Computational Statistics and Data Analysis 170 (2022) 107439

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

www.elsevier.com/locate/csda

# Robust modeling of multivariate longitudinal data using modified Cholesky and hypersphere decompositions

Anbin Rhee [a], Min-Sun Kwak [b], Keunbaik Lee [c,*]

[a] Department of Statistics, Virginia Tech., USA
[b] Department of Internal Medicine, Healthcare Research Institute, Healthcare System Gangnam Center, Seoul National University Hospital, Republic of Korea
[c] Department of Statistics, Sungkyunkwan University, Republic of Korea

A R T I C L E I N F O

*Article history:*
Received 25 May 2021
Received in revised form 18 January 2022
Accepted 25 January 2022
Available online 8 February 2022

*Keywords:*
Autoregressive
Correlation matrix
Heterogeneity
Innovation variance
t distribution
Positive definite

A B S T R A C T

Due to the convenience of the statistical interpretation and parameter estimation, a normal distribution is typically assumed for multivariate longitudinal data analysis. However, this assumption may be questionable in practice, because it is possible that outliers exist or that the underlying data will show heavy tails. In addition, since the covariance matrix should explain complex correlation structures, it must be positive-definite, and as it is also high-dimensional, the modeling of the covariance matrix is not easy. To solve these problems, we propose the robust modeling of multivariate longitudinal data by considering multivariate $t$ distribution, and we exploit modified Cholesky and hypersphere decompositions to model the covariance matrix. The estimation of the models is shown to be robust when the data include outliers and exhibit heavy tails. The performance of our proposed model and algorithm is illustrated using a nonalcoholic fatty liver disease data set and several simulation studies.



## 1. Introduction

Multivariate longitudinal data are collected in various fields, including epidemiology, biomedicine, and public health science. The main issue in such studies is modeling a proper covariance matrix, as multiple responses are repeatedly measured from each subject over time. However, unlike univariate longitudinal data, it is not easy to model the covariance matrix for multivariate longitudinal data because the covariance matrix should consider three complex correlations: the correlation within separate responses over time, the cross-correlation between different responses at different times, and the correlation between different responses at each time point. To analyze these data while considering all relevant factors, multivariate linear models have been developed with multivariate normal errors (Kim and Zimmerman, 2012). Since then multivariate linear models assuming a multivariate normal distribution have continued to be used in multivariate data analysis (Xu and Mackenzie, 2012, Feng et al., 2016, Kohli et al., 2016, Lee et al., 2020, Lee et al., 2021), because the multivariate normality assumption leads to easier parameter estimation. However, the normality assumption may be questionable in practice - such as when the data include outliers or exhibit heavy tails - and there may be bias in parameter estimation. In this situation, robust estimation for multivariate linear models using the multivariate $t$ distribution can be useful.

* Corresponding author.
*E-mail address:* keunbaik@skku.edu (K. Lee).

https://doi.org/10.1016/j.csda.2022.107439
0167-9473/

Statistical literature has proposed a variety of linear models using a $t$ error distribution. For example, He et al. (2004) proposed a robust estimator of linear regression, while Pinheiro et al. (2001) proposed linear mixed effects models using multivariate $t$ distributions, and these models were shown to be efficient for the multivariate $t$ distribution relative to the normal distribution. Lin and Wang (2009) proposed a robust approach involving the joint modeling of mean and scale covariance for univariate longitudinal data. Their approach exploits Pourahmadi (1999)'s modified Cholesky decomposition (MCD) and the multivariate $t$ distribution. Lin and Wang (2011) presented a fully Bayesian approach for Lin and Wang (2009)'s model. In this paper, we extend Lin and Wang (2009)'s models to accommodate multivariate longitudinal data and develop robust estimation in the case of outliers or heavy-tailed errors.

Multivariate $t$ error distributions are used for multivariate linear mixed models to analyze multivariate longitudinal data. For example, Wang and Fan (2010) and Wang and Fan (2011) proposed multivariate linear mixed models with an autoregressive $t$ error distribution. In another study, Wang et al. (2018) extended the models outlined in Wang and Fan (2010) and Wang and Fan (2011) to accommodate censored heavy-tailed multivariate longitudinal data. Sequential papers extending these models to nonlinear mixed models have also been proposed (Wang and Lin, 2014, Lin and Wang, 2020). In the above papers, the scale matrix in the multivariate $t$ distribution has a Kronecker product (KP) structure. The KP structure makes the implicit assumption that, for response variables, the longitudinal correlation structure is the same, and that the covariance between response variables at the same time does not depend on time and instead remains constant over all time points. However, this is often too strong an assumption. In this paper, we consider the MCD to allow for a more flexible structure.

For univariate longitudinal data analysis, the MCD decomposes the inverse covariance matrix into unconstrained parameters (generalized autoregressive parameters (GARPs) and log innovation variances (IVs)), and the covariance matrix from the MCD is guaranteed to be positive-definite (Pourahmadi, 1999). However, the MCD cannot directly be exploited to model the covariance matrix for multivariate longitudinal data analysis because there are complex correlated structures to consider (the correlation within separate responses over time, the cross-correlation between different responses at different times, and the correlation between responses at each time point). In particular, the MCD cannot be used to explain the correlation between responses at each time point. Therefore, statistical models have been developed to explain these correlations in multivariate longitudinal data analysis.

Various studies have attempted to explain these complex correlations. All of the models described below use the MCD to decompose a covariance matrix into a generalized autoregressive matrix (GARM) and an innovation covariance matrix (ICM). The GARM explains the serial correlations (the correlation within separate responses over time and the cross-correlation between different responses at different times), while the ICM explains the correlations between responses at each time point. Further, the estimated covariance matrix using the MCD is positive-definite. Various models have been proposed to describe the correlations between responses at each time point. For example, Kim and Zimmerman (2012) exploited another MCD to model the ICM, but they imposed ordering of the responses, which is unnatural in most practical situations. Xu and Mackenzie (2012) used matrix logarithmic covariance modeling (Chiu et al., 1996) to model the ICM. However, the parameters from the matrix logarithmic covariance modeling are difficult to estimate and interpret. Kohli et al. (2016) used the enhanced Anderson (1973)'s linear covariance models to model the ICM, and the estimation of the ICM requires a complicated algorithm. Recently, Lee et al. (2020) decomposed the ICM into innovation standard deviations (ISDs) and correlation matrices, and they modeled the correlation matrices using hypersphere decomposition (HD). In this paper, we extend Lee et al. (2020)'s models to accommodate multivariate longitudinal data with outliers or heavy-tailed errors.

The rest of this paper is organized as follows. Section 2 begins with brief literature reviews related to the proposed model, while Section 3 proposes robust multivariate linear models with covariance matrix for multivariate longitudinal data using the MCD and HD. Section 4 presents the maximum likelihood estimations from our proposed models. Section 5 provides the simulation results to illustrate the performance of the proposed models, and Section 6 presents the analysis of the motivating data using our proposed models. Finally, Section 7 concludes this paper.

## 2. Literature review

In this section, we elaborate upon modelings of the covariance matrix in linear models for univariate longitudinal data. As longitudinal data is obtained by repeatedly measuring from the same subjects, there are inevitable correlations that must be considered in properly analyzing the data. However, there are still some difficulties in modeling the covariance matrix for longitudinal data: the constraints of positive-definiteness and the high dimensionality of the covariance matrix, $\Sigma$. To solve these problems, several methods have been proposed. In this section, we review two recently developed methods for modeling the covariance matrix: MCD and HD. Both decompositions are exploited for linear models with a normal error distribution.

### 2.1. Modified Cholesky decomposition

Pourahmadi (1999) proposed the MCD to model the covariance matrix in linear models for univariate longitudinal data. Through this process, the covariance matrix is decomposed into unconstrained parameters.

Let $y_i = (y_{i1}, \ldots, y_{it}, \ldots, y_{in_i})^T$ be the response vector for subject $i$ ($i = 1, 2, \ldots, N$) measured at time $t$ ($t = 1, 2, \ldots, n_i$) and $x_{it}$ be a $p \times 1$ vector of covariates corresponding to the response $y_{it}$. We then have

$$
\begin{aligned}
y_{i1} &= x_{i1}^T\beta + e_{i1},\\
y_{it} &= x_{it}^T\beta + \sum_{j=1}^{t-1}\phi_{itj}(y_{ij} - x_{ij}^T\beta) + e_{it},\\
e_i &= (e_{i1}, e_{i2}, \ldots, e_{in_i})^T \sim N(0, D_i),
\end{aligned} \tag{1}
$$

where $\beta$ is a $p \times 1$ vector of unknown mean parameters, $\phi_{itj}$ is a generalized autoregressive parameter (GARP), and $D_i = diag(\sigma_{i1}^2, \ldots, \sigma_{it}^2, \ldots, \sigma_{in_i}^2)$. Note that the $\sigma_{it}^2$'s are called innovation variances (IVs).

After rearranging both sides of (1), we obtain

$$
y_{it} - x_{it}^T\beta - \sum_{j=1}^{t-1}\phi_{itj}(y_{ij} - x_{ij}^T\beta) = e_{it}. \tag{2}
$$

Then, we have the following matrix form from (2):

$$
T_i(Y_i - X_i^T\beta) = e_i, \tag{3}
$$

where

$$
T_i = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0\\ -\phi_{i21} & 1 & 0 & \cdots & 0\\ \vdots & \vdots & 1 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ -\phi_{in1} & -\phi_{in2} & \cdots & -\phi_{in,n-1} & 1 \end{pmatrix}, \quad X_i = \begin{pmatrix} x_{i1}^T\\ x_{i2}^T\\ \vdots\\ x_{in_i}^T \end{pmatrix}.
$$

Since there may be differences in the number of times each subject has been measured, the number of repeated measurements is written as $n_i$.

Taking covariances on both sides of (3), we obtain:

$$
\begin{aligned}
&T_i\Sigma T_i^T = D_i,\\
&\Sigma_i = T_i^{-1}D_iT_i^{-T} \iff \Sigma_i^{-1} = T_i^T D_i^{-1} T_i.
\end{aligned} \tag{4}
$$

To reduce the dimensionality and satisfy the positive-definiteness constraint of covariance matrix $\Sigma$, the GARPs and IVs can be modeled as:

$$
\phi_{itj} = \omega_{itj}^T\alpha, \quad \log(\sigma_{it}^2) = h_{it}^T\lambda,
$$

where $\alpha$ and $\lambda$ are $a \times 1$ vector and $b \times 1$ vector of unknown parameters, respectively. Further, $\omega_{itj}$ and $h_{it}$ are $a \times 1$ and $b \times 1$ design vectors, respectively, and they are time and/or subject specific covariates to model GARPs and IVs, respectively. Through the reparameterization of the covariance matrix $\Sigma$, the number of parameters is reduced to $a + b$. Further, all positive IVs guarantee the positive-definiteness of the estimated covariance matrix $\Sigma$. This also allows for easier interpretations and simpler computation.

### *2.2. Hypersphere decomposition*

Zhang et al. (2015) proposed a joint mean-variance-correlation modeling approach for longitudinal data using hypersphere decomposition (HD). In this approach, the covariance matrix is first decomposed as $\Sigma_i = S_iR_iS_i$ where $S_i = diag\left\{\sigma_{i1}, \ldots, \sigma_{in_i}\right\}$ and $R_i$ is the correlation matrix. HD is exploited to account for the correlation between repeated responses over time in longitudinal studies.

To make $\Sigma_i$ be positive-definite, $R_i$ should also be positive-definite. However, modeling $R_i$ is not easy for three reasons: 1) the diagonal elements should be 1's; 2) the off-diagonal elements of $R_i$ should be in the range $[-1, 1]$; and 3) $R_i$ must be positive-definite. HD can be used to solve each of these difficulties.

Using HD, $R_i$ is reparameterized as $R_i = F_iF_i^T$ where $F_i$ is a lower triangular matrix with the (1,1)th element being one and the other elements determined as

$$
f_{ilm} = \begin{cases} cos(\omega_{ilm}), & \text{for } m = 1,\ l = 2, \ldots, K;\\ cos(\omega_{ilm})\prod_{r=1}^{m-1} sin(\omega_{ilr}), & \text{for } 2 \le m < l \le K;\\ \prod_{r=1}^{m-1} sin(\omega_{ilr}), & \text{for } l = m;\ m = 2, \ldots, K, \end{cases}
$$

with $\omega_{ilm} \in (0, \pi)$. Note that $f_{ilm}$ are trigonometric functions of $\omega_{ilr}$ called hypersphere parameters (HPs). Using HD ensures the positive-definiteness of $R_i$ and enables the identification of the correlation between responses over time. However, no studies have used HD to model the covariance matrix for multivariate $t$ longitudinal data. Therefore, we aim to address this gap.

## 3. Robust modeling of multivariate longitudinal data

For the analysis of multivariate longitudinal data, we commonly use multivariate linear models that assume a multivariate normal distribution. In these models, three correlations are considered to estimate covariate effects on the responses: the correlation within separate responses over time, the cross-correlation between different responses at different times, and the correlations between different responses at different times (Kim and Zimmerman, 2012, Kohli et al., 2016, Lee et al., 2020). However, when the data include outliers or exhibit heavy tails, the multivariate normality assumption may be questionable. In this section, we overcome this limitation by proposing multivariate $t$ linear models (MTLM).

### 3.1. The proposed models

Let $y_i = (y_{i1}^T, \ldots, y_{it}^T, \ldots, y_{in_i}^T)^T$ be the response vector for the $i$th subject where $y_{it}^T = (y_{it1}, \ldots, y_{itK})$ is $K$ continuous responses at time period $t$ ($i = 1, \ldots, N$; $t = 1, \ldots, n_i$). Also let $x_{it}$ be the covariate vector corresponding to $y_{it}$. Assume that $y_i$, for $i = 1, \ldots, N$, are independent. Then we specify our proposed models as follows, for $k = 1, \ldots, K$,

$$\begin{aligned} y_{i1k} &= x_{i1}^T \beta_k + e_{i1k}, \\ y_{itk} &= x_{it}^T \beta_k + \sum_{j=1}^{t-1} \sum_{g=1}^{K} \phi_{itj,kg} (y_{ijg} - x_{ij}^T \beta_g) + e_{itk}, \end{aligned} \tag{5}$$

where $\beta_k$ is a $p \times 1$ unknown mean parameter vector, $\phi_{itj,kg}$'s are generalized autoregressive parameters (GARPs), and $e_{itk}$'s are prediction errors.

Note that this model exhibits an AR structure and that $\phi_{itj,kg}$ is the coefficient of the $j$th previous residual of outcome $g$. In this model, $y_{it}$ is affected by all of the previous responses and the GARPs ($\phi_{itj,kg}$) account for two serial correlations: the correlation within each response over time and the cross-correlation between different responses at different times.

We assume that

$$e_i \overset{indep}{\sim} t_\nu(0, D_i), \tag{6}$$

where $e_i = (e_{i1}^T, \ldots, e_{in_i}^T)^T$ with $e_{it} = (e_{it1}, \ldots, e_{itK})^T$, and $t_\nu(0, D_i)$ is the multivariate $t$-distribution with degrees of freedom (d.f.) $\nu$, location vector 0 and scale covariance matrix $D_i$.

The multivariate $t$-distribution can be represented as the following two-level hierarchical form

$$e_i \overset{indep}{\sim} N(0, u_i^{-1} D_i), \tag{7}$$

$$u_i \overset{iid}{\sim} Gamma(\nu/2, \nu/2), \tag{8}$$

where $D_i = diag\left\{D_{i1}, \ldots, D_{in_i}\right\}$ is called the innovation covariance matrix (ICM) with $D_{it} = var(e_{it})$.

We can rewrite (5) in matrix form as follows:

$$T_i(y_i - X_i\beta) = e_i, \tag{9}$$

where

$$T_i = \begin{pmatrix} I & 0 & \cdots & 0 \\ -\Phi_{i21} & I & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ -\Phi_{in_i1} & -\Phi_{in_i2} & \cdots & I \end{pmatrix}, \quad X_i = \begin{pmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{in_i} \end{pmatrix},$$

with

$$\Phi_{itj} = \begin{pmatrix} \phi_{itj,11} & \cdots & \phi_{itj,1K} \\ \phi_{itj,21} & \cdots & \phi_{itj,2K} \\ \vdots & \ddots & \vdots \\ \phi_{itj,K1} & \cdots & \phi_{itj,KK} \end{pmatrix},$$

and $X_{it} = diag\left\{x_{it}^T, \ldots, x_{it}^T\right\}$ being a $K \times Kp$ matrix. Here, $\Phi_{itj}$ is called the GARP matrix (GARPM). Note that $T_i$ and $D_i$ uniquely exist, and that $T_i$ is nonsingular (Lee et al. (2017), Lee et al. (2020)).

Lee et al. (2020) proposed the modeling of ICM in the multivariate linear models (MLMs) using the modified Cholesky decomposition and hypersphere decomposition. Recall that the purposes of these two decompositions are to solve the following difficulties: 1) the constraint of the positive-definiteness of $\Sigma$; 2) high-dimensionality; and 3) heteroscedasticity of $\Sigma$.

The conditional mean of $y_i$ in (9) is given by:

$$E(y_i; u_i) = X_i\beta.$$

To calculate the conditional covariance of $y_i$, we take the conditional covariances on both sides of (9). We then obtain the following result:

$$T_i var(y_i; u_i) T_i^T = u_i^{-1} D_i \quad \Leftrightarrow \quad var(y_i; u_i) = u_i^{-1} T_i^{-1} D_i T_i^{-T} \overset{let}{=} u_i^{-1} \Sigma_i, \tag{10}$$

where $\Sigma_i = T_i^{-1} D_i T_i^{-T}$. The marginal mean and variance of $y_i$ are respectively given by:

$$E(y_i) = X_i\beta,$$
$$var(y_i) = \frac{\nu}{\nu - 2}\Sigma_i.$$

From Peel and McLachlan (2000), the marginal distribution of $y_i$ is a multivariate $t$-distribution with degrees of freedom (d.f.) $\nu$, mean vector $X_i\beta$, scale covariance matrix $\Sigma_i$ and the probability density function of:

$$p(y_i|\beta, \Sigma_i, \nu) = \frac{\Gamma\left(\frac{\nu+n_iK}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\pi)^{\frac{n_iK}{2}}} \nu^{-\frac{n_iK}{2}} |\Sigma_i|^{-\frac{1}{2}} \left(1 + \frac{(y_i - X_i\beta)^T \Sigma_i^{-1}(y_i - X_i\beta)}{\nu}\right)^{-\frac{\nu+n_iK}{2}}. \tag{11}$$

Note that the heavy tail of the multivariate $t$-distribution shows the robustness of the parameter estimation for the multivariate linear models.

The parameters in the matrix $T_i$ are not identifiable without knowledge of any of the covariance matrix structure. The matrix $T$ in (10) has $nK(nK-1)/2$ parameters where $n = max(n_i)$. However, with a specific structure of the covariance matrix such as those based on the MCD structure in (10), the matrix can be identified and the identifiability can easily be assessed by checking the invertibility of the Hessian matrix in Section 4.

Note that $\Sigma_i$ in (10) is positive-definite if and only if all of the diagonal matrices of $D_i$ are positive-definite (Lee et al., 2020). Therefore, we need to model $D_{it}$ as being positive-definite. Remember that $D_{it}$ presents the correlation between responses at time $t$ for subject $i$, and that the responses at the same time cannot be ordered. Therefore, the modeling of $D_{it}$ cannot use MCD, unlike Kim and Zimmerman (2012). Instead, we first consider the variance-correlation decomposition as follows:

$$D_{it} = S_{it} R_i S_{it}, \tag{12}$$

where $S_{it} = diag\{\sigma_{it1}, \ldots, \sigma_{itK}\}$ and $R_i$ is the $K \times K$ correlation matrix for $e_{it}$. Here $\sigma_{itk}$'s are called innovation standard deviations (ISDs).

From (12), the positive-definiteness of $D_{it}$ is equivalent to both the positiveness of all $\sigma_{itk}$'s and the positive-definiteness of $R_i$. However, the modeling of the positive-definite $R_i$ is not easy, because the diagonal elements of $R_i$ should be ones. To overcome this problem, we use hypersphere decomposition (Lee et al., 2020) through the following decomposition:

$$R_i = F_i F_i^T,$$

where

$$F_i = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ f_{i21} & f_{i22} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_{iK1} & f_{iK2} & f_{iK3} & \cdots & f_{iKK} \end{pmatrix},$$

with

$$f_{ilm} = \begin{cases} cos(\omega_{ilm}), & \text{for } m = 1, l = 2, \cdots, K; \\ cos(\omega_{ilm}) \prod_{r=1}^{m-1} sin(\omega_{ilr}), & \text{for } 2 \le m < l \le K; \\ \prod_{r=1}^{m-1} sin(\omega_{ilr}), & \text{for } l = m;\ m = 2, \cdots, K, \end{cases}$$

being trigonometric functions of angles $\omega_{ilm} \in (0, \pi)$, which are called the hypersphere parameters (HPs) under the triangular angles parametrization (Rapisarda et al., 2007).

Note that for the direct modeling of the correlation matrix, the triangular angles parametrization is used for unconstrained correlation parametrization, and this parametrization can be directly interpreted for the correlations.

### 3.2. Modeling of GARPs, ISDs, and HPs

The parameters of the GARPs, ISDs, and HPs can be modeled using time and/or subject-specific covariate vectors $w_{itj}$ by setting:

$$\phi_{itj,kg} = w_{itj}^T \alpha_{kg}, \tag{13}$$

$$\log \sigma_{itk} = h_{it}^T \lambda_k, \tag{14}$$

$$\log\left(\frac{\omega_{ilm}}{\pi - \omega_{ilm}}\right) = g_{ilm}^T \delta, \tag{15}$$

where $\alpha_{kg}$, $\lambda_k$, and $\delta$ are respectively $a \times 1$, $b \times 1$, and $c \times 1$ vectors of unknown parameters, while $w_{itj}$, $h_{it}$, and $g_{ilm}$ are the corresponding time and/or subject-specific covariate vectors (Lee and Chen, 2019).

Note that the number of parameters in the covariance matrix can be reduced by considering models (13)-(15). These design vectors include subject-specific covariates for the heteroscedastic covariance matrix. For example, the time lag, $|Time_{it} - Time_{ij}|$ in the design vector $w_{itj}$ specifies higher lag models. $Time_{it}$ in the design vector $h_{it}$ specifies the linearity in $Time_{it}$. $g_{ilm}$ is also a subject-specific response-dependent vector. However, we consider a simple structure such as $g_{ilm}^T \nu = \nu_l + \nu_m$ for $l, m = 1, \ldots, K$. We also note that the model (14) makes $\sigma_{itk}$ positive while the model (15) makes $R_i$ positive-definite through the HD. As a result, the covariance matrix is guaranteed to be positive-definite.

## 4. Maximum likelihood estimation

In this section, we present the derivation of the maximum likelihood estimation for our proposed models. Let $\theta = (\beta^T, \alpha^T, \lambda^T, \nu^T)^T$. The loglikelihood function is given by:

$$\begin{aligned}
\log L(\theta; y) &= \sum_{i=1}^{N}\Bigg[\log\Gamma\left(\frac{n_iK+\nu}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) - \frac{n_iK}{2}\log(\pi\nu) - \sum_{t=1}^{n_i}\sum_{k=1}^{K} h_{it}^T\lambda_k \\
&\qquad - n_i \log|F_i| - \frac{n_iK+\nu}{2}\log\left(1 + \frac{1 + (y_i - X_i\beta)^T \Sigma_i^{-1}(y_i - X_i\beta)}{\nu}\right)\Bigg] \\
&= \sum_{i=1}^{N}\Bigg[\log\Gamma\left(\frac{n_iK+\nu}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) - \frac{n_iK}{2}\log(\pi\nu) - \sum_{t=1}^{n_i}\sum_{k=1}^{K} h_{it}^T\lambda_k \\
&\qquad - n_i \log|F_i| - \frac{n_iK+\nu}{2}\log\left(1 + \frac{1 + (r_i - C_i\alpha)^T D_i^1 - 1(r_i - C_i\alpha)}{\nu}\right)\Bigg],
\end{aligned}$$

where $r_i = y_i - X_i\beta$,

$$C_i = \begin{pmatrix} 0 \\ C_{i2} \\ \vdots \\ C_{in_i} \end{pmatrix}, \quad C_{it} = \begin{pmatrix} \sum_{j=1}^{t-1}\sum_{k=1}^{K}(y_{ijk} - x_{ij}^T\beta_k)W_{itj}(1,k) \\ \vdots \\ \sum_{n_i=1}^{t-1}\sum_{k=1}^{K}(y_{ijk} - x_{ij}^T\beta_k)W_{itj}(K,k) \end{pmatrix}.$$

Here, 0 in $C_i$ is a $K \times 1$ zero vector and for $l = 1, \ldots, K$, $W_{itj}(l,k)$ is a $1 \times aK^2$ vector such that

$$W_{itj}(l,k)\alpha = w_{itj}^T \alpha_{lk},$$

where $\alpha = (\alpha_{11}^T, \alpha_{12}^T, \ldots, \alpha_{KK}^T)^T$ and $W_{itj}(l,k) = (0^T \cdots 0^T w_{itj}^T 0^T \cdots 0^T)$ with $w_{itj}^T$ being located for $\alpha_{lk}$.

Maximizing the log-likelihood with respect to $\theta$ yields the score functions that are given by:

$$\frac{\partial \log L(\theta; y)}{\partial \beta} = \sum_{i=1}^{N} \tau_i X_i^T \Sigma_i^{-1}(y_i - X_i\beta), \tag{16}$$

$$\frac{\partial \log L(\theta; y)}{\partial \alpha} = \sum_{i=1}^{N} \tau_i C_i^T D_i^{-1}(r_i - C_i\alpha), \tag{17}$$

$$\frac{\partial \log L(\theta; y)}{\partial \lambda_{kg}} = -\sum_{i=1}^{N}\left\{\sum_{t=1}^{n_i} h_{itg} + \frac{1}{2}\tau_i r_i^T T_i^T \frac{\partial D_i^{-1}}{\partial \lambda_{kg}} T_i r_i\right\}, \quad k = 1, \ldots, K; g = 1, \ldots, b, \tag{18}$$

$$\frac{\partial \log L(\theta; y)}{\partial \delta_g} = -\sum_{i=1}^{N}\left\{ n_i \sum_{k=2}^{K} \frac{\frac{\partial f_{ikk}}{\partial \delta_g}}{f_{ikk}} + \frac{1}{2}\tau_i r_i^T T_i^T \frac{\partial D_i^{-1}}{\partial \delta_g} T_i r_i \right\}, \quad g = 1, \ldots, c, \tag{19}$$

$$\frac{\partial \log L(\theta; y)}{\partial \nu} = \frac{1}{2}\sum_{i=1}^{N}\left\{ \gamma\left(\frac{\nu + n_i K}{2}\right) - \gamma\left(\frac{\nu}{2}\right) - \frac{n_i K}{\nu} - \log\left(1 + \frac{r_i^T \Sigma_i^{-1} r_i}{\nu}\right) + \frac{1}{\nu}\tau_i r_i^T \Sigma_i^{-1} r_i \right\}, \tag{20}$$

where

$$\tau_i = \frac{n_i K + \nu}{\nu + r_i^T \Sigma_i^{-1} r_i},$$

$$\frac{\partial D_i^{-1}}{\partial \lambda_{kg}} = diag\left\{ \frac{\partial S_{i1}^{-1}}{\partial \lambda_{kg}} R_i^{-1} S_{i1}^{-1} + S_{i1}^{-1} R_i^{-1} \frac{\partial S_{i1}^{-1}}{\partial \lambda_{kg}}, \ldots, \frac{\partial S_{in_i}^{-1}}{\partial \lambda_{kg}} R_i^{-1} S_{in_i}^{-1} + S_{in_i}^{-1} R_i^{-1} \frac{\partial S_{in_i}^{-1}}{\partial \lambda_{kg}} \right\},$$

$$\frac{\partial D_i^{-1}}{\partial \delta_g} = diag\left\{ -S_{i1}^{-1} R_i^{-1} \frac{\partial R_i}{\partial \delta_g} R_i^{-1} S_{i1}^{-1}, \ldots, -S_{in_i}^{-1} R_i^{-1} \frac{\partial R_i}{\partial \delta_g} R_i^{-1} S_{in_i}^{-1} \right\},$$

$\frac{\partial S_{it}^{-1}}{\partial \lambda_{kg}}$ is a $K \times K$ diagonal matrix with $-\frac{h_{itg}}{\sigma_{itk}}$ for the $(k, k)$th element and zeros for the other elements, and $\frac{\partial R_i}{\partial \delta_g} = \frac{\partial F_i}{\partial \delta_g} F_i^T + F_i \frac{\partial F_i^T}{\partial \delta_g}$ having

$$\frac{\partial f_{ilm}}{\partial \delta_g} = \begin{cases} -f_{ilm} \tan(\omega_{ilm}) \frac{\partial \omega_{ilm}}{\partial \delta_g}, & \text{for } m = 1, l = 2, \ldots, K; \\ f_{ilm} \sum_{r=1}^{m-1} \frac{1}{\tan(\omega_{ilr})} \frac{\partial \omega_{ilr}}{\partial \delta_g}, & \text{for } l = m; m = 2, \ldots, K; \\ f_{ilm} \left\{ -\tan(\omega_{ilm}) \frac{\partial \omega_{ilm}}{\partial \delta_g} + \sum_{r=1}^{m-1} \frac{1}{\tan(\omega_{ilr})} \frac{\partial \omega_{ilr}}{\partial \delta_g} \right\}, & \text{for } 2 \le m < l \le K, \end{cases}$$

and $\gamma(x) = \frac{\partial \log \Gamma(x)}{\partial x}$.

We now have the necessary ingredients to present the Fisher information in terms of the blocks of a partitioned $5 \times 5$ matrix corresponding to $\beta$, $\alpha$, $\lambda$, $\delta$, and $\nu$. Let $\psi = (\alpha^T, \lambda^T, \delta^T, \nu)^T$ and $\theta_0 = (\beta_0^T, \psi_0^T)^T$ be the true value of $\theta$. Also let $I(\theta_0) = diag\{I(\beta_0), I(\psi_0)\}$, where

$$I(\beta) = \sum_{i=1}^{N} \frac{\nu + n_i K}{\nu + n_i K + 2} X_i^T \Sigma_i^{-1} X_i,$$

$$I(\psi) = \begin{pmatrix} I(\alpha) & I(\alpha, \lambda) & I(\alpha, \delta) & I(\alpha, \nu) \\ I(\alpha, \lambda)^T & I(\lambda) & I(\lambda, \delta) & I(\lambda, \nu) \\ I(\alpha, \delta)^T & I(\lambda, \delta)^T & I(\delta) & I(\delta, \nu) \\ I(\alpha, \nu)^T & I(\lambda, \nu)^T & I(\delta, \nu)^T & I(\nu) \end{pmatrix},$$

and the elements of the information matrix $I(\psi)$ are given in (21)-(30) in the Appendix.

To obtain the elements of the Fisher information matrix, we need the following Proposition:

**Proposition.** *For the multivariate t distribution in (11), we have the following results:*
*(a)* $E(\tau_i) = \left(\frac{\nu+1}{\nu+2}\right)^{\frac{n_i K}{2}}$, *(b)* $E(\tau_i r_i) = 0$, *(c)* $E(\tau_i^2 r_i) = 0$,
*(d)* $E(\tau_i r_i r_i^T) = \Sigma_i$, *(e)* $E(\tau_i^2 r_i r_i^T) = \frac{n_i K + \nu}{n_i K + \nu + 2}\Sigma_i$.

**Proof.** The detailed proof is given in the Appendix.

Using the score function for $\beta$ from (16), we obtain the maximum likelihood estimate (MLE) for $\beta$, which is given by:

$$\hat{\beta} = \left(\sum_{i=1}^{N} \tau_i X_i^T \hat{\Sigma}_i^{-1} X_i\right)^{-1} \left(\sum_{i=1}^{N} \tau_i X_i^T \hat{\Sigma}_i^{-1} y_i\right),$$

where $\hat{\Sigma}_i = \Sigma_i(\hat{\alpha}, \hat{\lambda}, \hat{\delta})$.

Since the solutions of the score functions for $\alpha$, $\lambda$, and $\delta$ are not available, a numerical method must be used. Once the information matrix is computed, the iterative Fisher-scoring algorithm can be used to compute the MLE of the parameters. To this end, the working estimates $(\hat{\alpha}, \hat{\lambda}, \hat{\delta})$ are updated on the $o$th iteration as follows:

$$\begin{pmatrix} \alpha^{(o+1)} \\ \lambda^{(o+1)} \\ \delta^{(o+1)} \\ \nu^{(o+1)} \end{pmatrix} = \begin{pmatrix} \alpha^{(o)} \\ \lambda^{(o)} \\ \delta^{(o)} \\ \nu^{(o)} \end{pmatrix} + \left[ \begin{pmatrix} I(\alpha) & I(\alpha,\lambda) & I(\alpha,\delta) & I(\alpha,\nu) \\ I(\alpha,\lambda)^T & I(\lambda) & I(\lambda,\delta) & I(\lambda,\nu) \\ I(\alpha,\delta)^T & I(\lambda,\delta)^T & I(\delta) & I(\delta,\nu) \\ I(\alpha,\nu)^T & I(\lambda,\nu)^T & I(\delta,\nu)^T & I(\nu) \end{pmatrix} \begin{pmatrix} \frac{\partial \log L(\theta;y)}{\partial \alpha} \\ \frac{\partial \log L(\theta;y)}{\partial \lambda} \\ \frac{\partial \log L(\theta;y)}{\partial \delta} \\ \frac{\partial \log L(\theta;y)}{\partial \nu} \end{pmatrix} \right]_{(\alpha,\lambda,\delta,\nu)=(\alpha^{(o)},\lambda^{(o)},\delta^{(o)},\nu^{(o)})}.$$

This procedure is iterated until convergence.

## 5. Simulation study

The main purpose of the simulation study is to investigate the performance and robustness of the proposed models through inference on the marginal mean regression coefficients. We considered two scenarios in the simulations: The first simulation study was conducted to verify the performance of the estimation for the parameters in the proposed model. The second simulation study was conducted to examine the robustness of the proposed models compared to the multivariate linear model based on the assumption of normality.

### 5.1. Performance of the proposed model

The parameter estimation performance of the proposed model was verified using various sample sizes and degrees of freedom.

We designed 500 random multivariate longitudinal data sets from a $t$-distribution. The simulation studies were conducted under proposed models (5)-(8) with two covariates: group and time. For the sake of convenience, the true parameter values are chosen to be the same as those of the multivariate linear models in Lee et al. (2020). For $k = 1, 2, 3$ and $t = 1, \ldots, n_i = 10$,

$$\begin{aligned}
y_{itk} &= x_{it}^T\beta_k + \sum_{j=1}^{t-1}\sum_{g=1}^{K} \phi_{itj,kg}(y_{ijg} - x_{ij}^T\beta_g) + e_{itk}, \\
x_{it}^T\beta_k &= \beta_{k0} + \beta_{k1}Group_i + \beta_{k2}Time_{ij} + \beta_{k3}Group_i \times Time_{it}, \\
\phi_{itj,kg} &= \alpha_{lm0} I_{|t-j|=1}, \\
\log\sigma_{itk} &= \lambda_{k0}, \\
\log\left(\frac{\omega_{ilm}}{\pi - \omega_{ilm}}\right) &= \delta_l + \delta_m,
\end{aligned}$$

where $Time_{it} \sim N(0, 1)$ and $Group_i$ equals 0 or 1 with each group having approximately the same sample size. The true parameters in the simulations are as follows:

$$\begin{aligned}
&\beta_1 = (0.3, -0.1, 0.2, 0.3)^T, \beta_2 = (0.2, -0.1, 0.2, 0.3)^T, \beta_3 = (0.2, -0.2, 0.2, 0.4)^T, \\
&(\alpha_{110}, \alpha_{120}, \ldots, \alpha_{330}) = (0.3, 0.4, 0.1, 0.1, 0.3, 0.1, 0.1, 0.3, 0.2), \\
&(\lambda_{10}, \lambda_{20}, \lambda_{30}) = (0.1, 0.2, 0.2), \quad (\delta_1, \delta_2, \delta_3) = (-0.5, -0.4, -0.3).
\end{aligned}$$

To demonstrate the consistency of the proposed model, each table presents the percent relative bias (PRB), coverage probability (CP), mean of standard errors (SE), standard deviation of 500 estimates (SD), mean of estimated mean parameters (Mean) and degrees of freedom. We also present Frobenius norms (Frob) of estimated scale matrices, which involve squaring the difference between aspects of the estimator and the target. We conducted simulations with sample sizes of 100, 300, and 500 as well as various degrees of freedom ($\nu = 3, 5, 7$).

Means, PRBs, SEs, SDs, CPs, and Frob for degrees of freedom 3, 5, and 7 are respectively listed in Tables 1, 2, and 3. As the sample size increases, PRBs, SEs, SDs, and Frob all decrease. $|PRB|$ at the bottom of the table indicates the sum of the absolute PRB values of all estimates of $\beta$. A higher value of $|PRB|$ indicates a more biased and inaccurate estimation. Further, as shown in all the tables, SEs and SDs are approximately the same for all estimated parameters, thus demonstrating that the calculation of SEs using the Fisher information matrix works well.

### 5.2. Evaluation of robustness

Two simulations were conducted to compare the robustness of the proposed models and multivariate linear models (MLMs) with varying numbers and degree of outliers. Note that MLMs assume the multivariate normal distribution which is more vulnerable to outliers than the $t$-distribution. We generated 500 multivariate longitudinal data sets from the multivariate normal distribution with the data sets contaminated by a spot of outliers. To evaluate the robustness, the estimated mean, PRB, CP, SE, SD, and $|PRB|$ are presented in each table.

**Table 1**
Simulation results for parameter estimates of MTLM. 500 data sets were generated from a $t$ distribution with degree of freedom 3. The average regression coefficient estimate (Mean), percent relative bias (PRB), average standard error (SE) and standard deviation (SD) of 500 estimates, as well as coverage probability (CP), total absolute value of PRB ($|PRB|$), and Frobenius norm (Frob) are displayed.

| Parameter (True) | N=100 | | N=300 | | N=500 | |
|---|---|---|---|---|---|---|
| | Mean<br>SE (SD) | PRB<br>CP(%) | Mean<br>SE (SD) | PRB<br>CP(%) | Mean<br>SE (SD) | PRB<br>CP(%) |
| $\beta_{10}$ | 0.311 | 3.52 | 0.309 | 3.01 | 0.304 | 1.41 |
| (0.3) | 0.157 (0.162) | 93.6 | 0.090 (0.091) | 94.2 | 0.070 (0.076) | 90.8 |
| $\beta_{11}$ | -0.115 | 15.35 | -0.109 | 9.29 | -0.101 | 0.82 |
| (-0.1) | 0.222 (0.220) | 95.0 | 0.128 (0.134) | 94.0 | 0.099 (0.107) | 91.8 |
| $\beta_{12}$ | 0.192 | -3.86 | 0.198 | -1.24 | 0.199 | -0.39 |
| (0.2) | 0.067 (0.070) | 94.0 | 0.038 (0.039) | 94.0 | 0.028 (0.028) | 94.6 |
| $\beta_{13}$ | 0.311 | 3.72 | 0.302 | 0.78 | 0.299 | -0.49 |
| (0.3) | 0.092 (0.099) | 93.6 | 0.053 (0.055) | 93.6 | 0.040 (0.042) | 94.8 |
| $\beta_{20}$ | 0.209 | 4.56 | 0.204 | 2.06 | 0.202 | 1.02 |
| (0.2) | 0.131 (0.139) | 93.4 | 0.075 (0.073) | 96.0 | 0.059 (0.060) | 94.0 |
| $\beta_{21}$ | -0.109 | 9.25 | -0.104 | 4.18 | -0.100 | 0.16 |
| (-0.1) | 0.185 (0.184) | 95.4 | 0.107 (0.109) | 94.4 | 0.083 (0.087) | 92.6 |
| $\beta_{22}$ | 0.196 | -2.21 | 0.199 | -0.40 | 0.197 | -1.70 |
| (0.2) | 0.070 (0.071) | 93.8 | 0.041 (0.044) | 91.0 | 0.030 (0.031) | 93.8 |
| $\beta_{23}$ | 0.307 | 2.20 | 0.300 | 0.13 | 0.303 | 1.10 |
| (0.3) | 0.096 (0.100) | 93.6 | 0.056 (0.060) | 94.2 | 0.043 (0.045) | 94.2 |
| $\beta_{30}$ | 0.203 | 6.53 | 0.203 | 1.46 | 0.203 | 1.66 |
| (0.2) | 0.140 (0.144) | 93.8 | 0.081 (0.080) | 95.4 | 0.062 (0.065) | 93.8 |
| $\beta_{31}$ | -0.215 | 7.49 | -0.203 | 1.28 | -0.204 | 1.88 |
| (-0.2) | 0.197 (0.189) | 95.4 | 0.114 (0.117) | 93.8 | 0.088 (0.093) | 93.2 |
| $\beta_{32}$ | 0.198 | -1.00 | 0.202 | 1.06 | 0.198 | -1.14 |
| (0.2) | 0.074 (0.072) | 95.0 | 0.043 (0.045) | 94.8 | 0.032 (0.033) | 94.2 |
| $\beta_{33}$ | 0.406 | 1.44 | 0.399 | -0.22 | 0.401 | 0.14 |
| (0.4) | 0.102 (0.102) | 95.0 | 0.060 (0.064) | 93.6 | 0.046 (0.047) | 94.0 |
| $\nu$ | 3.046 | 1.54 | 3.028 | 0.924 | 3.009 | 0.40 |
| (3) | 0.458 (0.480) | 94.6 | 0.262 (0.267) | 94.6 | 0.201 (0.201) | 95.4 |
| $\|PRB\|$ | | 61.14 | | 25.11 | | 11.91 |
| Frob | 0.075 | | 0.019 | | 0.016 | |

*5.2.1. Study 1*

Outliers were provided for all response values at random time points (see Table 4). The simulation data sets respectively included 1% values of responses contaminated by outliers in simulations of sample sizes 100 and 500, and the outliers were three times the remaining values. The results indicate that even a single outlier (in a sample size of 100) can seriously affect the mean parameter estimates, since all mean parameter estimates in MLM are biased with higher absolute PRB values. For all sample sizes, the total absolute PRBs in MTLM were approximately half of those in MLM. In addition, when the sample size was 500, the results show total absolute PRBs similar to the results listed in Tables 1, 2, and 3. These findings indicate that our proposed estimation is robust to the outliers.

*5.2.2. Study 2*

We considered that 5% values of specific response variables ($y_{it3}$) were contaminated by outliers at random time points. That is, we provided outliers for $y_{it3}$ by adding covariates by the maximum values of the covariates. As presented in Table 5, the sums of absolute PRBs are much larger in MLM than they are in MTLM. Specifically, $\beta_{30}$, $\beta_{31}$, $\beta_{32}$, and $\beta_{33}$ were severely affected by outliers in MLM. On the other hand, the results in MTLM show reduced biases regardless of the existence of outliers. For all sample sizes, the total absolute biases of the estimates were not much different from those without outliers, as presented in the tables (Table 1, 2, 3). Thus, when the data sets had outliers, the estimation using MTLM was robust to the outliers in all cases.

## 6. Analysis of NAFLD data

In this section, we described the application of our proposed models to data collected from a nonalcoholic fatty liver disease (NAFLD) study which aimed to determine the effect of NAFLD on lung function in the general Korean population from October 2003 to December 2016 (Lee et al., 2018). In addition to lung function (forced vital capacity (FVC) and forced expiratory volume in 1 second (FEV1)), Lee et al. (2020) included body mass index (BMI) as a response variable and analyzed the multivariate longitudinal outcomes using multivariate linear models with a multivariate normal distribution. We analyze these three responses (FVC, FEV1, and BMI) while including five explanatory variables from the NAFLD data: group (denoted by Arm, Arm=1 for NAFLD; =0 for non-NAFLD), Sex (male=1; female=0), Age, metabolic syndrome (denoted

**Table 2**
Simulation results for parameter estimates of MTLM. 500 data sets were generated from a $t$distribution with degree of freedom 5. The average regression coefficient estimate (Mean), percent relative bias (PRB), average standard error (SE) and standard deviation (SD) of 500 estimates, as well as coverage probability (CP), total absolute value of PRB ($|PRB|$), and Frobenius norm (Frob) are displayed.

| Parameter (True) | N=100 | | N=300 | | N=500 | |
|---|---|---|---|---|---|---|
| | Mean<br>SE (SD) | PRB<br>CP(%) | Mean<br>SE (SD) | PRB<br>CP(%) | Mean<br>SE (SD) | PRB<br>CP(%) |
| $\beta_{10}$ | 0.304 | 1.34 | 0.302 | 0.59 | 0.302 | 0.77 |
| (0.3) | 0.131 (0.133) | 95.4 | 0.076 (0.081) | 91.4 | 0.059 (0.061) | 93.6 |
| $\beta_{11}$ | -0.090 | -9.75 | -0.098 | -1.74 | -0.102 | 1.59 |
| (-0.1) | 0.185 (0.183) | 95.6 | 0.107 (0.111) | 95.4 | 0.083 (0.092) | 91.6 |
| $\beta_{12}$ | 0.200 | 0.17 | 0.199 | -0.31 | 0.199 | -0.40 |
| (0.2) | 0.056 (0.057) | 93.8 | 0.032 (0.034) | 93.4 | 0.023 (0.023) | 95.6 |
| $\beta_{13}$ | 0.302 | 0.56 | 0.300 | 0.02 | 0.300 | 0.14 |
| (0.3) | 0.077 (0.078) | 95.4 | 0.045 (0.044) | 94.8 | 0.033 (0.034) | 95.0 |
| $\beta_{20}$ | 0.205 | 2.40 | 0.199 | -0.71 | 0.200 | 0.14 |
| (0.2) | 0.109 (0.110) | 95.2 | 0.063 (0.064) | 94.4 | 0.049 (0.050) | 95.0 |
| $\beta_{21}$ | -0.094 | -5.55 | -0.096. | -3.74 | -0.100 | -0.03 |
| (-0.1) | 0.154 (0.148) | 96.8 | 0.089 (0.090) | 94.6 | 0.069 (0.075) | 94.6 |
| $\beta_{22}$ | 0.203 | 1.48 | 0.200 | 0.02 | 0.198 | -0.30 |
| (0.2) | 0.058 (0.062) | 93.2 | 0.034 (0.035) | 94.4 | 0.025 (0.025) | 93.2 |
| $\beta_{23}$ | 0.297 | -0.87 | 0.300 | -0.11 | 0.304 | -0.87 |
| (0.3) | 0.080 (0.085) | 91.6 | 0.047 (0.046) | 95.6 | 0.035 (0.035) | 94.2 |
| $\beta_{30}$ | 0.204 | 1.97 | 0.200 | 0.14 | 0.201 | 1.47 |
| (0.2) | 0.117 (0.119) | 94.6 | 0.067 (0.069) | 94.2 | 0.052 (0.053) | 94.4 |
| $\beta_{31}$ | -0.193 | -3.51 | -0.196 | -1.79 | -0.200 | 0.47 |
| (-0.2) | 0.165 (0.162) | 95.0 | 0.095 (0.100) | 94.8 | 0.074 (0.079) | 95.4 |
| $\beta_{32}$ | 0.203 | 1.53 | 0.201 | 0.68 | 0.200 | 0.19 |
| (0.2) | 0.062 (0.065) | 95.0 | 0.036 (0.037) | 93.2 | 0.026 (0.025) | 93.6 |
| $\beta_{33}$ | 0.402 | 0.45 | 0.400 | -0.11 | 0.402 | 0.08 |
| (0.4) | 0.085 (0.089) | 93.2 | 0.050 (0.051) | 94.4 | 0.037 (0.038) | 94.8 |
| $\nu$ | 5.160 | 3.208 | 5.048 | 0.96 | 5.020 | 0.40 |
| (5) | 0.896 (1.107) | 94.0 | 0.499 (0.516) | 95.4 | 0.383 (0.393) | 93.8 |
| $\|PRB\|$ | | 29.59 | | 9.98 | | 6.86 |
| Frob | 0.058 | | 0.033 | | 0.026 | |

by Meta, Meta=1 for metabolic syndrome; =0 for non-metabolic syndrome), and follow-up duration (Duration). Age and Duration were respectively rescaled as log(age) and (duration $-$ *mean*(duration))/100.

### 6.1. Data description

We first considered the outcomes of the first 10 visits in the NAFLD data (Lee et al., 2020) and we assumed that data were missing at random (MAR). Fig. 1 shows the trend of each mean response variable of the subjects for the first 10 visits. In FVC and FEV1, the mean differences between the two arms were small, and as the number of visits increased, the means of FVC and FEV1 decreased linearly. However, the means of BMI had a non-linear trend with visits, and the mean difference between the two arms was significantly different.

Table 6 presents a brief description of the explanatory variables including the proportion, mean, and standard deviation of each group. These were almost equal proportions of subjects in non-NAFLD and NAFLD. However, the proportion of males was higher than that of females, and the proportion of subjects without any metabolic syndrome was much higher than that of subjects with metabolic syndrome.

Table 7 presents the outcomes of two sample $t$ tests with the categorical explanatory variables used in the analysis. Except for the p-value of Arm for FVC, all variables were significant under the significance level of 0.05. Thus, there were differences in all responses depending on sex and the presence of metabolic syndromes. There were also differences in FEV1 and BMI depending on Arm (NAFLD and non-NAFLD groups). Table 8 lists the correlations of the response variables. FVC and FEV1 show a particularly strong positive correlation (0.955).

We first fit the NAFLD data using a typical MLM, as described in Lee et al. (2020). Using the standardized residuals from the model, we conducted diagnostics to check the adequacy of the MLM. Figs. 2, 3, and 4 show boxplots of the standardized residuals against Arm, Sex, and Meta. These plots show that the residuals varied depending on the level of each predictor variable. There are also a number of outliers. These results show that a heavy-tailed distribution is needed and that the covariance matrix must have a structure that varies with the predictor variables. To elaborate, there are extreme outliers for which the absolute value of the residuals is over 6 in FVC and FEV1 when considering the female group, arm group, and metabolic syndrome group. As indicated in the previous simulation results (Study 1), MLMs have proved vulnerable to few

**Table 3**
Simulation results for parameter estimates of MTLM. 500 data sets were generated from a $t$distribution with degree of freedom 7. The average regression coefficient estimate (Mean), percent relative bias (PRB), average standard error (SE) and standard deviation (SD) of 500 estimates, as well as coverage probability (CP), total absolute value of PRB ($|PRB|$), and Frobenius norm (Frob) are displayed.

| Parameter (True) | N=100 | | N=300 | | N=500 | |
|---|---|---|---|---|---|---|
| | Mean<br>SE (SD) | PRB<br>CP(%) | Mean<br>SE (SD) | PRB<br>CP(%) | Mean<br>SE (SD) | PRB<br>CP(%) |
| $\beta_{10}$ | 0.299 | -0.08 | 0.304 | -0.01 | 0.299 | -0.40 |
| (0.3) | 0.131 (0.134) | 94.8 | 0.076 (0.082) | 96.2 | 0.059 (0.059) | 95.4 |
| $\beta_{11}$ | -0.108 | 7.70 | -0.106 | 1.52 | -0.098 | -1.53 |
| (-0.1) | 0.186 (0.188) | 93.4 | 0.107 (0.107) | 96.4 | 0.084 (0.081) | 95.4 |
| $\beta_{12}$ | 0.204 | 2.00 | 0.204 | 0.54 | 0.201 | 0.31 |
| (0.2) | 0.065 (0.067) | 94.8 | 0.039 (0.039) | 93.8 | 0.029 (0.027) | 95.8 |
| $\beta_{13}$ | 0.293 | -2.50 | 0.294 | -0.82 | 0.300 | -0.03 |
| (0.3) | 0.094 (0.094) | 94.4 | 0.053 (0.050) | 94.2 | 0.040 (0.038) | 95.0 |
| $\beta_{20}$ | 0.200 | 0.07 | 0.201 | -0.88 | 0.200 | -0.05 |
| (0.2) | 0.113 (0.115) | 94.2 | 0.066 (0.068) | 95.0 | 0.051 (0.052) | 94.8 |
| $\beta_{21}$ | -0.109 | 9.25 | -0.104 | -0.60 | -0.099 | -1.33 |
| (-0.1) | 0.161 (0.162) | 95.2 | 0.093 (0.092) | 95.6 | 0.072 (0.072) | 95.8 |
| $\beta_{22}$ | 0.203 | 1.26 | 0.205 | -1.13 | 0.200 | -0.07 |
| (0.2) | 0.070 (0.073) | 95.0 | 0.041 (0.044) | 94.0 | 0.031 (0.030) | 96.0 |
| $\beta_{23}$ | 0.297 | -0.95 | 0.295 | 0.55 | 0.301 | 0.26 |
| (0.3) | 0.100 (0.105) | 93.6 | 0.057 (0.056) | 93.4 | 0.043 (0.043) | 95.4 |
| $\beta_{30}$ | 0.206 | 3.00 | 0.202 | 0.18 | 0.199 | -0.06 |
| (0.2) | 0.121 (0.126) | 93.6 | 0.070 (0.073) | 94.4 | 0.055 (0.055) | 95.0 |
| $\beta_{31}$ | -0.219 | 9.38 | -0.203 | -1.05 | -0.200 | -0.62 |
| (-0.2) | 0.171 (0.177) | 93.2 | 0.100 (0.098) | 96.2 | 0.077 (0.074) | 95.8 |
| $\beta_{32}$ | 0.204 | 1.99 | 0.203 | -0.82 | 0.200 | 0.02 |
| (0.2) | 0.074 (0.076) | 95.2 | 0.043 (0.046) | 94.4 | 0.032 (0.032) | 95.4 |
| $\beta_{33}$ | 0.394 | -1.57 | 0.396 | 0.16 | 0.399 | -0.12 |
| (0.4) | 0.105 (0.110) | 94.4 | 0.060 (0.058) | 95.6 | 0.046 (0.045) | 96.0 |
| $\|PRB\|$ | | 39.75 | | 8.26 | | 4.80 |
| $\nu$ | 7.431 | 6.16 | 7.098 | 1.40 | 7.072 | 1.03 |
| (7) | 1.460 (1.512) | 97.2 | 0.785 (0.798) | 94.8 | 0.604 (0.604) | 94.8 |
| Frob | 0.060 | | 0.019 | | 0.019 | |

extreme outliers. The existence of extreme outliers implies that MLMs are not appropriate for fitting NAFLD data. Therefore, these boxplots indicate that more robust models are needed, as there are a number of outliers.

### 6.2. Model fit

The analysis of the NAFLD study in Lee et al. (2020) suggested an AR(4) covariance matrix with ISDs depending on Arm, Sex, and Meta in MLMs. Fig. 2 also shows the dependence of the predictor variables. We therefore considered the structure of this covariance matrix.

For further analysis, we considered five MTLMs with various covariance matrices depending on Arm, Sex, and Meta (see Table 9). Models 1 - 5 respectively correspond to heteroscedastic AR(1), AR(2), AR(3), AR(4), and AR(5) covariance matrices with ISDs depending on (Arm, Sex, and Meta). We also considered a multivariate normal linear model with an AR(4) covariance matrix depending on Arm, Sex, and Meta (Model 0).

To reduce the number of iterations until convergence, we used the MLE estimated from the model in (Lee et al., 2020) as initial values. The convergence criterion was $\sum_l |\hat{\theta}^{old} - \hat{\theta}^{new}| \le 10^{-5}$ where $\hat{\theta}^{old}$ and $\hat{\theta}^{new}$ are the previous and current fitted values of the parameters, respectively.

Table 10 provides the maximized log-likelihoods and Akaike information criterion (AIC). We first conducted likelihood ratio tests (LRTs) to compare nested models (Model 1 versus Model 2 ($\chi^2 = 2487.30$, $p < 0.0001$); Model 2 versus Model 3 ($\chi^2 = 290.24$, $p < 0.0001$); Model 3 versus Model 4 ($\chi^2 = 41.98$, $p < 0.0001$); and Model 4 versus Model 5 ($\chi^2 = 14.78$, $p = 0.0972$)). The results showed Model 4 was better than the other models. Since Models 0 and 4 were not nested, we compared Models 4 and 0 using AICs; this comparison indicated that Model 4 also outperformed Model 0 (AIC= 51281.60 for Model 0; 46099.92 for Model 4).

Table 11 presents the maximum likelihood estimates of the mean parameters for Models 0, 3, 4, and 5. The estimated degrees of freedom were similar (8.773, 8.763, and 8.763 for Models 3, 4, and 5, respectively). The MLEs of the mean parameters for the four models were similar due to the orthogonality of the mean and the other parameters. Since Model 4 showed the best performance among the models considered, we focus on the estimated mean parameters in that model.

**Table 4**
Simulation results of multivariate linear model (MLM) and multivariate $t$ linear model (MTLM) with outliers. 500 data sets from normal distribution were generated and 1% outliers of the data with all responses at random time points were put into the dataset.

| Parameter (True) | N=100 | | | | N=500 | | | |
|---|---|---|---|---|---|---|---|---|
| | MLM | | MTLM | | MLM | | MTLM | |
| | Mean<br>SE (SD) | PRB<br>CP(%) | Mean<br>SE (SD) | PRB<br>CP(%) | Mean<br>SE (SD) | PRB<br>CP(%) | Mean<br>SE (SD) | PRB<br>CP(%) |
| $\beta_{10}$ | 0.312 | 4.01 | 0.305 | 1.67 | 0.311 | 3.82 | 0.302 | 0.52 |
| (0.3) | 0.110 (0.112) | 94.4 | 0.106 (0.107) | 94.2 | 0.049 (0.050) | 94.8 | 0.048 (0.049) | 94.4 |
| $\beta_{11}$ | -0.114 | 13.82 | -0.106 | 6.43 | -0.105 | 4.59 | -0.095 | -5.50 |
| (-0.1) | 0.155 (0.158) | 93.8 | 0.150 (0.156) | 93.4 | 0.070 (0.070) | 94.8 | 0.067 (0.070) | 94.2 |
| $\beta_{12}$ | 0.209 | 4.45 | 0.203 | 1.42 | 0.206 | 2.77 | 0.200 | 0.10 |
| (0.2) | 0.044 (0.044) | 93.0 | 0.042 (0.046) | 93.0 | 0.020 (0.020) | 93.2 | 0.019 (0.020) | 94.0 |
| $\beta_{13}$ | 0.289 | -3.57 | 0.299 | -0.18 | 0.304 | 1.25 | 0.300 | 0.08 |
| (0.3) | 0.063 (0.064) | 94.4 | 0.059 (0.062) | 93.6 | 0.029 (0.029) | 94.8 | 0.027 (0.029) | 92.2 |
| $\beta_{20}$ | 0.209 | 4.71 | 0.207 | 3.33 | 0.209 | 4.54 | 0.201 | 0.45 |
| (0.2) | 0.091 (0.095) | 95.0 | 0.089 (0.091) | 94.4 | 0.041 (0.041) | 94.4 | 0.040 (0.041) | 93.8 |
| $\beta_{21}$ | -0.108 | 8.80 | -0.107 | 7.21 | -0.105 | 4.63 | -0.095 | -5.49 |
| (-0.1) | 0.129 (0.131) | 94.8 | 0.125 (0.131) | 94.0 | 0.058 (0.057) | 96.6 | 0.056 (0.057) | 94.0 |
| $\beta_{22}$ | 0.206 | 3.10 | 0.201 | 0.73 | 0.206 | 3.04 | 0.201 | 0.38 |
| (0.2) | 0.047 (0.048) | 95.0 | 0.045 (0.046) | 94.2 | 0.021 (0.022) | 93.2 | 0.020 (0.021) | 92.6 |
| $\beta_{23}$ | 0.296 | -1.35 | 0.298 | -0.55 | 0.303 | 0.92 | 0.299 | 0.22 |
| (0.3) | 0.067 (0.069) | 94.0 | 0.064 (0.064) | 94.8 | 0.030 (0.030) | 95.6 | 0.028 (0.030) | 94.0 |
| $\beta_{30}$ | 0.209 | 4.48 | 0.206 | 3.19 | 0.210 | 4.91 | 0.201 | 0.59 |
| (0.2) | 0.098 (0.102) | 94.0 | 0.094 (0.097) | 93.8 | 0.044 (0.046) | 93.4 | 0.042 (0.044) | 93.8 |
| $\beta_{31}$ | -0.209 | 4.52 | -0.209 | 4.35 | -0.207 | 3.36 | -0.192 | -3.77 |
| (-0.2) | 0.138 (0.139) | 94.6 | 0.134 (0.138) | 95.0 | 0.062 (0.063) | 94.8 | 0.060 (0.063) | 93.6 |
| $\beta_{32}$ | 0.206 | 3.20 | 0.201 | 0.70 | 0.206 | 3.18 | 0.201 | 0.74 |
| (0.2) | 0.049 (0.049) | 94.6 | 0.048 (0.049) | 94.4 | 0.022 (0.022) | 93.6 | 0.021 (0.021) | 94.2 |
| $\beta_{33}$ | 0.392 | -1.86 | 0.394 | -1.58 | 0.404 | 1.02 | 0.400 | -0.08 |
| (0.4) | 0.071 (0.073) | 94.0 | 0.067 (0.069) | 94.0 | 0.032 (0.031) | 95.6 | 0.030 (0.030) | 95.8 |
| $\|PRB\|$ | | 57.87 | | 31.34 | | 38.03 | | 17.92 |

**Table 5**
Simulation results of multivariate linear model (MLM) and multivariate $t$ linear model (MTLM) with outliers. 500 data sets from multivariate normal distribution were generated and 5% outliers of the data with specific responses at random time points were put into the dataset.

| Parameter (True) | N=100 | | | | N=500 | | | |
|---|---|---|---|---|---|---|---|---|
| | MLM | | MTLM | | MLM | | MTLM | |
| | Mean<br>SE (SD) | PRB<br>CP(%) | Mean<br>SE (SD) | PRB<br>CP(%) | Mean<br>SE (SD) | PRB<br>CP(%) | Mean<br>SE (SD) | PRB<br>CP(%) |
| $\beta_{10}$ | 0.308 | 2.81 | 0.299 | -0.22 | 0.313 | 4.33 | 0.301 | 0.18 |
| (0.3) | 0.106 (0.108) | 94.2 | 0.105 (0.110) | 92.0 | 0.048 (0.048) | 93.6 | 0.047 (0.050) | 93.6 |
| $\beta_{11}$ | -0.102 | 2.44 | -0.095 | -4.85 | -0.104 | 3.98 | -0.095 | -5.44 |
| (-0.1) | 0.149 (0.156) | 93.0 | 0.149 (0.158) | 93.2 | 0.067 (0.067) | 95.2 | 0.067 (0.072) | 94.2 |
| $\beta_{12}$ | 0.197 | -1.27 | 0.197 | -1.28 | 0.203 | 1.35 | 0.200 | -0.03 |
| (0.2) | 0.043 (0.043) | 94.6 | 0.041 (0.042) | 93.0 | 0.019 (0.020) | 93.8 | 0.018 (0.020) | 92.8 |
| $\beta_{13}$ | 0.305 | 1.75 | 0.300 | 0.01 | 0.298 | -0.65 | 0.300 | -0.16 |
| (0.3) | 0.062 (0.063) | 94.0 | 0.060 (0.064) | 93.4 | 0.027 (0.028) | 95.8 | 0.027 (0.030) | 90.8 |
| $\beta_{20}$ | 0.206 | 2.83 | 0.198 | -1.22 | 0.212 | 6.15 | 0.200 | 0.09 |
| (0.2) | 0.088 (0.090) | 94.2 | 0.088 (0.092) | 94.2 | 0.040 (0.040) | 93.4 | 0.039 (0.042) | 92.2 |
| $\beta_{21}$ | -0.100 | -0.03 | -0.093 | -7.12 | -0.104 | 3.54 | -0.094 | -5.85 |
| (-0.1) | 0.125 (0.128) | 94.6 | 0.124 (0.131) | 92.6 | 0.056 (0.056) | 95.4 | 0.056 (0.059) | 92.8 |
| $\beta_{22}$ | 0.199 | -0.55 | 0.199 | -0.61 | 0.203 | 1.58 | 0.200 | 0.20 |
| (0.2) | 0.046 (0.048) | 94.0 | 0.043 (0.047) | 93.2 | 0.020 (0.019) | 96.2 | 0.020 (0.022) | 91.4 |
| $\beta_{23}$ | 0.305 | 1.51 | 0.301 | 0.35 | 0.297 | -1.11 | 0.299 | -0.44 |
| (0.3) | 0.066 (0.069) | 93.8 | 0.064 (0.071) | 92.0 | 0.029 (0.028) | 96.8 | 0.028 (0.031) | 92.8 |
| $\beta_{30}$ | 0.362 | 81.17 | 0.201 | 0.74 | 0.349 | 74.47 | 0.201 | 0.27 |
| (0.2) | 0.102 (0.090) | 65.6 | 0.096 (0.098) | 93.4 | 0.046 (0.044) | 9.2 | 0.043 (0.046) | 95.0 |
| $\beta_{31}$ | -0.248 | 23.75 | -0.196 | -1.80 | -0.215 | 7.26 | -0.192 | -4.00 |
| (-0.2) | 0.145 (0.131) | 96.6 | 0.136 (0.141) | 94.6 | 0.065 (0.061) | 95.4 | 0.061 (0.065) | 93.6 |
| $\beta_{32}$ | 0.203 | 1.60 | 0.198 | -0.78 | 0.170 | -14.68 | 0.201 | 0.45 |
| (0.2) | 0.063 (0.050) | 98.2 | 0.049 (0.048) | 94.2 | 0.028 (0.021) | 88.0 | 0.022 (0.022) | 95.8 |
| $\beta_{33}$ | 0.357 | -10.78 | 0.400 | 0.04 | 0.443 | 10.70 | 0.399 | -0.20 |
| (0.4) | 0.091 (0.072) | 96.8 | 0.072 (0.071) | 96.0 | 0.040 (0.029) | 88.6 | 0.032 (0.031) | 96.0 |
| $\|PRB\|$ | | 130.49 | | 19.02 | | 129.80 | | 17.31 |

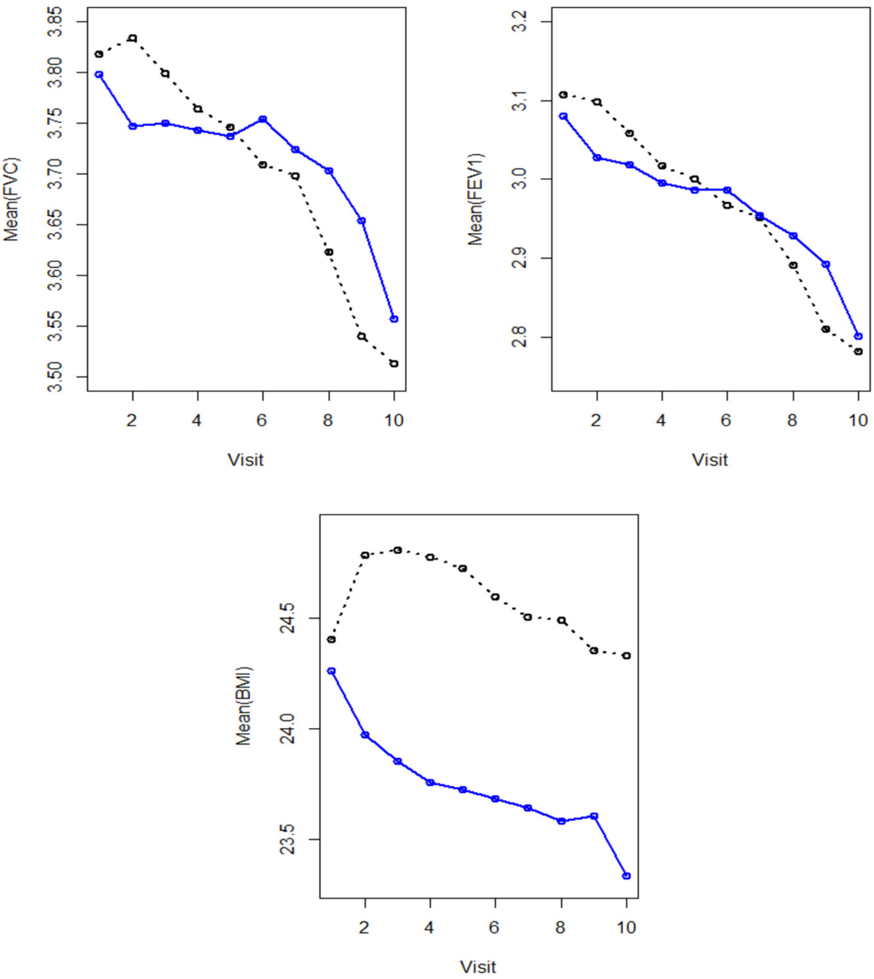


**Fig. 1.** Plots of mean FVC, mean FEV1, and mean BMI of subjects over visit for two arms (NAFLD (solid line) and non-NAFLD (dashed line)).

**Table 6**
Description of explanatory variables (Arm=0 for non-NAFLD, Arm=1 for NAFLD, Sex=0 for female, Sex=1 for male, Meta =0 for non-metabolic syndrome, Meta=1 for metabolic syndrome).

| Variable | | Proportion (count)/ Mean | SD |
|---|---|---|---|
| Arm | 0 | 0.49 (12980) | – |
| | 1 | 0.51 (13555) | – |
| Sex | 0 | 0.36 (9519) | – |
| | 1 | 0.64 (17016) | – |
| Meta | 0 | 0.83 (22142) | – |
| | 1 | 0.17 (4393) | – |
| Duration | | 42.30 | 33.32 |
| Age | | 53.12 | 9.10 |

The fitted models using Model 4 were given by the following equations:

$$\hat{FVC}_{it} = 8.283^{*} - 0.018^{*} Arm_i + 0.007 Duration_{it} - 1.352^{*} Age_{it} + 1.205^{*} Sex_i - 0.008 Meta_{it},$$
$$\hat{FEV}1_{it} = 8.096^{*} - 0.011^{*} Arm_i + 0.004 Duration_{it} - 1.438^{*} Age_{it} + 0.906^{*} Sex_i - 0.010^{*} Meta_{it},$$
$$\hat{BMI}_{it} = 24.267^{*} + 0.306^{*} Arm_i - 0.220^{*} Duration_{it} - 0.224^{*} Age_{it} + 0.845^{*} Sex_i + 0.431^{*} Meta_{it},$$

where * indicates significance with 95% confidence level.

**Table 7**
Two sample t-test with Arm, Sex, Meta (Arm=0 for non-NAFLD, Arm=1 for NAFLD, Sex=0 for female, Sex=1 for male, Meta =0 for non-metabolic syndrome, Meta=1 for metabolic syndrome).

| Explanatory Variables | Responses | Test statistics (P-value) |
|---|---|---|
| Arm | FVC | 0.125 |
| | FEV1 | 0.017 |
| | BMI | $< 0.01$ |
| Sex | FVC | $< 0.01$ |
| | FEV1 | $< 0.01$ |
| | BMI | $< 0.01$ |
| Meta | FVC | $< 0.01$ |
| | FEV1 | $< 0.01$ |
| | BMI | $< 0.01$ |

**Table 8**
Correlation of response variables.

| | FVC | FEV1 | BMI |
|---|---|---|---|
| FVC | 1 | 0.955 | 0.173 |
| FEV1 | 0.955 | 1 | 0.172 |
| BMI | 0.173 | 0.172 | 1 |

**Table 9**
Models for $\phi_{itj,lm}$, $l_{itj,lm}$ and $\log \sigma_{it}$ in the NAFLD data. $A_i$ is $Arm_i$; $S_i$ is $Sex_i$; $M_i$ is $Meta_i$.

| Model | Distribution | GARP | ISD |
|---|---|---|---|
| Model 0 | Normal | $\phi_{itj,lm} = \sum_{g=0}^{3} \alpha_{lmg} I_{(\lvert t-j \rvert = g+1)}$ | $\log \sigma_{itk} = \lambda_{k0} + \lambda_{k1} A_i + \lambda_{k2} S_i + \lambda_{k3} M_{it}$ |
| Model 1 | $t$ | $\phi_{itj,lm} = \alpha_{lm0} I_{(\lvert t-j \rvert = 1)}$ | $\log \sigma_{itk} = \lambda_{k0} + \lambda_{k1} A_i + \lambda_{k2} S_i + \lambda_{k3} M_{it}$ |
| Model 2 | $t$ | $\phi_{itj,lm} = \sum_{g=0}^{1} \alpha_{lmg} I_{(\lvert t-j \rvert = g+1)}$ | $\log \sigma_{itk} = \lambda_{k0} + \lambda_{k1} A_i + \lambda_{k2} S_i + \lambda_{k3} M_{it}$ |
| Model 3 | $t$ | $\phi_{itj,lm} = \sum_{g=0}^{2} \alpha_{lmg} I_{(\lvert t-j \rvert = g+1)}$ | $\log \sigma_{itk} = \lambda_{k0} + \lambda_{k1} A_i + \lambda_{k2} S_i + \lambda_{k3} M_{it}$ |
| Model 4 | $t$ | $\phi_{itj,lm} = \sum_{g=0}^{3} \alpha_{lmg} I_{(\lvert t-j \rvert = g+1)}$ | $\log \sigma_{itk} = \lambda_{k0} + \lambda_{k1} A_i + \lambda_{k2} S_i + \lambda_{k3} M_{it}$ |
| Model 5 | $t$ | $\phi_{itj,lm} = \sum_{g=0}^{4} \alpha_{lmg} I_{(\lvert t-j \rvert = g+1)}$ | $\log \sigma_{itk} = \lambda_{k0} + \lambda_{k1} A_i + \lambda_{k2} S_i + \lambda_{k3} M_{it}$ |

**Table 10**
Maximized log likelihoods and AICs for models.

| Model | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Max. loglik. | -25571.80 | -24389.72 | -23146.07 | -23000.95 | -22979.96 | -22972.57 |
| AIC | 51281.60 | 48951.44 | 46396.14 | 46123.90 | 46099.92 | 46103.14 |

**Table 11**
Maximum likelihood estimates of the mean parameters ($\beta$) for Models 0, 3, 4 and 5. * indicates significance with 95% confidence level.

| | Model 0 | Model 3 | Model 4 | Model 5 |
|---|---|---|---|---|
| Response 1: FVC | | | | |
| Int. | 8.435* (0.088) | 8.282* (0.082) | 8.283* (0.082) | 8.284* (0.082) |
| Arm | −0.018* (0.004) | −0.018* (0.004) | −0.018* (0.004) | −0.018* (0.004) |
| Duration | 0.019* (0.009) | 0.007 (0.008) | 0.007 (0.008) | 0.008* (0.008) |
| Age | −1.385* (0.022) | −1.352* (0.021) | −1.352* (0.021) | −1.352* (0.021) |
| Sex | 1.226* (0.008) | 1.205* (0.007) | 1.205* (0.007) | 1.205* (0.007) |
| Meta | −0.008 (0.005) | −0.008* (0.004) | −0.008* (0.004) | −0.008* (0.004) |
| Response 2: FEV1 | | | | |
| Int. | 8.253* (0.071) | 8.093* (0.067) | 8.096* (0.067) | 8.097* (0.067) |
| Arm | −0.013* (0.003) | −0.011* (0.003) | −0.011* (0.003) | −0.011* (0.003) |
| Duration | 0.012 (0.007) | 0.004 (0.007) | 0.004 (0.007) | 0.004 (0.007) |
| Age | −1.474* (0.018) | −1.438* (0.017) | −1.438* (0.017) | −1.439* (0.017) |
| Sex | 0.917* (0.006) | 0.905 (0.006) | 0.906* (0.006) | 0.906* (0.006) |
| Meta | −0.010* (0.004) | −0.010* (0.003) | −0.010* (0.003) | −0.010* (0.003) |
| Response 3: BMI | | | | |
| Int. | 24.745* (0.356) | 24.266* (0.328) | 24.267* (0.328) | 24.266* (0.328) |
| Arm | 0.325* (0.016) | 0.306* (0.015) | 0.306* (0.015) | 0.306* (0.015) |
| Duration | −0.202* (0.039) | −0.221* (0.035) | −0.220* (0.035) | −0.220* (0.035) |
| Age | −0.317* (0.089) | −0.224* (0.082) | −0.224* (0.082) | −0.225* (0.082) |
| Sex | 0.775* (0.035) | 0.845* (0.033) | 0.845* (0.033) | 0.846* (0.033) |
| Meta | 0.475* (0.019) | 0.431* (0.017) | 0.431* (0.017) | 0.431* (0.017) |
| Max. loglik. | −25571.80 | −23000.95 | −22979.96 | −22972.57 |
| AIC | 51281.60 | 46123.90 | 46099.92 | 46103.14 |

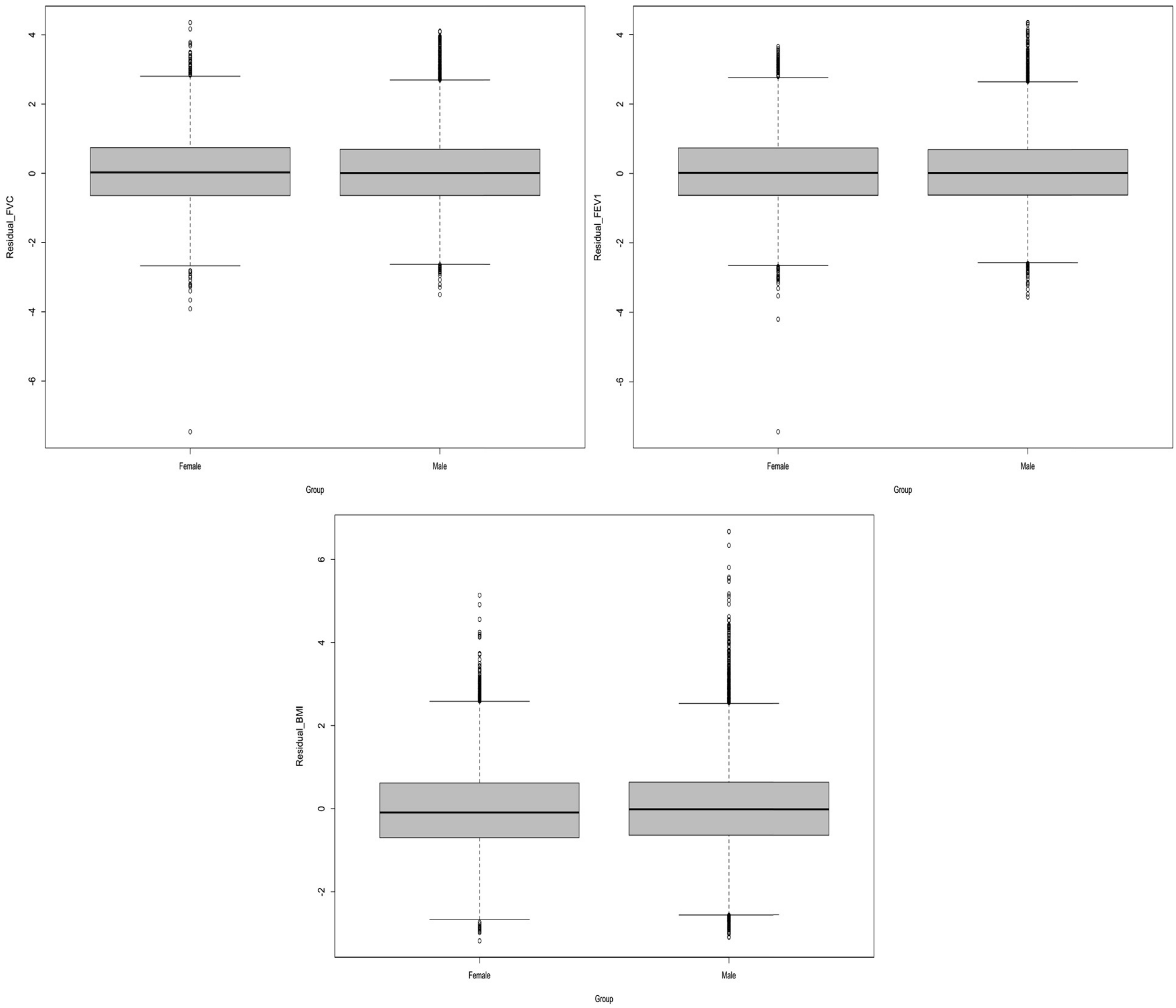


**Fig. 2.** The boxplots of standardized residuals for each response (FVC, FEV1, and BMI) against Sex.

In the response FVC, all coefficients of covariates except for Duration were significant. It can be seen that the estimated mean of FVC was lower for NAFLD than for non-NAFLD, and as the subject's age increased, the estimated mean of FVC decreased. It was also higher for males than for females. Lastly, estimated mean of FVC was lower for subjects without any metabolic syndromes than it was for those with a metabolic syndrome.

In the response FEV1, all coefficients of covariates except Duration were significant. This indicates that the estimated mean of FEV1 was lower for NAFLD than for non-NAFLD, and that as the subject's age increased, the estimated mean of FEV1 decreased. The estimated mean of FEV1 was higher for males than females while it was lower for subjects who had a metabolic syndrome than it was for those who did not have any metabolic syndromes.

In the response BMI, all coefficients of covariates were significant. This indicates that the estimated mean of BMI for NAFLD was higher than that for non-NAFLD, and that it decreased as the follow-up duration increased. Further, as the subject's age increased, the estimated mean of BMI decreased, and it was higher for males than females. The estimated mean of BMI was lower for subjects without any metabolic syndromes than it was for those with a metabolic syndrome.

Note that, in the above, each subject's age is the logarithm of the age and follow-up duration was also rescaled.

Table 12 presents the estimates of GARPs, $\alpha_{kg}$. The results indicate that in the first-, second-, and third-order, all responses were correlated with their own. In the fourth-order, only FEV1 was correlated with the same response. Overall, two lung functions (FVC and FEV1) were correlated, and when BMI was given in the first and second order, all responses were dependent.

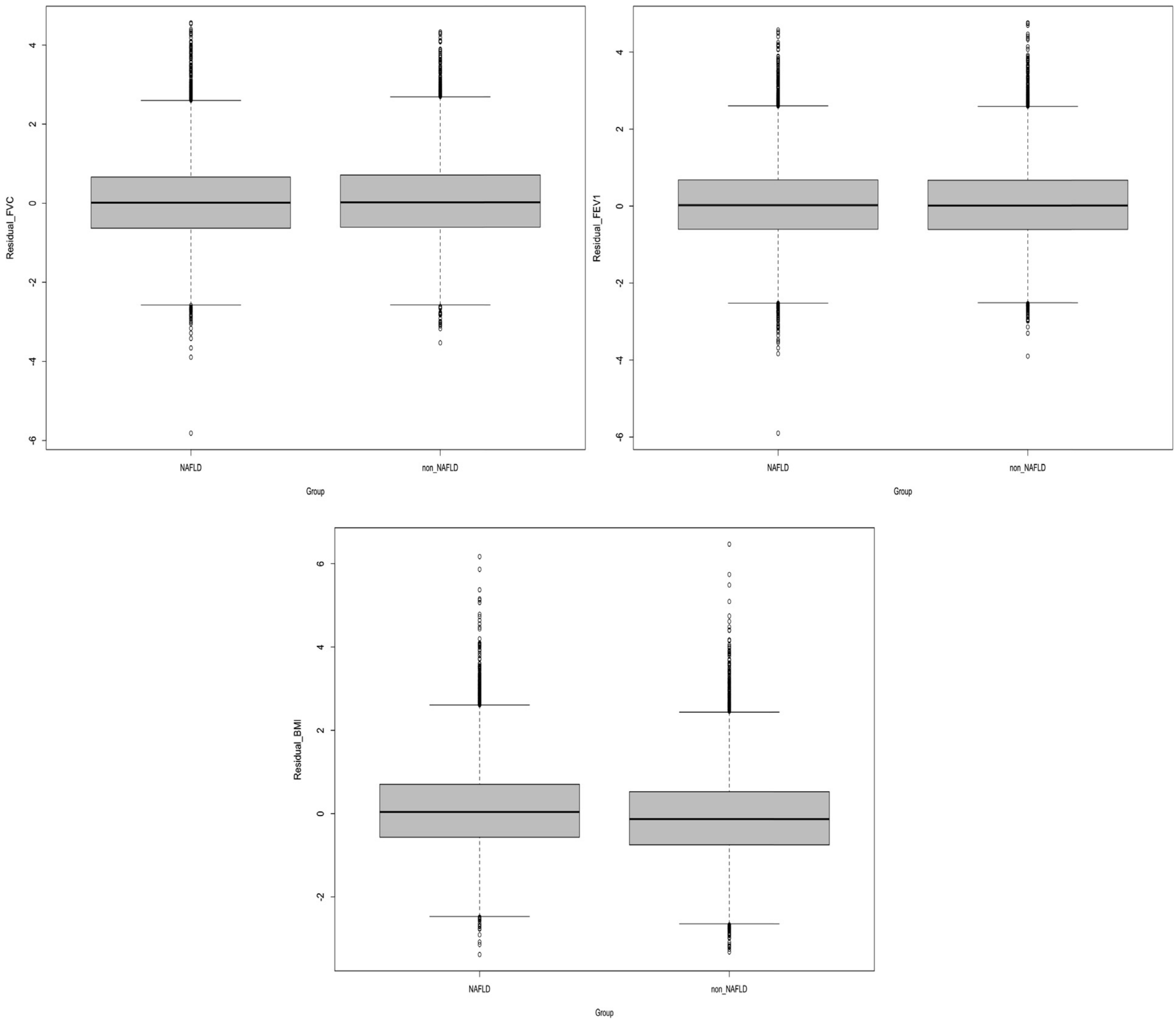


**Fig. 3.** The boxplots of standardized residuals for each response (FVC, FEV1, and BMI) against Arm.

**Table 12**
Maximum likelihood estimates of $\alpha_{kg}$ for Model 4. * indicates significance with 95% confidence level.

| $\alpha_{kg}$ | Estimate | | |
|---|---|---|---|
| $I_{\{\lvert t-j\rvert=1\}}$ | 0.604* | 0.208* | 0.008* |
| $\alpha_{kg1}$ | 0.047* | 0.716* | 0.008* |
| | 0.004 | 0.070 | 0.833* |
| $I_{\{\lvert t-j\rvert=2\}}$ | 0.242* | −0.109* | −0.005* |
| $\alpha_{kg2}$ | 0.005* | 0.135* | −0.003* |
| | 0.021 | −0.051 | 0.119* |
| $I_{\{\lvert t-j\rvert=3\}}$ | 0.079* | −0.069* | −0.002 |
| $\alpha_{kg3}$ | −0.025 | 0.060* | −0.001 |
| | −0.010 | −0.037 | 0.031* |
| $I_{\{\lvert t-j\rvert=4\}}$ | 0.020 | 0.010 | −0.001 |
| $\alpha_{kg4}$ | −0.025 | 0.056* | −0.003 |
| | −0.054 | 0.097 | 0.005 |

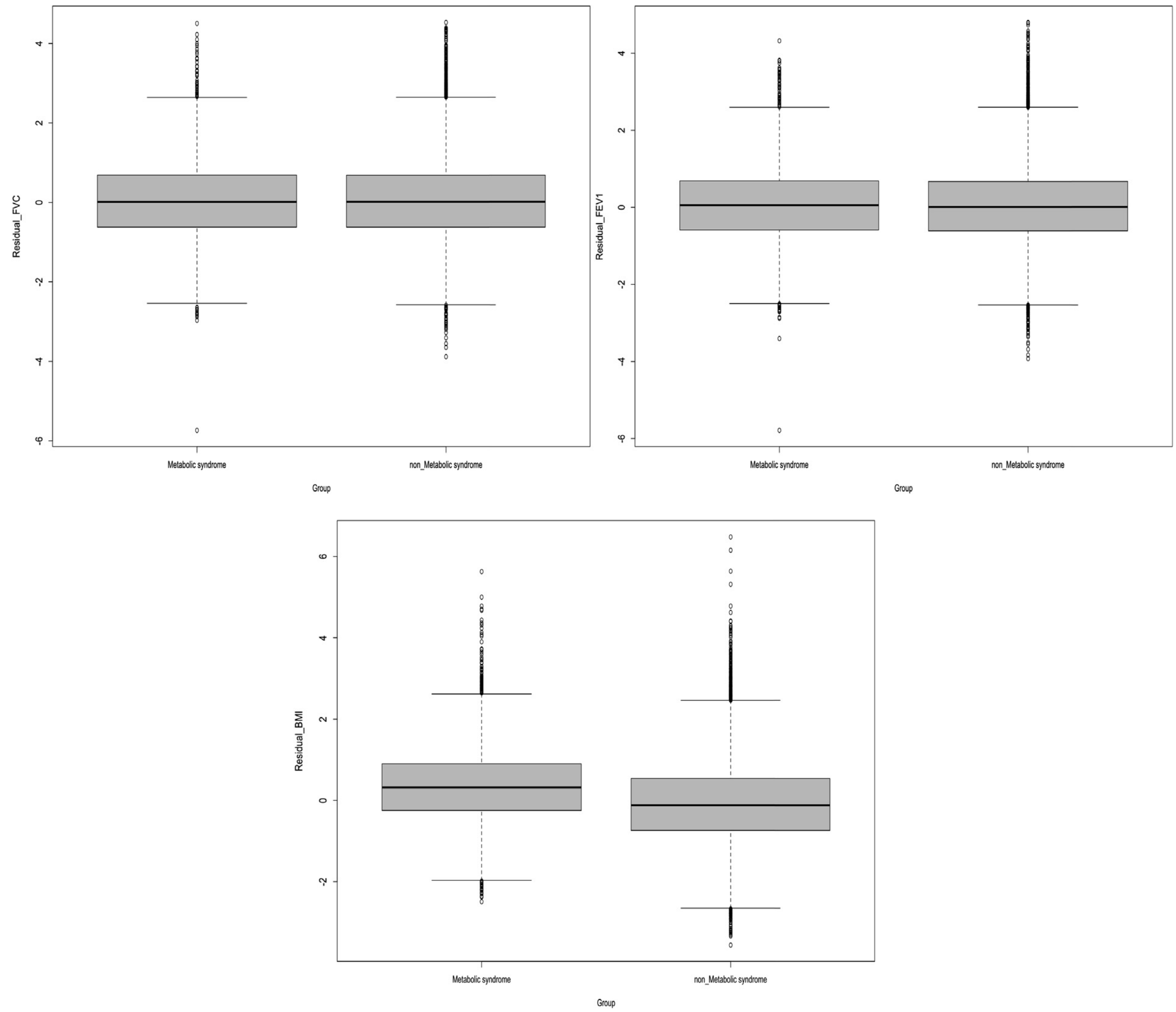


**Fig. 4.** The boxplots of standardized residuals for each response (FVC, FEV1, and BMI) against Meta.

The estimated ISDs were given by:

$$\log \hat{\sigma}_{it1} = -1.593^{*} - 0.026^{*} Arm_i + 0.347^{*} Sex_i - 0.036^{*} Meta_{it},$$

$$\log \hat{\sigma}_{it2} = -1.790^{*} - 0.021^{*} Arm_i + 0.333^{*} Sex_i - 0.050^{*} Meta_{it},$$

$$\log \hat{\sigma}_{it3} = 0.106^{*} - 0.034^{*} Arm_i - 0.162^{*} Sex_i - 0.019 Meta_{it},$$

where * indicates significance with 95% confidence level.

All the estimated log(ISD) ($\log \hat{\sigma}_{it1}$) for FVC, FEV1, and BMI were significant. This indicates that the estimated prediction SD differed depending on Arm, Sex, and Meta. This means that the estimated covariance matrix was heteroscedastic depending on the subject's treatment arm, gender, and the presence of metabolic syndrome.

The estimated correlation matrix between responses was given by:

$$\hat{R}_i = \begin{pmatrix} 1.000 & 0.833 & -0.052 \\ 0.833 & 1.000 & -0.053 \\ -0.052 & -0.053 & 1.000 \end{pmatrix}.$$

Thus, the estimated correlations of FVC versus FEV1, FVC versus BMI, and FEV1 versus BMI were 0.833, -0.052, and -0.053, respectively.

## 7. Conclusion

In this paper, we proposed multivariate $t$ linear models (MTLMs) for multivariate longitudinal data for which the normal assumption was not guaranteed. Unlike the prior literature using Kronecker product structured covariance matrix, we employed a flexible structured covariance matrix using modified Cholesky decomposition (MCD). The covariance matrix in MTLMs was decomposed into the generalized autoregressive parameter matrix (GARPM), correlation matrix, and diagonal matrix with innovation standard deviations (ISDs) using the MCD and hypersphere decompositions (HD) to explains three correlations: the correlation within separate responses over time, the correlation between different responses at the same time point, and the cross-correlation between different responses at different times. Using the two decompositions, we decompose the covariance matrix into unconstrained parameters, which this ensures that the covariance matrix is positive-definite and can be heteroscedastic with fewer parameters.

Through the simulation studies, we showed that the algorithm of MTLM works well, and that MTLM is robust when outliers exist and the data exhibit heavy tails. As a result, we conclude that the proposed MTLM is a promising model for multivariate longitudinal data.

The performance of the MTLM was obtained through an analysis of the NAFLD study, and we compared several models with different covariance matrices to identify the best model for NAFLD study. There were significant differences in each of FEV1, FVC, and BMI between the two arms. Controlling the other covariates (sex, age, duration, and meta), NAFLD had negative effects on FEV1 and FVC along with positive effects on BMI, respectively.

We can extend MTLMs to multivariate $t$ linear mixed models for multivariate longitudinal data. In the models, the random effects and within-subject variations are decomposed and the within-subject variations are also decomposed using the MCD and HD. Further, we consider multivariate $t$ linear models with an autoregressive moving-average (ARMA) structured scaled matrix to accommodate multivariate longitudinal data with many replications. Instead of using the high-order AR structure for the scale matrix, we utilize the ARMA structured scale matrix using autoregressive moving-average Cholesky decomposition (ARMACD) (Lee et al., 2021). Finally we also consider the modeling of multivariate longitudinal categorical data in multivariate generalized linear mixed models (MGLMMs). In the MGLMMs, the covariance matrix for latent variables has the three correlations that are modeled using the MCD and HD. These will continue to be explored in ongoing work.

## 8. Software

Software in the form of R code is available on request from the corresponding author (keunbaik@skku.edu).


## Acknowledgements

This project was supported by Basic Science Research Program through the National Research Foundation of Korea (KRF) funded by the Ministry of Education, Science and Technology (NRF-2019R1F1A1058553).


## Appendix A

Detailed calculations on the Fisher information matrix

$$I(\alpha) = E\left(-\frac{\partial^2 \log L(\theta; y)}{\partial\alpha\partial\alpha^T}\right) = \left(I_{\alpha_{l_1 m_1 g_1}, \alpha_{l_2 m_2 g_2}}\right),$$

$$I(\lambda) = E\left(-\frac{\partial^2 \log L(\theta; y)}{\partial\lambda\partial\lambda^T}\right) = \left(I_{\lambda_{k_1 g_1}, \lambda_{k_2 g_2}}\right),$$

$$I(\delta) = E\left(-\frac{\partial^2 \log L(\theta; y)}{\partial\delta\partial\delta^T}\right) = \left(I_{\delta_{g_1}, \delta_{g_2}}\right),$$

$$I(\alpha, \lambda) = E\left(-\frac{\partial^2 \log L(\theta; y)}{\partial\lambda\partial\alpha^T}\right) = \left(I_{\lambda_{k_1 g_1}, \alpha_{l_2 m_2 g_2}}\right),$$

$$I(\alpha, \delta) = E\left(-\frac{\partial^2 \log L(\theta; y)}{\partial\delta\partial\alpha^T}\right) = \left(I_{\delta_{g_1}, \alpha_{l_2 m_2 g_2}}\right),$$

$$I(\lambda, \delta) = E\left(-\frac{\partial^2 \log L(\theta; y)}{\partial\delta\partial\lambda^T}\right) = \left(I_{\delta_{g_1}, \lambda_{k_2 g_2}}\right),$$

where

$$I_{\alpha_{l_1 m_1 g_1}, \alpha_{l_2 m_2 g_2}} = \sum_{i=1}^{N} \frac{n_i K + \nu}{n_i K + \nu + 2} tr\left(\Sigma_i \frac{\partial T_i^T}{\partial \alpha_{l_1 m_1 g_1}} D_i^{-1} \frac{\partial T_i}{\partial \alpha_{l_2 m_2 g_2}}\right), \tag{21}$$

$$I_{\lambda_{k_1g_1},\lambda_{k_2g_2}} = \sum_{i=1}^{N} \frac{n_iK+\nu}{2(n_iK+\nu+2)} tr\left(\Sigma_i T_i^T \frac{\partial^2 D_i^{-1}}{\partial\lambda_{k_1g_1}\partial\lambda_{k_2g_2}} T_i\right) + \sum_{i=1}^{N} \frac{1}{n_iK+\nu+2} \left\{2\left(\sum_{t=1}^{n_i} h_{itg_1}\right)\left(\sum_{t=1}^{n_i} h_{itg_2}\right)\right.$$
$$\left. + \left(\sum_{t=1}^{n_i} h_{itg_2}\right) tr\left(\Sigma_i T_i^T \frac{\partial D_i^{-1}}{\partial\lambda_{l_1g_1}} T_i\right) + \left(\sum_{t=1}^{n_i} h_{itg_1}\right) tr\left(\Sigma_i T_i^T \frac{\partial D_i^{-1}}{\partial\lambda_{l_2g_2}} T_i\right)\right\}, \tag{22}$$

$$I_{\delta_{g_1},\delta_{g_2}} = \sum_{i=1}^{N} \left\{ \frac{n_i(n_iK+\nu)}{n_iK+\nu+2} \sum_{k=2}^{K} \left( \frac{\frac{\partial^2 f_{ikk}}{\partial\delta_{g_1}\partial\delta_{g_2}}}{f_{ikk}} - \frac{\frac{\partial f_{ikk}}{\partial\delta_{g_2}}\frac{\partial f_{ikk}}{\partial\delta_{g_1}}}{f_{ikk}^2} \right) + \frac{n_iK+\nu}{2(n_iK+\nu+2)} tr\left(\Sigma_i T_i^T \frac{\partial^2 D_i^{-1}}{\partial\delta_{g_1}\partial\delta_{g_2}} T_i\right)\right.$$
$$+ \frac{2n_i^2}{n_iK+\nu+2}\left(\sum_{k=2}^{K}\frac{\frac{\partial f_{ikk}}{\partial\delta_{g_1}}}{f_{ikk}}\right)\left(\sum_{k=2}^{K}\frac{\frac{\partial f_{ikk}}{\partial\delta_{g_2}}}{f_{ikk}}\right) + \frac{n_i}{n_iK+\nu+2}\left(\sum_{k=2}^{K}\frac{\frac{\partial f_{ikk}}{\partial\delta_{g_1}}}{f_{ikk}}\right) tr\left(\Sigma_i T_i^T \frac{\partial D_i^{-1}}{\partial\delta_{g_2}} T_i\right)$$
$$\left. + \frac{n_i}{n_iK+\nu+2}\left(\sum_{k=2}^{K}\frac{\frac{\partial f_{ikk}}{\partial\delta_{g_2}}}{f_{ikk}}\right) tr\left(\Sigma_i T_i^T \frac{\partial D_i^{-1}}{\partial\delta_{g_1}} T_i\right)\right\}, \tag{23}$$

$$I(\nu) = \frac{1}{4}\sum_{i=1}^{N}\left\{\psi\left(\frac{\nu+n_iK}{2}\right) - \psi\left(\frac{\nu}{2}\right) - 2\frac{n_iK(\nu+n_iK+4)}{\nu(\nu+n_iK)(\nu+n_iK+2)}\right\}, \tag{24}$$

$$I_{\lambda_{k_1g_1},\alpha_{l_2m_2g_2}} = \sum_{i=1}^{N}\left\{\frac{n_iK+\nu}{n_iK+\nu+2} tr\left(\Sigma_i T_i^T \frac{\partial D_i^{-1}}{\partial\lambda_{k_1g_1}}\frac{\partial T_i}{\partial\alpha_{l_2m_2g_2}}\right)\right.$$
$$\left. + \frac{2}{n_iK+\nu+2}\left(\sum_{t=1}^{n_i} h_{itg_1}\right) tr\left(\Sigma_i T_i^T D_i^{-1}\frac{\partial T_i}{\partial\alpha_{l_2m_2g_2}}\right)\right\}, \tag{25}$$

$$I_{\delta_{g_1},\alpha_{l_2m_2g_2}} = \sum_{i=1}^{N}\left\{\frac{n_iK+\nu}{n_iK+\nu+2} tr\left(\Sigma_i T_i^T \frac{\partial D_i^{-1}}{\partial\delta_{g_1}}\frac{\partial T_i}{\partial\alpha_{l_2m_2g_2}}\right)\right.$$
$$\left. - \frac{2n_i}{n_iK+\nu+2}\sum_{l=2}^{K}\frac{\frac{\partial f_{ill}}{\partial\delta_{g_1}}}{f_{ill}} tr\left(\Sigma_i T_i^T D_i^{-1}\frac{\partial T_i}{\partial\alpha_{l_2m_2g_2}}\right)\right\}, \tag{26}$$

$$I_{\nu,\alpha_{l_2m_2g_2}} = \sum_{i=1}^{N}\frac{1}{(n_iK+\nu)(n_iK+\nu+2)}\left\{tr\left(\frac{\partial T_i^T}{\partial\alpha_{l_2m_2g_2}}T_i^{-T}\right) + tr\left(T_i^{-1}\frac{\partial T_i}{\partial\alpha_{l_2m_2g_2}}\right)\right\}, \tag{27}$$

$$I_{\delta_{g_1},\lambda_{k_2g_2}} = \sum_{i=1}^{N}\left[\frac{n_iK+\nu}{2(n_iK+\nu+2)} tr\left(\Sigma_i T_i^T \frac{\partial^2 D_i^{-1}}{\partial\delta_{g_2}\partial\lambda_{k_2g_2}} T_i\right) + \frac{1}{n_iK+\nu+2}\left\{2n_i\left(\sum_{m=2}^{K}\frac{\frac{\partial f_{imm}}{\partial\delta_{g_1}}}{f_{imm}}\right)\left(\sum_{t=1}^{n_i} h_{itg_2}\right)\right.\right.$$
$$\left.\left. + \left(\sum_{t=1}^{n_i} h_{itg_2}\right) tr\left(\Sigma_i T_i^T \frac{\partial D_i^{-1}}{\partial\delta_{g_1}} T_i\right) + n_i\left(\sum_{m=2}^{K}\frac{\frac{\partial f_{imm}}{\partial\delta_{g_1}}}{f_{imm}}\right) tr\left(\Sigma_i T_i^T \frac{\partial D_i^{-1}}{\partial\lambda_{k_2l_2g_2}} T_i\right)\right\}\right], \tag{28}$$

$$I_{\nu,\lambda_{k_2g_2}} = \sum_{i=1}^{N}\frac{1}{(n_iK+\nu)(n_iK+\nu+2)} tr\left(\Sigma_i T_i^{-1}\frac{\partial D_i^{-1}}{\partial\lambda_{k_2g_2}} T_i\right), \tag{29}$$

$$I_{\nu,\delta_{g_1}} = \sum_{i=1}^{N}\frac{1}{(n_iK+\nu)(n_iK+\nu+2)} tr\left(D_i\frac{\partial D_i^{-1}}{\partial\delta_{g_1}}\right). \tag{30}$$

**Proof of Proposition.** (a) Proof of $E(\tau_i) = \left(\frac{\nu+1}{\nu+2}\right)^{\frac{n_iK}{2}}$:

$$E(\tau_i) = \int \frac{n_iK+\nu}{\nu+\Delta_i(\theta)} f(r_i;\nu,\Sigma_i)dr_i$$
$$= \left(\frac{\nu+1}{\nu+2}\right)^{\frac{n_iK}{2}} \int f\left(r_i;\nu+2,\frac{\nu}{\nu+2}\Sigma_i\right)dr_i$$

$$= \left(\frac{\nu+1}{\nu+2}\right)^{\frac{n_iK}{2}},$$

where $\Delta_i(\theta) = (y_i - X_i\beta)^T \Sigma_i^{-1}(y_i - X_i\beta)$, $f(r_i; \nu, \Sigma)$ is the $t$ distribution with mean vector 0, scale covariance matrix $\Sigma$, and degrees of freedom $\nu$.

(b) Proof of $E(\tau_i r_i) = 0$:

$$\begin{aligned} E(\tau_i r_i) &= \int \frac{\nu+n_iK}{\nu+\Delta_i(\theta)} r_i f(r_i; \nu, \Sigma_i) dr_i \\ &= \int r_i f\left(r_i; \nu+2, \frac{\nu}{\nu+2}\Sigma_i\right) dr_i \\ &= 0. \end{aligned}$$

(c) Proof of $E(\tau_i^2 r_i) = 0$:

$$\begin{aligned} E(\tau_i^2 r_i) &= \int \left(\frac{\nu+n_iK}{\nu+\Delta_i(\theta)}\right)^2 r_i f(r_i; \nu, \Sigma_i) dr_i \\ &= \left(\frac{\nu+n_iK}{\nu}\right)^2 \frac{(\nu+2)\nu}{(n_iK+\nu+2)(n_iK+\nu)} \int r_i f\left(r_i; \nu+4, \frac{\nu}{\nu+4}\Sigma_i\right) dr_i \\ &= 0. \end{aligned}$$

(d) Proof of $E(\tau_i t_i r_i^T) = \Sigma_i$:

$$\begin{aligned} E(\tau_i r_i r_i^T) &= \int \frac{\nu+n_iK}{\nu+\Delta_i(\theta)} r_i r_i^T f(r_i; \nu, \Sigma_i) dr_i \\ &= \int r_i r_i^T f\left(r_i; \nu+2, \frac{\nu}{\nu+2}\Sigma_i\right) dr_i \\ &= \Sigma_i. \end{aligned}$$

(e) Proof of $E(\tau_i^2 r_i r_i^T) = \Sigma_i$:

$$\begin{aligned} E(\tau_i^2 r_i r_i^T) &= \int \left(\frac{\nu+n_iK}{\nu+\Delta_i(\theta)}\right)^2 r_i r_i^T f(r_i; \nu, \Sigma_i) dr_i \\ &= \frac{(n_iK+\nu)(\nu+2)}{(n_iK+\nu+2)\nu} \int r_i r_i^T f\left(r_i; \nu+4, \frac{\nu}{\nu+4}\Sigma_i\right) dr_i \\ &= \frac{\nu+n_iK}{n_iK+\nu+2}\Sigma_i. \end{aligned}$$

## References


Anderson, T.W., 1973. Asymptotically efficient estimation of covariance matrices with linear structure. Ann. Stat. 1, 135–141.

Chiu, T.Y.M., Leonard, T., Tsui, K.W., 1996. The matrixlogarithm covariance matrix. J. Am. Stat. Assoc. 91, 198–210.

Feng, S., Lian, H., Xue, L., 2016. A new nested cholesky decomposition and estimation for the covariance matrix of bivariate longitudinal data. Comput. Stat. Data Anal. 102, 98–109.

He, X., Cui, H., Simpson, D.G., 2004. Longitudinal data analysis using t-type regression. J. Stat. Plan. Inference 122, 253–269.

Kim, C., Zimmerman, D.L., 2012. Unconstrained models for the covariance structure of multivariate longitudinal data. J. Multivar. Anal. 107, 104–118.

Kohli, P., Garcia, T.P., Pourahmadi, M., 2016. Modeling the cholesky factors of covariance matrices of multivariate longitudinal data. J. Multivar. Anal. 145, 87–100.

Lee, C.-H., Choi, S.H., Chung, G.E., Park, B., Kwak, M.-S., 2018. Nonalcoholic fatty liver disease is associated with decreased lung function. Liver Int. 38 (11), 2091–2100.

Lee, K., Cho, H., Kwak, M.-S., Jang, E.J., 2020. Estimation of covariance matrix of multivariate longitudinal data using modified choleksky and hypersphere decompositions. Biometrics 76 (1), 75–86.

Lee, K., Lee, C.-H., Kwak, M.-S., Jang, E.J., 2021. Analysis of multivariate longitudinal data using arma cholesky and hypersphere decompositions. Comput. Stat. Data Anal. 107144.

Lee, K.-J., Chen, R.-B., 2019. Bayesian variable selection in a finite mixture of linear mixed-effects models. J. Stat. Comput. Simul. 89 (13), 2434–2453.

Lee, K., Baek, C., Daniels, M.J., 2017. ARMA Cholesky factor models for the covariance matrix of linear models. Comput. Stat. Data Anal. 115, 267–280.

Lin, T.-I., Wang, W.-L., 2020. Multivariate-t linear mixed models with censored responses, intermittent missing values and heavy tails. Stat. Methods Med. Res. 29 (5), 1288–1304.

Lin, T.-I., Wang, W.-L., 2011. Bayesian inference in joint modelling of location and scale parameters of the t distribution for longitudinal data. J. Stat. Plan. Inference 141, 1543–1553.

Lin, T.-I., Wang, Y.-J., 2009. A robust approach to joint modeling of mean and scale covariance for longitudinal data. J. Stat. Plan. Inference 139, 3013–3026.

Peel, D., McLachlan, G.J., 2000. Robust mixture modelling using the t distribution. Stat. Comput. 10, 339–348.

Pinheiro, J.C., Liu, C., Wu, Y.N., 2001. Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. J. Comput. Graph. Stat. 10 (2), 249–276.
Pourahmadi, M., 1999. Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. Biometrika 86 (3), 677–690.
Rapisarda, F., Brigo, D., Mercurio, F., 2007. Parameterizing correlations: a geometric interpretation. IMA J. Manag. Math. 18, 55–73.
Wang, W.-L., Fan, T.-H., 2010. Ecm-based maximum likelihood inference for multivariate linear mixed models with autoregressive errors. Comput. Stat. Data Anal. 54, 1328–1341.
Wang, W.-L., Fan, T.-H., 2011. Estimation in multivariate t linear mixed models for multiple longitudinal data. Stat. Sin. 21, 1857–1880.
Wang, W.-L., Lin, T.-I., 2014. Multivariate t nonlinear mixed-effects models for multi-outcome longitudinal data with missing values. Stat. Med. 33, 3029–3046.
Wang, W.-L., Lin, T.-I., Lachos, V.H., 2018. Extending multivariate-t linear mixed models for multiple longitudinal data with censored responses and heavy tails. Stat. Methods Med. Res. 27 (1), 48–64.
Xu, J., Mackenzie, G., 2012. Modelling covariance structure in bivariate marginal models for longitudinal data. Biometrika 99 (3), 649–662.
Zhang, W., Leng, C., Tang, C.Y., 2015. A joint modelling approach for longitudinal studies. J. R. Stat. Soc. B 77, 219–238.