{\Lambda})$, $\gamma_i|u_i \sim TN(0,u_i^{-1})$ and $U_i = u_i \sim H(u_i|\mathbf{v})$ is

$$f(\mathbf{y}_i|\boldsymbol{\theta}) = \int\int_0^\infty\int_{\mathbb{R}^q}\left[\phi_{s_i}\left(\mathbf{y}_i|(\mathbf{b}_i,\gamma_i,u_i)\right)\phi_q\left(\mathbf{b}_i|(\gamma_i,u_i)\right)\phi(\gamma_i|u_i)h(u_i|\mathbf{v})\right]d\mathbf{b}_i d\gamma_i du_i,$$

$$= \int\int_0^\infty\int_{\mathbb{R}^q}\left[\phi_{s_i}\left(\mathbf{y}_i|\boldsymbol{\mu}_i(\boldsymbol{\beta},\mathbf{b}_i),u_i^{-1}\mathbf{R}_i\right)\phi_q\left(\mathbf{b}_i|\boldsymbol{\alpha}\gamma_i,u_i^{-1}\boldsymbol{\Lambda}\right)\phi\left(\gamma_i|0,u_i^{-1}\right)h\left(u_i|\mathbf{v}\right)\right]d\mathbf{b}_i d\gamma_i du_i,$$

$$= \int\int_0^\infty\int_{\mathbb{R}^q}(2\pi)^{\frac{-(s_i+q+1)}{2}}|\mathbf{R}_i|^{-\frac{1}{2}}|\boldsymbol{\Lambda}|^{-\frac{1}{2}}u_i^{\frac{(s_i+q+1)}{2}}$$

$$\exp\left(-\frac{u_i}{2}\left(\left(\mathbf{y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta},\mathbf{b}_i)\right)^T\mathbf{R}_i^{-1}\left(\mathbf{y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta},\mathbf{b}_i)\right)+(\mathbf{b}_i-\boldsymbol{\alpha}\gamma_i)^T\boldsymbol{\Lambda}^{-1}\left(\mathbf{b}_i-\boldsymbol{\alpha}\gamma_i\right)+\gamma_i^2\right)+\log\left(h\left(u_i|\mathbf{v}\right)\right)\right)d\mathbf{b}_i d\gamma_i du_i. \tag{7}$$

The marginal density of $\mathbf{y}_i$, when $\mathbf{y}_i|(\mathbf{b}_i,u_i) \sim N_{s_i}\left(\boldsymbol{\mu}_i(\boldsymbol{\beta},\mathbf{b}_i),u_i^{-1}\mathbf{R}_i\right)$ and $\mathbf{b}_i|u_i \sim SNI_q(0,u_i^{-1}\mathbf{F}^T\mathbf{F},\boldsymbol{\lambda}_b,H)$ is

$$f(\mathbf{y}_i|\boldsymbol{\theta}) = \int_{\mathbb{R}^q}\int_0^\infty\phi_{s_i}\left(\mathbf{y}_i|\boldsymbol{\mu}_i(\boldsymbol{\beta},\mathbf{b}_i),u_i^{-1}\mathbf{R}_i\right)\phi_q(\mathbf{b}_i|0,u_i^{-1}\mathbf{F}^T\mathbf{F},\boldsymbol{\lambda}_b,H)dH(u_i|\mathbf{v})d\mathbf{b}_i,$$

$$= 2\int_0^\infty\int_{\mathbb{R}^q}(2\pi)^{-(s_i+q)/2}|\mathbf{R}_i|^{-1/2}|\mathbf{F}|^{-1}u_i^{(s_i+q)/2}$$

$$\exp\left(-\frac{u_i}{2}\left(\left(\mathbf{y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta},\mathbf{b}_i)\right)^T\mathbf{R}_i^{-1}\left(\mathbf{y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta},\mathbf{b}_i)\right)+\mathbf{b}_i^T(\mathbf{F}\mathbf{F}^T)^{-1}\mathbf{b}_i\right)\right)$$

$$\Phi\left(\frac{u_i^{\frac{1}{2}}\boldsymbol{\delta}^T\mathbf{F}^{-1}\mathbf{b}_i}{\sqrt{1-\boldsymbol{\delta}^T\boldsymbol{\delta}}}\right)d\mathbf{b}_i dH\left(u_i|\mathbf{v}\right). \tag{8}$$

From Equation (7), we can obtain the marginal distribution of $\mathbf{y}_i$ by integrating over $(\mathbf{b}_i^T,\gamma_i^T,u_i^T)$ and the joint distribution of $(\mathbf{y}_i^T,\mathbf{b}_i^T,\gamma_i^T,u_i^T)$, where $\phi_{s_i}\left(\mathbf{y}_i|(\mathbf{b}_i,\gamma_i,u_i)\right)$, $\phi_q\left(\mathbf{b}_i|(\gamma_i,u_i)\right)$, $\phi(\gamma_i|u_i)$, $h(u_i|\mathbf{v})$ distributed as a multivariate normal, a multivariate normal, a truncated normal and a pdf of cdf $H(u_i|\mathbf{v})$, respectively. Also from Equation (8), we can obtain the marginal distribution of $\mathbf{y}_i$ by integrating over $(\mathbf{b}_i,u_i)$ the joint distribution of $(\mathbf{y}_i^T,\mathbf{b}_i^T,u_i^T)$, where $\phi_{s_i}\left(\mathbf{y}_i|(\mathbf{b}_i,u_i)\right)$, $\phi_q(\mathbf{b}_i|u_i)$ distributed as a multivariate normal, a multivariate skew normal independent, respectively.

The log-likelihood function for the observed data $\mathbf{Y}=\{\mathbf{y}_i\}_{i=1}^m$ of Equations (7) and (8), respectively, is

$$l(\boldsymbol{\theta}|\mathbf{Y}_C) = A_i - \frac{1}{2}\sum_{i=1}^m\left[-2\log\int_0^\infty B_i\, dH(u_i|\mathbf{v})\right], \tag{9}$$

where

$$A_i = \log\frac{1}{4}+(s_i+q)\log(2\pi)+\log|\mathbf{R}_i|+\log|\mathbf{F}\mathbf{F}^T|,$$

$$B_i = u_i^{(s_i+q)/2}\exp\left(-\frac{u_i}{2}\left(\left(\mathbf{y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta},\mathbf{b}_i)\right)^T\mathbf{R}_i^{-1}\left(\mathbf{y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta},\mathbf{b}_i)\right)+\mathbf{b}_i^T(\mathbf{F}\mathbf{F}^T)^{-1}\mathbf{b}_i\right)+\log\Phi\left(\frac{u_i^{\frac{1}{2}}\boldsymbol{\delta}^T\mathbf{F}^{-1}\mathbf{b}_i}{\sqrt{1-\boldsymbol{\delta}^T\boldsymbol{\delta}}}\right)\right)$$

$$l\left(\boldsymbol{\theta}|\mathbf{Y}_C\right) = -\frac{1}{2}\sum_{i=1}^m\left[A_i+B_i\right]. \tag{10}$$

where

$$A_i = \log\frac{1}{4} + (s_i + q + 1)\log(2\pi) + \log|\mathbf{R}_i| + \log|\boldsymbol{\Lambda}| - 2\log h(u_i|\nu) - (s_i + q + 1)\log u_i$$

$$B_i = u_i\left((\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta},\mathbf{b}_i))^T \mathbf{R}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta},\mathbf{b}_i)) + (\mathbf{b}_i - \boldsymbol{\alpha}\gamma_i)^T \boldsymbol{\Lambda}^{-1}(\mathbf{b}_i - \boldsymbol{\alpha}\gamma_i) + \gamma_i^2\right).$$

3.2. Maximum likelihood estimation via the PNLS-MLME procedure of MSNI-NLMM

The procedure consists of two steps: a penalized nonlinear least squares (PNLS) step and a multivariate LME (MLME) step. The basic idea is to estimate the unobservable random effects $\boldsymbol{b}_i$ via the PNLS step and then update the ML estimates of parameters $\boldsymbol{\theta}$ based on the formulation of MLMM for the pseudo-data. The proposed PNLS-MLME procedure is described as follows.

The PNLS step: According to Equation (9), we first define:

$$g(\mathbf{y}_i,\mathbf{b}_i,u_i,\boldsymbol{\theta}) = -2\log\int_0^\infty u_i^{\frac{(s_i+q)}{2}} \exp(A_i)dH(u_i|\boldsymbol{\nu}), \tag{11}$$

$$A_i = -\frac{u_i}{2}\left((\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta},\mathbf{b}_i))^T \mathbf{R}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta},\mathbf{b}_i)) + \mathbf{b}_i^T(\mathbf{F}\mathbf{F}^T)^{-1}\mathbf{b}_i\right) + \log\Phi\left(\frac{u_i^{1/2}\boldsymbol{\delta}^T\mathbf{F}^{-1}\mathbf{b}_i}{\sqrt{1-\boldsymbol{\delta}^T\boldsymbol{\delta}}}\right),$$

where $\mu_i(\boldsymbol{\beta},\mathbf{b}_i) = \boldsymbol{\mu}_i(\phi_i, \mathbf{X}_i)$, for $i = 1,\ldots,m$, is a function of fixed effects and random effects $\mathbf{b}_i$. Fixing the current estimates of parameters $\hat{\boldsymbol{\theta}}^{(c)} = (\hat{\boldsymbol{\beta}}^{(c)},\hat{\mathbf{F}}^{(c)},\hat{\mathbf{W}}^{(c)},\hat{\xi}^{(c)},\hat{\psi}^{(c)},\hat{\boldsymbol{\lambda}}^{(c)},\hat{\mathbf{v}}^{(c)})$ and $\left\{\hat{u}_i^{(c)}\right\}_{i=1}^m$, the conditional modes of random effects $\mathbf{b}_i$ are obtained through minimizing a penalized nonlinear least-squares objective function:

$$\left\{\hat{\mathbf{b}}_i^{(c)}\right\}_{i=1}^m = \arg\min\sum_{i=1}^m g(\mathbf{y}_i,\mathbf{b}_i,\hat{u}_i^{(c)},\hat{\boldsymbol{\theta}}^{(c)}). \tag{12}$$

In practice, solving over $\hat{\mathbf{b}}_i^{(c)}$ for each subject can be implemented by minimizing $g(\mathbf{y}_i, \mathbf{b}_i, \hat{u}_i^{(c)}, \hat{\boldsymbol{\theta}}^{(c)})$ in Equation (11) with respect to q -dimensional random effects.

The MLME step: The parameter estimates are updated by utilizing the first-order Taylor expansion of Equation (3) around the current estimates $\hat{\phi}_i^{(c)} = \mathbf{X}_i\hat{\boldsymbol{\beta}}^{(c)} + \mathbf{Z}_i\hat{\mathbf{b}}_i^{(c)}$. The pseudo-data is denoted by $\tilde{y}_{ij,k} = y_{ij,k} - \mu_j(\hat{\phi}_i^{(c)},\mathbf{x}_{ij,k}) + \tilde{\mathbf{x}}_{ij,k}\hat{\boldsymbol{\beta}}^{(c)} + \tilde{\mathbf{Z}}_{ij,k}\hat{\mathbf{b}}_i^{(c)}$, where $\mathbf{x}_{ij,k} = \dot{\mu}_j(\hat{\phi}_i^{(c)},\mathbf{x}_{ij,k})^T\mathbf{X}_i$ and $\mathbf{Z}_{ij,k} = \dot{\mu}_j(\hat{\phi}_i^{(c)},\mathbf{x}_{ij,k})^T Z_i$. Consequently, the model for the super vector of the pseudo-data for the i^{th} subject is

$$\tilde{\mathbf{y}}_i = \tilde{\mathbf{X}}_i\boldsymbol{\beta} + \tilde{\mathbf{Z}}_i\mathbf{b}_i + \mathbf{e}_i, \tag{13}$$

where $\tilde{\mathbf{y}}_i$ is a $n_i r\times 1$ vector composed of r pseudo-response vectors $\mathbf{y}_{ij} = (y_{ij,1},\ldots,y_{ij,n_i})^T$, $\tilde{\mathbf{X}}_i$ is a $n_i r\times p$ matrix with rows made up of $p\times 1$ vector $\tilde{\mathbf{x}}_{ij,k}$ and $\tilde{\mathbf{Z}}_i$ is a $n_i r\times q$ matrix with rows made up of $q\times 1$ vector $\tilde{\mathbf{z}}_{ij,k}$. According to Equation (13), it is easy to verify that

$$\tilde{\mathbf{y}}_i \sim SNI_{s_i}(\tilde{\mathbf{X}}_i\boldsymbol{\beta},\tilde{\boldsymbol{\Sigma}}_i,\tilde{\boldsymbol{\lambda}}_i,H), \tag{14}$$

where $\tilde{\boldsymbol{\Sigma}}_i = \tilde{\mathbf{Z}}_i\mathbf{F}\mathbf{F}^T\tilde{\mathbf{Z}}_i^T + \mathbf{W}\otimes\mathbf{C}_i$ and $\tilde{\boldsymbol{\lambda}}_i = \dfrac{\Sigma_i^{-1/2}\tilde{\mathbf{Z}}_i\mathbf{F}\mathbf{F}^T\boldsymbol{\tau}_i}{(1+\boldsymbol{\tau}_i^T\tilde{\boldsymbol{\Gamma}}_i\boldsymbol{\tau}_i)^{1/2}}$ with $\boldsymbol{\tau}_i = \mathbf{F}^{-1}\boldsymbol{\lambda}$, $\tilde{\boldsymbol{\Gamma}}_i = (\mathbf{F}\mathbf{F}^T + \tilde{\mathbf{Z}}_i\mathbf{R}_i\tilde{\mathbf{Z}}_i^T)^{-1}$. It follows immediately from Equations (2) and (14) that

$$\mathrm{f}\left(\tilde{\mathbf{y}}_{\mathrm{i}}\right)=2\int_{0}^{\infty}\phi_{\mathrm{s_i}}\left(\tilde{\mathbf{y}}_{i}\mid\tilde{\mathbf{X}}_{\mathrm{i}}\boldsymbol{\beta},\mathrm{u}_{\mathrm{i}}^{-1}\tilde{\boldsymbol{\Sigma}}_{\mathrm{i}}\right)\Phi\left(\mathrm{u}_{\mathrm{i}}^{1/2}\tilde{\mathrm{E}}_{\mathrm{i}}\right)\mathrm{dH}\left(\mathrm{u}_{\mathrm{i}}|\boldsymbol{\nu}\right), \tag{15}$$

where $\tilde{\mathrm{E}}_{\mathrm{i}}=(1+\boldsymbol{\tau}_{\mathrm{i}}^{\mathrm{T}}\tilde{\boldsymbol{\Gamma}}_{\mathrm{i}}\boldsymbol{\tau}_{\mathrm{i}})^{-1/2}\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{F}\tilde{\mathbf{Z}}_{\mathrm{i}}\tilde{\boldsymbol{\Sigma}}_{\mathrm{i}}^{-1}(\tilde{\mathbf{y}}_{\mathrm{i}}-\tilde{\mathbf{X}}_{\mathrm{i}}\boldsymbol{\beta})=\dfrac{\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{F}\tilde{\mathbf{Z}}_{\mathrm{i}}\tilde{\boldsymbol{\Sigma}}_{\mathrm{i}}^{-1}(\tilde{\mathbf{y}}_{\mathrm{i}}-\tilde{\mathbf{X}}_{\mathrm{i}}\boldsymbol{\beta})}{(1+\boldsymbol{\tau}_{\mathrm{i}}^{\mathrm{T}}\tilde{\boldsymbol{\Gamma}}_{\mathrm{i}}\boldsymbol{\tau}_{\mathrm{i}})^{1/2}}.$

From Equation (14), the three level hierarchy of MSNI-NLMM are

$$\tilde{\boldsymbol{y}}_i\left|(\gamma_i,U_i=u_i)\right)\sim N_{s_i}(\tilde{\boldsymbol{X}}_i\boldsymbol{\beta}+\tilde{\boldsymbol{d}}_i\gamma_i,u_i^{-1}\tilde{\boldsymbol{\Psi}}_i^{-1}),\ \gamma_i\left|U_{i=}u_i\sim TN\left((0,u_i^{-1})|(0,\infty)\right),\ U_{i=}u_i\sim H\left(u_i|\boldsymbol{\nu}\right).\right. \tag{16}$$

The parameter vector $\boldsymbol{\nu}$ is assumed to be known. In practice, the optimum value of $\boldsymbol{\nu}$ can be determined using the profile likelihood and the Schwarz information criterion. Note that $U_i=u_i\sim h(u_i\mid\nu)$, the conditional distribution of $\tilde{\boldsymbol{y}}_i$ given u_i follows $\tilde{\boldsymbol{y}}_i\left|U_i=u_i\sim SN_{s_i}(\tilde{\boldsymbol{X}}_i\boldsymbol{\beta},u_i^{-1}\tilde{\boldsymbol{\Sigma}}_i,\boldsymbol{\lambda})\right.$. In order to update $\hat{\boldsymbol{\beta}}^{(\mathrm{c})}$, we first set an initial guess of $\{u_i\}_{i=1}^m$ as

$$\hat{u}_i^{(c)}=\arg\min_{u_i}\left\{-\frac{u_i}{2}(\tilde{\boldsymbol{y}}_i-\tilde{\boldsymbol{X}}_i\boldsymbol{\beta})^T\tilde{\boldsymbol{\Sigma}}^{-1}(\tilde{\boldsymbol{y}}_i-\tilde{\boldsymbol{X}}_i\boldsymbol{\beta})+\log\Phi\left(u_i^{1/2}\tilde{E}_i\right)+\log h(u_i|\boldsymbol{\nu})\right\},$$

which maximizes the log-likelihood function for the complete-data $\left\{\tilde{\boldsymbol{y}}_i,u_i\right\}_{i=1}^m$. In this step, we update $\hat{\beta}^{(c)}$ by a generalized least-squares approach. The solution of $\left\{\hat{u}_i^{(\mathrm{c})}\right\}_{i=1}^m$ can be obtained by exploiting the *nlminb* optimization routine available in R package. The initial guess is generated from $h(u_i|\boldsymbol{\nu}^{(0)})$. Next, we perform a generalized least squares step

$$\hat{\boldsymbol{\beta}}^{(c+1)}=\sum_{i=1}^m\left[\hat{u}_i^{(\mathrm{c})}\tilde{\boldsymbol{X}}_i^T\tilde{\boldsymbol{\Sigma}}^{-1(\mathrm{c})}\tilde{\boldsymbol{X}}_i\right]^{-1}\sum_{i=1}^m\left[\hat{u}_i^{(\mathrm{c})}\tilde{\boldsymbol{X}}_i^T\tilde{\boldsymbol{\Sigma}}^{-1(\mathrm{c})}\tilde{\boldsymbol{y}}_i-\hat{u}_i^{1/2(\mathrm{c})}\eta^{(\mathrm{c})}\varsigma^{(\mathrm{c})}\right], \tag{17}$$

where $\varsigma^{(\mathrm{c})}=\dfrac{\tilde{\boldsymbol{x}}_i^{T(\mathrm{c})}\tilde{\boldsymbol{\Psi}}^{-1(\mathrm{c})}\tilde{\boldsymbol{d}}_i^{(\mathrm{c})}}{\tilde{a}_i^{(c)}}$, $\tilde{a}_i^{(\mathrm{c})}=(1+\tilde{\boldsymbol{d}}_i^T\tilde{\boldsymbol{\Psi}}_i^{-1}\tilde{\boldsymbol{d}}_i)^{1/2}$ and $\eta^{(\mathrm{c})}=\dfrac{\phi(u_i^{1/2}\tilde{E}_i)}{\Phi(u_i^{1/2}\tilde{E}_i)}$.

Given the current estimate $\hat{\beta}^{(\mathrm{c}+1)}$, we update $\hat{\boldsymbol{\omega}}^{(\mathrm{c})}=\left(\mathrm{vech}(\hat{\boldsymbol{F}}^{(\mathrm{c})}),\mathrm{vech}(\hat{\boldsymbol{W}}^{(\mathrm{c})}),\ \hat{\xi}^{(\mathrm{c})},\hat{\psi}^{(\mathrm{c})},\hat{\boldsymbol{\lambda}}^{(\mathrm{c})},\hat{\boldsymbol{\nu}}^{(\mathrm{c})}\right)$ by the Newton-Raphson method:

$$\hat{\boldsymbol{\omega}}^{(c+1)}=\hat{\boldsymbol{\omega}}^{(c)}-\hat{\boldsymbol{H}}_{\boldsymbol{\omega}\boldsymbol{\omega}}^{(c+1/2)^{-1}}\hat{\boldsymbol{s}}_{\boldsymbol{\omega}}^{(c+1/2)}, \tag{18}$$

where $\hat{\boldsymbol{S}}_{\boldsymbol{\omega}}^{(\mathrm{c}+1/2)}$ and $\boldsymbol{H}_{\boldsymbol{\omega}\boldsymbol{\omega}}^{(\mathrm{c}+1/2)}$ are the score vectors $\boldsymbol{s}_{\boldsymbol{\omega}}$ and Hessian matrix $\boldsymbol{H}_{\boldsymbol{\omega}\boldsymbol{\omega}}$ evaluated at $\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}^{(c+1)}$ and $\boldsymbol{\omega}=\hat{\boldsymbol{\omega}}^{(\mathrm{c})}$. Explicit expressions for elements in $\boldsymbol{s}_{\boldsymbol{\omega}}$ and $\boldsymbol{H}_{\boldsymbol{\omega}\boldsymbol{\omega}}$ can be derived as usual. Iterations of Equations (12), (17) and (18) continue until either the maximum number of iterations or the user-specified convergence tolerance has been achieved.

3.3. Pseudo-ECM algorithm of MSNI-NLMM

According to the pseudo-data model specified in Equation (13), treating the random effects $\{\boldsymbol{b}_i\}_{i=1}^m$, $\{u_i\}_{i=1}^m$ and $\{\gamma_i\}_{i=1}^m$ as latent data, we establish a complete data framework of the model in Equation (22). Given the pseudo-complete data $\{\boldsymbol{y}_i\}_{i=1}^m$, $\{\boldsymbol{b}_i\}_{i=1}^m$, $\{u_i\}_{i=1}^m$ and $\{\gamma_i\}_{i=1}^m$, the log-likelihood function of $\boldsymbol{\theta}$ is

$$\mathrm{l}\left(\boldsymbol{\theta}|\mathbf{Y}_{\mathrm{C}}\right)=-\frac{1}{2}\sum_{i=1}^m\left[\log\left|\boldsymbol{W}\otimes\boldsymbol{C}_i\right|+\log\left|\boldsymbol{\Lambda}\right|+tr\left((\boldsymbol{R}_i)^{-1}\boldsymbol{\Theta}_{1i}\right)+tr(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Theta}_{2i})-2\log h(u_i|\boldsymbol{\nu})\right]+C, \tag{19}$$

where C is a constant that is independent of the parameter vector $\boldsymbol{\theta}$ and $h(u_i|\boldsymbol{\nu})$ is a function that depends on $\boldsymbol{\theta}$ only through ν, $\boldsymbol{\Theta}_{1i} = u_i\tilde{\boldsymbol{e}}_i\tilde{\boldsymbol{e}}_i^T = u_i(\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\beta} - \tilde{\boldsymbol{Z}}_i\boldsymbol{b}_i)(\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\beta} - \tilde{\boldsymbol{Z}}_i\boldsymbol{b}_i)^T$, and $\boldsymbol{\Theta}_{2i} = u_i\left(\boldsymbol{b}_i - \boldsymbol{\alpha}\gamma_i\right)\left(\boldsymbol{b}_i - \boldsymbol{\alpha}\gamma_i\right)^T$. Letting

$$\hat{u}_i^{(c)} = E(u_i|\tilde{\boldsymbol{y}}_i, \hat{\boldsymbol{\theta}}^{(c)}),\ (\widehat{\mathrm{u}\boldsymbol{b}})_i^{(c)} = E(u_i\mathbf{b}_i|\tilde{\boldsymbol{y}}_i, \hat{\boldsymbol{\theta}}^{(c)}),\ (\widehat{\mathrm{u}\boldsymbol{b}\boldsymbol{b}^T})_i^{(c)} = E(u_i\mathbf{b}_i\boldsymbol{b}_i^T|\tilde{\boldsymbol{y}}_i, \hat{\boldsymbol{\theta}}^{(c)}),\ (\widehat{\mathrm{u}\gamma})_i^{(c)} = E(u_i\gamma_i|\tilde{\boldsymbol{y}}_i, \hat{\boldsymbol{\theta}}^{(c)}),$$

$$(\widehat{\mathrm{u}\gamma^2})_i^{(c)} = E(u_i\gamma_i^2|\tilde{\boldsymbol{y}}_i, \hat{\boldsymbol{\theta}}^{(c)}),\ (\widehat{\mathrm{u}\gamma\boldsymbol{b}})_i^{(c)} = E(u_i\gamma_i\mathbf{b}_i|\tilde{\boldsymbol{y}}_i, \hat{\boldsymbol{\theta}}^{(c)}),$$

and $\hat{V}_{1i}^{(c)} = E\left(\left(u_i^{1/2}\eta(u_i^{1/2}\tilde{\mu}_{\gamma_i}^{(c)}/\tilde{\sigma}_{\gamma_i}^{(c)})\right)|\tilde{\boldsymbol{y}}_i, \hat{\boldsymbol{\theta}}^{(c)}\right)$ with $\hat{\eta}_i^{(c)} = \phi(\hat{\hat{E}}_i)/\Phi(\hat{\hat{E}}_i)$ and $\hat{\hat{E}}_i = u_i^{1/2}\tilde{\mu}_{\gamma_i}^{(c)}/\tilde{\sigma}_{\gamma_i}^{(c)}$. It can be proved that

$$(\widehat{\mathrm{u}\gamma})_i^{(c)} = \hat{u}_i^{(c)}\hat{\tilde{\mu}}_{\gamma_i}^{(c)} + \hat{V}_{1i}^{(c)}\hat{\tilde{\sigma}}_{\gamma_i}^{(c)}, \tag{20}$$

$$(\widehat{\mathrm{u}\gamma^2})_i^{(c)} = \hat{u}_i^{(c)}\hat{\tilde{\mu}}_{\gamma_i}^{2\,(c)} + \hat{\tilde{\sigma}}_{\gamma_i}^{2\,(c)} + \hat{V}_{1i}^{(c)}\hat{\tilde{\mu}}_{\gamma_i}^{(c)}\hat{\tilde{\sigma}}_{\gamma_i}^{(c)}, \tag{21}$$

$$(\widehat{\mathrm{u}\boldsymbol{b}})_i^{(c)} = \hat{\tilde{\boldsymbol{\mu}}}_{\boldsymbol{b}_i}^{(c)}(\widehat{\mathrm{u}\gamma})_i^{(c)} + \hat{u}_i^{(c)}\hat{\tilde{\boldsymbol{\upsilon}}}_{\boldsymbol{b}_i}^{(c)}, \tag{22}$$

$$\begin{aligned}(\widehat{\mathrm{u}\boldsymbol{b}\boldsymbol{b}^T})_i^{(c)} &= \hat{\tilde{\boldsymbol{\Sigma}}}_{\boldsymbol{b}_i}^{(c)} + (\widehat{\mathrm{u}\boldsymbol{b}})_i^{(c)}(\widehat{\mathrm{u}\boldsymbol{b}})_i^{(c)T} = \hat{\tilde{\boldsymbol{\Sigma}}}_{\boldsymbol{b}_i}^{(c)} + \hat{\tilde{\boldsymbol{\mu}}}_{\boldsymbol{b}_i}^{(c)}\hat{\tilde{\boldsymbol{\mu}}}_{\boldsymbol{b}_i}^{(c)T}(\widehat{\mathrm{u}\gamma^2})_i^{(c)} \\ &+ \left(\hat{\tilde{\boldsymbol{\mu}}}_{\boldsymbol{b}_i}^{(c)}\hat{\tilde{\boldsymbol{\upsilon}}}_{\boldsymbol{b}_i}^{(c)T} + \hat{\tilde{\boldsymbol{\upsilon}}}_{\boldsymbol{b}_i}^{(c)}\hat{\tilde{\boldsymbol{\mu}}}_{\boldsymbol{b}_i}^{(c)}\right)\left(\widehat{\mathrm{u}\gamma}\right)_i^{(c)} + u_i^{(c)}\hat{\tilde{\boldsymbol{\upsilon}}}_{\boldsymbol{b}_i}^{(c)}\hat{\tilde{\boldsymbol{\upsilon}}}_{\boldsymbol{b}_i}^{(c)T},\end{aligned} \tag{23}$$

$$(\widehat{\mathrm{u}\gamma\boldsymbol{b}})_i^{(c)} = \hat{\tilde{\boldsymbol{\mu}}}_{\boldsymbol{b}_i}^{(c)}(\widehat{\mathrm{u}\gamma^2})_i^{(c)} + \hat{\tilde{\boldsymbol{\upsilon}}}_{\boldsymbol{b}_i}^{(c)}(\widehat{\mathrm{u}\gamma})_i^{(c)}, \tag{24}$$

where $\tilde{\mu}_{\gamma_i} = (1 + \tilde{\boldsymbol{d}}_i^T\tilde{\boldsymbol{\Psi}}_i^{-1}\tilde{\boldsymbol{d}}_i)^{-1}\tilde{\boldsymbol{d}}_i^T\tilde{\boldsymbol{\Psi}}_i^{-1}(\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\beta})$ and scale parameter $\tilde{\sigma}_{\gamma_i}^2 = (1 + \tilde{\boldsymbol{d}}_i^T\tilde{\boldsymbol{\Psi}}_i^{-1}\tilde{\boldsymbol{d}}_i)^{-1}$ in the truncated range $(0,\infty)$ are the mean and variance of γ_i, $\tilde{\boldsymbol{\mu}}_{\boldsymbol{b}_i} = (\boldsymbol{I}_q - \tilde{\boldsymbol{\Sigma}}_{b_i}\tilde{\boldsymbol{Z}}_i^T\boldsymbol{R}_i^{-1}\tilde{\boldsymbol{Z}}_i)\boldsymbol{\alpha} = \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{b}_i}\boldsymbol{\Lambda}^{-1}\boldsymbol{\alpha}$, and $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{b}_i} = (\tilde{\boldsymbol{Z}}_i^T\boldsymbol{R}_i^{-1}\tilde{\boldsymbol{Z}}_i + \boldsymbol{\Lambda}^{-1})^{-1}$ are the mean vector and the variance/covariance matrix of the conditional distribution of $\boldsymbol{b}_i$ given $\tilde{\boldsymbol{y}}_i$ and γ_i, respectively, $\tilde{\boldsymbol{\upsilon}}_{\boldsymbol{b}_i} = \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{b}_i}\tilde{\boldsymbol{Z}}_i^T\boldsymbol{R}_i^{-1}(\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\beta})$.

The expected complete-data log-likelihood function (Q-function) can be evaluated as follows:

E step: Evaluate the expected complete-data log-likelihood Function Equation (19) conditioning on the current estimates $\hat{\boldsymbol{\theta}}^{(c)}$ and the pseudo- responses $\tilde{\boldsymbol{y}}_i = \tilde{\boldsymbol{y}}_i(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{b}}_i^{(c-1)})$, which linearize the regression function around the previous estimates of mixed effects $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{b}}_i^{(c-1)})$ and should be updated at each iteration. This gives rise to the so-called Q-function:

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(c)}) = E_{u,b,\gamma}\left(l_c^p\left(\boldsymbol{\theta}|\tilde{\boldsymbol{y}}_c\right)|\widetilde{\boldsymbol{y}}, \hat{\boldsymbol{\theta}}^{(c)}\right)$$

$$= -\frac{1}{2}\sum_{i=1}^{m}\left(\log\left|\boldsymbol{W}\otimes\boldsymbol{C}_i\right| + \log\left|\boldsymbol{\Lambda}\right| + tr\left((\boldsymbol{W}\otimes\boldsymbol{C}_i)^{-1}\hat{\boldsymbol{\Theta}}_{1i}^{(c)}\right) + tr(\Lambda^{-1}\hat{\Theta}_{2i}^{(c)}) + (\widehat{\mathrm{u}\gamma^2})_i^{(c)} - 2\hat{\Theta}_{3i}^{(c)}\right), \tag{25}$$

where

$$\hat{\boldsymbol{\Theta}}_{1i}^{(c)} = E(u_i\tilde{\boldsymbol{e}}_i\tilde{\boldsymbol{e}}_i^T|\tilde{\boldsymbol{y}}_i, \hat{\boldsymbol{\theta}}^{(c)})$$

$$= \hat{u}_i(\tilde{\boldsymbol{y}}_i^{(c)} - \tilde{\boldsymbol{X}}\hat{\boldsymbol{\beta}}^{(c)})(\tilde{\boldsymbol{y}}_i^{(c)} - \tilde{\boldsymbol{X}}\hat{\boldsymbol{\beta}}^{(c)})^T + \tilde{\boldsymbol{Z}}_i^T(\widehat{\mathrm{u}\boldsymbol{b}\boldsymbol{b}^T})_i^{(c)}\tilde{\boldsymbol{Z}}_i - \tilde{\boldsymbol{Z}}_i(\widehat{\mathrm{u}\boldsymbol{b}})_i^{(c)}(\tilde{\boldsymbol{y}}_i^{(c)} - \tilde{\boldsymbol{X}}\hat{\boldsymbol{\beta}}^{(c)})^T - (\tilde{\boldsymbol{y}}_i^{(c)} - \tilde{\boldsymbol{X}}\hat{\boldsymbol{\beta}}^{(c)})(\widehat{\mathrm{u}\boldsymbol{b}})_i^{(c)}\tilde{\boldsymbol{Z}}_i^T,$$

$$\hat{\boldsymbol{\Theta}}_{2i}^{(c)} = \hat{u}_i(\hat{\boldsymbol{b}}_i^{(c)} - \hat{\boldsymbol{\alpha}}^{(c)}\hat{\gamma}_i^{(c)})(\hat{\boldsymbol{b}}_i^{(c)} - \hat{\boldsymbol{\alpha}}^{(c)}\hat{\gamma}_i^{(c)})^T = (\widehat{\mathrm{u}\boldsymbol{b}\boldsymbol{b}^T})_i^{(c)} + (\widehat{\mathrm{u}\gamma^2})_i^{(c)}\hat{\boldsymbol{\alpha}}^{(c)}\hat{\boldsymbol{\alpha}}^{(c)T} - (\widehat{\mathrm{u}\gamma\boldsymbol{b}})_i^{(c)}\hat{\boldsymbol{\alpha}}^{(c)T} - \hat{\boldsymbol{\alpha}}^{(c)}(\widehat{\mathrm{u}\gamma\boldsymbol{b}})_i^{(c)T}$$

$\hat{\Theta}_{3i}^{(c)} = E\left(\log h(u_i|\boldsymbol{v})\right)$, $\hat{R}_i^{(c)} = \hat{\boldsymbol{W}}^{(c)} \otimes \boldsymbol{C}_i(\hat{\xi}^{(c)},\hat{\psi}^{(c)})$, $\hat{\tilde{\boldsymbol{e}}}_i^{(c)} = E\left[(\tilde{\boldsymbol{e}}_i|\tilde{\boldsymbol{y}}_i,\hat{\boldsymbol{\theta}}^{(c)})\right] = \tilde{\boldsymbol{y}}_i^{(c)} - \tilde{\boldsymbol{X}}\hat{\boldsymbol{\beta}}^{(c)} - \tilde{\boldsymbol{Z}}_i\hat{\boldsymbol{b}}_i^{(c)}$ and $\tilde{\boldsymbol{y}}_i = \tilde{\boldsymbol{y}}_i(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{b}}_i^{(c-1)})$ represents the update pseudo-responses.

CM step: update the current estimates $\hat{\boldsymbol{\beta}}^{(c)},\hat{\boldsymbol{F}}^{(c)},\boldsymbol{W}^{(c)},\hat{\xi}^{(c)},\hat{\psi}^{(c)},\hat{\lambda}^{(c)}$ and $\hat{\boldsymbol{v}}^{(c)}$ by maximization the Q function in Equation (25):

$$\hat{\boldsymbol{\beta}}^{(c+1)} = \sum_{i=1}^{m}\left[\hat{u}_i^{(c)}\tilde{\boldsymbol{X}}_i^T\mathbf{R}_i^{-1(c)}\tilde{\boldsymbol{X}}_i\right]^{-1}\sum_{i=1}^{m}\left[\tilde{\boldsymbol{X}}_i^T\mathbf{R}_i^{-1(c)}\left(\hat{u}_i^{(c)}\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{Z}}_i\ \widehat{(\mathrm{u}\boldsymbol{b})}_i^{(c)}\right)\right],\quad \hat{\boldsymbol{\alpha}}^{(c+1)} = \frac{\sum_{i=1}^{m}\widehat{(\mathrm{u}\gamma\boldsymbol{b})}_i^{(c)}}{\sum_{i=1}^{m}\widehat{(\mathrm{u}\gamma^2)}_i^{(c)}},$$

$$\hat{\boldsymbol{\Lambda}}^{(c+1)} = \frac{1}{m}\sum_{i=1}^{m}\hat{\boldsymbol{\Theta}}_{2i}^{(c+1)},\quad \hat{\mathbf{D}}^{(c+1)} = \hat{\boldsymbol{\Lambda}}^{(c+1)} + \hat{\boldsymbol{\alpha}}^{(c+1)}\hat{\boldsymbol{\alpha}}^{(c+1)T},\quad \hat{\boldsymbol{\delta}}^{(c+1)} = \hat{\boldsymbol{F}}^{(c+1)-1}\hat{\boldsymbol{\alpha}}^{(c+1)},$$

where $\hat{\boldsymbol{F}}^{(c+1)}$ is the square root matrix of $\hat{\mathbf{D}}^{(c+1)}$. The updated estimates of $\boldsymbol{\lambda}$ can be calculated as $\hat{\boldsymbol{\lambda}}^{(c+1)} = \hat{\boldsymbol{\delta}}^{(c+1)}(1-\hat{\boldsymbol{\delta}}^{(c+1)T}\hat{\boldsymbol{\delta}}^{(c+1)})^{-1/2}$. For updating $\hat{\mathbf{W}}^{(c)}$, we define $\boldsymbol{W}^{-1} = \left[\sigma^{jl}\right]$ and $\boldsymbol{W} = \left[\sigma_{jl}\right]$, $j,l = 1,\ldots,r$. Given $\hat{\boldsymbol{\Theta}}_{1i}^{(c+1)}$, we update the diagonal-elements in $\hat{\mathrm{W}}^{(c)}$ which are defined as

$$\hat{\sigma}_{jj}^2 = \left(\sum_{i=1}^{m}n_i\right)^{-1}\sum_{i=1}^{m}tr\left(\mathbf{C}_i^{-1(c)}\hat{\boldsymbol{\Theta}}_{1\mathrm{ijl}}^{(c+1/2)}\right) \text{ for } j=l \text{ and}$$

$$\hat{\sigma}_{jj}^2 = \left(2\sum_{i=1}^{m}n_i\right)^{-1}\sum_{i=1}^{m}tr\left(\mathbf{C}_i^{-1(c)}\left(\hat{\boldsymbol{\Theta}}_{1\mathrm{ijl}}^{(c+1)} + \hat{\boldsymbol{\Theta}}_{1\mathrm{ilj}}^{(c+1/2)}\right)\right),\ \text{for } j \neq l,$$

where $\hat{\boldsymbol{\Theta}}_{1\mathrm{ijl}}^{(c+1/2)}$ is an n_i - dimensional square matrix, given by $\hat{\boldsymbol{\Theta}}_{1\mathrm{ijl}}^{(c+1/2)} = \tilde{\boldsymbol{e}}_{ij}^{(c)}\tilde{\boldsymbol{e}}_{il}^{T(c)} + \tilde{\boldsymbol{Z}}_i^T\widehat{(\mathrm{u}\gamma\boldsymbol{b})}_i^{(c)}\tilde{\boldsymbol{Z}}_i$, where $\tilde{\boldsymbol{e}}_{ij}^{(c)} = \tilde{\boldsymbol{y}}_{ij} - \tilde{\boldsymbol{X}}\hat{\boldsymbol{\beta}}_j^{(c+1)} - \tilde{\boldsymbol{Z}}_i\hat{\boldsymbol{b}}_{ij}^{(c)}$ and $\tilde{\boldsymbol{e}}_{il}^{(c)} = \tilde{\boldsymbol{y}}_{il} - \tilde{\boldsymbol{X}}\hat{\boldsymbol{\beta}}_l^{(c+1)} - \tilde{\boldsymbol{Z}}_i\hat{\boldsymbol{b}}_{il}^{(c)}$ with $\hat{\boldsymbol{b}}_{ij}^{(c)}$ and $\widehat{(\mathrm{u}\gamma\boldsymbol{b})}_{ijl}^{(c)}$ being a $q_j \times 1$ sub vector of $\tilde{\boldsymbol{\mu}}_{b_i}$ and a $q_j \times q_l$ sub matrix of $\widehat{(\mathrm{u}\gamma\boldsymbol{b})}_i^{(c)}$ respectively, evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(c+1/2)}$. Unfortunately, equating the first derivatives of Equation (31) with zero does not result in closed form. We use the *nlminb* routine in R Package to perform a numerical search for updating the $\left(\hat{\xi}^{(c+1)},\hat{\psi}^{(c+1)}\right)$. That is,

$$\left(\hat{\xi}^{(c+1)},\hat{\psi}^{(c+1)}\right) = \arg\max_{(\xi,\psi)}\left(r\sum_{i=1}^{m}\left|\boldsymbol{C}_i\right| - tr\left(\left(\boldsymbol{W}^{(c+1)}\otimes\boldsymbol{C}_i\right)^{-1}\hat{\boldsymbol{\Theta}}_{1\mathrm{i}}^{(c+1/2)}\right)\right).$$

The $\hat{v}^{(c)}$ is updated by optimizing the constrained log-likelihood function $\hat{\boldsymbol{v}}^{(c+1)} = \arg\max_{v} f\left(\tilde{\boldsymbol{y}}|\hat{\boldsymbol{\theta}}^{(c+1)},\boldsymbol{v}\right)$, where $f\left(\boldsymbol{y}|\boldsymbol{\theta}\right)$ is as in Equation (16).

3.4. Estimation of random effects and prediction of future values for MSNI-NLMM

The posterior density of $\boldsymbol{b}_i$ given $\left(\boldsymbol{Y}_i,\boldsymbol{U}_i\right) = \left(\boldsymbol{y}_i,\boldsymbol{u}_i\right)$ belong to the extended multivariate skew-normal (EMSN) family (Azzalini and Capitanio 1999)

$$f(\boldsymbol{b}_i|\tilde{\boldsymbol{y}}_i) = \frac{f(\tilde{\boldsymbol{y}}_i|\boldsymbol{b}_i)f(\boldsymbol{b}_i)}{\int f(\tilde{\boldsymbol{y}}_i|\boldsymbol{b}_i)f(\boldsymbol{b}_i)d\boldsymbol{b}_i} = \phi_q\left(\boldsymbol{\mu}_{b_i|\tilde{y}_i}, u_i^{-1}\boldsymbol{\Sigma}_{\boldsymbol{b}_i|\tilde{y}_i}\right)\Phi\left(u_i^{1/2}\boldsymbol{\lambda}^T\boldsymbol{D}^{-1/2}\boldsymbol{b}_i\right)\Phi\left(u^{1/2}\tilde{E}_i\right),$$

where $\boldsymbol{\mu}_{b_i|\tilde{y}_i} = \boldsymbol{D}\tilde{\boldsymbol{Z}}_i^T\tilde{\boldsymbol{\Sigma}}_i^{-1}\left(\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\beta}\right)$ and $\boldsymbol{\Sigma}_{\boldsymbol{b}_i|\tilde{y}_i} = \left[\boldsymbol{D}^{-1} + \tilde{\boldsymbol{Z}}_i^T\left(\boldsymbol{W}\otimes\boldsymbol{C}_i\right)\tilde{\boldsymbol{Z}}_i\right]^{-1}$. It is straightforward to show that the estimator that minimizes the overall mean square error (MSE) is

$\hat{\boldsymbol{b}}_i(\boldsymbol{\theta}) = \min_{\boldsymbol{b}_i^*} E\left[(\boldsymbol{b}_i - \boldsymbol{b}_i^*)(\boldsymbol{b}_i - \boldsymbol{b}_i^*)^T\right]$. A simple way to obtain the minimum MSE estimator $\hat{\boldsymbol{b}}_i$ is to use the conditional mean of $\boldsymbol{b}_i$ given observed data $\boldsymbol{y}_i$. After some algebraic manipulations, we obtain

$$\hat{\boldsymbol{b}}_i(\boldsymbol{\theta}) = E\left(\boldsymbol{b}_i | \tilde{\boldsymbol{y}}_i, \boldsymbol{\theta}\right) = \tilde{\boldsymbol{\mu}}_{\boldsymbol{b}_i|\tilde{y}_i} + \frac{\tilde{V}_{-1i}}{\left(1 + \boldsymbol{\lambda}_{\boldsymbol{b}}^T \boldsymbol{D}^{1/2} \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{b}_i|\tilde{y}_i} \boldsymbol{D}^{1/2} \boldsymbol{\lambda}_{\boldsymbol{b}}\right)^{1/2}} \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{b}_i|\tilde{y}_i} \boldsymbol{D}^{1/2} \boldsymbol{\lambda}_{\boldsymbol{b}},$$

where $\tilde{V}_{-1i} = E\left[\left(u_i^{-1/2} \eta(u_i^{1/2} \tilde{E}_i) | \tilde{\boldsymbol{y}}_i, \boldsymbol{\theta}\right)\right]$, with $\eta(u_i^{1/2} \tilde{E}_i) = \dfrac{\phi(u_i^{1/2} \tilde{E}_i)}{\Phi(u_i^{1/2} \tilde{E}_i)}$. Substituting $\boldsymbol{\theta}$ with ML estimates $\hat{\boldsymbol{\theta}}$, the estimates $\hat{\boldsymbol{b}}_i(\boldsymbol{\theta}) = \hat{\boldsymbol{b}}_i(\hat{\boldsymbol{\theta}})$ is called the empirical Bayes estimates. Therefore, the fitted values of responses can be calculated as

$$\hat{\boldsymbol{y}}_i = \mu_i(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{b}}_i), i = 1, \ldots, m. \tag{26}$$

4. Application

The proposed approaches are applied to real data concerns the HIV/AIDS. The University Hospitals of Cleveland, Rush Presbyterian St. Luke's Medical Center and University of Colorado Health Science Center have recruited 53 HIV-1 infected patients. The plasma RNA viral load and CD4 T cells of patients were repeatedly measured at days 0, 2, 7, 10, 14, 28, 56, 84, 168, and 196 after the start of ARV treatment. Five patients are excluded as four patients dropped out of the study prematurely and a patient left due to a problem with the study therapy. So, the analysis depends on 48 patients. The data have been analyzed by many authors such as Wu and Ding (1999).

To verify the existence of skewness in the random effects, we start by fitting a traditional MN-NLMM model as in Pinheiro and Bates (2000). Figure 1 describes the normal Q-Q plots of the empirical Bayes estimates of $\mathrm{b_i}$ and shows that there are some non-normal patterns on the random effects, including outliers and possibly skewness. This supports the use of thick-tailed distributions. Additionally, patients are not regularly measured in the study and the lengths of their follow-up times are distinct. These conclusions encourage applying a model which can flexibly handle asymmetric, unbalanced and irregularly observed multi-outcome longitudinal data.

Let $y_{i1,k}$ and $y_{i2,k}$ be $\log_{10}$ RNA and $\mathrm{CD4}^{0.5}$ markers, respectively, at the k^{th} occasion for the i^{th} patient. We use the bivariate functions for $y_{i1,k}$ and $y_{i2,k}$:

$$y_{i1,k} = \log_{10}\left(\exp\left(\phi_{1i} - \phi_{2i} t_{ik}\right) + \exp\left(\phi_{3i}\right)\right) + e_{i1,k},$$

$$y_{i2,k} = \frac{\phi_{4i}}{\left(1 + \exp\left(\dfrac{(\phi_{5i} - t_{ik})}{\phi_{6i}}\right)\right)} + e_{i2,k}, \tag{27}$$

where $\phi_{1i} = \beta_1 + \mathrm{b_{i1}}$, $\phi_{2i} = \beta_2$, $\phi_{3i} = \beta_3 \mathrm{RNA_i}$, $\phi_{4i} = \beta_4 + \mathrm{b_{i2}}$, $\phi_{5i} = \beta_5$, $\phi_{6i} = \beta_6$ and $\mathrm{t_{ik}} = \mathrm{day_{ik}} / 7$ is the k^{th} visit for patient i with RNA_i being hisher $\log_{10}(\mathrm{RNA})$ levels at the start of the study, $(\mathrm{b_{i1}}, \mathrm{b_{i2}})$ are the bivariate skew normal independent distribution of random effects; and $\left(e_{\mathrm{i1}}, e_{\mathrm{i2}}\right)^{\mathrm{T}} =$ $(e_{\mathrm{i1,1}}, \ldots, e_{\mathrm{i1,n_i}}, e_{\mathrm{i2,1}}, \ldots, e_{\mathrm{i2,n_i}})$ are the within-subject errors that follow a multivariate normal independent distribution with zero mean and variance-covariance $\mathrm{u_i^{-1}}(\mathbf{W} \otimes \mathbf{C}_\mathrm{i})$. Lin and Wang (2013) incorporated base RNA to the analysis because it is significant covariate.

(a) (b)

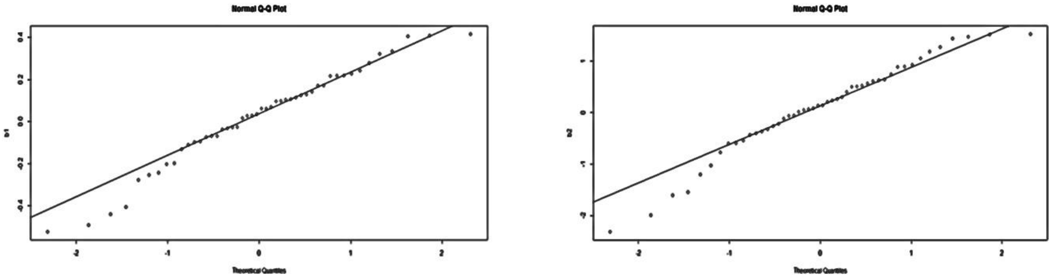

Figure 1. Normal Q-Q plots of estimated random effects based on MNLMM with
(a) $\mathrm{b}_{1\mathrm{i}}$ for $\log_{10}$ RNA and (b) $\mathrm{b}_{2\mathrm{i}}$ for $\mathrm{CD4}^{0.5}$

We fit the models MSNNLMM, MST-NLMM, MSS-NLMM and MSCN-NLMM with specific nonlinear mean functions. Three covariance structures have been considered for within-subject dependence; the uncorrelated (UNC) structure, a continuous-type autoregressive model with order1 AR(1), and the damped exponential correlation (DEC). The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are used for independent variables selection.

Table 1 summarizes the values of l_{max}, AIC and BIC for the different models under different covariance structures. The results show that the model with DEC covariance structure performs better than the one with UNC errors or AR (1) errors. The results of the l_{max}, AIC and BIC suggest that the best model is the multivariate skew slash-nonlinear mixed effect models (MSS-NLMM) with DEC dependence. The l_{max}, AIC and BIC criteria indicate that the multivariate skew normal independent (MSNI) model presents the best fit than the multivariate skew normal (MSN-NLMM) model, suggesting a departure of the data from normality.

The ML estimates and their standard errors (SE) of the four best fit models are presented in Table 2. The significance of the fixed effects parameters (the estimate relative to 2 SE) are similar for the MSN-NLMM, MST-NLMM, MSS-NLMM and MSCN-NLMM models. Under the fitted MSN-NLMM, MST-NLMM, MSS-NLMM and MSCN-NLMM, the estimated correlation coefficients of random effects $\hat{D}_{12} / \sqrt{\hat{D}_{11}\hat{D}_{22}}$ are 0.9097, 0.9232, 0.8827 and 0.8701, respectively. This means that there is a highly positive correlation between patient-specific variabilities of RNA viral load and CD4-T cells. The estimated correlation coefficients of within-subject errors $\hat{W}_{12} / \sqrt{\hat{W}_{11}\hat{W}_{22}}$ are -0.1381, -0.1261, -0.1216 and -0.1302, respectively. Hence, the relationship between the two responses, CD4-T cells and HIV-1 RNA levels (viral load), is negative and weak. The estimates of autocorrelation and damping parameters ξ and ψ show serial correlations among occasions. For testing $H_0: \psi = 1$ (AR(1) structure) versus $H_1: \psi \neq 1$ (DEC structure), the likelihood ratio test (LRT) statistic for the fitted MSN-NLMM, MST-NLMM, MSSNLMM and MSC-NLMM are 73.122, 71.964, 66.626 and 62.684, respectively, which is highly significant at any reasonable significance level. Hence, the null hypothesis is not supported by the data meaning that the DEC structure is more reasonable to data. The estimates of degrees of freedom of multivariate skew the t-distribution, the multivariate skew slash distribution and the multivariate skew contaminated normal distribution are $\hat{\nu} = 10$, $\hat{\nu} = 2.8533$, $\hat{\nu}_1 = 0.7725$, $\hat{\nu}_2 = 0.5234$, respectively. These estimates support the presence of heavy-tailed behavior of between and within patient variability. It is noticed that the estimate of re- scaled skewness parameters are significantly which support the appropriateness of multivariate SNI distribution.

Table 1 Comparison of fitting performances for the ACTG 315 data under Pseudo ECM. Bold values indicate the smallest value from each

C_i	Criteria	MSN	MST	MSS	MSC
UNC	n_{par}	14	15	15	16
	l_{max}	−1193.392	−1184.640	−1179.817	−1189.025
	AIC	2414.583	2399.296	2389.634	2410.050
	BIC	2440.780	2427.364	2417.702	2439.989
AR(1)	n_{par}	15	16	16	17
	l_{max}	−1106.288	−1102.378	−1093.172	−1097.072
	AIC	2242.576	2236.756	2218.344	2228.144
	BIC	2270.644	2266.695	2248.283	2259.954
DEC	n_{par}	16	17	17	18
	l_{max}	−1069.727	−1066.396	−1059.859	−1065.730
	AIC	**2171.454**	**2166.792**	**2153.718**	**2167.460**
	BIC	**2201.393**	**2198.602**	**2185.528**	**2201.142**

MSN=MSN-NLMM; MST= ST-NLMM; MSS=MSS-NLMM; MSC=MSC-NLMM; UNC=uncorrelated structure; AR(1)=continuous time autoregressive of order 1; DEC=damped exponential correlation; n_{par} =number of parameters; l_{max} = the maximum log-likelihood value.

Also, the results show that estimates of the fixed effects parameters are similar for the four fitted models, however the standard errors of the MST-NLMM, MSS-NLMM and MSCN-NLMM are smaller than those of the MSN-NLMM model for the most parameter of the model, indicating that the three models with longer tails than MSN seem to produce more accurate maximum likelihood estimates. The standard error approximation for the skewness parameter (δ_1) seems to be poor.

We next compare the fitted values, calculated by using Equation (26), based on the four best models selected in terms of AIC and BIC in each of MSN-NLMM, MST-NLMM, MSS-NLMM and MSC-NLMM with DEC dependence frameworks. The best model is MSSNLMM with DEC errors. To measure the discrepancy of the fitted values relative to original responses, we use the mean squared deviation $\left(\text{MSD} = \sum_{ij,k}(y_{ij,k} - \hat{y}_{ij,k})^2 / M\right)$ and mean absolute relative deviation $\left(\text{MARD} = \sum_{ij,k}\left|(y_{ij,k} - \hat{y}_{ij,k}) / y_{ij,k}\right| / M\right)$, where $M = \sum_{i=1}^{m} s_i$ is the number of entire observations. The quantities of MSD for MSN-NLMM, MST-NLMM, MSS-NLMM and MSC-NLMM with DEC dependence are 6.344, 5.828, 5.486 and 5.599, respectively. The quantities of MARD for MSN-NLMM, MST-NLMM, MSS-NLMM and MSC-NLMM with DEC dependence are 0.195, 0.189, 0.186 and 0.187, respectively. This justifies that the MSNI-NLMM provides a considerably better fitting performance than the MSN-NLMM.

Table 2 ML estimates and standard errors under the DEC-MSN, MST, MSS, MSCN-NLMM for ACTG315 data set, where $F = D^{1/2}$.

Parameters	MSN	MST	MSS	MSC
		$\nu = 10$	$\nu = 2.8533$	$\nu_1 = 0.7725$, $\nu_2 = 0.5234$
$\hat{\beta}_1$	11.4176 (0.2392)	11.3001 (0.2217)	11.3795 (0.2314)	11.3527 (0.2253)
$\hat{\beta}_2$	2.1253 (0.1205)	2.0593 (0.1101)	2.0925 (0.1164)	2.0807 (0.1145)
$\hat{\beta}_3$	1.2934 (0.0399)	1.2756 (0.0373)	1.2905 (0.0387)	1.2822 (0.0375)
$\hat{\beta}_4$	16.8892 (0.3294)	16.9068 (0.3092)	16.9383 (0.3159)	16.9139 (0.3074)
$\hat{\beta}_5$	−1.4524 (0.2387)	−1.6272 (0.2644)	−1.5225 (0.2487)	−1.5499 (0.2523)
$\hat{\beta}_6$	1.2017 (0.1940)	1.2851 (0.2073)	1.2208 (0.1980)	1.2516 (0.2010)
$\hat{F}_{11}$	0.1960 (0.4727)	0.2117 (0.3376)	0.1816 (0.4766)	0.1782 (0.4971)
$\hat{F}_{12}$	0.2775 (1.0274)	0.3156 (0.6535)	0.2304 (1.0530)	0.2155 (1.1009)
$\hat{F}_{22}$	1.4672 (0.7095)	1.5864 (0.3815)	1.2789 (0.6263)	1.2144 (0.6741)
$\hat{W}_{11}$	0.5770 (0.0410)	0.4899 (0.0350)	0.3875 (0.0278)	0.3329 (0.0238)
$\hat{W}_{12}$	−0.2951 (0.0382)	−0.2299 (0.0322)	−0.1724 (0.0254)	−0.1597 (0.0219)
$\hat{W}_{22}$	7.9108 (0.1874)	6.7845 (0.1731)	5.1877 (0.1408)	4.5189 (0.1140)
$\hat{\delta}_1$	−0.0801 (2.7080)	−0.1201 (0.9724)	0.3180 (1.1310)	0.2956 (1.2421)
$\hat{\delta}_2$	−0.9689 (2.229)	−0.9782 (0.1191)	−0.4620 (0.9280)	−0.3880 (1.1015)
$\hat{\xi}$	0.7715 (0.0135)	0.7695 (0.0136)	0.7745 (0.0134)	0.7615 (0.0141)
$\hat{\psi}$	0.5231 (0.0355)	0.5142 (0.0336)	0.5225 (0.0345)	0.5215 (0.0354)

5. Simulation Study

The aim of this simulation study is to evaluate the performance of proposed models. The data were generated from the MSN-NLMM, MST-NLMM, MSS-NLMM and MSC-NLMM with nonlinear mean curves as in Equation (27). The model parameters are fixed as

$$\beta = (12,3,1,17,-2,1)^T, \quad D = \begin{bmatrix} 0.1 & 0.42 \\ 0.42 & 2 \end{bmatrix}, \quad W = \begin{bmatrix} 1 & -0.26 \\ -0.26 & 7 \end{bmatrix}.$$

The $\boldsymbol{C}_i$ is 10×10 DEC dependence matrix, with $\xi=0.8$ and $\psi=0.5$, $\delta_1=0.3$ and $\delta_2=-0.48$. For simplicity the degrees of freedom $\boldsymbol{\nu}$ are fixed at its true value. The sample sizes are fixed at N = 25 (small sample) and N = 100 (relatively large sample).

For each sample size, 100 replications from the MSN-NLMM model and the MSNI–NLME model in Equation (27) were used under four scenarios: under the multivariate skew-normal model (MSN-NLME), the multivariate skew-t with ν = 2 (MST-NLME), the multivariate skew–slash with ν = 2.8 (MSS-NLME), and the multivariate skew-contaminated normal model with $\boldsymbol{\nu}$ = (0.1,0.5) (MSCN-NLME). The values of $\boldsymbol{\nu}$ were chosen to yield a highly skewed and heavy-tailed distribution for the random effects.

For each replication the empirical average ML estimates (EST) of parameters, empirical bias (bias) and empirical mean square error (MSE) over all samples were calculated. The computational procedures were implemented using the R software.

The results for N=25 are presented in Table 3 and for N=100 in Table 4. The results show that the bias and the mean square errors (MSE) for most of the parameters, decrease when the sample size increases, indicating an asymptotic convergence for true parameter values as the sample size increases. The results show that the proposed approximate ML estimates based on the EM algorithm provide good asymptotic properties. In general, the mean square errors (MC-MSE) provide results close to the empirical ones, and the closeness improves as the number of subjects increases. Under considered sample sizes, the fixed effect parameters have lower MSE under the multivariate skew normal (MSN), the MSE of the random components are smaller under the multivariate skew slash (MSS) distribution, the MSE of the variance-covariance components within subject errors and skew parameters are smaller under the multivariate skew contaminated normal (MSC) distribution. For considered sample sizes, multivariate skew normal independent-nonlinear mixed effects models (MSNI-NLMM) tend to give lower MSE values than multivariate skew normal-nonlinear mixed effects model (MSN-NLMM) for all parameter of models except fixed effect parameter.

Table 3 Simulation results (N = 25, 100 Replications): Mean estimates, bias, and MSE of parameters

Parameter (True)	Criteria	MSN	MSS	MST	MSCN
			$\nu=2.8$	$\nu=2$	$\nu_1=0.1$, $\nu_2=0.5$
β_1 (12)	EST	12.1282	12.0921	12.0581	11.9896
	bias	0.1282	0.0921	0.0581	0.0104
	MSE	0.2557	0.2572	0.2260	0.2377
β_2 (3)	EST	2.9946	3.0227	3.0066	3.0018
	bias	−0.0054	0.0227	0.0066	0.0018
	MSE	0.0422	0.0495	0.0315	0.0362
β_3 (1)	EST	1.0099	1.0079	1.0178	0.9946
	bias	0.0099	0.0079	0.0178	−0.0054
	MSE	0.0051	0.0074	0.0063	0.0057
β_4 (17)	EST	16.9928	17.1762	17.1572	16.9506
	bias	−0.0072	0.1762	0.1572	−0.0494
	MSE	0.1607	0.3291	0.4869	0.2314

Table 3 (Continued)

Parameter (True)	Criteria	MSN	MSS $\nu = 2.8$	MST $\nu = 2$	MSCN $\nu_1 = 0.1$, $\nu_2 = 0.5$
β_5 (-2)	EST	−1.9526	−2.2014	−2.1509	−2.0713
	bias	0.0474	−0.2014	−0.1509	−0.0713
	MSE	0.3843	1.0119	0.6413	0.6151
β_6 (1)	EST	0.9812	1.1235	1.0767	1.0526
	bias	−0.0188	0.1235	0.0767	0.0526
	MSE	0.1160	0.2893	0.1863	0.1888
D_{11} (0.1)	EST	0.0991	0.0874	0.0936	0.0960
	bias	−0.0009	−0.0126	−0.0064	−0.0040
	MSE	0.0021	0.0010	0.0018	0.0021
D_{12} (0.42)	EST	0.3895	0.3441	0.3723	0.3757
	bias	−0.0129	−0.0583	−0.0301	−0.0367
	MSE	0.0335	0.0202	0.0321	0.0267
D_{22} (2)	EST	1.9117	1.7319	1.8659	1.8600
	bias	−0.0883	−0.2681	−0.1341	−0.1400
	MSE	0.5907	0.4353	0.6085	0.4637
W_{11} (1)	EST	0.9544	0.8912	0.9559	0.9389
	bias	−0.0456	−0.1088	−0.0440	−0.0611
	MSE	0.0162	0.0265	0.0305	0.0222
W_{12} (−0.26)	EST	−0.2510	−0.2636	−0.2652	−0.2552
	bias	0.0135	0.0009	−0.0008	0.0093
	MSE	0.0225	0.0282	0.0262	0.0203
W_{22} (7)	EST	6.6581	6.2456	6.6681	6.6959
	bias	-0.3419	−0.7544	−0.3319	−0.3041
	MSE	0.6336	1.2880	1.4110	0.8570
δ_1 (0.3)	EST	0.3026	0.3145	0.2461	0.2903
	bias	0.0026	0.0145	−0.0539	−0.0097
	MSE	0.0026	0.0115	0.0496	0.0011
δ_2 (-0.48)	EST	−0.4616	−0.3492	−0.4325	−0.4711
	bias	0.0184	0.1308	0.0475	0.0089
	MSE	0.0199	0.1032	0.2141	0.0028
ξ (0.8)	EST	0.7877	0.7875	0.7839	0.7846
	bias	−0.0123	−0.0125	−0.0161	−0.0154
	MSE	0.0008	0.0008	0.0009	0.0009
ψ (0.5)	EST	0.4818	0.4860	0.4867	0.4974
	bias	−0.0182	−0.0140	−0.0035	−0.0026
	MSE	0.0035	0.0029	0.0030	0.0026

Table 4 Simulation results (N = 100, 100 Replications): Mean estimates, bias, and MSE of parameters

Parameter (True)	Criteria	MSN	MSS $\nu = 2.8$	MST $\nu = 2$	MSCN $\nu_1 = 0.1$, $\nu_2 = 0.5$
β_1 (12)	EST	11.9813	12.001	11.9464	12.0365
	bias	−0.0187	−0.001	−0.0537	0.0365
	MSE	0.0376	0.0694	0.0551	0.0626
β_2 (3)	EST	2.9971	2.9884	2.9895	2.9903
	bias	−0.0029	-0.0116	−0.0105	−0.0097
	MSE	0.0088	0.0118	0.0101	0.0112
β_3 (1)	EST	0.9960	1.0096	0.9938	1.0088
	bias	-0.0040	0.0097	−0.0062	0.0088
	MSE	0.0010	0.0017	0.0018	0.0013
β_4 (17)	EST	17.0121	16.9705	17.0498	17.0194
	bias	0.0121	-0.0295	0.0498	0.0194
	MSE	0.0544	0.0901	0.1245	0.0555
β_5 (-2)	EST	−1.9499	−2.0713	−1.9215	−2.0490
	bias	0.0501	−0.0713	0.0785	−0.0490
	MSE	0.1268	0.2415	0.1652	0.1351
β_6 (1)	EST	0.9675	1.0385	1.0246	1.0328
	bias	−0.0325	0.0385	−0.0299	0.0328
	MSE	0.0371	0.0667	0.0440	0.0394
D_{11} (0.1)	EST	0.1029	0.1006	0.0994	0.1029
	bias	0.0029	0.0006	-0.0006	0.0029
	MSE	0.0013	0.0007	0.0015	0.0012
D_{12} (0.42)	EST	0.4088	0.4012	0.4154	0.4110
	bias	0.0064	−0.0012	−0.0046	0.0086
	MSE	0.0241	0.0133	0.0270	0.0227
D_{22} (2)	EST	1.9997	1.9781	1.9719	2.0236
	bias	−0.0003	−0.0219	−0.0281	0.0236
	MSE	0.4539	0.2603	0.4903	0.4523
W_{11} (1)	EST	0.9879	0.9830	0.9932	0.9767
	bias	-0.0121	−0.0170	−0.0068	−0.0233
	MSE	0.0066	0.0070	0.0162	0.0058
W_{12} (−0.26)	EST	−0.2846	−0.2607	−0.2573	−0.2649
	bias	−0.0201	0.0038	0.0027	0.0004
	MSE	0.0068	0.0068	0.0095	0.0059
W_{22} (7)	EST	6.9366	6.9194	6.9581	6.8504
	bias	−0.0634	−0.0806	−0.0419	−0.1497
	MSE	0.3603	0.3739	0.8058	0.3284
δ_1 (0.3)	EST	0.2979	0.2837	0.2723	0.2959
	bias	−0.0021	−0.0163	−0.0277	−0.0041
	MSE	0.0037	0.0171	0.0136	0.0006

Table 4 (Continued)

Parameter (True)	Criteria	MSN	MSS	MST	MSCN
			$\nu = 2.8$	$\nu = 2$	$\nu_1 = 0.1$, $\nu_2 = 0.5$
δ_2	EST	−0.4666	−0.4515	−0.4500	−0.4795
(−0.48)	bias	0.0134	0.0285	0.0300	0.0005
	MSE	0.0111	0.0699	0.0805	0.0018
ξ	EST	0.7949	0.7937	0.7967	0.7932
(0.8)	bias	−0.0051	-0.0063	−0.0032	−0.0068
	MSE	0.0003	0.0002	0.0003	0.0002
ψ	EST	0.4934	0.4921	0.4938	0.4917
(0.5)	bias	−0.0066	−0.0079	−0.0062	−0.0068
	MSE	0.0005	0.0009	0.0008	0.0009

6. Conclusions

A robust extension of MNLMM by using multivariate skew normal independent (MSNI) distribution for the random effects and the multivariate normal independent (MNI) distribution for the within-subject errors has been introduced. It is assumed that the relationship between the response and the covariates to be nonlinear in parameters. The proposed model capable of handling a broader range of multivariate longitudinal data especially in the presence of outliers or heavy-tailed noises. The proposed model includes the MSN-NLMM, MST-NLMM, MSS-NLMM and MSCN-NLMM as special cases. We also consider the scenario where only a subset of the multiple responses can be collected at any occasion. The autocorrelation for responses at irregular time points is described by a parsimonious DEC function. This work generalizes the results of Schumacher et al. (2021a) and by developing some additional tools and making robust inferences in practical data analysis. We have described two flexible hierarchies for MSNI-NLMM. We developed computationally tractable PNLS-MLME procedure and Pseudo-ECM algorithm to obtain the ML estimates. We have created the pseudo data by using the first-order Taylor approximation and then implement the ECM algorithm to obtain the ML estimates. We also have created the pseudo data by using the first-order Taylor approximation and then implement multivariate linear mixed effects models (MLMM) to update the estimates of fixed effect by a generalized least-squares approach and estimates of the variance component by the Newton-Raphson method. The likelihood information-based method for approximating the standard errors of parameter estimates is also defined.

The proposed techniques are applied to the ACTG 315 data. This application supported flexibility of the MSS-NLMM among the robust distributions in terms of likelihood-based model selection criteria. The model with DEC dependence which takes into account the autocorrelation among occasions also performs better than the models with UNC errors and AR(1) errors. The analysis showed high positive correlation between patient-specific variabilities of RNA viral load and CD4-T cells. Also, the relationship between the two responses, CD4 cells and HIV-1 RNA levels (viral load) is negative. Furthermore, the simulation study showed that the proposed approximate ML estimates for fitted models based on the EM algorithm provide good asymptotic properties. The bias and the mean square error of the estimates generally decrease with the increase of the sample size.

Different venues of future research are possible. These include generalizing the multivariate SN distribution and SNI distribution depending on broader families of distributions, such as the multivariate skew t-distribution, the multivariate extended skew t-distribution, and the multivariate skew-elliptical distribution. Also, an imputation method to handle incomplete multiple repeated measures is possible. This can be done by adopting an extension of the multivariate nonlinear mixed effects model using multivariate skew normal and multivariate skew normal independent distributions for random effects and multivariate normal and multivariate normal independent distributions for the within-subject errors, taking the censoring information of multiple responses into account. A study to compare the proposed techniques with other methods such as the Monte Carlo EM (MCEM), the importance sampling EM (ISEM) and stochastic approximate EM (SAEM) algorithm to obtain the ML estimates of the multivariate version of skew-family nonlinear mixed models. This can be done by considering a more general structure for the within-subject covariance matrix, such as an AR(p) dependency structure (Schumacher et al. 2017).

References

Arellano-Valle RB, Bolfarine H, Lachos VH. Skew-normal linear mixed models. J Data Sci. 2005; 3(4): 415-438.

Azzalini A, Capitanio A. Statistical applications of the multivariate skew normal distribution. J R Stat Soc Ser B. 1999; 61(3): 579-602.

Lachos VH, Ghosh P, Arellano-Valle RB. Likelihood based inference for skew normal/independent linear mixed models. Stat Sin. 2010; 20(1): 303-322.

Lin TI, Lee JC. Estimation and prediction in linear mixed models with skew-normal random effects for longitudinal data. Stat Med. 2008; 27(9): 1490-1507.

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

Marshall G, De la Cruz-Mesía R, Barón AE, Rutledge JH, Zerbe GO. Non-linear random effects model for multivariate responses with missing data. Stat Med. 2006; 25(16): 2817-2830.

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.

Pereira MAA, Russo CM. Nonlinear mixed effects models with scale mixture of skew-normal distributions. J Appl Stat. 2019; 46(9): 1602-1620.

Pinheiro JC, Bates DM. Mixed-Effects Models in S and S-PLUS. New York: Springer; 2000.

Schumacher FL, Lachos VH, Dey DK. Censored regression models with autoregressive errors: A likelihood-based perspective. Can J Stat. 2017; 45(4): 375-392.

Schumacher FL, Dey DK, Lachos VH. Approximate inferences for nonlinear mixed effects models with scale mixtures of skew-normal distributions. J Stat Theory Pract. 2021a; 15(3): 1-26.

Schumacher FL, Lachos VH, Matos LA. Scale mixture of skew-normal linear mixed models with within-subject serial dependence. Stat Med. 2021b; 40(7): 1790-1810.

Wang WL. Approximate methods for maximum likelihood estimation of multivariate nonlinear mixed-effects models. Entropy. 2015; 17(8): 5353-5381.

Wang WL, Lin TI. Multivariate t nonlinear mixed-effects models for multi-outcome longitudinal data with missing values. Stat Med. 2014; 33(17): 3029-3046.

Wang WL, Lin TI. Multivariate-t nonlinear mixed models with application to censored multi-outcome AIDS studies. Biostatistics. 2017; 18(4): 666-681.

Wu H, Ding AA. Population HIV-1 dynamics in vivo: applicable models and inferential tools for virological data from AIDS clinical trials. Biometrics. 1999; 55(2): 410-418.