Multivariate *t* semiparametric mixed-effects model for longitudinal data with multiple characteristics

M. Taavoni, M. Arashi, Wan-Lun Wang & Tsung-I Lin

Published online: 01 Sep 2020.

Submit your article to this journal ↗

Article views: 292

View related articles ↗

View Crossmark data ↗

Citing articles: 5 View citing articles ↗

Taylor & Francis
Taylor & Francis Group

Check for updates

Multivariate *t* semiparametric mixed-effects model for longitudinal data with multiple characteristics

M. Taavoni[a], M. Arashi [ORCID][a], Wan-Lun Wang[b] and Tsung-I Lin [ORCID][c,d]

[a]Department of Statistic, Faculty of Mathematical Sciences, Shahrood University of Technology, Shahrood, Iran; [b]Department of Statistics, Graduate Institute of Statistics and Actuarial Science, Feng Chia University, Taichung, Taiwan; [c]Institute of Statistics, National Chung Hsing University, Taichung, Taiwan; [d]Department of Public Health, China Medical University, Taichung, Taiwan

ABSTRACT

Semiparametric mixed-effects models (SMM) have received increasing attention in recent years because of the greater flexibility in analysing longitudinal trajectories. However, the normality assumption of SMM may be unrealistic when outliers occur in the data. This paper presents a semiparametric extension of the multivariate *t* linear mixed-effects model (MtLMM), called the multivariate *t* semiparametric mixed model (MtSMM). To be specific, the MtSMM incorporates a parametric linear function related to the fixed covariate effects and random effects which have a joint multivariate *t* distribution together with an arbitrary nonparametric smooth function to capture the unexpected patterns. A computationally analytical EM-based algorithm is developed for carrying out maximum likelihood estimation of the MtSMM. Simulation studies and a real example concerning the analysis of PBCseq data are used to investigate the empirical behaviour of the proposed methodology.

1. Introduction

Analysis of longitudinal data, which aims to characterize the dependence of repeated measures between and within-subjects over time, has received growing attention in a variety of applications such as biological, economic, psychological, and clinical studies. A vast number of statistical models and methods have been proposed to handle such tasks. For examples, the linear mixed-effects model (LMM; Laird and Ware [1]) which utilizes the random component to take into account the variabilities of subjects is a commonly used analytical tool. Nevertheless, the linear relationship between the response and the covariates of LMM may be too restrictive to model any particular time trends using a simple parametric function, especially when the variety of response over time is in a complicated manner. To release such a limitation, Zeger and Diggle [2] proposed the semiparametric mixed models (SMM) in which the time effect is modelled nonparametrically. The parametric component in SMM provides a simple summary of covariates effects which are of

---

CONTACT M. Arashi ✉ m_arashi_stat@yahoo.com

main scientific interest, while the baseline function is included to enrich model flexibility. With both parametric and nonparametric components, SMM could be more flexible than the LMM. There exist general methods for appropriating the nonparametric functions such as the profile kernel, backfitting, smoothing spline, penalized spline and local linear regression methods. Recent work by Liu and Tu [3] explored the use of bivariate smooth functions in the semiparametric model for a pair of continuous longitudinal outcomes. Earlier developments of SMM can be found, for instance, in [4–9].

In longitudinal studies, it is common to have data with more than one response variables repeatedly measured on the same subject over a certain period of time leading to multivariate longitudinal data. A number of modelling methods have been developed specifically for dealing with multivariate longitudinal data. Shah et al. [10] originally proposed the multivariate linear mixed model (MLMM) and developed an iterative expectation maximization (EM) algorithm [11] to estimate the model parameters. For further developments of MLMM along with its applications, one can refer to Sammel et al. [12], Song et al. [13], Roy [14] and Wang and Fan [15], among others. To the best of our knowledge, however, the work discussing semiparametric models for multivariate longitudinal data is relatively rare up to now.

In the aforementioned models, the random effects and the within subject errors are routinely assumed to be normally distributed for desirable mathematical properties and computational tractability. However, such an assumption is not always tenable, especially when data contain influential and/or outlying observations. For a robust modelling of longitudinal data in the presence of potential outliers or atypical observations, a large amount of work has been done in the literature by adopting a more flexible class of distributions as a substitute for the routine use of normality assumption. For example, Pinheiro et al. [16] proposed a robust extension of LMMs by considering a joint multivariate-t distribution [17] for the random effects and within-subject errors, known as t linear mixed models (tLMM). Further developments along this line can be found in [18–21]. Under the framework of SMM, Qin and Zhu [7] and Sinha and Sattar [9] developed robust methods for maximum likelihood (ML) estimation of model parameters, which can be used to down weight the influence of potential outliers. However, the application of the referred approaches is limited to a single outcome longitudinal data. For robust inference against potential outliers in multi-outcome longitudinal data, Wang and Fan [22] have extended the tLMM to a multivariate setting, called the multivariate t linear mixed model (MtLMM). A comprehensive study that covers the ML and Bayesian methodologies, computational strategies and applications of the MtLMMs can be found in [23–26].

In spite of having robustness against outliers or atypical observations, the linearity scenario of MtLMMs can limit the practical use of the model to link the relationship between the responses and the covariates. Recently, Wang and Lin [27] and Wang and Castro [28] studied the multivariate t nonlinear mixed-effects model, which allows for the nonlinear fashion of the MtLMM and the robust extension of a multivariate nonlinear mixed-effects model (MNLMM; Marshall and Zerbe [29]), from the ML and Bayesian perspectives. Although the MtNLMM as well as the MNLMM can provide a wider range of practical uses than the linear models, they are considerably more complicated and computationally intensive. The main objective of this paper is to extend the existing SMM to a multivariate and robust settings, called multivariate t semiparametric mixed models (MtSMM). The proposed MtSMM, which incorporates the linear function in the fixed and random effects

to explore the effects of time-independent covariates and the time-dependent smooth function, providing a compromise between the MtLMM and MtNLMM approach. The proposed model indeed provides robustness in handling multivariate longitudinal data evolving potential outliers or heavy-tailed noises by considering a joint multivariate t distribution [17] for the random effects and within-subject errors, and adds extra flexibility by imposing an arbitrary smooth function of time in the framework of MtLMM. To carry out ML inference of the MtSMM, we consider a spline method for appropriating the nonparametric function and develop a computationally feasible expectation conditional maximization (ECM) algorithm [30] for estimating the unknown parameters.

The remainder of this paper is organized as follows. In Section 2, we establish the notation, formulate the MtSMM and describe the implementation of the smoothing spline method for estimation of the nonparametric function. An iterative ECM algorithm for computing ML estimates is described in Section 3. We proceed with a simulation study in Section 4 to examine the effectiveness of the proposed methodology and report the analysis results of a real data set. Finally, we conclude with a brief discussion in Section 5. All technical proofs are deferred to the Appendix.

2. Model specification

Suppose there are a sample of n subjects where the ith subject has r outcome variables with s_i repeatedly observations over time. Let $Y_i = [y_{i1} : \ldots : y_{ir}] = [y_{i,1}^\top : \ldots : y_{i,s_i}^\top]^\top$ be a $s_i \times r$ matrix of responses for the ith subject ($i = 1, \ldots, n$), where $y_{ij} = (y_{ij,1}, \ldots, y_{ij,s_i})^\top$ is a column vector of responses for the jth outcome ($j = 1, \ldots, r$) and $y_{i,k} = (y_{i1,k}, \ldots, y_{ir,k})$ is a row vector of responses collected at the kth occasion ($k = 1, \ldots, s_i$). Let $E_i = [e_{i1} : \ldots : e_{ir}] = [e_{i,1}^\top : \ldots : e_{i,s_i}^\top]^\top$ be a $s_i \times r$ matrix of within-subject errors, where $e_{ij} = (e_{ij,1}, \ldots, e_{ij,n_i})^\top$ is a column vector of error terms corresponding to y_{ij} and $e_{i,k} = (e_{i1,k}, \ldots, e_{ir,k})$ is a row vector of error terms corresponding to $y_{i,k}$. For notational convenience, we employ the vec($\cdot$) operator, which strings out all columns of a matrix vertically, to obtain stacked vectors, that is, $y_i = \text{vec}(Y_i)$ and $\varepsilon_i = \text{vec}(E_i)$ for the response variables and error terms, respectively, which are of $n_i = s_i r$ dimension.

The MtSMM for the ith subject can be expressed as

$$y_i = X_i \beta + g(t_i) + Z_i b_i + \varepsilon_i \tag{1}$$

along with the distributional assumption

$$\begin{bmatrix} b_i \\ \varepsilon_i \end{bmatrix} \sim t_{q+n_i} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} D & 0 \\ 0 & R_i \end{bmatrix}, v \right), \tag{2}$$

where $t_d(\mu, \Omega, v)$ denotes the multivariate t distribution with dimension d, location vector μ, scale-covariance matrix Ω, and degrees of freedom (DOF) v. Herein, $X_i = \text{diag}\{X_{i1}, \ldots, X_{ir}\}$ and $Z_i = \text{diag}\{Z_{i1}, \ldots, Z_{ir}\}$, where X_{ij} is a $s_i \times p_j$ full rank design matrix for fixed effects corresponding to the jth outcome of the ith subject, and Z_{ij}, formed usually by a subset of X_{ij}, is a $s_i \times q_j$ design matrix for random effects. Besides, $g(t_i) = (g_1(t_i)^\top, \ldots, g_r(t_i)^\top)^\top$, where $t_i = (t_{i,1}, \ldots, t_{i,si})^\top$ is measurement time points of the ith subject, $g_j(t_i) = (g_j(t_{i,1}), \ldots, g_j(t_{i,s_i}))^\top$ with $g_j(t_{i,k})$ being an unknown twice differentiable smooth function for the jth outcome measured at occasion $t_{i,k}$. The block diagonal

structures of X_i and Z_i allow the analysts to link the grand and subject specific relationships between covariates and each response, which is collected repeatedly at unequally spaced occasions for each subject, via distinct design matrices for each response. If the total dimensions of fixed effects and random effects are denoted by $p = \sum_{j=1}^{r} p_j$ and $q = \sum_{j=1}^{r} q_j$, respectively, then $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_r^\top)^\top$ is a $p \times 1$ vector of fixed effects with each $p_j \times 1$ subvector $\boldsymbol{\beta}_j$ used to describe the component mean profile of outcome j, and $\boldsymbol{b}_i = (\boldsymbol{b}_{i1}^\top, \ldots, \boldsymbol{b}_{ir}^\top)^\top$ is a $q \times 1$ vector of unobservable random effects with each $q_j \times 1$ subvector $\boldsymbol{b}_{ij}$ corresponding to subject-specific features on $\boldsymbol{y}_{ij}$. In the distributional assumption (2), $\boldsymbol{D}$ and $\boldsymbol{R}_i$ are scale-covariance matrices for random effects and within-subject errors, respectively, and ν is the DOF, which is a tuning parameter to harmonize the fatness of the tails of a distribution. Besides, $\boldsymbol{D} = [D_{jj'}]$ is a $q \times q$ symmetric and positive-definite matrix with $D_{jj'}$ being a partition matrix, in particular for $j = j'$, D_{jj} is a scale-covariance structure of random effects for the jth outcome, and for $j \neq j'$, $D_{jj'}$ is that for a pair of outcome variables.

For the sake of parsimony in scale-covariance matrix $\boldsymbol{R}_i$, we assume that $\boldsymbol{e}_{ij} \sim t_{s_i}(0, \sigma_{jj}\boldsymbol{\Omega}_i, \nu)$ for $j = 1, \ldots, r$, and $\boldsymbol{e}_{i,k} \sim t_r(0, \boldsymbol{\Sigma}, \nu)$ for $k = 1, \ldots, s_i$, where $\boldsymbol{\Sigma} = [\sigma_{jj'}]$ is used to describe the variances and covariances among r outcome variables, and $\boldsymbol{\Omega}_i$ is a time dependence correlation matrix used to address possibly serial correlation among s_i irregularly observed occasions. Accordingly, the within-subject error matrix $\boldsymbol{E}_i$ follows the matrix-t distribution [31], and thereby the stacked $n_i \times 1$ vector $\boldsymbol{\varepsilon}_i$ follows the multivariate t distribution with the DOF ν, location vector zero, and scale covariance matrix of having a Kronecker product (KP) structure, written as $\boldsymbol{R}_i = \boldsymbol{\Sigma} \otimes \boldsymbol{\Omega}_i$, which helps us to estimate $\boldsymbol{R}_i$ more accurately. To avoid the nonidentifiability problem resulting from nonunique solutions of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}_i$ in estimating $\boldsymbol{R}_i$ with a KP structure, we need to specify $\boldsymbol{\Omega}_i$ as a correlation matrix rather than a covariance matrix. To make estimation of $\boldsymbol{\Omega}_i$ to be more precise, we could choose a parsimonious structure on this correlation matrix, which can be a function of parameters $\boldsymbol{\phi}$ as well as time points $\boldsymbol{t}_i$, denoted by $\boldsymbol{\Omega}_i = \boldsymbol{\Omega}_i(\boldsymbol{\phi}, \boldsymbol{t}_i)$, based on the characteristics of the data at hand. Under model (1), it can be easily verified that $\boldsymbol{y}_i \sim t_{n_i}(X_i\boldsymbol{\beta} + \boldsymbol{g}(\boldsymbol{t}_i), \boldsymbol{\Lambda}_i, \nu)$, where $\boldsymbol{\Lambda}_i = Z_i D Z_i^\top + \boldsymbol{R}_i$.

As recommended by a number of literature related to the work of SMMs, see, for example, [5,7,9,32], the unspecified smooth function can be approximated sufficiently well by a spline basis function. Let $t_{i,1} = t_{ij}^{(1)} < \cdots < t_{ij}^{(L_j)} = t_{i,s_i}$ be a partition of the interval $[t_{i,1}, t_{i,s_i}]$ corresponding to the jth outcome of the ith subject, where $t_{ij}^{(l)}$ is the lth knot for $l = 1, \ldots, L_j$. Using a truncated polynomial spline basis function of order $d_j (\geq 1)$ for the jth outcome, $g_j(t_{i,k})$ can be represented as

$$g_j(t_{i,k}) = \alpha_{j,0} + \alpha_{j,1}t_{i,k} + \cdots + \alpha_{j,d_j}t_{i,k}^{d_j} + \sum_{l=1}^{L_j} \alpha_{(d_j+1)+l}(t_{i,k} - t_{ij}^{(l)})_+^{d_j} = \boldsymbol{B}_j(t_{i,k})\boldsymbol{\alpha}_j,$$

where $(a)_+ = \max(0, a)$, $\boldsymbol{B}_j(t_{i,k}) = (1, t_{i,k}, \ldots, t_{i,k}^{d_j}, (t_{i,k} - t_{ij}^{(1)})_+^{d_j}, \ldots, (t_{i,k} - t_{ij}^{(L_j)})_+^{d_j})$ is a $h_j \times 1$ vector of basis functions, and $\boldsymbol{\alpha}_j = (\alpha_{j,0}, \ldots, \alpha_{j,d_j}, \alpha_{j,d+1}, \ldots, \alpha_{j,h_j},)^\top$ is the spline coefficient vector of dimension $h_j = d_j + 1 + L_j$. Regression splines have some desirable properties in approximating a smooth function. It often provides good approximations with a small number of knots. The spline approach also treats a nonparametric function as

a linear one with the basic functions as pseudo design variables, and thus it linearizes our model (1) as

$$y_i = X_i\beta + B(t_i)\alpha + Z_i b_i + \varepsilon_i, \tag{3}$$

where $B(t_i) = \text{diag}(B_1(t_i), \ldots, B_r(t_i))$ with $B_j(t_i) = (B_j(t_{i,1})^\top, \ldots, B_j(t_{i,s_i})^\top)^\top$ being a $s_i \times h_i$ design matrix corresponding to the spline portion of the model. If the total dimension of spline effects for r outcomes is denoted by $h = \sum_{j=1}^r h_j$, then $\alpha = (\alpha_1^\top, \ldots, \alpha_r^\top)^\top$ is a $h \times 1$ vector of spline effects with each $h_j \times 1$ subvector α_j used for outcome j. For convenience, model (3) can be rewritten in the following form:

$$y_i = \tilde{X}_i\theta + Z_i b_i + \varepsilon_i, \tag{4}$$

where $\tilde{X}_i = \text{diag}\{\tilde{X}_{i1}, \ldots, \tilde{X}_{ir}\}$ with $\tilde{X}_{ij} = (X_{ij}, B_j(t_i))$ being a $s_i \times (p_j + h_j)$ design matrix combining the fixed-effects and spline-effects design matrices for the jth outcome of the ith subject, and $\theta = (\beta^\top, \alpha^\top)^\top$ is a $(p_j + h_j) \times 1$ combined regression parameter vector to be estimated. According to the linearization of MtSMM (1) coupled with (2) through a spline approach, the computational algorithm developed for fitting the MtLMM can be straightforwardly implemented for the MtSMM. Following Wang and Fan [22], we introduce a set of scaling variables τ_i, and then the MtSMM in (4) can be expressed as a three-level hierarchical form

$$y_i|(b_i, \tau_i) \sim N_{n_i}(\tilde{X}_i\theta + Z_i b_i, \tau_i^{-1}R_i),$$
$$b_i|\tau_i \sim N_q(0, \tau_i^{-1}D), \tag{5}$$
$$\tau_i \sim \text{Gamma}(\nu/2, \nu/2).$$

Recall from (4) that $b_i|\tau_i$ and $\varepsilon_i|\tau_i$ are implicitly assumed to be independent. Integrating out the b_i in (5), we can express model (4) in a two-level hierarchical form

$$y_i|\tau_i \sim N_{n_i}(\tilde{X}_i\theta, \tau_i^{-1}\Lambda_i), \quad \tau_i \sim \text{Gamma}(\nu/2, \nu/2). \tag{6}$$

3. ECM procedure for ML estimation

To carry out ML estimation of unknown parameters of model (4), we develop the EM-type algorithm [11] which is commonly used in handling the missing-data problem or the model with incomplete data. The EM algorithm, which consists of the expectation (E) step and the maximization (M) step at each iteration, has several appealing features such as simplicity of implementation and monotone increase of the likelihood at each iteration. Toward this end, we let $\Theta = (\theta, D, \Sigma, \phi, \nu)$ be the set of entire model parameters. Treating the unobservable random effects b_i and scaling variables τ_i as the missing data, we have the complete data $\{(y_i, b_i, \tau_i), i = 1, \ldots, n\}$. The complete-data log-likelihood function of Θ, omitting constant terms, is

$$\ell_c(\Theta) = \sum_{i=1}^n \ell_c^{[i]}(\Theta) = \sum_{i=1}^n \frac{1}{2}\left\{\log|R_i^{-1}| + \log|D^{-1}| - \tau_i[\varepsilon_i^\top R_i^{-1}\varepsilon_i + b_i^\top D^{-1}b_i]\right.$$
$$\left. + \nu\log\left(\frac{\nu}{2}\right) - 2\log\frac{\nu}{2} + \nu(\log\tau_i - \tau_i)\right\}. \tag{7}$$

For evaluating the conditional expectation of (7) given the observed data $\boldsymbol{y} = \{\boldsymbol{y}_i, i = 1, \ldots, n\}$ and the current estimates of parameters $\widehat{\boldsymbol{\Theta}}^{(r)} = (\widehat{\boldsymbol{\theta}}^{(r)}, \widehat{\boldsymbol{D}}^{(r)}, \widehat{\boldsymbol{\Sigma}}^{(r)}, \widehat{\boldsymbol{\phi}}^{(r)}, \widehat{\boldsymbol{\nu}}^{(r)})$, we utilize the results of Theorem 3 in Wang and Fan [22] to get the following proposition.

Proposition 3.1: *According to model* (4) *under the spline approach and the three-level hierarchical form* (5), *we have*

$$
\begin{bmatrix} \boldsymbol{y}_i \\ \boldsymbol{b}_i \end{bmatrix} | \tau_i \sim N_{n_i+q}\left(\begin{bmatrix} \tilde{X}_i\boldsymbol{\theta} \\ \mathbf{0} \end{bmatrix}, \quad \tau_i^{-1}\begin{bmatrix} \boldsymbol{\Lambda}_i & Z_i\boldsymbol{D} \\ \boldsymbol{D}Z_i^\top & \boldsymbol{D} \end{bmatrix} \right),
$$

$$
\boldsymbol{b}_i|(\boldsymbol{y}_i, \tau_i) \sim N_q\left(\boldsymbol{D}Z_i^\top \boldsymbol{\Lambda}_i^{-1}(\boldsymbol{y}_i - \tilde{X}_i\boldsymbol{\theta}), \tau_i^{-1}(\boldsymbol{D}^{-1} + Z^\top R_i^{-1}Z_i)^{-1} \right),
$$

$$
\tau_i|\boldsymbol{y}_i \sim \text{Gamma}\left((\nu + n_i)/2, (\nu + \Delta_i)/2 \right),
$$

where $\Delta_i = \Delta_i(\boldsymbol{\theta}, \boldsymbol{D}, \boldsymbol{\Sigma}, \phi, \nu) = \tilde{\boldsymbol{\epsilon}}_i^\top \boldsymbol{\Lambda}_i^{-1}\tilde{\boldsymbol{\epsilon}}_i$ *denotes the Mahalanobis distance between* $\boldsymbol{y}_i$ *and* $\tilde{X}_i\boldsymbol{\theta}$ *with* $\tilde{\boldsymbol{\epsilon}}_i = \boldsymbol{y}_i - \tilde{X}_i\boldsymbol{\theta}$.

The main steps of EM algorithm proceed as follows:

E-step: Compute the so-called Q function, $Q(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(r)}) = \mathrm{E}(\ell_c(\boldsymbol{\Theta})|\boldsymbol{y}, \widehat{\boldsymbol{\Theta}}^{(r)})$ which is the sum of $Q_i(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(r)}) = \mathrm{E}(\ell_c^{[i]}(\boldsymbol{\Theta})|\boldsymbol{y}_i, \widehat{\boldsymbol{\Theta}}^{(r)})$, for all $i = 1, \ldots, n$, given by

$$
Q_i(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(r)}) = \frac{1}{2}\Big\{ \log|R_i^{-1}| + \log|\boldsymbol{D}^{-1}| - \mathrm{tr}(\boldsymbol{D}^{-1}\widehat{\boldsymbol{B}}_i^{(r)}) - \mathrm{tr}(R_i^{-1}\widehat{\boldsymbol{\Psi}}_i^{(r)}(\boldsymbol{\theta}))
$$
$$
+ \nu\big(\log(\nu/2) + \widehat{\kappa}_i^{(r)} - \widehat{\tau}_i^{(r)} \big) \Big\} - \log(\nu/2),
$$

where

$$
\widehat{\tau}_i^{(r)} = \mathrm{E}(\tau_i|\boldsymbol{y}_i, \widehat{\boldsymbol{\Theta}}^{(r)}) = (\widehat{\nu}^{(r)} + n_i)/(\widehat{\nu}^{(r)} + \widehat{\Delta}_i^{(r)}),
$$

$$
\widehat{\kappa}_i^{(r)} = \mathrm{E}(\log\tau_i|\boldsymbol{y}_i, \widehat{\boldsymbol{\Theta}}^{(r)}) = \mathcal{D}_g\left(\frac{\widehat{\nu}^{(r)} + n_i}{2} \right) - \log\left(\frac{\widehat{\nu}^{(r)} + \widehat{\Delta}_i^{(r)}}{2} \right),
$$

$$
\widehat{\boldsymbol{B}}_i^{(r)} = \mathrm{E}(\tau_i\boldsymbol{b}_i\boldsymbol{b}_i^\top|\boldsymbol{y}_i, \widehat{\boldsymbol{\Theta}}^{(r)}) = \widehat{\tau}_i^{(r)}\widehat{\boldsymbol{b}}_i^{(r)}\widehat{\boldsymbol{b}}_i^{(r)\top} + \widehat{\boldsymbol{V}}_{b_i}^{(r)},
$$

$$
\widehat{\boldsymbol{\Psi}}_i^{(r)}(\boldsymbol{\theta}) = \mathrm{E}(\tau_i\boldsymbol{e}_i\boldsymbol{e}_i^\top|\boldsymbol{y}_i, \widehat{\boldsymbol{\Theta}}^{(r)})
$$
$$
= \widehat{\tau}_i^{(r)}(\boldsymbol{y}_i - \tilde{X}_i\boldsymbol{\theta} - Z_i\widehat{\boldsymbol{b}}_i^{(r)})(\boldsymbol{y}_i - \tilde{X}_i\boldsymbol{\theta} - Z_i\widehat{\boldsymbol{b}}_i^{(r)})^\top + Z_i\widehat{\boldsymbol{V}}_{b_i}^{(r)}Z_i^\top,
$$

with $\widehat{\Delta}_i^{(r)} = (\boldsymbol{y}_i - \tilde{X}_i\widehat{\boldsymbol{\theta}}^{(r)})^\top\widehat{\boldsymbol{\Lambda}}_i^{(r)^{-1}}(\boldsymbol{y}_i - \tilde{X}_i\widehat{\boldsymbol{\theta}}^{(r)})$, $\widehat{\boldsymbol{\Lambda}}_i^{(r)} = Z_i\widehat{\boldsymbol{D}}^{(r)}Z_i^\top + \widehat{\boldsymbol{R}}_i^{(r)}$, $\widehat{\boldsymbol{R}}_i^{(r)} = \widehat{\boldsymbol{\Sigma}}^{(r)} \otimes \boldsymbol{\Omega}_i(\widehat{\boldsymbol{\phi}}^{(r)})$, $\widehat{\boldsymbol{b}}_i^{(r)} = \mathrm{E}(\boldsymbol{b}_i|\boldsymbol{y}_i, \widehat{\boldsymbol{\Theta}}^{(r)}) = \widehat{\boldsymbol{D}}^{(r)}Z_i^\top\widehat{\boldsymbol{\Lambda}}_i^{(r)^{-1}}(\boldsymbol{y}_i - \tilde{X}_i\widehat{\boldsymbol{\theta}}^{(r)})$, $\widehat{\boldsymbol{V}}_{b_i}^{(r)} = \widehat{\tau}_i^{(r)}\mathrm{cov}$ $(\boldsymbol{b}_i|\boldsymbol{y}_i, \widehat{\boldsymbol{\Theta}}^{(r)}) = (\widehat{\boldsymbol{D}}^{(r)} + Z_i^\top\widehat{\boldsymbol{R}}_i^{(r)^{-1}}Z_i)^{-1}$.

M-step: Update $\widehat{\boldsymbol{\Theta}}^{(r)}$ by $\widehat{\boldsymbol{\Theta}}^{(r+1)} = \max_{\boldsymbol{\Theta}} Q(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(r)})$.

However, the M-step of EM algorithm for fitting the MtSMM is computationally intractable and thereby does not yield closed-form solutions for parameter estimators. To

make the estimation procedure more stable and effective, we develop an ECM algorithm for the proposed MtSMM. The ECM algorithm [30] is a variant of EM with its M-steps replaced by several computationally simpler conditional maximization (CM) steps.

In our proposed MtSMM model, the parameter vector $\boldsymbol{\theta}$ is expanded to include the fixed effects $\boldsymbol{\beta}$ and the set of B-spline coefficients $\boldsymbol{\alpha}$. Estimation of MtSMM model is essentially conducted in the same parametric maximum likelihood framework as MtLMM [22] via the ECM algorithm, which is an iterative procedure for ML estimation of models containing missing data or latent variables. As shown by Meng and Rubin [30], the adoption of ECM algorithm can determine a sequence of consistent roots of the likelihood function, which are exactly the maximum likelihood estimates (MLEs).

To obtain analytical expressions for parameter estimators at CM-steps, we let $\widehat{\boldsymbol{e}}_{il}^{(r)} = \boldsymbol{y}_{il} - \boldsymbol{X}_{il}\boldsymbol{\theta}_l - \boldsymbol{Z}_{il}\widehat{\boldsymbol{b}}_{il}^{(r)}$ and $\widehat{\boldsymbol{e}}_{is}^{(r)} = \boldsymbol{y}_{is} - \boldsymbol{X}_{is}\boldsymbol{\theta}_s - \boldsymbol{Z}_{is}\widehat{\boldsymbol{b}}_{is}^{(r)}$, where $\widehat{\boldsymbol{b}}_{il}^{(r)}$ is a $q_l \times 1$ subvector consisting of the $(\sum_{j=1}^{l-1} q_j + 1)$th to $(\sum_{j=1}^{l} q_j)$th entries of $\widehat{\boldsymbol{b}}_i^{(r)}$. It follows that $\widehat{\boldsymbol{\Psi}}^{(r)}(\boldsymbol{\theta}) = [\widehat{\boldsymbol{\psi}}_{ils}^{(r)}(\boldsymbol{\theta})]$, where $\widehat{\boldsymbol{\psi}}_{ils}^{(r)}(\boldsymbol{\theta}) = E(\tau_i \boldsymbol{e}_{il}\boldsymbol{e}_{is}^{\top}|\boldsymbol{y}_i, \widehat{\boldsymbol{\Theta}}^{(r)}) = \widehat{\tau}_i^{(r)}\widehat{\boldsymbol{e}}_{il}^{(r)}\widehat{\boldsymbol{e}}_{is}^{(r)\top} - \boldsymbol{Z}_{il}\widehat{\boldsymbol{V}}_{b_{ils}}^{(r)}\boldsymbol{Z}_{is}^{\top}$ is a square matrix of order s_i with $\widehat{\boldsymbol{V}}_{b_{ils}}^{(r)}$ a $q_l \times q_s$ submatrix consisting of the $(\sum_{j=1}^{l-1} q_j + 1)$th to $(\sum_{j=1}^{l} q_j)$th rows and the $(\sum_{j=1}^{s-1} q_j + 1)$th to $(\sum_{j=1}^{s} q_j)$th columns of $\widehat{\boldsymbol{V}}_{b_i}^{(r)}$, for $l, s = 1, \ldots, r$. In summary, the CM-steps of ECM algorithm proceed as Algorithm 1. In CM-step 2 of Algorithm 1, the first partial derivatives of Q-function with respect to ϕ and ν, respectively, to zero cannot deduce the updated estimators in closed forms. We adopted the nlminb routine in R to perform a numerical search of updated $\widehat{\phi}^{(r)}$ and $\widehat{\nu}^{(r)}$. The nlminb, like other optimization methods, may often suffer from computational difficulties such as slow or nonconvergence. In particular, a poor choice of initial values of ϕ and ν may lead to the convergence in the boundary of the parameter space. To overcome such potential problems, a default recommendation of obtaining reasonable initial values for ϕ and ν is recommended. The global optimal solution is obtained by choosing the one with the largest log-likelihood value. In simulation studies, if the method did not converge for a particular dataset or the method get trapped in one of many local maxima of the log-likelihood function, one can regenerate an additional dataset in the procedure.

To study the asymptotic properties of the proposed estimates $\widehat{\boldsymbol{\Xi}} = (\widehat{\boldsymbol{\beta}}^{\top}, \boldsymbol{\omega}^{\top}, \widehat{\nu})$ and $\widehat{\boldsymbol{g}}(t_i) = \boldsymbol{B}(t_i)\widehat{\boldsymbol{\alpha}}$, we give regular conditions in Appendix. Note that $\boldsymbol{\omega} = (vech(\boldsymbol{D})^{\top}, vech(\boldsymbol{\Sigma})^{\top}, \boldsymbol{\phi}^{\top})^{\top}$; here, $vech()$ is the half-vectorization operator. If $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{p} 0$, and $\sup_t|\boldsymbol{B}(t_i)\widehat{\boldsymbol{\alpha}} - \boldsymbol{g}(t_i)| \xrightarrow{p} 0$ as $n \to \infty$, $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{g}}(t_i))$ are said to be consistent estimators. Under the regularity conditions sketched in Appendix, we state the following theorem, which shows the asymptotic properties of the estimates $\widehat{\boldsymbol{\Xi}}$ and $\widehat{\boldsymbol{g}}(t_i)$.

Theorem 3.1: *Under the conditions* (A.1)–(A.5) *in Appendix, if the number of knots L_j approaches to $n^{1/(2m_j+1)}$, then*

(i) $1/n \sum_{i=1}^{n} \sum_{k=1}^{s_i} (\widehat{g}_j(t_{i,k}) - g_j(t_{i,k}))^2 = O_p(n^{2m_j/(2m_j+1)}), \quad j = 1, \ldots, r.$

(ii) $\widehat{\boldsymbol{\Xi}} \xrightarrow{p} \boldsymbol{\Xi}$, $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{I}_{\boldsymbol{\beta\beta}}^{-1})$, *and* $\sqrt{n}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{I}_{\boldsymbol{\eta\eta}}^{-1})$, *where* $N = \sum_{i=1}^{n} s_i$ *is the total number of observations,* $\boldsymbol{\eta} = (\boldsymbol{\omega}^{\top}, \nu)^{\top}$, *the matrices $\boldsymbol{I}_{\boldsymbol{\beta\beta}}$ and $\boldsymbol{I}_{\boldsymbol{\eta\eta}}$ are*

Algorithm 1 ECM algorithm for numerical estimation.

CM-step 1. Fix $\phi = \widehat{\phi}^{(r)}$ and $\nu = \widehat{\nu}^{(r)}$, and update $\widehat{\boldsymbol{\theta}}^{(r)}$, $\widehat{\boldsymbol{D}}^{(r)}$ and $\widehat{\boldsymbol{\Sigma}}^{(r)}$ by maximizing $Q(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(r)})$ to get

$$\widehat{\boldsymbol{\theta}}^{(r+1)} = \left(\sum_{i=1}^n \widehat{\tau}_i^{(r)} \tilde{\boldsymbol{X}}_i^\top \widehat{\boldsymbol{R}}_i^{(r)^{-1}} \tilde{\boldsymbol{X}}_i \right)^{-1} \sum_{i=1}^n \widehat{\tau}_i^{(r)} \tilde{\boldsymbol{X}}_i^\top \widehat{\boldsymbol{R}}_i^{(r)^{-1}} (\boldsymbol{y}_i - \boldsymbol{Z}_i \widehat{\boldsymbol{b}}_i^{(r)}),$$

$$\widehat{\boldsymbol{D}}^{(r+1)} = n^{-1} \sum_{i=1}^n \widehat{\boldsymbol{B}}_i^{(r)},$$

$$\widehat{\sigma}_{lm}^{(r+1)} = \begin{cases} \left(\sum_{i=1}^n s_i \right)^{-1} \sum_{i=1}^n \mathrm{tr} \left(\boldsymbol{\Omega}_i^{-1}(\widehat{\phi}^{(r)}) \widehat{\boldsymbol{\psi}}_{ils}^{(r)}(\widehat{\boldsymbol{\theta}}^{(r+1)}) \right) & \text{for } l = s, \\ \left(2 \sum_{i=1}^n s_i \right)^{-1} \sum_{i=1}^n \mathrm{tr} \left(\boldsymbol{\Omega}_i^{-1}(\widehat{\phi}^{(r)}) \left(\widehat{\boldsymbol{\psi}}_{ils}^{(r)}(\widehat{\boldsymbol{\theta}}^{(r+1)}) + \widehat{\boldsymbol{\psi}}_{isl}^{(r)}(\widehat{\boldsymbol{\theta}}^{(r+1)}) \right) \right) & \text{for } l \neq s, \end{cases}$$

which are the updated estimates of the distinct elements in $\boldsymbol{\Sigma}$.

CM-step 2. Given $\widehat{\boldsymbol{\theta}}^{(r+1)}$, $\widehat{\boldsymbol{D}}^{(r+1)}$, and $\widehat{\boldsymbol{\Sigma}}^{(r+1)}$, we calculate $(\widehat{\phi}^{(r+1)}, \widehat{\nu}^{(r+1)})$ by maximizing the constrained Q-function:

$$(\widehat{\phi}^{(r+1)}, \widehat{\nu}^{(r+1)}) = \underset{(\phi, \nu)}{\arg\max} \left\{ \sum_{i=1}^n \left(r \log |\boldsymbol{\Omega}_i^{-1}(\phi)| - \mathrm{tr} \left(\left(\widehat{\boldsymbol{\Sigma}}^{(r+1)^{-1}} \otimes \boldsymbol{\Omega}_i^{-1}(\phi) \right) \times \widehat{\boldsymbol{\Psi}}_i^{(r+1/2)} \right) \right) \right.$$
$$\left. + \nu \log \left(\frac{\nu}{2} \right) - 2 \log \Gamma \left(\frac{\nu}{2} \right) + \nu (\widehat{\kappa}_i^{(r)} - \widehat{\tau}_i^{(r)}) \right\},$$

where $\widehat{\boldsymbol{\Psi}}_i^{(r+1/2)} = \widehat{\boldsymbol{\Psi}}_i^{(r)}(\widehat{\boldsymbol{\theta}}^{(r+1)})$.

Convergence step. Iterating the E-step and CM-steps until the user's specified tolerance or the default maximum number of iterations is met, we obtain the ML estimates, denoted by $\widehat{\boldsymbol{\Theta}} = (\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{D}}, \widehat{\boldsymbol{\Sigma}}, \widehat{\phi}, \widehat{\nu})$.

---

defined in condition (A.5), $\overset{p}{\to}$ *and* $\overset{d}{\to}$ *denote convergence in probability and convergence in distribution, respectively.*

Under general conditions (Stone [33][Lemmas 8 and 9]), Theorem 3.1(i) implies that $\int (\widehat{\boldsymbol{g}}_j(t) - \boldsymbol{g}_j(t))^2 dt = O_p(n^{2m_j/(2m_j+1)})$, which is the optimal rate of convergence for estimating $\boldsymbol{g}_j(t_i)$ under the smoothness condition in (A.2). Theorem 3.1(ii) is useful for making large-sample inference on $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ such as construction of the confidence band and hypothesis testing. To this end, we need to obtain the estimate of the variance-covariance matrix for $\widehat{\boldsymbol{\Xi}}$. Since $\widehat{\boldsymbol{\Xi}}$ is consistent and asymptotically normal, the inverse of the Fisher information matrix evaluated at $\boldsymbol{\Xi} = \widehat{\boldsymbol{\Xi}}$ can be used to approximate the asymptotic variance–covariance matrix of $\widehat{\boldsymbol{\Xi}}$.

Selection of knots is generally an important aspect of spline smoothing. In this paper, our main focus is inference on the parameter $\boldsymbol{\beta}$. He et al. [34] found that knot selection is less critical for the estimate of $\boldsymbol{\beta}$ than for the estimate of $\boldsymbol{g}(\cdot)$. They also pointed out that in most applications, the primary focus is inference on the parameter $\boldsymbol{\beta}$, along with the understanding of some basic features of $\boldsymbol{g}(\cdot)$. Therefore, they are more concerned with the efficiency of the parameter estimate. For those reasons, and for the sake of simplicity, He et al. [34] opt for convenient choices of knot placement. More specifically, they

use the sample quantiles of $\{t_{i,k}, i = 1, \ldots, n; k = 1, \ldots, s_i\}$ as knots. For example, in the case of three internal knots, these are taken to be the three quartiles of the observed $t_{i,k}$. They use cubic splines (i.e. splines of order 4) and take the number of internal knots to be the integer part of $M^{1/5}$, where M is the number of distinct values in $t_{i,k}$. Also, Qin and Zhu [5] noted that the number of distinct knots has to increase with sample size for asymptotic consistency. On the other hand, too many knots would increase the variance of estimators. Therefore, the number of knots must be properly chosen to balance between the bias and variance. When n goes to ∞, the number of knots should be increasing at the rate of $n^{1/(2m+1)}$. Consequently, in this paper, the number of internal knots is taken to be $L_j \approx n^{1/(2m_j+1)}$ for $j = 1, \ldots, r$ where m_j is an integer, and select the internal knots equally spaced in the percentile ranks of $t_{i,k}$. For $j = 1, \ldots, r$, the m_j as assumed to be fixed at $m_j = 2$, therefore we choose $L_j \approx n^{1/5}$. This particular choice is consistent with the asymptotic theory, but it is mainly based on the empirical experience and desire for simplicity, and by no means it is an optimal choice. Data-adaptive choices for the number or the placement of knots can be made using leave-one-cluster-out, but we do not pursue this direction here.

4. Illustration

4.1. PBC sequential data analysis

We consider a subset of longitudinal data on the Primary Biliary Cirrhosis sequential (PBCseq) cohort study in which 312 patients were recruited from the Mayo Clinic between January 1974 and May 1984, and participated in either of two double-blind, placebo-controlled, randomized trials with D-penicillamine for treating primary biliary cirrhosis until April 1988. A clinical laboratory database which comprised ID number, five time-dependent variables (age and total number of follow-up days), eight categorical variables (sex, drug and status), two censoring indicators for events, and seven continuous measurement variables (natural logarithm scale of bili and albumin), was established on each patient who was collected repeatedly and prospectively at yearly intervals under standardized forms, definitions, and study protocols. The data set is composed entirely of 1945 visit rows and 38 columns (variables) on the 312 randomized patients. It can be freely available from the R mixAK package and electronically at http://lib.stat.cmu.edu/datasets/pbcseq. For a more detailed description of the data, one can refer to [35–39].

Orthotopic liver transplantation can be treated as a potentially life-saving alternative for patients with advanced or end-stage primary biliary cirrhosis. Serum bilirubin and serum albumin are two of the primary indicators to help evaluate and track the absence of liver diseases. An extremely higher level than the standards that bilirubin is excreted in bile and urine can indicate certain diseases. Serum albumin may be harmful to humans having too high or too low circulating serum albumin levels. Typically, it is believed that there exist some relationships between serum bilirubin and serum albumin levels, and thus a joint analysis of the longitudinally collected bilirubin and albumin has received increasing emphasis in diagnosing liver diseases. As pointed out by Wang [35], there exist some atypical observations or outliers in the data, and a separate analysis of the two markers can lose important information about evolutional relationships among multiple responses
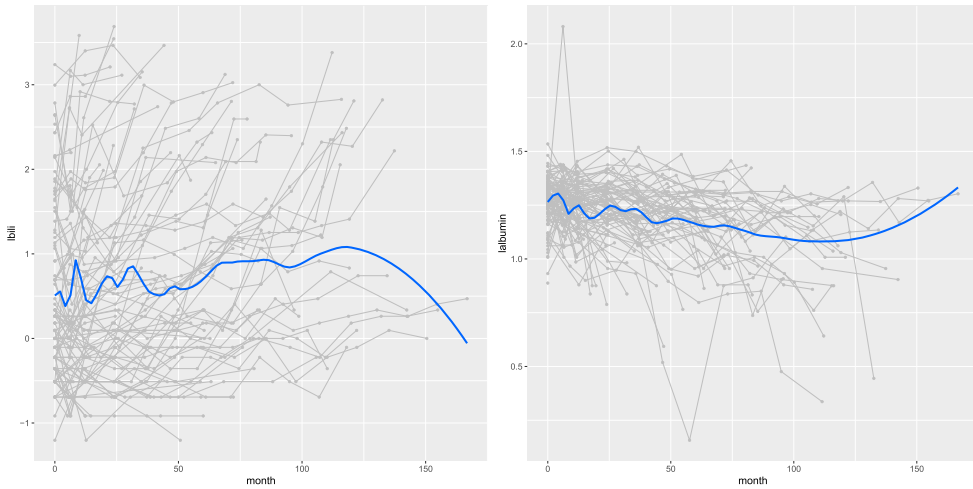
Figure 1. Trajectories plots for the PBCseq data. Observed evolution (in grey) of lbili and lalbumin markers for 100 randomly selected patients against time (in month). Solid lines show the smoothed mean profiles of responses for all patients.

across time. As a consequence, we concentrate on modelling the dependence of the longitudinal profiles of two markers say natural logarithm of serum bilirubin (lbili) and the natural logarithm of serum albumin (lalbumin), on time (visited years) and other covariates of interest (e.g. sex, drug, age). The issues that we have to investigate include (i) how the bilirubin/albumin levels evolve over time; (ii) how the evolution of bilirubin is related to the evolution of albumin; and (iii) how the association between bilirubin and albumin evolves over time. Figure 1 displays the trajectories of 100 randomly selected patients for exploring the evolution of lbili and lalbumin markers. It can be observed that the level of two markers varies over time in a complicated manner and it is difficult to model the time trend using a single parametric function. Hence, it is of potential interest in the study to model the time effect through a nonparametric scheme and to account for the outliers simultaneously.

We now apply the proposed MtSMM approach to analysing the PBCseq data and comparing with the multivariate normal semiparametric mixed-effects model (MNSMM). Let $\boldsymbol{y}_i = (\boldsymbol{y}_{i1}^\top, \boldsymbol{y}_{i2}^\top)^\top$ be the response vector for the ith patient, where $\boldsymbol{y}_{i1}$ and $\boldsymbol{y}_{i2}$ represent lbili and lalbumin levels, respectively. Apart from the time effect, it is particular of interest to take into account the relationship between the longitudinal evolutions of the two markers and the covariates of interest, including gender, drug treatment, and age. Thus, the design matrix for fixed effects $\boldsymbol{\beta} = (\beta_{10}, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}, \beta_{20}, \beta_{21}, \beta_{22}, \beta_{23}, \beta_{24})$ can be specified as

$$\boldsymbol{X}_i = \boldsymbol{I}_2 \otimes [\boldsymbol{1}_{s_i} : \boldsymbol{t}_i : \text{sex}_i \boldsymbol{1}_{s_i} : \text{drug}_i \boldsymbol{I}_{s_i} : \text{age}_i \boldsymbol{1}_{s_i}],$$

where $\boldsymbol{1}_{s_i}$ is a $s_i \times 1$ vector of ones, $\boldsymbol{t}_i = (t_{i1}, \ldots, t_{is_i})^\top$ with $t_{i,k} = \text{month}_{i,k}/12$ (years), sex_i is a gender indicator ($0 = \text{male}$ and $1 = \text{female}$), drug_i is a drug treatment indicator ($0 = \text{patient treated with placebo, and } 1 = \text{patient treated with D-penicillamine}$); and age_i is the age of patient i at entry in years. The design matrix for random intercept plus slope is

$Z_i = I_2 \otimes [1_{s_i} : t_i]$. We consider three dependence structures for Ω_i, namely uncorrelated (UNC), exchangeable (EX) and first-order autoregressive (AR(1)) processes.

To best identify a model supported by the data, we adopt the Akaike information criterion (AIC; [40]), the Bayesian information criterion (BIC; [41]) and mean squared prediction errors (MSPE) for the fitted responses. They are defined as

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

$$\text{MSPE} = \frac{1}{n} \sum_i (Y_{test}^{(-i)} - \widehat{Y}_{test}^{(-i)})^\top (Y_{test}^{(-i)} - \widehat{Y}_{test}^{(-i)}),$$

where $\ell_{\max}$ is the maximized log-likelihood value, m is the number of free parameters in the model, and $Y_{test}^{(-i)}$ and $\widehat{Y}_{test}^{(-i)}$, respectively, are test sets and their predicted values after leaving out the ith subject.

Table 1 summarizes the values of $\ell_{\max}$, AIC, and BIC together with MSPE scores under the fitted MtSMM and MNSMM models. In light of all criteria, the MtSMM with AR(1) errors is preferred among all candidates. Table 2 presents the summary of the ML estimates of parameters along with the standard errors of fixed effects for the MtSMM and MNSMM with AR(1) errors. We utilize the nonparametric bootstrapping technique [42,43] to evaluate the standard errors of estimators. The idea of the bootstrap is to mimic the process of randomly sampling from an assumed infinite population. It takes C samples of the same size drawing with replacement from the original data. For each of these C samples, we calculate the sample standard deviation of these estimates as an estimate of standard error. Denote by $\widehat{\theta}_c$, for $c = 1, \ldots, C$, the C estimates of θ obtained from the bootstrap samples. The bootstrap standard error for $\widehat{\theta}$ is then given by

$$SE(\widehat{\theta}) = \sqrt{\frac{1}{C-1} \sum_{c=1}^{C} (\widehat{\theta}_c - \bar{\theta})^2},$$

where $\bar{\theta} = (1/C) \sum_{c=1}^{C} \widehat{\theta}_c$ is the mean of the estimates across the C bootstrap samples.

Judging from Table 2, the ML estimates of fixed effects under the MtSMM and MNSMM are somewhat similar, but the standard errors under the MtSMM are moderately smaller than those under the MNSMM, signifying that the proposed approach enables the practitioners to provide more precise estimates.

Table 1. Summary of model selection criteria for the PBCseq data.

Model	Ω_i	Criteria			
		$\ell_{\max}$	AIC	BIC	MSPE
MtSMM	UNC	5090.49	−10142.98	−10054.63	6.40
	EX	5251.24	−10462.48	−10376.13	6.38
	AR(1)	5358.41	−10672.82	−10590.47	6.13
MNSMM	UNC	3817.94	−7593.88	−7515.28	7.24
	EX	3938.50	−7835.00	−7756.40	7.18
	AR(1)	4018.88	−7995.77	−7917.16	7.10

Note: MtSMM: multivariate t semiparametric mixed-effects model; MNSMM: multivariate normal semiparametric mixed-effects model; AIC: Akaike information criterion; BIC: Bayesian information criterion; MSPE: mean squared prediction errors for the fitted responses.

Table 2. Summary of parameter estimates along with standard errors of fixed effects (in parentheses) for the PBCseq data.

Parameters		MtSMM	MNSMM
Fixed effects	β_{10} (Intercept)	1.432 (0.018)	1.428 (0.112)
	β_{11} (Sex)	−0.542 (0.010)	−0.538 (0.042)
	β_{12} (Drug)	−0.015 (0.007)	−0.019 (0.013)
	β_{13} (Age)	−0.010 (3.3×10^{-4})	−0.0102 (0.002)
	β_{20} (Intercept)	1.171 (0.002)	1.172 (0.007)
	β_{21} (Sex)	−0.004 (0.001)	-0.005 (0.003)
	β_{22} (Drug)	0.010 (8.9×10^{-4})	0.009 (0.001)
	β_{23} (Age)	−0.002 (4.1×10^{-5})	−0.001 (1.1×10^{-4})
Random effects	d_{11}	0.807	0.738
	d_{21}	−0.012	−0.021
	d_{22}	0.317	0.309
	d_{31}	−0.022	−0.019
	d_{32}	−0.003	−0.002
	d_{33}	0.310	0.309
	d_{41}	−0.007	−0.006
	d_{42}	−0.005	−0.004
	d_{43}	−0.095	−0.095
	d_{44}	0.256	0.256
Within-subject errors	σ_{11}	−0.085	−0.114
	σ_{21}	−0.003	−0.002
	σ_{22}	−0.222	−0.224
	ν	3.120	–
	ϕ	0.407 (0.002)	0.220 (0.287)

Note: MtSMM: multivariate t semiparametric mixed-effects model; MNSMM: multivariate normal semiparametric mixed-effects model.
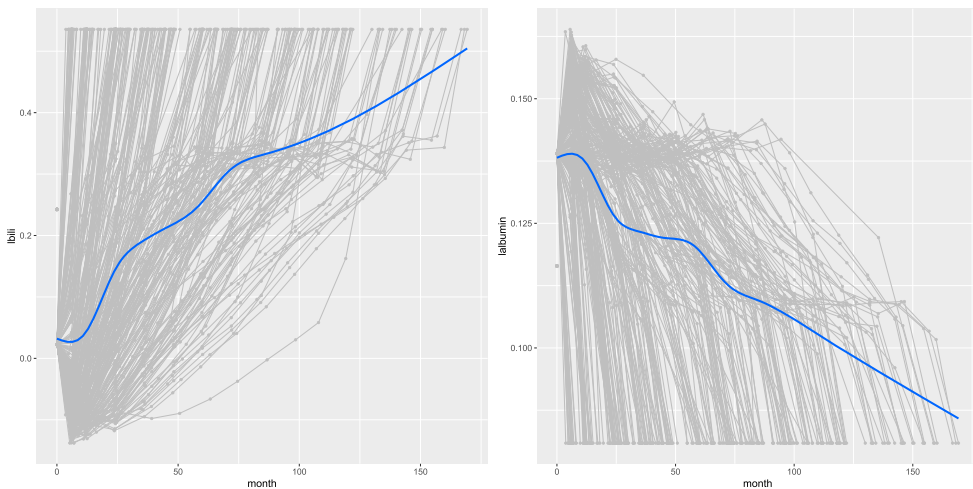
Figure 2. The estimated nonparametric function over time (in grey) of lbili and lalbumin markers for 312 patients against time (in month). Solid lines show the smoothed mean profiles for all patients.

The estimated nonparametric functions over time depicted in Figure 2 indicate an apparent nonlinearity, and it can be responded to basic issues concerning how the bilirubin/albumin levels evolve over time, how the evolution of bilirubin is related to the evolution of albumin, and how the association between bilirubin and albumin evolves over time. In the first year (first 12 months), the patients have decreasing and increasing trends

for lbili and lalbumin, respectively, and then followed by the reversed growth patterns for two markers. Among the first to sixth years (around the 75th month), an irregular increasing of lbili is associated with an irregular decreasing of lalbumin, and after the sixth year the growth of lbili and the decline of lalbumin are strict.

4.2. Simulation studies

In this section, we conduct two simulation studies to examine the performance of the proposed MtSMM in comparison with other existing models under different degrees of fat-tailed behaviour and nonlinear patterns of responses across time.

The first simulation is conducted to explore the robustness against outliers and the effect of ignoring nonlinear trend of responses over time, and thus we compare the MtSMM, MNSMM and MtLMM approaches. The simulated data with $r = 2$ outcome variables are generated from the MtSMM (1) along with assumption (2) and the nonlinear functions, given by

$$g_1(t_{i,k}) = 2.5 - 1.5t_{i,k} + 0.4t_{i,k}^2,$$

$$g_2(t_{i,k}) = 2\sin(2\pi t_{i,k}) + 2\cos(2\pi t_{i,k}),$$

where $k = 1, \ldots, s_i$, the number of observations per subject assumed to be fixed at $s_i = 5$ and the measurement time points $t_{i,k}$s are drawn from uniform distribution on $(1, 10)$ interval and sorted in ascending order. The design matrix for fixed effects of the first outcome X_{i1} is a $s_i \times 2$ matrix including an intercept and scheduled visits of time (1 to s_i), and that of the second outcome X_{i2} is a $s_i \times 2$ matrix whose first column including an intercept and second column represents the values of a continuous covariate drawn from a $N(0, 1)$ distribution. The design matrix for random effects Z_i contains an intercept only for each outcome. The presumed values of model parameters are $\boldsymbol{\beta} = (\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22})^\top = (1, 2, -2, 4,)^\top$,

$$D = \begin{bmatrix} 1 & 0.25 \\ 0.25 & 1 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 2 & 0.75 \\ 0.75 & 1 \end{bmatrix}, \quad \text{and} \quad \Omega = I_{s_i}.$$

Besides, two values of DOF, say $\nu = 5$ and 50, are considered to capture a fat-tailed behaviour and represent near normality, respectively. A quadratic spline with $d_1 = 2$ and a cubic spline with $d_2 = 3$ are used for the first and second outcomes, respectively. To examine the finite-sample property of ML estimation, the sample sizes are set to be a small value, $n = 25$, and a relatively large value $n = 100$. Each simulated data set is fitted by the MtSMM, MNSMM and MtLMM approach, separately, and a total of 100 replications are run for each combination of ν and n.

Estimation accuracy is measured by the empirical biases (BIAS), the standard deviations (SD), and the empirical mean squared errors (MSE) of the estimated parameters under the three fitted models. In addition, model selection is made by the computing average values of AIC and BIC. The results for the cases of $n = 25$ and $n = 100$ are given in Tables 3 and 4, respectively. To present a more comprehensive assessment, we also use other criteria for performance evaluation of the estimated nonparametric functions $g_1(\cdot)$ and $g_2(\cdot)$. We use the mean integrated square error (MRME) in Tables 3 and 4 where the rows are labelled

Table 3. Simulation results for the cases of sample size $n = 25$ and DOF $v = 5$ and 50.

			$n = 25$				
		$v = 5$			$v = 50$		
Parameters		MtSMM	MtLMM	MNSMM	MtSMM	MtLMM	MNSMM
β_{11}	BIAS	0.641	7.966	0.645	0.746	7.906	0.746
	SD	0.366	0.489	0.418	0.368	0.393	0.368
	MSE	0.543	63.701	0.589	0.690	62.658	0.690
β_{12}	BIAS	0.030	0.555	0.064	0.058	0.648	0.064
	SD	0.224	0.778	0.265	0.271	0.730	0.270
	MSE	0.050	0.908	0.074	0.076	0.947	0.076
β_{21}	BIAS	0.987	1.074	0.955	0.892	1.067	0.879
	SD	0.377	0.417	0.377	0.375	0.391	0.373
	MSE	1.116	1.324	1.053	0.935	1.289	0.910
β_{22}	BIAS	0.312	1.292	0.215	0.173	1.287	0.151
	SD	0.484	0.512	0.414	0.314	0.435	0.300
	MSE	0.329	1.930	0.216	0.128	1.844	0.112
d_{11}	BIAS	0.849	0.954	0.965	0.131	0.957	0.110
	SD	0.379	0.261	1.325	0.644	0.259	0.737
	MSE	0.862	0.977	2.670	0.427	0.983	0.550
d_{12}	BIAS	0.036	0.160	0.303	0.050	0.194	0.086
	SD	0.338	0.165	0.990	0.502	0.166	0.562
	MSE	0.114	0.053	1.062	0.251	0.065	0.320
d_{22}	BIAS	0.553	1.650	1.703	0.043	1.734	0.302
	SD	0.546	0.152	2.070	0.828	0.100	0.939
	MSE	0.601	2.744	7.143	0.680	3.017	0.964
σ_{11}	BIAS	1.226	3.135	0.188	0.770	3.247	0.595
	SD	0.177	0.511	0.734	0.190	0.614	0.223
	MSE	1.533	10.085	0.569	0.627	10.913	0.403
σ_{12}	BIAS	0.453	0.768	0.083	0.227	0.770	0.156
	SD	0.157	0.127	0.432	0.240	0.096	0.270
	MSE	0.230	0.606	0.192	0.108	0.602	0.100
σ_{22}	BIAS	0.398	0.639	2.168	1.556	0.646	1.883
	SD	0.222	0.057	0.507	0.310	0.060	0.344
	MSE	0.207	0.411	4.953	2.515	0.420	3.661
v	BIAS	0.023	3.262	–	1.157	48.999	–
	SD	0.132	0.439	–	0.979	0.008	–
	MSE	0.018	108	–	2.288	2400.884	–
	ℓ_{max}	−215.329	−232.374	−327.253	−265.923	−228.757	−287.089
	AIC	452.658	486.748	674.506	553.846	479.515	594.177
	BIC	466.065	500.156	686.695	567.254	492.922	606.366
	MISE1	0.710	–	0.789	0.661	–	0.714
	MISE2	4.057	–	4.229	3.993	–	4.004

Note: MtSMM: multivariate t semiparametric mixed-effects model; MtLMM: multivariate t linear mixed-effects model; MNSMM: multivariate normal semiparametric mixed-effects model; BIAS: difference between the mean of the estimates and the true value of the parameter; SD: standard deviation of the estimates of parameter; MSE: mean squared errors for the parameters estimates; AIC: Akaike information criterion; BIC: Bayesian information criterion; MISE1: mean integrated square error for $g_1(\cdot)$; MISE2: mean integrated square error for $g_2(\cdot)$.

by MISE1 and MISE2 for $g_1(\cdot)$ and $g_2(\cdot)$, respectively. By the simulation design described above, we define the MISE1 of a procedure as

$$\mathrm{E}\left[\int_T \left(\widehat{g}_1(t) - g_1(t)\right)^2 \, \mathrm{d}t\right],$$

where T specifies the range of t, and MISE1 is approximated by averaging the calculated integral over the simulated samples. The MISE2 is defined similarly for $g_2(\cdot)$. In terms of the MISE1 and MISE2 criteria, the MtSMM outperforms MNSMM. Although, in the case of data with larger sample size, the performance of both MtSMM and MNSMM improved

Table 4. Simulation results for the cases of sample size $n = 100$ and DOF $\nu = 5$ and 50.

| Parameters | | $n = 100$ | | | | | |
| | | $\nu = 5$ | | | $\nu = 50$ | | |
		MtSMM	MtLMM	MNSMM	MtSMM	MtLMM	MNSMM
β_{11}	BIAS	0.724	8.187	0.727	0.740	8.194	0.740
	SD	0.190	0.255	0.219	0.162	0.1766	0.163
	MSE	0.560	67.085	0.576	0.573	67.173	0.573
β_{12}	BIAS	0.033	0.869	0.023	0.006	0.8257	0.006
	SD	0.120	0.272	0.138	0.100	0.247	0.103
	MSE	0.015	0.828	0.019	0.010	0.742	0.011
β_{21}	BIAS	0.018	0.162	0.014	0.061	0.179	0.056
	SD	0.223	0.239	0.239	0.148	0.214	0.147
	MSE	0.050	0.083	0.057	0.025	0.077	0.025
β_{22}	BIAS	0.708	0.568	0.622	0.643	0.509	0.632
	SD	0.230	0.286	0.218	0.155	0.211	0.157
	MSE	0.552	0.404	0.434	0.437	0.303	0.424
d_{11}	BIAS	0.562	1.028	1.570	0.114	1.044	0.212
	SD	0.277	0.167	0.975	0.339	0.158	0.394
	MSE	0.392	1.084	3.404	0.127	1.114	0.199
d_{12}	BIAS	0.044	0.244	0.253	0.021	0.237	0.021
	SD	0.208	0.087	0.747	0.2708	0.079	0.311
	MSE	0.045	0.067	0.616	0.073	0.062	0.096
d_{22}	BIAS	0.332	1.631	1.697	0.102	1.683	0.458
	SD	0.293	0.080	0.861	0.351	0.058	0.403
	MSE	0.195	2.668	3.612	0.133	2.836	0.370
σ_{11}	BIAS	1.137	4.009	0.298	0.768	4.320	0.542
	SD	0.123	0.623	0.403	0.106	0.580	0.123
	MSE	1.308	16.452	0.250	0.601	18.998	0.309
σ_{12}	BIAS	0.423	0.741	0.121	0.247	0.780	0.156
	SD	0.079	0.058	0.223	0.114	0.056	0.130
	MSE	0.185	0.552	0.064	0.074	0.611	0.041
σ_{22}	BIAS	0.8078	0.569	2.695	1.921	0.567	2.391
	SD	0.207	0.054	0.260	0.183	0.052	0.175
	MSE	0.695	0.326	7.327	3.723	0.325	5.745
ν	BIAS	0.014	3.295	–	0.760	48.998	–
	SD	0.143	0.080	–	0.976	0.006	–
	MSE	0.020	10.864	–	1.520	2400.846	–
	ℓ_{max}	−994.808	−1034.491	−1400.377	−1135.948	−1038.500	−1226.631
	AIC	2011.617	2090.982	2820.753	2293.896	2098.999	2473.262
	BIC	2040.274	2119.639	2846.805	2322.552	2127.656	2499.314
	MISE1	0.639	–	0.566	0.534	–	0.574
	MISE2	3.911	–	3.941	3.887	–	3.925

MtSMM: multivariate t semiparametric mixed-effects model; MtLMM: multivariate t linear mixed-effects model; MNSMM: multivariate normal semiparametric mixed-effects model; BIAS: difference between the mean of the estimates and the true value of the parameter; SD: standard deviation of the estimates of parameter; MSE: mean squared errors for the parameter estimates; AIC: Akaike information criterion; BIC: Bayesian information criterion; MISE1: mean integrated square error for $g_1(\cdot)$; MISE2: mean integrated square error for $g_2(\cdot)$.

as expected. Nevertheless, the proposed method has smaller MISE1 and MISE2, which guarantees the quality of the estimation procedure for the nonparametric function in the model. When comparing the MtSMM with the MtLMM, it is also worth mentioning that ignoring the nonlinerity time trend may cause large biases in estimation and produce misleading results in subsequent analyses. Consequently, the proposed MtSMM can provide greater flexibility in model fitting and superiority for outlier resistance, especially when the variety of responses over time is in a complicated manner.

The second simulation study is conducted to compare the performance of the proposed MtSMM with the MNSMM in the attendance of other heavy tailed distributions

for random effects and errors. We consider the following four types of distributions for random effects and errors.

(i) Multivariate t (MVT) distribution: $\boldsymbol{\varepsilon}_i \sim t_{n_i}(\mathbf{0}, \boldsymbol{R}_i, \nu)$, $\boldsymbol{b}_i \sim t_q(\mathbf{0}, \boldsymbol{D}, \nu)$ with $\nu = 5$.

(ii) Multivariate slash (MSL) distribution: $\boldsymbol{\varepsilon}_i \sim SL_{n_i}(\mathbf{0}, \boldsymbol{R}_i, \nu)$, $\boldsymbol{b}_i \sim SL_q(\mathbf{0}, \boldsymbol{D}, \nu)$ with $\nu = 2$.

(iii) Multivariate contaminated normal (MCN) distribution: $\boldsymbol{\varepsilon}_i \sim 0.5N_{n_i}(\mathbf{0}, \boldsymbol{R}_i) + 0.5N_{n_i}(\mathbf{0}, 0.3\boldsymbol{R}_i)$, $\boldsymbol{b}_i \sim 0.5N_q(\mathbf{0}, \boldsymbol{D}) + 0.5N_q(\mathbf{0}, 0.3\boldsymbol{D})$.

(vi) Multivariate normal (MVN) distribution: $\boldsymbol{\varepsilon}_i \sim N_{n_i}(\mathbf{0}, \boldsymbol{R}_i)$, $\boldsymbol{b}_i \sim N_q(\mathbf{0}, \boldsymbol{D})$

Table 5 reports the absolute bias of parameter estimates and the relative efficiencies of our proposed method as compared to its normal counterpart. Obviously, the biases of the parameter estimates of fixed effects for both models are similar, whereas our proposed approach works very well and consistently outperforms its normal counterpart when the random effects and errors exhibit heavy tails or are contaminated by outliers. Even when the random effects and errors follow the MVN distribution, the performance of the

Table 5. Estimation accuracy in terms of bias (BIAS) and efficiency for the MtSMM (T) and MNSMM (N) approaches when the random effects and errors are generated from the multivariate t with DOF $\nu = 5$ (t_5), multivariate slash (MSL), multivariate contaminated normal (MCN) and multivariate normal (MVN) distributions.

Parameters		MVT	MSL	MCN	MVN
		Distributions of random effects and errors			
β_{11}	BIAS(T)	0.770	0.684	0.644	0.687
	BIAS(N)	0.703	0.643	0.615	0.695
	Efficiency	1.293	1.020	1.103	1.020
β_{12}	BIAS(T)	0.194	0.192	0.151	0.109
	BIAS(N)	0.214	0.176	0.172	0.105
	Efficiency	1.201	1.159	1.015	0.946
β_{21}	BIAS(T)	0.302	0.210	0.232	0.177
	BIAS(N)	0.178	0.139	0.138	0.137
	Efficiency	1.151	0.801	0.905	0.926
β_{22}	BIAS(T)	0.049	0.158	0.153	0.242
	BIAS(N)	0.223	0.264	0.289	0.299
	Efficiency	0.695	0.509	0.513	0.808
d_{11}	BIAS(T)	0.789	0.622	0.577	0.323
	BIAS(N)	1.348	2.199	2.553	0.027
	Efficiency	21.566	22.682	13.391	0.982
d_{12}	BIAS(T)	0.062	0.039	0.020	0.095
	BIAS(N)	0.348	0.357	0.555	0.035
	Efficiency	13.586	11.576	10.687	1.414
d_{22}	BIAS(T)	0.226	0.264	0.230	0.322
	BIAS(N)	2.267	2.724	3.169	0.791
	Efficiency	15.317	16.466	10.238	1.407
σ_{11}	BIAS(T)	1.247	1.185	1.156	0.887
	BIAS(N)	0.322	0.655	0.991	0.626
	Efficiency	29.242	14.794	11.639	0.470
σ_{12}	BIAS(T)	0.438	0.403	0.426	0.312
	BIAS(N)	0.244	0.331	0.411	0.189
	Efficiency	20.845	8.842	10.025	1.344
σ_{22}	BIAS(T)	0.413	0.258	0.166	1.327
	BIAS(N)	2.357	2.451	2.559	1.847
	Efficiency	8.942	2.559	5.063	0.334

MtSMM model is similar with that of the MNSMM. This is essentially because the latter method can be treated as a special case of our proposed model, and thus the efficiency loss is minimal when no outlier is present in the data. When the data are generated from the MVT, MSL and MCN scenarios, the relative efficiency of estimates obtained from the MtSMM may be very high. This is because the MNSMM may fail miserably when applied to heavy-tailed distributions.

5. Conclusion

This paper is devoted to proposing a novel robust generalization of the SMM, called as the MtSMM, in which the Gaussian distributions for the random effects and the within-subject errors are replaced by the MVT distribution. For analysing multivariate longitudinal data in the presence of atypical observations or outliers. The formulation of MtSMM assumes that a single weight is associated with random effects and within-subject errors for each subject. We have utilized the regression spline method to estimate the nonparametric function of time and provided a feasible ECM algorithm for ML estimation of fixed effects, random effects, smoothing effects and variance components simultaneously. A key feature of MtSMM is that it allows us to make systematic inference on all model parameters by representing a semiparametric model as a modified parametric linear mixed model. Furthermore, the asymptotic properties of the parameter estimators of the MtSMM, denoted by $(\widehat{\boldsymbol{g}}(\boldsymbol{t}_i), \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{D}}, \widehat{\boldsymbol{\Sigma}}, \widehat{\phi}, \widehat{\nu})$, have been described in a theorem.

Numerical results in simulations and the PBCseq data example have demonstrated that the proposed method outperforms the MNSMM counterparts on the provision of likelihood-based model selection and estimating the regression coefficients, nonparametric functions, and the variance components. In addition, the MtSMM can provide substantially better fitting performance than the linear or parametric models when the growth curves of the data exhibit an nonlinear pattern over time, especially in a complicated manner.

A large number of works have been proposed in the literature by adopting a more flexible class of distributions when the data are contaminated with outliers. The class of elliptically contoured distributions [45–47] is a multivariate unified framework including many of the most common multivariate distributions such as normal, *t*, exponential power, Kotz-type, Pearson-type VII, slash, contaminated normal, logistic, and so on. For future work, one can consider a joint elliptical distribution for the random effects and within-subject errors. Another possible future direction is to consider multivariate skew distributions [48,49] rather than symmetric ones. In medical experiments, there commonly exist missing observations in the responses or covariates and censored responses due to a quantification limit of assay which should be taken account [50–53]. Therefore, it is of interest to extend the current approach to accommodate possibly missing values.

As stressed by Harville [54], one potential disadvantage of the ML estimates of variance components is that they are biased downwards in finite samples. Instead, the REML estimates produce unbiased estimating equations for the variance components by adjusting the loss of degrees of freedom incurred in estimating the fixed effects. It could also be interesting to develop a feasible procedure to compute reliable t-REML estimates in the proposed model. Besides, it deserves further investigation to pursue some modified procedures, such as the parameter-expanded EM algorithm [16,55] via the covariance adjustment technique

or the multi-cycle scheme operated in the alternating expectation conditional expectation algorithm [56], to speed up the parent ECM for estimating MtSMM.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

M. Arashi http://orcid.org/0000-0002-5881-9241
Tsung-I Lin http://orcid.org/0000-0002-3992-1128

References

[1] Laird NM, Ware JH. Random effects models for longitudinal data. Biometrics. 1982;38: 963–974.
[2] Zeger SL, Diggle PJ. Semi-parametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. Biometrics. 1994;50:689–699.
[3] Liu H, Tu W. A semiparametric regression model for paired longitudinal outcomes with application in childhood blood pressure development. Ann Appl Stat. 2012;6:1861–1882.
[4] Zhang D. Generalized linear mixed models with varying coefficients for longitudinal data. Biometrics. 2004;60:8–15.
[5] Qin GY, Zhu ZY. Robust estimation in generalized semiparametric mixed models for longitudinal data. J Multivariate Anal. 2007;98:1658–1683.
[6] Liang H. Generalized partially linear mixed-effects models incorporating mismeasured covariates. Ann Inst Stat Math. 2009;61:27–46.
[7] Qin GY, Zhu ZY. Robustified maximum likelihood estimation in generalized partial linear mixed model for longitudinal data. Biometrics. 2009;65:52–59.
[8] Li Z, Zhu L. On variance components in semiparametric mixed models for longitudinal data. Scand J Stat. 2010;37:442–457.
[9] Sinha SK, Sattar A. Inference in semi-parametric spline mixed models for longitudinal data. METRON. 2015;73:377–395.
[10] Shah A, Laird N, Schoenfeld D. A random-effects model for multiple characteristics with possibly missing data. J Am Stat Assoc. 1997;92:775–779.
[11] Dempster AP, Laird NM, Rubin DB. Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). J R Stat Soc B. 1977;39:1–38.
[12] Sammel M, Lin X, Ryan L. Multivariate linear mixed models for multiple outcomes. Stat Med. 1999;18:2479–2492.
[13] Song X, Davidian M, Tsiatis AA. An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. Biostatistics. 2002;3:511–528.

[14] Roy A. Estimating correlation coefficient between two variables with repeated observations using mixed effects model. Biom J. 2006;48:286–301.

[15] Wang WL, Fan TH. ECM-based maximum likelihood inference for multivariate linear mixed models with autoregressive errors. Comput Stat Data Anal. 2010;54:1328–1341.

[16] Pinheiro JC, Liu CH, Wu YN. Efficient algorithms for robust estimation in linear mixed-effects model using the multivariate *t* distribution. J Comput Graph Stat. 2001;10:249–276.

[17] Kotz S, Nadarajah S. Multivariate *t* distributions and their applications. Cambridge (UK): Cambridge University Press; 2004.

[18] Rosa GJM, Gianola D, Padovani CR. Bayesian longitudinal data analysis with mixed models and thick-tailed distributions using MCMC. J Appl Stat. 2004;31:855–873.

[19] Lin TI, Lee JC. A robust approach to *t* linear mixed models applied to multiple sclerosis data. Stat Med. 2006;25:1397–1412.

[20] Lin TI, Lee JC. Bayesian analysis of hierarchical linear mixed modeling using the multivariate *t* distribution. J Stat Plann Infer. 2007;137:484–495.

[21] Song PXK, Zhang P, Qu A. Maximum likelihood inference in robust linear mixed-effects models using multivariate *t* distributions. Stat Sinica. 2007;17:929–943.

[22] Wang WL, Fan TH. Estimation in multivariate *t* linear mixed models for multiple longitudinal data. Stat Sinica. 2011;21:1857–1880.

[23] Wang WL. Multivariate *t* linear mixed models for irregularly observed multiple repeated measures with missing outcomes. Biom J. 2013;55:554–571.

[24] Wang WL, Fan TH. Bayesian analysis of multivariate *t* linear mixed models using a combination of IBF and Gibbs samplers. J Multivariate Anal. 2012;105:300–310.

[25] Wang WL, Lin TI. Bayesian analysis of multivariate *t* linear mixed models with missing responses at random. J Stat Comput Simul. 2015;85:3594–3612.

[26] Wang WL, Lin TI, Lachos VH. Extending multivariate-*t* linear mixed models for multiple longitudinal data with censored responses and heavy tails. Stat Meth Med Res. 2018;27(1):48–64.

[27] Wang WL, Lin TI. Multivariate *t* nonlinear mixed-effects models for multi-outcome longitudinal data with missing values. Stat Med. 2014;33:3029–3046.

[28] Wang WL, Castro LM. Bayesian inference on multivariate-*t* nonlinear mixed-effects models for multiple longitudinal data with missing values. Stat Interface. 2018;11:251–264.

[29] Marshall G, Zerbe GO. Non-linear random effects model for multivariate responses with missing data. Stat Med. 2006;25:2817–2830.

[30] Meng XL, Rubin DB. Maximum likelihood estimation via the ECM algorithm: a general framework. Biometrika. 1993;80:267–278.

[31] Kibria BMG. The matrix-*t* distribution and its applications in predictive inference. J Multivariate Anal. 2006;97:785–795.

[32] He XM, Zhu ZY, Fung WK. Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. Biometrika. 2002;89:579–590.

[33] Stone C. Additive regression and other nonparametric models. Ann Stat. 1985;13:689–705.

[34] He X, Fung WK, Zhu Z. Robust estimation in generalized partial linear models for clustered data. J Am Stat Assoc. 2005;100:1176–1184.

[35] Wang WL. Mixture of multivariate *t* linear mixed models for multi-outcome longitudinal data with heterogeneity. Stat Sinica. 2017;27:733–760.

[36] Dickson ER, Grambsch PM, Fleming TR, et al. Prognosis in primary biliary-cirrhosis model for decision-making. Hepatology. 1989;10:1–7.

[37] Markus BH, Dickson ER, Grambsch PM, et al. Efficacy of liver transplantation in patients with primary biliary-cirrhosis. New Engl J Med. 1989;320:1709–1713.

[38] Fleming TR. Harrington DP. Counting processes and survival analysis. New York (NY): Wiley; 1991.

[39] Murtaugh PA, Dickson ER, Van Dam GM, et al. Primary biliary cirrhosis: prediction of short-term survival based on repeated patient visits. Hepatology. 1994;20:126–134.

[40] Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F, editors, 2nd Int. Symp. on Information Theory. Budapest: Akademiai Kiado; 1973. p. 267–281.

[41] Schwarz G. Estimating the dimension of a model. Ann Stat. 1978;6:461–464.

[42] Efron B, Tibshirani R. An introduction to the bootstrap. New York (NY): Chapman and Hall; 1993.

[43] Davison AC, Hinkley DV. Bootstrap methods and their application. Cambridge (UK): Cambridge University Press; 1997.

[44] Schumaker LL. Spline functions. New York (NY): Wiley; 1981.

[45] Fang KT, Anderson TW. Statistical inference in elliptically contoured and related distributions. New York (NY): Allerton Press Inc; 1990.

[46] Fang KT, Kotz S, Ng KW. Symmetric multivariate and related distributions. London: Chapman and Hall; 1990.

[47] Fang KT, Zhang YT. Generalized multivariate analysis. Berlin: Springer-Verlag; 1990.

[48] Azzalini A, Capitano A. The Skew-Normal and related families. London: Cambridge University Press; 2014.

[49] Genton MG. Skew-elliptical distributions and their applications: a journey beyond normality. USA: Chapman and Hall; 2004.

[50] Lin TI, Lachos VH, Wang WL. Multivariate longitudinal data analysis with censored and intermittent missing responses. Stat Med. 2018;37(19):2822–2835.

[51] Castro LM, Wang WL, Lachos VH, et al. Bayesian semiparametric modeling for HIV longitudinal data with censoring and skewness. Stat Meth Med Res. 2019;28(5):1457–1476.

[52] Lin TI, Wang WL. Multivariate-t linear mixed models with censored responses, intermittent missing values and heavy tails. Stat Meth Med Res. 2020;29(5):1288–1304.

[53] Wang WL. Bayesian analysis of multivariate linear mixed models with censored and intermittent missing responses. Stat Med. 2020;39(1930):2518–2535.

[54] Harville DA. Bayesian inference for variance components using only error contracts. Biometrika. 1977;61:383–385.

[55] Liu C, Rubin DB, Wu YN. Parameter expansion to accelerate EM: the PX-EM algorithm. Biometrika. 1998;85:755–770.

[56] Meng XL, van Dyk D. The EM algorithm - an old folk-song sung to a fast new tune. J R Stat Soc B. 1997;59:511–567.

Appendix

A.1 Score vector and Fisher information matrix

The first partial derivatives of log-likelihood function (7) with respect to each element of parameters $\mathbf{\Xi}$ yield the score vector $\mathbf{S}_{\mathbf{\Xi}}$, which contains the following entries:

$$\mathbf{S}_{\boldsymbol{\beta}} = \sum_{i=1}^{n} \left(\frac{\nu + n_i}{\nu + \Delta_i} \right) \mathbf{X}_i^{\top} \mathbf{\Lambda}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}),$$

$$[\mathbf{S}_{\boldsymbol{\omega}}]_l = -\frac{1}{2} \sum_{i=1}^{n} \left\{ \operatorname{tr}(\mathbf{\Lambda}_i^{-1} \overset{\circ}{\mathbf{\Lambda}}_{il}) - \left(\frac{\nu + n_i}{\nu + \Delta_i} \right) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^{\top} \mathbf{\Lambda}_i^{-1} \overset{\circ}{\mathbf{\Lambda}}_{il} \mathbf{\Lambda}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right\},$$

$$s_{\nu} = \frac{1}{2} \sum_{i=1}^{n} \left\{ \mathcal{D}_g \left(\frac{\nu + n_i}{2} \right) - \mathcal{D}_g \left(\frac{\nu}{2} \right) - \frac{n_i}{\nu} - \log \left(1 + \frac{\Delta_i}{\nu} \right) \right.$$

$$\left. + \frac{(\nu + n_i)\Delta_i}{(\nu + \Delta_i)\nu} \right\},$$

for $l = 1, \ldots, g$, $g = q(q+1)/2 + r(r+1)/2 + 2$, where $\mathcal{D}_g(x) = \partial \log \Gamma(x)/\partial x$ is a digamma function, and

$$
\overset{\circ}{\boldsymbol{\Lambda}}_{il} = \frac{\partial \Delta_i(\boldsymbol{D}, \boldsymbol{\Sigma}, \boldsymbol{\phi})}{\partial \boldsymbol{\omega}_l} =
\begin{cases}
\boldsymbol{Z}_i \dfrac{\partial \boldsymbol{D}}{\partial \boldsymbol{\omega}_l} \boldsymbol{Z}_i^\top \\
\qquad\qquad \text{if } \boldsymbol{\omega}_l = \mathrm{vec}(\boldsymbol{D}), \\[2mm]
\dfrac{\partial \boldsymbol{\Sigma}}{\partial \boldsymbol{\omega}_l} \otimes \boldsymbol{\Omega}_i \\
\qquad\qquad \text{if } \boldsymbol{\omega}_l = \mathrm{vec}(\boldsymbol{\Sigma}), \\[2mm]
\boldsymbol{\Sigma} \otimes \dfrac{\partial \boldsymbol{\Omega}_i}{\partial \boldsymbol{\omega}_l} \\
\qquad\qquad \text{if } \boldsymbol{\omega}_l = \boldsymbol{\phi}.
\end{cases}
$$

Here, $\partial \boldsymbol{\Sigma}/\partial \boldsymbol{\omega}_l$ is one in the (j,l)th and (l,j)th elements of $\boldsymbol{\Sigma}$ as $\boldsymbol{\omega}_l = \sigma_{jl}$, and zero otherwise. Similarly for $\partial \boldsymbol{D}/\partial \boldsymbol{\omega}_l$ when $\boldsymbol{\omega}_l$ is the distinct element of $\boldsymbol{D}$, and $\partial \boldsymbol{\Omega}_i/\partial \boldsymbol{\phi}$ depends on the given correlation structure.

Moreover, taking the expectation of negative of the second partial derivatives of log-likelihood function with respect to each element of $\boldsymbol{\Xi}$ leads to the Fisher information matrix, given by

$$
\boldsymbol{J}_{\boldsymbol{\Xi}\boldsymbol{\Xi}} =
\begin{bmatrix}
\boldsymbol{J}_{\beta\beta} & \boldsymbol{J}_{\beta\eta} \\
\boldsymbol{J}_{\beta\eta}^\top & \boldsymbol{J}_{\eta\eta}
\end{bmatrix},
$$

where

$$
\boldsymbol{J}_{\beta\beta} = \sum_{i=1}^n \frac{(\nu + n_i)}{(\nu + n_i + 2)} \boldsymbol{X}_i^\top \boldsymbol{\Lambda}_i^{-1} \boldsymbol{X}_i, \quad \boldsymbol{J}_{\beta\eta} = 0, \quad \text{and} \quad \boldsymbol{J}_{\eta\eta} =
\begin{bmatrix}
\boldsymbol{J}_{\omega\omega} & \boldsymbol{J}_{\omega\nu} \\
\boldsymbol{J}_{\omega\nu}^\top & \boldsymbol{J}_{\nu\nu}
\end{bmatrix}
$$

has the following elements:

$$
[\boldsymbol{J}_{\omega\omega}]_{ls}
$$
$$
= \frac{1}{2} \sum_{i=1}^n \frac{1}{\nu + n_i + 2} \left((\nu + n_i)\mathrm{tr}\left(\boldsymbol{\Lambda}_i^{-1} \overset{\circ}{\boldsymbol{\Lambda}}_{il} \boldsymbol{\Lambda}_i^{-1} \overset{\circ}{\boldsymbol{\Lambda}}_{is} \right) - \mathrm{tr}\left(\boldsymbol{\Lambda}_i^{-1} \overset{\circ}{\boldsymbol{\Lambda}}_{il} \right) \mathrm{tr}\left(\boldsymbol{\Lambda}_i^{-1} \overset{\circ}{\boldsymbol{\Lambda}}_{is} \right) \right),
$$

$$
[\boldsymbol{J}_{\omega\nu}]_l = -\sum_{i=1}^n \frac{1}{(\nu + n_i)(\nu + n_i + 2)} \mathrm{tr}\left(\boldsymbol{\Lambda}_i^{-1} \overset{\circ}{\boldsymbol{\Lambda}}_{il} \right),
$$

$$
\boldsymbol{J}_{\nu\nu} = \frac{1}{4} \sum_{i=1}^n \left(\mathcal{T}_g\left(\frac{\nu}{2}\right) - \mathcal{T}_g\left(\frac{\nu + n_i}{2}\right) - \frac{2n_i(\nu + n_i + 4)}{\nu(\nu + n_i)(\nu + n_i + 2)} \right),
$$

for $l, s = 1, \ldots, g$, with $\mathcal{T}_g(x) = \partial^2 \log \Gamma(x)/\partial x^2$ being a trigamma function.

A.2   Conditions for the asymptotic property and proof of Theorem 3.1

To study the asymptotic properties of ML estimators, the following regularity conditions are required.

(A.1)   Model (1) is correct, n_i is a bounded sequence of positive integers, and the distinct values of t_{ij} form a quasi-uniform sequence that grows dense on $[0, 1]$.

(A.2)   The kth-order derivative of $\boldsymbol{g}_j(\boldsymbol{t}_i)$ is bounded for some $k \geq 2$.

(A.3)   The parameter spaces for $\boldsymbol{\beta}$, $\boldsymbol{\omega}$, and ν are compact sets of $\mathcal{R}^p$, $\mathcal{R}^g$, and $\mathcal{R}^+$, and the true value of $\widehat{\boldsymbol{\Xi}}$ is in the interior of the parameter space of $\widehat{\boldsymbol{\Xi}}$.

(A.4)   As $n \to \infty$, $n^{-1}J_{\beta\beta} \to I_{\beta\beta}$, $n^{-1}J_{\eta\eta} \to I_{\eta\eta}$, and for any $p \times r$ vector $a \neq 0$.

$$\frac{\max_i \left\{ a^\top X_i^\top X_i a \right\}}{a^\top \left(\sum_{i=1}^n X_i^\top X_i \right) a} \to 0. \tag{A1}$$

To verify Theorem 3.1, we need the following lemma.

Lemma A.1: *Under conditions* (A.1) *and* (A.2), *for $j = 1, \ldots, r$, there exists a constant C_j depending only on d_j such that*

$$\sup_{t \in [0,1]} |g_j(t) - B_j(t)\alpha_j| \leq C_j L_j^{-m_j}.$$

The proof of this lemma follows readily from Theorem 12.7 of Schumaker [44].

Proof of Theorem 3.1: By Lemma A.1, we approximate $g_j(t_i)$ by $B_j(t_i)\alpha_j$, then by choosing $L_j \approx N^{1/(2m_j+1)}$ we have

$$
\begin{aligned}
\left(\widehat{g}_j(t) - g_j(t) \right)^2 \\
&= |\widehat{g}_j(t) - g_j(t)||\widehat{g}_j(t) - g_j(t)| \\
&\leq \sup_{t \in [0,1]} |\widehat{g}_j(t) - g_j(t)| \sup_{t \in [0,1]} |\widehat{g}_j(t) - g_j(t)| \\
&\leq C_j L_j^{-m_j} C_j L_j^{-m_j} = C_j^2 N^{2m_j/(2m_j+1)} = O_p(N^{2m_j/(2m_j+1)})
\end{aligned}
$$

which proves part (i) of Theorem 3.1. Part (ii) of Theorem 3.1 follows directly from the weak law of large number, the central limit theorem, and Slutsky's Theorem. ∎